

Wilhelm Wall

Energetisch vergleichbare Städtegruppen -
eine gesamtheitliche Clusteranalyse und
Clusterauswahl deutscher kreisfreier Städte
auf Basis der typischen Verbrauchssektoren
und sozio-energetischer Indikatoren

Bochum 2016

W. Wall: Energetisch vergleichbare Städtegruppen

Energetisch vergleichbare Städtegruppen – Eine gesamtheitliche Clusteranalyse und Clusterauswahl deutscher kreisfreier Städte auf Basis der typischen Verbrauchssektoren und sozio-energetischer Indikatoren

Wilhelm Wall

Ruhr-Universität Bochum

Juli 2016

© 2016 Selbstverlag des Lehrstuhls Energiesysteme und Energiewirtschaft,
Ruhr-Universität Bochum, 44780 Bochum

ISBN 978-3-934951-41-9

Das dieser Untersuchung zugrunde liegende Vorhaben „Eneff Stadt BF II: Begleitforschung zum Wettbewerb Energieeffiziente Stadt“ wird vom Bundesministerium für Bildung und Forschung (BMBF) unter dem Förderkennzeichen 03SF0415A gefördert.

Inhaltsverzeichnis

1	Einführung	1
2	Methodik, Arten der Clusteranalyse und Datengrundlagen	3
2.1	Methodik - Clusteranalyse	3
2.2	Wichtige Begriffe der Faktor- und Clusteranalyse	4
2.2.1	Cluster	4
2.2.2	Objekt	5
2.2.3	Variable	5
2.2.4	Ähnlichkeitsmaß, Distanzmaß	7
2.2.5	Fusionsalgorithmus - Fusionsmethode	9
2.3	Clusteranalyseverfahren und -methoden	9
2.3.1	Einteilung der Clusteranalyseverfahren	9
2.3.2	Erläuterung der Clusteranalyseverfahren	12
2.3.3	Verwendete Software	18
2.4	Auswahl einiger clusteranalytischer Untersuchungen	19
2.5	Untersuchungsrahmen	26
2.5.1	Voruntersuchung - Methodenscreening	26
2.5.2	Untersuchung kreisfreier Städte	27
2.5.3	Verbrauchssektoren	31
3	Voruntersuchung – Anwendung und Analyse diverser Clusteranalysemethoden	35
3.1	Datenbasis für die Voruntersuchung	35
3.2	Clusteranalyse mit gängigen, hierarchischen agglomerativen Algorithmen	36
3.2.1	Single-Linkage-Verfahren (nächster Nachbar)	36
3.2.2	Complete-Linkage-Verfahren (entferntester Nachbar)	39
3.2.3	Average-Linkage-Verfahren (zwischen den Gruppen)	41
3.2.4	Ward-Verfahren	42
3.3	Gegenüberstellung, Analyse und Auswertung der Clusterergebnisse	44
4	Datenerhebung für betrachtete kreisfreie Städte	47
4.1	Definition, Formen, Konzept und Datenbasis von Indikatoren	47
4.1.1	Formen von Indikatoren	47
4.1.2	Konzept und Datenbasis	49
4.2	Indikatoren des Verbrauchssektors private Haushalte	50
4.3	Indikatoren des Verbrauchssektors Verkehr	60
4.4	Indikatoren der Verbrauchssektoren Industrie und GHD	67
5	Clusteranalyse der kreisfreien Städte	74
5.1	Datenvorbereitung	74
5.1.1	Datenlücken	74
5.1.2	Z-Transformation der Indikatorwerte	74
5.2	Faktorenanalyse	76
5.2.1	Vorgehen der Faktorenanalyse	76
5.2.2	Ergebnisse der Faktorenanalyse	77
5.3	Clusteranalyse und Auswahl der Clusteranzahl	83
5.3.1	Durchführung des zweiten Clusteranalysedurchlaufs	84
5.3.2	Analyse und Auswertung der Clusterergebnisse der kreisfreien Städte	88
6	Zusammenfassung und Ausblick	98
	Quellenverzeichnis	100
	Anhang	108

Abbildungsverzeichnis

Abbildung 2-1:	Übersicht der Clusteranalyseverfahren aus der Literatur	11
Abbildung 2-2:	Übersicht der Grundvorstellungen der agglomerativen Clusteranalyse	12
Abbildung 3-1:	Single-Linkage (nächster Nachbar) – schematische Darstellung	37
Abbildung 3-2:	Dendrogramm der Fusionsschritte – Single-Linkage	37
Abbildung 3-3:	Complete-Linkage (entferntester Nachbar) – schematische Darstellung	39
Abbildung 3-4:	Dendrogramm der Fusionsschritte – Complete-Linkage	40
Abbildung 3-5:	Average-Linkage (zwischen den Gruppen) – schematische Darstellung	41
Abbildung 3-6:	Dendrogramm der Fusionsschritte – Average-Linkage	42
Abbildung 3-7:	Dendrogramm der Fusionsschritte – Ward Verknüpfung	44
Abbildung 5-1:	Schematische Struktur des gewählten Vorgehens	75
Abbildung 5-2:	Clusteranalyse – Schematische Clusterzusammenführung	84
Abbildung 5-3:	Clusteranalyse – Dendrogramm der Fusionsschritte (103 Städte)	87
Abbildung 5-4:	Clusteranalyse – Mittelwerte der 8 Cluster bezogen auf die 11 Faktoren ...	90

Tabellenverzeichnis

Tabelle 2-1:	Kriterien zur Systematisierung und Einteilung der Clusteranalyseverfahren	10
Tabelle 2-2:	Übersicht einiger kostenpflichtiger und –freier Statistikprogramme	19
Tabelle 2-3:	Themenfeldübergreifende Übersicht publizierter Clusteranalysen	20
Tabelle 2-4:	Verschiedene Städtegrößen und ihre Einwohnerzahlen	26
Tabelle 2-5:	Betrachtete Städte und deren Einordnung - Voruntersuchung	27
Tabelle 2-6:	Alle deutschen kreisfreien Städte und deren Einordnung	28
Tabelle 2-7:	Bundesländer nach Anzahl der kreisfreien Städte	30
Tabelle 2-8:	Gegenüberstellung der beiden Untersuchungen	31
Tabelle 2-9:	Endenergieverbrauch in Haushalten - 2012	32
Tabelle 3-1:	Für das Methodenscreening verwendete Indikatoren	35
Tabelle 3-2:	7-Clusterlösung mit Single-Linkage	38
Tabelle 3-3:	7-Clusterlösung mit Complete-Linkage	40
Tabelle 3-4:	7-Clusterlösung mit Average-Linkage	42
Tabelle 3-5:	Clusterlösung mit Ward	43
Tabelle 3-6:	Gegenüberstellung der Ergebnisse der untersuchten Fusionsalgorithmen	45
Tabelle 4-1:	Indikatorkenndaten – „Bevölkerungsentwicklung der letzten elf Jahre“	51
Tabelle 4-2:	Indikatorkenndaten – „Wohnfläche pro Person“	51
Tabelle 4-3:	Indikatorkenndaten – „Anteil Einpersonenhaushalte“	52
Tabelle 4-4:	Indikatorkenndaten – „Durchschnittliche Haushaltsgröße“	52
Tabelle 4-5:	Entwicklung der mittleren Haushaltsgröße	53
Tabelle 4-6:	Indikatorkenndaten – „Haushaltsdichte“	53
Tabelle 4-7:	Indikatorkenndaten – „Pro-Kopf Stromverbrauch“	54
Tabelle 4-8:	Indikatorkenndaten – „Erdgasverbrauch“	54
Tabelle 4-9:	Indikatorkenndaten – „Anteil Wohnungen in 1 und 2-Familienhäusern“	55
Tabelle 4-10:	Indikatorkenndaten – „Verfügbares Einkommen“	56
Tabelle 4-11:	Indikatorkenndaten – „Anteil über 65-Jähriger“	56
Tabelle 4-12:	Indikatorkenndaten – „Arbeitslosenquote“	57
Tabelle 4-13:	Indikatorkenndaten – „Anteil Siedlungs- und Verkehrsfl. an Gesamtfläche“	58
Tabelle 4-14:	Indikatorkenndaten – „Heiztage Mittel der Jahre 2009-2013“	58
Tabelle 4-15:	Indikatorkenndaten – „Heizgradtage Mittel der Jahre 1970-2013“	59
Tabelle 4-16:	Indikatorkenndaten – „Gradtagzahlen Mittel der Jahre 1970-2013“	59
Tabelle 4-17:	Indikatorkenndaten – „Motorisierungsgrad“	61
Tabelle 4-18:	Indikatorkenndaten – „Personenkraftwagen pro 1.000 Einwohner“	62
Tabelle 4-19:	Indikatorkenndaten – „Dieselfahrzeuge pro 1.000 Einwohner“	62
Tabelle 4-20:	Indikatorkenndaten – „Benzinfahrzeuge pro 1.000 Einwohner“	62
Tabelle 4-21:	Indikatorkenndaten – „Elektrofahrzeuge pro 1.000 Einwohner“	63
Tabelle 4-22:	Indikatorkenndaten – „Bevölkerungsdichte“	63
Tabelle 4-23:	Indikatorkenndaten – „Anteil 18 – 65 Jähriger“	64
Tabelle 4-24:	Indikatorkenndaten – „Anteil Auspendler an Beschäftigten am Wohnort“	64
Tabelle 4-25:	Indikatorkenndaten – „ÖPNV Fahrten pro Einwohner“	65
Tabelle 4-26:	Indikatorkenndaten – „Personenkilometer pro Einwohner“	66
Tabelle 4-27:	Indikatorkenndaten – „Mittlere Reiseweite“	67
Tabelle 4-28:	Indikatorkenndaten – „Beschäftigtenanteil im jeweiligen Sektor“	68
Tabelle 4-29:	Indikatorkenndaten – „Energieproduktivität im jeweiligen Sektor“	70
Tabelle 4-30:	Indikatorkenndaten – „Energieintensität im jeweiligen Sektor“	70
Tabelle 4-31:	Indikatorkenndaten – „Produktivitätsniveau im jeweiligen Sektor“	70

Tabelle 4-32:	Indikatorkenndaten – „Spezifischer Energieverbrauch im jeweiligen Sektor“	71
Tabelle 4-33:	Indikatorkenndaten – „Umsatzsteuer des jeweiligen Sektors an der Gesamtumsatzsteuer“	71
Tabelle 4-34:	Zuordnung der WZ08 zu den Sektoren (eigene Darstellung [DES2008])	72
Tabelle 4-35:	Indikatorkenndaten – „Entwicklung der Beschäftigtenanteile des jeweiligen Sektors“	72
Tabelle 4-36:	Indikatorkenndaten – „Veränderung Energieintensität 2007-2011 Industrie“	73
Tabelle 5-1:	Faktorenanalyse – Erklärte Varianz der untersuchten 41 Indikatoren	78
Tabelle 5-2:	Faktorenanalyse – Übersicht der Zuordnung der Indikatoren (Durchlauf 1)	79
Tabelle 5-3:	Faktorenanalyse – Übersicht der fünf Durchläufe	80
Tabelle 5-4:	Faktorenanalyse – Übersicht der Zuordnung der Indikatoren (Durchlauf 5)	81
Tabelle 5-5:	Clusteranalyse – Ein Stadt Cluster des ersten Durchlaufs (107 Städte)	85
Tabelle 5-6:	Clusteranalyse – Auswahl aus der Zuordnungsübersicht der Fusionsschritte (103 Städte)	85
Tabelle 5-7:	Clusteranalyse – Die acht ermittelten Cluster sowie die zugeordneten Städte	103 88
Tabelle 5-8:	Clusteranalyse – Clustermittelwerte der Faktoren	89
Tabelle 5-9:	Clusteranalyse – Clustermittelwerte der Ausgangsindikatoren und jeweiligen Faktorladungen	91
Tabelle 5-10:	Clusteranalyse – Städte des ersten Clusters	92
Tabelle 5-11:	Clusteranalyse – Städte des zweiten Clusters	93
Tabelle 5-12:	Clusteranalyse – Städte des dritten Clusters	93
Tabelle 5-13:	Clusteranalyse – Städte des vierten Clusters	94
Tabelle 5-14:	Clusteranalyse – Städte des fünften Clusters	95
Tabelle 5-15:	Clusteranalyse – Städte des sechsten Clusters	96
Tabelle 5-16:	Clusteranalyse – Städte des siebten Clusters	97
Tabelle 5-17:	Clusteranalyse – Städte des achten Clusters	97
Tabelle A-1:	Indikatoren des Verbrauchssektor private Haushalte	108
Tabelle A-2:	Indikatoren des Verbrauchssektors Verkehr	110
Tabelle A-3:	Indikatoren der Verbrauchssektoren Industrie und GHD	112

Verzeichnis der Abkürzungen und Formelzeichen

1.2Fam	Anteil Wohnungen in 1 und 2 Familienhäusern
A18.65J	Anteil der Bevölkerung zwischen 18 und 65 Jahren
A65	Anteil über 65-Jähriger
ARB	Arbeitslosenquote
ASV	Anteil der Siedlungs- und Verkehrsfläche an Gesamtfläche
BA	Beschäftigtenanteil
BeD	Bevölkerungsdichte
Bev	Bevölkerungsentwicklung
BHKW	Blockheizkraftwerk
CO ₂	Kohlenstoffdioxid
D-Kfz, B-Kfz, E-Kfz/TEW	Diesel-, Benzin-, Elektrofahrzeuge pro tausend Einwohner
DWD	Deutscher Wetterdienst
EBA	Entwicklung Beschäftigtenanteil
ED	Euklidische Distanz
EI	Energieintensität
Eink	verfügbares Einkommen der Haushalte
EP	Energieproduktivität
EW	Einwohner
GHD	Gewerbe, Handel und Dienstleistungen
GTZ	Gradtagzahlen
GTZ	Gradtagzahlen
HAT	Heiztage
HDI	Human Development Index
HGT	Heizgradtage
HGT	Heizgradtage
HHDi	Haushaltsdichte
HT	Heiztage
IU/GU – GHDU/GU	Industrie/GHD Umsatzsteuer an Gesamtumsatzsteuer
k	Clusteranzahl
KDD	Knowledge Discovery in Databases
Kfz/TEW	Motorisierungsgrad
Krfr.	Kreisfrei
KWK	Kraft-Wärme-Kopplung
MIV	Motorisierter Individualverkehr
MRW	Mittlere Reiseweite im ÖPNV
ÖPNV	Öffentlicher Personennahverkehr
ÖPNV/EW	Anzahl Fahrten im ÖPNV pro Einwohner und Jahr
P.km./EW	Anzahl gefahrener Personenkilometer im ÖPNV pro Person und Jahr
PKS	Pro-Kopf-Stromverbrauch
Pkw	Personenkraftwagen
Pkw/TEW	Pkw pro tausend Einwohner
PN	Produktivitätsniveau
QED	Quadrierte euklidische Distanz
R ²	Zweidimensionaler Raum
R ³	Dreidimensionaler Raum
SEV	Spezifischer Energieverbrauch
VDV	Verband Deutscher Verkehrsunternehmen e. V.
WF	Wohnfläche pro Person
WZ	Wirtschaftszweige

1 Einführung

In unserer Gesellschaft ist die Nutzung der Energie zum Grundbedürfnis geworden. Die Entwicklung als auch das Überleben der Gesellschaft, wie wir sie kennen, hängt unverkennbar damit zusammen. Aktuell lässt sich der wesentliche Teil der verwendeten Energie den fossilen Energieträgern zuordnen. Damit einher geht der Ausstoß von Treibhausgasen, allen voran des Kohlenstoffdioxids (CO_2), das maßgeblich am anthropogenen Treibhauseffekt beteiligt ist. In ökonomisch entwickelten Ländern, zu denen auch Deutschland zählt, ist der CO_2 -Ausstoß als auch der Anteil der in Städten lebenden Bevölkerung besonders hoch. In Deutschland liegt der CO_2 -Ausstoß im Jahr 2013 bei 842,8 Mio. t [BP2014] und ca. dreiviertel aller Menschen leben in Kommunen mit stadtähnlichen Strukturen [BPB2010]. Diese Beziehung verdeutlicht, dass ein großer Teil des CO_2 -Ausstoßes in Städten anfällt. Der Trend zur Verstädterung in den sich entwickelnden Ländern wird auch Zukunft anhalten. So wird in den nächsten Jahren ein viel größerer Anteil der Weltbevölkerung in Städten leben als es heute der Fall ist [BPB2010]. Das bedeutet auch, dass ein dementsprechend hoher Anteil der Energieverwendung auf das System Stadt als Energieverbraucher entfällt.

Um dem Absenken des Energieverbrauchs und des CO_2 -Ausstoßes zu begegnen werden viele verschiedene Maßnahmen zeitparallel angegangen. Es hat sich eine Gemengelage aus Erneuerbaren Energien, Energieeffizienz und in letzter Zeit immer verstärkter sogar mit Ansätzen zur Suffizienz in der Energieverwendung gebildet [KOP2014]. Alles wirkt bei der Energiewende mit, die als Begriff für eine Sammlung an Maßnahmen dient, um die Klimaschutzziele, die von der Bundesregierung im Jahr 2010 erarbeitet wurden, zu erreichen [BMU2014].

In naher Zukunft ist die Verflechtung von Statistik und Fragestellungen aus dem Bereich der Energie denkbar, denn bereits jetzt wird die Statistik in der Wirtschaft und der Wissenschaft immer wichtiger. Vorausschauende Prognosen und Analysen lassen sich mit deren Anwendung auf Daten erwirken. Eine Vielzahl von Programmiersprachen existiert bereits auf diesem Feld, die verschiedenste statistische Analysen ermöglichen. [VDI2015] Eine Analysemöglichkeit aus der Statistik stellt die in dieser Arbeit thematisierte Clusteranalyse dar.

Die Clusteranalyse als statistische, multivariante Analysemethode könnte auf dem Weg zu einem nachhaltigeren Umgang mit der Energie beitragen, wenn z. B. Hilfestellung bei der Übertragbarkeit von Maßnahmen auf Städte durch errechnete Cluster gegeben oder Potenziale betrachteter Maßnahmen überschlägig, basierend auf erhobenen statistischen Daten, für betrachtete Maßnahmen beziffert werden können.

Ziel dieser Arbeit ist es neben der Schaffung eines Überblicks der vorhandenen Anwendungsbereiche der Clusteranalyse, die theoretische Methodik der Clusteranalyse in groben Zügen zu erläutern. Darüber hinaus soll in einer Voruntersuchung ein Methodenscreening durchgeführt werden, um eine geeignete Clusteranalysemethode für die weiterführende Untersuchung

des Gesamtsystems deutscher kreisfreier Städte, bestehend aus der Betrachtung der vier Verbrauchssektoren Haushalte, Verkehr, Industrie und Gewerbe, Handel und Dienstleistungen (GHD), zu wählen. Ein ausführlich dokumentiertes Vorgehen ist in dieser Arbeit unumgänglich, um die Konsistenz der Daten und die Nachvollziehbarkeit des Vorgehens für weitere Arbeiten sicherzustellen. Abschließend sollen auf Basis der erhobenen Indikatoren energetisch sich ähnelnde und somit vergleichbare Städtegruppen benannt werden, die bei der eingangs erwähnten Hilfestellung zur Übertragbarkeit beitragen können.

2 Methodik, Arten der Clusteranalyse und Datengrundlagen

Daten stellen heute mehr denn je ein immer wichtiger werdendes Gut dar. Durch die Digitalisierung und die Nutzung von Computern wird auch in Zukunft der Umfang von Daten erheblich steigen. Mit dem Anstieg der Datenmenge geht oft die Übersichtlichkeit verloren. Zudem ermöglichen große Datenmengen unbekannte Erkenntnisse, die im immer größer werdenden Feld des *Data-Minings* durch die Anwendung mathematisch statistischer Methoden auf Datenmengen Muster und Zusammenhänge erkannt werden, die aufgrund des enormen Umfangs in der Vergangenheit oft nicht oder nur mit enormem Aufwand in Erfahrung gebracht werden konnten. In diesem Kapitel wird auf die Grundlagen zu Methodik sowie die Datengrundlagen der durchgeführten Arbeit eingegangen. Die theoretische Methodik der Clusteranalyse sowie deren Einordnung in das Feld der multivariaten Analysemethoden, die Probleme und die Anwendungsfelder werden beschrieben. Ebenfalls wird der Untersuchungsrahmen für die durchgeführten Analysen erläutert.

2.1 Methodik - Clusteranalyse

Die Basis der heutigen empirischen Forschung stellen in großem Maße die multivariaten Analysemethoden dar. Die multivariaten Analysemethoden umfassen eine Reihe an Methoden, die je nach Anwendungsfall oft kombiniert ihren Einsatz finden. Eine Auswahl ist nachfolgend aufgelistet [BAC2013]:

- Regressionsanalyse
- Varianzanalyse
- Diskriminanzanalyse
- Logistische Regression
- Conjoint-Analyse
- **Faktorenanalyse**
- **Clusteranalyse**

Multivariate Analysemethoden zeichnen sich von z. B. *Univariaten* Analysemethoden dadurch ab, dass gleich mehrere Variablen zeitgleich untersucht werden, die nicht selten sogar Zusammenhänge aufweisen. [UZH2015] Die Analysemethoden lassen sich zudem in zwei Verfahrensarten einteilen. Es wird unterschieden zwischen Strukturen-prüfenden Verfahren und Strukturen-entdeckenden Verfahren. Erstere zielen darauf ab zu überprüfen wie die Zusammenhänge der Variablen untereinander sind. Deren Beziehung und einige Zusammenhänge untereinander können bereits im Vorfeld bekannt sein. Strukturen-entdeckende Verfahren (explorative Verfahren) hingegen haben das Ziel, Zusammenhänge zwischen unterschiedlichen Variablen herauszustellen bzw. zu „entdecken“. Dies soll auch in der hier durchgeführten Untersuchung erfolgen. Die Faktoren- und die Clusteranalyse werden zu den explorativen Verfahren gezählt, da beide Strukturen erkennen. Die Faktorenanalyse extrahiert Faktoren aus

Variablen, wodurch die Datenzahl reduziert wird und die Clusteranalyse ermittelt Zusammenhänge zwischen Objekten, wodurch diese Gruppen zugeordnet werden können. [BAC2006] Die Zuordnung zu Gruppen ist bei der Clusteranalyse mit einer Reduktion der Objekte zu einer geringeren Anzahl an Clustern verbunden.

Das Ziel einer Clusteranalyse ist es eine Menge von zu untersuchenden bzw. vorhandenen Objekten (z. B. Städten) homogenen Gruppen (den Clustern) zuzuordnen [BAC2006]. Um Cluster möglichst genau zu beschreiben kann als Definition die Aussage herangezogen werden, dass die Objekte in einem Cluster zueinander möglichst ähnlich sein bzw. homogen die Cluster selbst zueinander jedoch möglichst keine Ähnlichkeit aufweisen bzw. heterogen sein sollen, wodurch eine eindeutige Clusterstruktur erzeugt werden kann [BAC2006], [BAC2010]. Die Multivarianz der Clusteranalyse wird besonders hier sehr deutlich, da für eine belastbare Clusteranalyse möglichst eine große Objektanzahl mit vielen Variablen zeitgleich betrachtet werden muss.

2.2 Wichtige Begriffe der Faktor- und Clusteranalyse

Dieses Kapitel umfasst die zum Verstehen der Clusteranalyse und des statistischen Vorgehens notwendigen Begriffe und Bezeichnungen. Fortan in dieser Arbeit werden diese Begriffe verwendet, um die Sachverhalte bei der Clusterung der betrachteten Städte zu beschreiben.

2.2.1 Cluster

Als Cluster wird eine Menge an Objekten bezeichnet, die aufgrund ihrer Homogenität zu einer Gruppe zusammengefasst werden. Gleichbedeutend werden in diesem Zusammenhang für den Begriff Cluster auch die Begriffe Gruppen, Klassen sowie Typen bezeichnet, wodurch die Clusterung auch als Gruppierung bzw. Typisierung bezeichnet wird. Schon die Brüder Grimm haben in ihrem *Deutschen Wörterbuch* das althochdeutsche Wort „*Kluster*“ aufgeführt, was schon damals zur Beschreibung von etwas „*was dicht und dick zusammensitzt*“ [Gri1961: Bd. 11, Sp. 1308 bis 1309]“ verwendet wurde [GRI1961], [GRI1999]. Somit weisen die einem Cluster angehörenden Objekte in ihren Ausprägungen (Variablen) Ähnlichkeiten auf, die dieses Cluster stellvertretend repräsentieren. [BAC2006], [BAC2010], [POP2015]

Die in einer mathematisch-statistischen Clusteranalyse ermittelten Cluster sind nicht immer gleichbedeutend mit der Verwendung des Begriffes Cluster und dürfen nicht verwechselt werden. Deshalb an dieser Stelle eine kurze Abgrenzung, die die Begrifflichkeit Cluster in die zwei Bereiche aufteilt. In der Clusteranalyse ermittelte Cluster sind statistisch sinnvolle Cluster, denen eine mathematisch-statistische Analyse vorweg geht. Der Begriff Cluster wird, besonders in letzter Zeit, häufig zur Bezeichnung von Zusammenschlüssen verwendet (z. B. Exzellenzcluster, Wissenschaftscluster, Wirtschaftcluster, etc.), die inhaltlich ähnliche wirtschaftliche,

unternehmerische oder wissenschaftliche Ausrichtungen in einem oft geografischen bzw. infrastrukturell kompakten Raum zusammenfassen [BMB2015], jedoch keine mathematisch-statistische Untersuchung im Vorfeld erfahren haben. Ziel ist hierbei Akteure aus bestimmten wirtschaftlichen Bereichen (Chemie, Ernährung, Maschinenbau, Medizin, Logistik, etc.) untereinander zu vernetzen um Synergieeffekte nutzen zu können. Derzeit gibt es in Nordrhein Westfalen 16 Wirtschaftscluster. [Exz2014] Ähnliche Exzellenzcluster gibt es auch im wissenschaftlichen Bereich, wo Universitäten und Forschungseinrichtungen zusammengeschlossen werden, um weltweite Spitzenpositionen in der Forschung zu erreichen [DFG2013]. [POP2015]

2.2.2 Objekt

Als Objekt wird in der Clusteranalyse das zu untersuchende Element bezeichnet [BAC2010]. In dem hier durchgeführten Analysefall sind es deutsche kreisfreie Städte. Ein Objekt ist ein einzelnes Element eines Clusters, das eigene Ausprägungen in Form von Variablenwerten aufweist. Objekte werden demnach mit Hilfe der Ausprägungen der Variablenwerte vergleichbar und somit „clusterbar“. [STE1977]

2.2.3 Variable

Als Variablen werden veränderbare Größen (Ausprägungen, Eigenschaften, Merkmale) von Objekten bezeichnet. Variablen bilden durch deren Größe die Merkmale der zu untersuchenden Objekte, wodurch diese clusteranalytisch nutzbar werden. Variablen sind für die Clusteranalyse essenziell, da diese neben der Beschreibung der zu untersuchenden Objekte deren mathematischen Vergleich ermöglichen. Jede Variable benötigt zunächst Daten, die aus verschiedenen Quellen stammen können. [STE1977], [POP2015] Variablen werden in der hier durchgeführten Analyse in zwei Gruppen geteilt, die Indikatoren und die Faktoren. Die beiden verwendeten Variablen sowie die damit zusammenhängenden Begriffe Rohdaten und Skalenniveau werden hier erläutert.

Rohdaten

Die Daten mit denen die zu klassifizierenden Objekte beschrieben werden, entstammen in der hier durchgeführten Analyse im Wesentlichen den frei zugänglichen Datenbanken und den in regelmäßigen Abständen veröffentlichten Erhebungen der statistischen Landes- und Bundesämter, Unternehmensverbänden, Zensus- und Unternehmensdaten sowie kommunalen Daten (s. Kapitel 4). Die dort vorliegenden Daten sind Rohdaten und daher heterogen in ihrer Struktur und Aufbereitung [BAC2006]. Sie sind oft in Listen dargestellt bzw. abrufbar und haben keine große Datenbearbeitung erfahren, da Rohdaten prinzipiell als reine Daten aus Messungen oder Erhebungen entstammen. Das ist auch der Grund, weshalb sich die Rohdaten als solche nicht für die direkte Nutzung in der hier durchgeführten Clusteranalyse eignen. Die Nutzung der Rohdaten setzt eine Datenvorbereitung sowie Datenbearbeitung voraus, um die reinen

Rohdaten in für eine Clusteranalyse nutzbare Kennzahlen (Indikatoren) umzuwandeln, die einen Vergleich erlauben und die Durchführung der clusteranalytischen Betrachtung deutscher Städte ermöglichen.

Skalenniveau

Das Skalenniveau beschreibt die Maßeinheit der verwendeten Daten. Vier wesentliche Skalen werden in der Statistik im Bereich der Clusteranalyse angewendet, die nachfolgend aufgelistet sind [SPR2016], [ABO2007]:

- Nominalskala
- Ordinalskala
- Intervallskala
- Verhältnisskala

Die *Nominalskala* dient einer simplen Klassifikation von Objekten [SPR2016]. Damit könnte beispielsweise bezeichnet werden, ob ein Objekt ein bestimmtes Merkmal vorweist oder nicht [SPR2016], [ABO2007] (z. B. gibt es eine Universität in einer Stadt; Ja=1, Nein=0). Die Auswertung die daran angeknüpft werden kann, beschränkt sich jedoch nur darauf die Objekte auf diese Häufigkeit zu überprüfen [SPR2016].

Die *Ordinalskala* erlaubt Objekte zu ordnen. Jedem Objekt können Größen zugeteilt werden, die es ermöglichen das Objekt nach gewissen Kriterien zu sortieren (das Bilden einer Reihenfolge) nicht aber die Dimension sowie die Unterschiede zu beziffern (z. B. Bewertungen auf einem Fragebogen; 1=sehr Gut, 6=sehr schlecht). Die Gründe hierbei sind jedoch nicht quantifizierbar. [SPR2016], [ABO2007], [POP2015]

Die *Intervallskala* ist die für diese Arbeit wichtigste Skala. Die Werte aller erarbeiteten Indikatoren (siehe Kapitel 4) sind intervallskaliert, wodurch deren Größen qualifizierte Rückschlüsse auf quantifizierbare Sachverhalte erlauben (z. B. Unterschiede im Energieverbrauch). Ein natürlicher Nullpunkt, an dem die Variablenwerte bemessen werden können, ist bei der Intervallskala nicht existent. [SPR2016], [ABO2007], [POP2015]

Die *Verhältnisskala* hat zu der Intervallskala einen festgelegten Nullpunkt (z. B. *Höchstgeschwindigkeit bei einem Fahrzeug* [SPR2016]). [SPR2016], [ABO2007], [POP2015] Die beiden letzten Skalen werden auch als metrische Skalen bezeichnet, da zwischen zwei Werten immer eine beliebig große Anzahl an Zwischenwerten vorhanden ist. Für den letzten Schritt der Clusteranalyse sind die Faktorwerte auf einer Verhältnisskala dargestellt, um auf einen Mittelwert (Nullpunkt) bezogene Daten vergleichen zu können.

Die Clusteranalyse erfordert kein besonderes Skalenniveau, kann Variablen mit verschiedenen Skalenniveaus verarbeiten und liefert Ergebnisse. Es sollte jedoch vermieden werden

metrische Skalen (Intervall- und Verhältnisskala) mit Nominal- und Ordinalskalen in einer Analyse zu kombinieren, da dadurch Probleme mit der Vergleichbarkeit der Variablen entstehen und die Ergebnisse nur schwer auszuwerten sind. [BAC2006]

Indikator

Die verschiedenen Formen von Indikatoren werden in Abschnitt 4.1.1 detailliert auf den hier angewendeten Analysefall diskutiert. An dieser Stelle wird der Indikator als Basis-Variable erläutert. Auf den jeweiligen Fall bearbeitete Rohdaten, z. B. das Bilden von Verhältniszahlen, um eine Vergleichbarkeit zwischen den betrachteten Objekten zu generieren, werden als Indikatoren bezeichnet. Indikatoren basieren, anders als Faktoren, auf direkt messbaren Daten und können in der hier durchgeführten Untersuchung als „natürliche“ Variablen bezeichnet werden.

Faktor

Faktoren werden als „künstliche“ Variablen bezeichnet. Diese Bezeichnung begründet sich dadurch, dass die Faktorwerte, anders als bei den Indikatoren, nicht gemessen oder direkt erhoben werden können. Faktoren werden durch die Faktorenanalyse ermittelt und vereinen Indikatoren, die statistisch sinnvoll und hinreichend genau von einem Faktor abgebildet werden können. Die Faktoren müssen dabei die Aussage der zusammengefassten Indikatoren enthalten, wodurch die Indikatoren auch im Nachhinein abgeleitet und nachvollzogen werden können [WAL2013].

Die extrahierten Faktoren korrelieren nicht miteinander, wodurch sich deren Aussagen gegenseitig nicht beeinflussen. Dadurch sind die Faktoren unabhängig voneinander, was bedeutet, dass kein kausaler und inhaltlicher Zusammenhang zwischen den Faktoren besteht. [HÖP1997] [POP2015]

2.2.4 Ähnlichkeitsmaß, Distanzmaß

Durch Variablen beschriebene Objekte bilden die Basis einer Clusteranalyse. Um die zu klassifizierenden Objekte miteinander in Bezug setzen zu können, bedarf es einer Möglichkeit die multivariat abgebildeten Objekte zu vergleichen. Bei der Clusteranalyse basiert der Vergleich auf der Betrachtung der Ähnlichkeit der zu klassifizierenden Objekte, die durch die mathematische Distanz dieser zueinander ausgedrückt wird. [BAC2006], [BAC2010], [BUE2012], [SCH2010], [STE2011] Durch die mathematische Distanz sind die Ähnlichkeiten bzw. Unähnlichkeiten quantifizierbar [BAC2006].

Erfolgt die Quantifizierung der Distanz bzgl. des Ähnlichkeitsmaßes (maximale Ähnlichkeit) bzw. des Distanzmaßes (minimale Unähnlichkeit), so sind die beiden Objekte am ähnlichsten, deren Ähnlichkeitswert am größten ist. Die Distanz der Objekte ist bei einer Unähnlichkeit am größten. [SCH2010: 27], [POP2015] Verschiedene Formeln zur Berechnung des Ähnlichkeits-

und Distanzmaße werden in der Statistik verwendet, wobei sich jedoch nicht alle Maße in beliebiger Kombination mit den möglichen Clusteranalyseverfahren eignen, da dadurch die Belastbarkeit der Clusterergebnisse beeinflusst und die Ergebnisse geschönt dargestellt werden können [SCH2010: 27].

Die simple Betrachtung einer einfachen Distanz ist Teil der linearen Algebra und wird durch die Berechnung der Entfernung zwischen zwei Punkten in einem definierten Raum bestimmt. Für die Clusteranalyse hingegen eignet sich die simple Bestimmung der Entfernung nicht, da die zu bestimmenden Distanzen der Objekte von vielen Variablen abhängen (Betrachtungsraum $> R^3$) und die zudem je nach Anwendungsfall und Skalenniveau differenziert betrachtet werden müssen. Um dieser Problematik zu begegnen wurden diverse Formeln zur Berechnung der Distanzmaße entwickelt, deren Einteilung im Wesentlichen nach der Art der verwendeten verarbeiteten Skalenniveaus der zu betrachtenden Objekte und Variablen erfolgt. Eine Clusteranalyse ist eine Methode, die kein bestimmtes Skalenniveau benötigt [BAC2006]. Darüber hinaus können in derselben Untersuchung zeitgleich Variablen unterschiedlicher Skalen betrachtet werden [BAC2006]. Das ist zwar möglich und liefert Ergebnisse, jedoch werden die Ergebnisse schwer zu interpretieren sein und die berechneten Distanzen können zudem Probleme generieren, die durch die gemischten Skalen der Variablen auftreten [BAC2006]. Die Formeln der Distanzbildung sind auf bestimmte Skalen ausgelegt, die diese gut abbilden. Gemischte Variablen bilden dabei Sonderfälle (verstärkt in der Markt- und Meinungsforschung), wo Anpassungen der Formeln zur Berechnung der Distanzmaße vorgenommen oder die Variablen zunächst bearbeitet werden müssen, um valide Ergebnisse zu bilden. Für diese Untersuchung sind die Sonderfälle jedoch nicht von Bedeutung, da die Variablen zur Beschreibung der Städte ausschließlich intervallskaliert (metrisch) sind. Aus den vielen Arten, die es erlauben das Proximitätsmaß (Distanz; Ähnlichkeit) zu quantifizieren wird aufgrund des Anwendungsfalls nur der für diese Untersuchung wichtige Ausschnitt (Ähnlichkeits- und Distanzmaße für intervallskalierte Variablen) dargestellt. Nachfolgend sind die wichtigsten Maße für die fallorientierte Clusterung aufgelistet [SCH2010]:

- **Euklidische Distanz**
- **Quadrierte euklidische Distanz**
- Kosinus
- Pearson-Korrelation
- Tschebyscheff
- Block
- Minkowski

Die beiden Fett gekennzeichneten Distanzmaße, sind die gängigen Distanzmaße für die Clusterzentrenanalysen, die sich für den vorliegenden Untersuchungsfall eignen. Es wird nur auf diese beiden Distanzen eingegangen. Die euklidische Distanz berechnet

„die Quadratwurzel der Summe der quadrierten Differenzen zwischen den Werten der Objekte [SCH2010: 28]“.

Die quadrierte euklidische Distanz hingegen berechnet

„die Summe der quadrierten Differenzen zwischen den Werten der Objekte [SCH2010: 28]“.

Im nachfolgenden Unterkapitel 2.3 werden die relevanten Clusteranalyseverfahren (Fusionsalgorithmen) diskutiert sowie die Clustermethoden erläutert und die Differenzen herausgestellt. Darüber hinaus erfolgt eine Zuordnung der Distanzmaße zu den einzelnen Methoden der hierarchisch-agglomerativen Clusteranalyse.

2.2.5 Fusionsalgorithmus - Fusionsmethode

Die Regeln nach welchen die statistisch zu betrachtenden Objekte fusioniert (vereint) werden beschreibt der Fusionsalgorithmus. Es gibt eine Vielzahl von Fusionsalgorithmen, die sich für verschiedene Clusteranalyseverfahren, die unterschiedlichen zu analysierenden Daten sowie dem Ziel einer Clusteranalyse unterscheiden und zur Anwendung kommen.

2.3 Clusteranalyseverfahren und -methoden

Unabhängig von der Art einer Clusteranalyse hat jedes der dort angewendeten Verfahren das Ziel die dem Untersuchungsrahmen angehörende Menge an zu klassifizierenden Objekten statistisch sinnvollen Gruppen zuzuordnen. Sehr wohl besteht jedoch ein Unterschied in den Methoden und Regeln nach denen die Clusterung erfolgt sowie wann diese angewendet werden kann und soll. Alle Clusteranalysen haben darüber hinaus eine weitere Gemeinsamkeit. Diese ist die multivariate, zeitgleiche Untersuchung aller Variablen des Untersuchungsdatensatzes. [BAC2010] [BAC2016] [SCH2010]

Aus einer über die Jahre erarbeiteten unüberschaubaren Anzahl an diversen Clusteranalyseverfahren, die für unterschiedliche Anwendungen vorgeschlagen werden, liegt hier der Fokus auf den großen und in Ihrer Anwendung vielfältigen, allgemeinen Klassifizierungsverfahren, die vorgestellt und in Ihren Merkmalen erläutert werden [BAC2010].

2.3.1 Einteilung der Clusteranalyseverfahren

Die Clusteranalyse als Oberbegriff für eine Vielzahl verschiedener Clusteranalyseverfahren fasst diese nur aufgrund deren Ziels zusammen. Die Verfahren werden hingegen auf verschiedene Weisen eingeteilt. So werden in Steinhausen et al. [STE1977: 69ff] die Clusteranalyseverfahren in drei Blöcke (Systeme) geteilt. Diese sind in nachfolgender Tabelle nach [STE1977: 69] aufgetragen. Dabei sind die Clusteranalysen in den einzelnen Gruppen (jeweils zwei Zeilen) kombinierbar. So kann beispielsweise eine hierarchische Clusteranalyse (Schritt für Schritt) disjunkt (nicht überlappend) realisiert werden, jedoch nicht erschöpfend sein. Das

„nicht erschöpfend“ bezieht sich in diesem Zusammenhang darauf, dass nicht alle zu klassifizierenden Objekte auch tatsächlich einem Cluster zugeordnet werden und somit nicht vereinte Objekte verbleiben. Die Einteilung nach dem jeweiligen Gruppierungsprozess erfolgt iterativ bzw. nicht iterativ. Dabei können in den einzelnen Schritten die Objekte durch schrittweises Zusammenfügen zu sich vergrößernden Clustern fusioniert bzw. ein zu Beginn bestehendes Großcluster zu kleineren Clustern reduziert wird. Die dritte Untergruppe wird in sequentiell und global unterteilt. Dabei ist die Differenzierung nur auf die Berechnung der Ähnlichkeits- und Distanzmaße bezogen. Sequentielle Gruppierungen betrachten die Objekte paarweise. Globale Verfahren prüfen die Ähnlichkeiten bzw. Distanzen aller Objekte. Ebenfalls sind andere Kombinationen denkbar und möglich. [STE1977]

Tabelle 2-1: Kriterien zur Systematisierung und Einteilung der Clusteranalyseverfahren

Gruppierungsergebnis	hierarchisch	disjunkt	erschöpfend
	nicht-hierarchisch	überlappend	nicht erschöpfend
Gruppierungsprozess	iterativ	divisiv	---
	nicht iterativ	agglomerativ	---
Gruppierungskriterium	sequentiell	---	---
	global	---	---

Zudem werden weitere Einteilungen angeführt. Diese sind in [STE1977] und [BAC2016] erwähnt, jedoch in beiden Fällen als nicht zielführend für eine multivariate Clusteranalyse deklariert. Sie sind die polythetischen und monothetischen Verfahren, von denen ausschließlich die polythetischen Verfahren eine vollständige Clusteranalyse großer Datenumfänge erlaubt, da bei der monothetischen Clusteranalyse lediglich eine Variable der Objekte betrachtet wird, was dem multivariaten Anspruch nicht gerecht wird. [STE1977] [BAC2016]

Ferner werden Clusteranalyseverfahren nach *Bacher et al.* [BAC2010: 18] in drei große Gruppen unterteilt auf die im weiteren Verlauf (Abschnitt 2.3.2) näher eingegangen wird.

- Unvollständige Clusteranalyseverfahren
- Deterministische Clusteranalyseverfahren
- Probabilistische Clusteranalyseverfahren

Diese drei Gruppen nach [BAC2010] weisen Überschneidungen mit der Einteilung der Clusteranalyseverfahren anderer Darstellungen auf. So weist zu den probabilistischen Clusteranalyseverfahren auch eine gesonderte Gruppe (possibilistische Verfahren, s. Abschnitt 2.3.2) Ähnlichkeiten auf. Hier werden ebenfalls überlappende Cluster erzeugt, jedoch erfolgt die Unterscheidung dadurch, dass hierbei zur Zuordnung keine Wahrscheinlichkeiten sondern Möglichkeiten der Zugehörigkeiten einbezogen werden.

Eine weitere sehr umfangreiche Einteilung der Verfahren wird von Backhaus et al. [BAC2016] vorgenommen. Diese ist anders strukturiert, schließt im Wesentlichen dieselben großen Clusteranalyseverfahren ein. Diese sind jedoch anderen ‘Verfahrensgruppen’ zugeordnet, die nachfolgen aufgelistet sind:

- Grafentheoretische Clusteranalyseverfahren
- Hierarchische Clusteranalyseverfahren (s. Abschnitt 2.3.2)
- Partitionierende Clusteranalyseverfahren (s. Abschnitt 2.3.2)
- Optimierungsclusteranalyseverfahren

Die grafentheoretischen Clusteranalyseverfahren sowie die Optimierungsverfahren stellen Sonderfälle der Clusteranalyse dar, deren Anwendungsbereich dementsprechend stark limitiert ist. Deshalb wird auf die beiden Verfahrensgruppen nicht näher eingegangen. Der Vollständigkeit halber werden die beiden Gruppen hier aufgeführt sowie in der Abbildung 2-1 abgebildet, um dennoch eine Einordnung in die Übersicht der Clusteranalyseverfahren zu ermöglichen. Die hierarchischen sowie die partitionierenden Verfahren werden im weiteren Verlauf detaillierter thematisiert. Ersteren kommt in der Anwendung der Clusteranalyse ein Großteil an Bedeutung zu. [BAC2016]

Die nachfolgende Grafik (Abbildung 2-1) vereint die in der Literatur am häufigsten auftretenden Clusteranalyseverfahren. Sie fasst die Einteilungen der Verfahren nach [BAC2016], [STE1977], [BAC2010], [HÖP1997] und weiterer Information aus [SCH2010], [EVE2011] zusammen und gibt einen umfassenden Überblick. In der Grafik sind zeitgleich auch die Gruppierungsprozesse aus [STE1977] für die hierarchische Clusteranalyse zu erkennen, die auch für die Clusteranalyse deutscher Städte als relevant gelten.

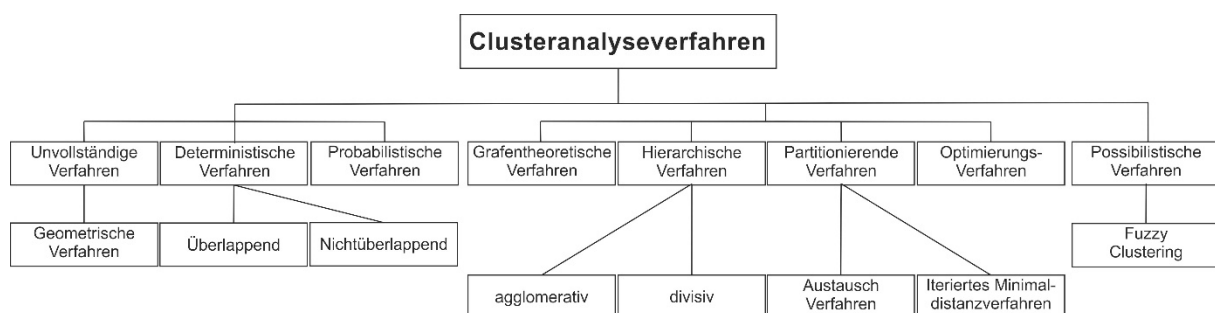


Abbildung 2-1: Übersicht der Clusteranalyseverfahren aus der Literatur¹

Der Fokus aus der Gesamtschau der Clusteranalyseverfahren wird für diese Analyse auf die hierarchischen Verfahren gelegt, da diese in der Anwendung die gängigste Verfahrensgruppe darstellen. Eine genauere Unterteilung der hierarchischen Verfahren ist in der folgenden Abbildung 2-2 dargestellt, die eingangs genannten Clusteranalyseverfahren wurden dazu der Übersichtlichkeit halber ausgeblendet, um so das Augenmerk auf die für diese Arbeit relevante

¹ Darstellung in Anlehnung an [BAC2010], [BAC2016], [STE1977], [SCH2010], [HÖP1997], [EVE2011]

Verfahrensvariante zu legen. Dort liegt der Fokus auf den agglomerativen Verfahren. Anders als bei den divisiven Verfahren werden bei den agglomerativen Verfahren zu Beginn einer Untersuchung alle Objekte separat betrachtet. Bei der divisiven Herangehensweise bilden zu Beginn alle Objekte eine Gruppe, die im Verlauf der Analyse in kleinere Gruppen aufgebrochen wird. [BAC2016]

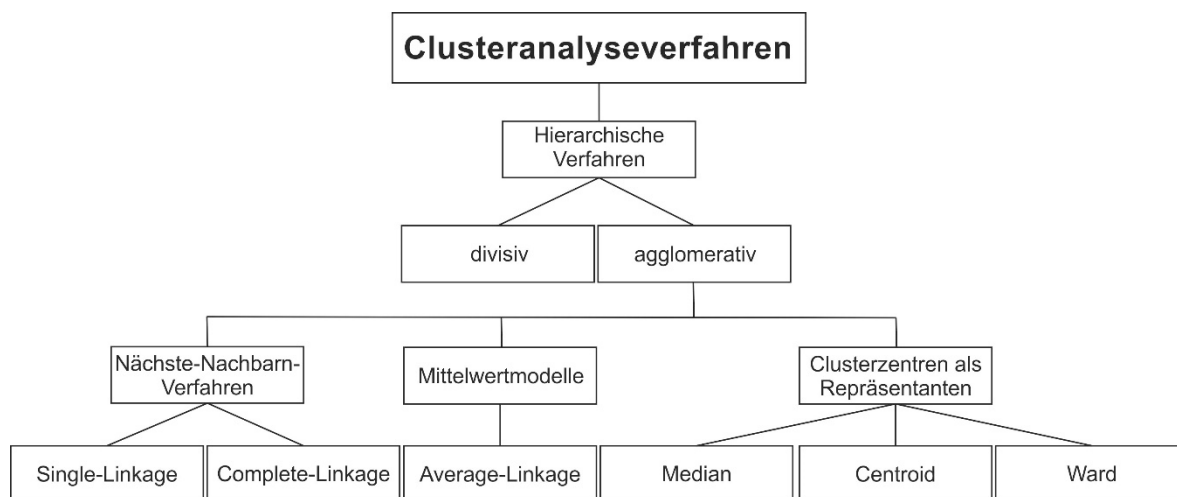


Abbildung 2-2: Übersicht der Grundvorstellungen der agglomerativen Clusteranalyse²

Die Vorgehensweise im Fusionierungsprozess kann als Clusteranalysemethode bezeichnet werden, da hier die Regeln festgelegt werden nach denen die Objekte den Gruppen zugeteilt werden. Bei den hierarchisch agglomerativen Verfahren werden sechs wesentliche Methoden (Fusionsalgorithmen) diskutiert und verstärkt verwendet. Die Abbildung 2-2 stellt eine Übersicht dar. Die Fusionsalgorithmen der agglomerativen Verfahren werden hier in drei Gruppen eingeteilt. Diese Gruppen werden in dem Abschnitt 2.3.2 detaillierter diskutiert. In den nachfolgenden Abschnitten werden die im Verlauf dieses Unterkapitels sowie in Abbildung 2-1 dargestellten Clusteranalyseverfahren separat beschrieben.

2.3.2 Erläuterung der Clusteranalyseverfahren

Unvollständige Clusteranalyseverfahren (geometrisch)

Als unvollständige Clusteranalyseverfahren werden häufig auch geometrische Clusteranalysen bezeichnet. Hier erfolgt das Ermitteln der Gruppen durch die Betrachtung der Punktwolken in einem zwei- bzw. dreidimensionalen Raum (R^2 oder R^3). Unvollständig sind diese Clusteranalysen, weil der Betrachter selbstständig die Punkte (Objekte) Gruppen zuordnen muss. [HÖP1997], [BAC2010] Geometrische Clusteranalysen können hingegen zur Unterstützung der vollständigen Clusteranalysen herangezogen werden, um die Wahl einer geeigneten Clusteranzahl zu validieren (s. [WAL2014]). Diese kommen nur zur Anwendung sofern die Anzahl der

² Die Darstellung der Grundvorstellungen erfolgt nach [BAC2016] und [BAC2010]

betrachteten Variablen grafisch darstellbar ist. Ein Problem bei der ausschließlichen Clusterung mithilfe der geometrischen Darstellung ist, dass durch das augenscheinliche Betrachten Zusammenhänge gefunden werden können, wo statistisch keine Zusammenhänge bestehen. Deshalb sollte die geometrische Clusteranalyse nur unterstützend und nicht als eigenständige Clusteranalyse verwendet werden. [EVE2011] [HÖP1997] [BAC2010]

Deterministische Clusteranalyseverfahren

Deterministische Clusteranalyseverfahren hingegen umfassen eine Vielzahl von Verfahren, die alle gemeinsam haben, dass die Objekte, die einem Cluster zugewiesen werden mit der Wahrscheinlichkeit von eins oder null einem oder mehreren Clustern zugeordnet werden. [BAC2010] Werden in den darin enthaltenen Verfahren einzelne Objekte ausschließlich einem Cluster zugeordnet, so werden diese Verfahren als disjunkt bezeichnet. Je größer der Anteil der Objekte ist, die nicht ausschließlich einem Cluster angehören, desto stärker entfernt sich das Ergebnis der Clusteranalyse von der Clustergrundidee, die besagt, dass die ermittelten Cluster zueinander heterogen sein sollen [BAC2010]. Wenn ein großer Teil der betrachteten Objekte Element verschiedener Cluster ist, so ist wird auch die eindeutige Auswertung der Cluster immer schwieriger [BAC2010]. Für die Clusteranalyse deutscher Städte eignen sich somit nur disjunkte Clusterfusionsalgorithmen, da diese die betrachteten Objekte ausschließlich einem einzigen Cluster zugeordnet werden.

Viele Fusionsalgorithmen, die von [BAC2010] den Fusionsalgorithmen der deterministischen Clusteranalyseverfahren zugewiesen werden, weisen Überschneidungen mit den Algorithmen der hierarchischen Clusteranalyse durch [BAC2016] auf. Auf eine doppelte Beschreibung der Fusionsalgorithmen der deterministischen Clusteranalyseverfahren wird verzichtet, da die Hauptalgorithmen auch in den hierarchischen Clusteranalyseverfahren ihre Anwendung finden (s. Abbildung 2-2). Diese werden folgend ausführlich thematisiert. Es gibt weitere Clusterverfahren, die zu dieser Verfahrensgruppe gezählt werden (z. B. Two-Step Clusteranalyse [SCH2010]) auf die hier jedoch nicht eingegangen wird.

Probabilistische Clusteranalyseverfahren

Probabilistische Clusteranalyseverfahren bieten die Möglichkeit Objekte mit einer Wahrscheinlichkeit zwischen null und eins einem Cluster zuzuordnen, sodass diese Cluster dann nicht überlappungsfrei (nicht disjunkt) sind. Ein Objekt kann und ist somit zeitgleich ein Element verschiedener Cluster, wobei die Summe der Zuordnungswahrscheinlichkeiten eines Objektes immer eins beträgt [BAC2010: 351]. Diese Art der Clusteranalyse bietet die Möglichkeit sehr große Datenmengen auszuwerten, bei denen die erwarteten Gruppen nicht eindeutig beschrieben sein müssen [BAC2010: 354]. Die Objekte können auch durch verschiedene Variablen skalen beschrieben sein, ohne dass die Qualität der Clusterlösung verzerrt wird [BAC2010]. Die Untergruppe, die auch als eigenständige Gruppe possibilistischer Verfahren beschrieben

wird, ist von den Grundsätzen ähnlich. Es müssen dort jedoch die Zuordnungswahrscheinlichkeiten durch Zugehörigkeiten dargestellt werden [HÖP1997].

Hierarchische Clusteranalyseverfahren

Die in der Anwendung wohl am weitesten verbreitete Verfahrensgruppe bilden die hierarchischen Clusteranalyseverfahren. In der Definition nach [BAC2016] werden die hierarchischen Clusteranalyseverfahren als Gruppe bezeichnet, wobei in [BAC2010] diese Verfahrensgruppe zu den deterministischen Verfahren gezählt wird. Unabhängig von der Zuordnung dieser Verfahrensgruppe gibt es bei den hierarchischen Verfahren zwei übliche Herangehensweisen, die agglomerative und die divisive. Während bei der agglomerativen Herangehensweise zu Beginn jedes der betrachteten Objekte einem einzelnen Cluster zugeordnet wird und erst im Verlauf der Analyse die Gruppen anwachsen „agglomerieren“, befinden sich bei der divisiven Analyse zu Beginn sämtliche zu betrachtende Objekte in einem Cluster, dass dann schrittweise in kleinere Gruppen aufgeteilt wird, bis jedes Objekt ein eigenes Cluster beschreibt [STE1977], [BAC2010], [BAC2016]. Unabhängig von dem Vorgehen ist, dass die Analyse schrittweise erfolgt. Das bedeutet, dass ein einem Cluster zugeordnetes Objekt Element dieses Clusters ist, bleibt und somit auch im weiteren Verlauf das Cluster nicht verlässt. In der Praxis finden die agglomerativen hierarchischen Verfahren ihre Anwendung am häufigsten [BAC2016], [BAC2010], [SCH2010]. Das zeigt sich auch daran, dass die sehr verbreitete Statistiksoftware SPSS von IBM für die hierarchischen Verfahren lediglich die agglomerativen beinhaltet [SCH2010], [BUE2012]. Aus diesem Grund werden die Wesentlichen Fusionsalgorithmen mit denen die hierarchische agglomerative Clusteranalyse durchgeführt werden kann nachfolgend erläutert. Die Abbildung 2-2 zeigt schematisch die genauere Aufschlüsselung der hierarchischen agglomerativen Verfahren. Im Folgenden werden die in der agglomerativen Clusteranalyse gängigen Fusionsalgorithmen vorgestellt. Eine vertiefte Darstellung sowie die tiefgehende Erläuterung stellen die Arbeiten von *Bacher et al.* [BAC2010], *Backhaus et al.* [BAC2016], *Steinhausen et al.* [STE1977] sowie *Schendera* [SCH2010] dar.

Single-Linkage

Die Single-Linkage (Nächstgelegener Nachbar) Methode bestimmt die Cluster über die Berechnung des geringsten Abstandes zwischen zwei beobachteten Objekten [Sch2010: 25]. Diese Methode bildet zusammen mit der Complete-Linkage Methode eine Basismethode in der Clusterentwicklung, auf der die weiteren Nächste-Nachbarn-Verfahren aufbauen. Die Variation besteht darin, dass versucht wird die Schwächen der Methoden zu eliminieren bzw. abzuschwächen. Die Schwäche des Single-Linkage ist, dass jedes Objekt mindestens einen nächsten Nachbarn haben muss und die Entfernung nicht das primäre Kriterium ist, wodurch entstehende Cluster dazu neigen Ketten zu bilden (*Kontraktionseffekt* [BAC2010], [STE1977]). Werden zwei Objekte zu einem Cluster zusammengefügt, so werden diese im weiteren Ver-

lauf, anders als bei Clusterzentrenanalysen, nicht als einzelnes Objekt betrachtet. Jedes Objekt des gebildeten Clusters wird auch weiterhin eigenständig bei der Berechnung der Distanzen berücksichtigt [BAC2016]. Das Single-Linkage bildet wie alle hier vorgestellten Methoden (Algorithmen) disjunkte Cluster. [BAC2010] Dieser Algorithmus ist einer der ersten und einer der einfachsten und eignet sich besonders dafür strukturelle Zusammenhänge zu entdecken als Cluster mit hoher interner Homogenität [STE1977].

Complete-Linkage

Das Complete-Linkage ist eine Weiterführung der Single-Linkage Methode. Hier definiert der Algorithmus, dass alle Objekte eines Clusters nächste Nachbarn zueinander sind. Das bedeutet, dass alle Objekte, die einem Cluster angehören, sich dadurch abgrenzen, dass sie zu den Objekten im Cluster eine geringere Distanz aufweisen, als zu Objekten außerhalb des Clusters. Wird ein Objekt mit einem neuen Objekt vereint, da beide zueinander nächste Nachbarn sind, so muss das neue Objekt auch zu den bereits im Cluster enthaltenen Objekten nächster Nachbar sein. [BAC2010], [STE1977] Bei dem Complete-Linkage wird der nächste Nachbar, anders als bei dem Single-Linkage, durch die Berechnung der Maximalabstände zwischen den Objekten berechnet. Die geringste Maximaldistanz zeigt den nächsten Nachbarn an. [SCH2010] Auch hier werden die einzelnen Objekte betrachtet und nicht gebildete Clusterzentren. Diese Regel ist konservativer als die des Single Linkage, hat jedoch auch zur Folge, dass dieser Algorithmus dazu tendiert viele zu kleine Gruppen zu bilden (*Dilatationseffekt* [BAC2010], [STE1977]), da nur auf diese Weise diese Prämisse erfüllt werden kann. [BAC2010] Dieser Algorithmus bildet auch überlappende Cluster [SCH2010], weshalb sich dieser Algorithmus nicht für die Clusterung deutscher Städte eignet, da überlappungsfreie Cluster für eine eindeutige Einteilung, Beschreibung und Interpretation der Cluster notwendig sind.

Average-Linkage

Das Average-Linkage ist der Hauptvertreter der Mittelwertmodelle. Mittelwertmodelle werden als Modifikationen der Nächste-Nachbarn Verfahren bezeichnet, da diese deren die Schwächen (Dilatations- und Kontraktionseffekt) eliminieren sollen. Dabei werden nicht die einzelnen Objekte eines Clusters zur Identifikation der als nächstes zu fusionierenden Objekte herangezogen, sondern der Mittelwert der Abstände zwischen den Objekten des bereits bestehenden Clusters. Dieser Algorithmus liefert überlappungsfreie Cluster. [BAC2010] [SCH2010] [STE1977] Es gibt noch weitere Ausführungen des Average-Linkage (Weighted-Average-Linkage, Within-Average-Linkage [BAC2010]) auf die an dieser Stelle jedoch nicht eingegangen wird, da diese eher Sonderfälle darstellen.

Median und Centroid

Der Median der Centroid sowie der Ward Fusionsalgorithmus bilden die dritte Gruppe der agglomerativen Fusionsalgorithmen. Hier wird anders als bei den vorherigen Algorithmen ein Repräsentant für das jeweilige Cluster ermittelt. Dieser Repräsentant soll dabei das Objekt

sein, das die meisten nächsten Nachbarn aufweist. Bei dem Median und dem Centroid Verfahren geschieht dies vereinfachend ausgedrückt durch das Bilden der Clusterzentren, basierend auf den kürzesten Abständen zwischen den Clustern (Median) und zwischen den Centroiden, die aus der mittleren Distanz zwischen den Elementen der Objekte eines Clusters bestehen (Centroid). [STE1977] [SCH2010] [BAC2010] Für diesen Algorithmus wird die quadrierte euklidische Distanz als Distanzmaß empfohlen. Diese beiden Algorithmen führen auch zu Clustern, wenn andere Distanzmaße verwendet werden. Sie lassen sich jedoch schwer interpretieren und müssen angepasst werden. [STE1977] Aus diesem Grund werden die beiden Fusionsalgorithmen für die clusteranalytische Betrachtung von Städten ausgeschlossen.

Ward

Der Ward-Algorithmus bildet zusammen mit den beiden zuvor erklärten Algorithmen ein Repräsentantenverfahren. Bei diesem Verfahren werden die Objekte vereint, die ein vorgegebenes Maß der Heterogenität in einem möglichst geringen Umfang erhöhen [SCH2010]. In Backhaus et al. [BAC2016: 485] wird das Zielkriterium für das Zusammenfassen der Objekte (Gruppen) definiert als -

„Vereinige diejenigen Objekte, die die Fehlerquadratsumme am wenigsten erhöhen“.

Ab besten lässt sich dieser Ausdruck verstehen, wenn die Fehlerquadratsumme als Streuung im Cluster betrachtet wird. Somit werden die Objekte vereint, die die Intracusterstreuung am geringsten erhöhen. Das Heterogenitätsmaß wird als Varianzkriterium bezeichnet und wird nach folgender Gleichung 2-1 berechnet [BAC2016: 485].

$$V_g = \sum_{k=1}^{K_g} \sum_{j=1}^J (x_{kjg} - \bar{x}_{jg})^2 \quad \text{Gleichung 2-1}$$

Für die Berechnung des Varianzkriteriums werden die jeweiligen Werte x der untersuchten Variablen j ($j=1, \dots, J$) für die einzelnen betrachteten Objekte (Gruppen) k ($k=1, \dots, K_g$ in der dazugehörenden Gruppe) herangezogen. Diese Werte stellen in Gleichung 2-1 den Ausdruck x_{kjg} dar. Der Ausdruck $\bar{x}_{jg}$ ist hingegen der Mittelwert der Variablenwerte der jeweiligen betrachteten Gruppe [BAC2016: 485], [WAL2013]. Dieser Mittelwert bildet zeitgleich den Repräsentanten des jeweiligen Clusters. Es erfolgt für jedes Objekt diese Untersuchung. Zeitgleich erhöht sich die Größe der Gruppe über die die Variablenwerte gemittelt werden, da auch in diesem Verfahren hierarchisch vorgegangen wird und die Gruppengröße mit dem Fortschritt der Untersuchung anwächst. Die so ermittelten Gruppen sind intern sehr homogen und durch die berechnete Varianz, die in jedem einzelnen Cluster möglichst gering ist, können diese Gruppen voneinander gut abgegrenzt werden [BAC2010], [BAC2016]. Dementsprechend weist dieser Algorithmus eine gute Eigenschaft auf, die die Clustergrundidee repräsentiert.

Auch der Ward-Algorithmus weist Schwächen auf. Dieser Algorithmus neigt dazu gleichmäßige Gruppen, auch auf die Gruppengröße bezogen, zu erzeugen. Ebenfalls kann dieser Algorithmus langgezogene Gruppen schwerer erkennen, als z. B. das Single-Linkage. Das führt auch dazu, dass stark heterogene Objekte als Ausreißer auftreten und unter Umständen dann sogar einzelne Cluster bilden [SCH2010]. Das geschieht durch die Betrachtung der Streuung im Cluster. Ausreißer erhöhen diese um ein Vielfaches, wodurch diese nicht mit Clustern fusioniert werden, damit die Fehlerquadratsumme gering bleibt. In Bacher et al. [BAC2010: 294] wurde eine exemplarische Anwendung des Ward Algorithmus auf 1.660 Objekte im Untersuchungsrahmen angewendet. Das bedeutet, dass dieser Algorithmus auch für Untersuchungsrahmen mit einer Objektanzahl im geringen vierstelligen Bereich angewendet werden, wobei für sehr große Betrachtungsrahmen von der Verwendung dieses Algorithmus abgeraten wird [SCH2010], [BAC2010]. Im nachfolgenden Unterkapitel 2.4 werden einige durchgeführten Untersuchungen in der Clusteranalyse in verschiedenen Wissenschaftsbereichen aufgezeigt. Der Ward Algorithmus bildet in vielen der Untersuchungen den gewählten Algorithmus.

Der Vorteil des Ward-Algorithmus ist, dessen „Anwenderfreundlich“ und gute Ergebnisse, die erreicht werden ohne einen großen mathematischen Sachverstand vorauszusetzen. Zeitgleich ist dieser Algorithmus auch nicht so genau wie manche anderen Verfahren, da in der gängigen Softwareumgebung keine vollständig modifizierbaren Softwareumsetzungen existieren. [BAC2010: 283]. Für die reine Anwendung der Clusteranalyse jedoch ist es die beste Wahl, da es einen guten Kompromiss zwischen Genauigkeit, Komplexität und Anwenderfreundlichkeit bietet. Für hochkomplexe statistische Untersuchungen, sowie die Statistikforschung eignen sich diese Verfahren nicht.

Partitionierende Clusteranalyseverfahren

Als partitionierende Verfahren werden Clusteranalyseansätze beschrieben, die die zu untersuchenden Objekte nach Möglichkeit ideal in eine zuvor festgelegte Clusteranzahl aufteilen [BAC2016]. Das Problem bei der alleinigen Nutzung der partitionierenden Verfahren ist genau die Wahl dieser anfänglichen Clusteranzahl bereits vor der Untersuchung, die ohne eine vorgeschaltete Hilfsanalyse nur durch das Erraten gefunden werden kann. Aus diesem Grund werden oftmals deterministische bzw. hierarchische Clusteranalysen vor der partitionierenden Clusteranalyse durchgeführt, um zunächst sinnvolle Clusteranzahlen zu ermitteln, die in einem weiteren Schritt durch eine Partitionierung verfeinert werden können [SCH2010: 133].

In [BAC2010] lassen sich auch partitionierende Verfahren unter den deterministischen Verfahren wiederfinden. Das liegt daran, dass auch partitionierende Verfahren deterministisch sind, deren clusteranalytisches Vorgehen jedoch partitionierend ist. Als am häufigsten diskutierter Vertreter wird unter den deterministischen partitionierenden Verfahren das k-Means Verfahren angeführt [STE2007]. Das k-Means Verfahren erlaubt es vermutete Cluster zu überprüfen. Dazu ist jedoch eine Vorgabe der Clusteranzahl (k Cluster) notwendig, die auch als einzelne

Lösung in einem Durchlauf berechnet wird [SCH2010]. Diese Clusterung wird oft in Ergänzung für z. B. hierarchische Clusteranalysen angewendet. Die so ermittelten oder vermuteten Cluster werden durch das k-Means Verfahren verfeinert. In [STE1977] wird das k-Means Verfahren deshalb auch als „Verfahren zur Verbesserung der anfänglichen Partition bezeichnet“. Die Cluster werden hierbei ebenfalls über die Konstruktion von Clusterzentren, ähnlich wie bei den Median, Centroid und Ward Algorithmen ermittelt. Lediglich die Herangehensweise ist, anders als bei den drei genannten Verfahren, partitionierend [BAC2010]. Zu Beginn wird die Menge der betrachteten Objekte auf die gegebene Clusteranzahl aufgeteilt, die über die Clusterzentren dann homogenisiert werden. [BAC2010] Dieses Verfahren wird für große Fallzahlen vorgeschlagen [SCH2010: V]. In [BAC2010: 302] wird als Größenordnung für Clusterlösungen, die fast überlappungsfrei sind die Objektanzahl von >500 als ausreichend angegeben.

Possibilistische Clusteranalyseverfahren

Ein Vertreter der possibilistischen Verfahren ist die Fuzzy-Clusteranalyse. Diese Art der Analyse bestimmt die Cluster ähnlich wie die probabilistischen Verfahren. Hier sind nicht die Wahrscheinlichkeiten der Betrachtungsgegenstand, sondern die Zugehörigkeit (Zugehörigkeitsgrade) [HÖP1997], [TIM2002]. Diese Algorithmen erfordern auch keine normierten Daten, weswegen die Zugehörigkeiten auch nicht in Summe eins ergeben müssen. Für diese Arbeit ist diese Clusteranalysemethode nicht relevant und wird deshalb nicht weiter ausgeführt. Die Fuzzy-Clusteranalyse findet häufig ihre Anwendung z. B. in der Bilderkennung sowie in der Regelungstechnik, wo eindeutige Zuordnungen nicht möglich bzw. sinnvoll sind und eine große Datenmenge vorhanden ist [TIM2002].

2.3.3 Verwendete Software

Die Nutzung und Anwendung dieser Daten erfordert die Unterstützung durch Software, die für dieses Anwendungsfeld abgestimmt ist. Viele der hier vorgestellten Verfahren und Algorithmen laufen im Hintergrund der Softwareumgebung ab und geben die Ergebnisse in der Form aus, wie es für die Analyse benötigt wird. Durch den Fortschritt in der Computertechnik ist die Statistiksoftware heute in der Lage in sehr kurzer Zeit große Datenumfänge zu verarbeiten. Durch diese Entwicklung ist es in manchen Feldern möglich Erkenntnisse zu erzielen, die früher nicht vorgelegen haben. Als Beispiel kann hier z. B. das Data-Mining oder das Knowledge-Discovery in Databases – KDD angeführt werden. Große Konzerne (z. B. Amazon, Wal-Mart, Automobilkonzerne, Finanzbranche, etc.) nutzen heutzutage in großem Maße Daten, die digital erhoben werden. Um diese Daten auszuwerten und darin enthaltene Datenmuster zu identifizieren werden heutzutage Softwareprogramme angewendet. Beispielsweise können so Kundengruppen segmentiert (Amazon), neue Finanzpakete für Kundengruppen entwickelt (Finanzbranche) oder Verkaufsdaten ausgewertet werden (Wal-Mart) [TIM2002]. Auch die softwaregestützte Clusteranalyse findet in diesem Bereich intensive Anwendung.

Die Statistiksoftware ist in der Lage in sehr kurzer Zeit große Datenmengen zu verarbeiten. In Bacher et al. [BAC2010] wird diverse Statistiksoftware vorgestellt, die zur Verfügung steht. Es gibt neben diverser Freeware Software, die durch die Kostenfreiheit und oftmals freien Quellcode modifiziert werden kann, eine Vielzahl an kostenpflichtiger Software, die neben Ihrer Benutzerfreundlichkeit eine standardisierte Anwendung erlaubt sowie validierte Verfahren beinhaltet. Der Nachteil dabei ist, dass in der standardisierten kostenpflichtigen Software nicht jede Funktion und jeder spezielle statistische Fall enthalten ist. Bei sehr speziellen Analysen wird die Nutzung der freien Software, die modifiziert werden kann, empfohlen. Für die klassische Anwendung auf den Untersuchungsrahmen deutscher kreisfreier Städte reichen jedoch die kostenpflichtigen Standardprogramme aus, sodass das Programm SPSS für die Analyse gewählt wurde. In nachfolgender Tabelle sind einige gängige kostenfreie und –pflichtige Programme aufgeführt.

Tabelle 2-2: Übersicht einiger kostenpflichtiger und –freier Statistikprogramme

Kostenpflichtige Software		Kostenfreie Software	
SPSS (IBM)	[IBM2016]	Softwarepaket „R“	[RPR2016]
SAS	[SAS2016]	AutoClass	[NAT2016]
STATA	[DPC2016]	Cluster 3.0	[INS2016]
SYSTAT	[Sys2016]	ALMO-Statistiksystem	[HOL2015]

2.4 Auswahl einiger clusteranalytischer Untersuchungen

In diesem Kapitel soll ein themenfeldübergreifender Überblick zu einigen publizierten Clusteranalysen gegeben werden. Es kann natürlich nur ein Ausschnitt aus dem Anwendungsfeld der Clusteranalyse, deren Bedeutung besonders mit der Digitalisierung in den meisten Wirtschafts- und Wissenschaftsfeldern einhergeht, gegeben werden. In der Tabelle 2-3 sind diese Clusteranalysen aufgetragen. Dabei wurde das Ziel verfolgt den Untersuchungsrahmen der Clusteranalysen sowie die verwendete Methodik zu identifizieren, nicht aber die wissenschaftlichen Ergebnisse der durchgeführten Clusteranalysen, da aufgrund der vielfältigen betrachteten Wissenschaftsfelder die Ergebnisse für die Clusteruntersuchung deutscher kreisfreier Städte nicht zielführend sind.

In der Tabelle 2-3 sind einige Clusteranalysen aus dem Zeitraum zwischen 1983 und 2014 dargestellt. Die Tabelle ist eingeteilt nach dem Untersuchungsrahmen, der Methodik, dem Ergebnis (Clusteranzahl) sowie sonstigen Information, wie dem Jahr der Durchführung der Clusteranalyse, dem Wissenschaftsfeld und der verwendeten Software. Zum Untersuchungsrahmen zählen die *Anzahl der betrachteten Objekte* sowie die *Anzahl betrachteter Variablen*. Zur Methodik gehören die Spalten *Verfahren*, *Algorithmus*, *Distanzmaß* sowie *Datenvorbereitung*.

Tabelle 2-3: Themenfeldübergreifende Übersicht publizierter Clusteranalysen

Fachbereich	Jahr der Durchführung	Anzahl betrachteter Objekte	Anzahl betrachteter Variablen	Clusteranzahl
Archäologie	2004	162	16 Variablen	2
Astronomie	1996	159	k. A.	6
Biologie	2003	77	25 Indikatoren	2
Biologie	2004	85	14 Indikatoren	4
Chemie	2004	24	210 Indikatoren	9
Epidemiologie	2014	3.076	5 Indikatoren	5
		3.737	5 Indikatoren	3
Geschichte	2010	3.994	85 Indikatoren	11-4
Ingenieurwissenschaften	2013	123	18 Indikatoren, 5 Faktoren	13
Ingenieurwissenschaften	1994	30	6 Indikatoren	3
Literaturwissenschaft	2003	19	23 Indikatoren	4
Medizin	2015	66-80	6 Indikatoren	3
Medizin	1994	337	9 Indikatoren	7
Medizin	1997	260	87 Indikatoren, 6 Faktoren	3
Ökonomie	1999	202/178	5 Indikatoren	8
Ökonomie	2004	25	8 Indikatoren	4
		25	8 Indikatoren	5
Psychologie	2004	36	30 Indikatoren	k. A.
Psychologie	2004	103	12 Indikatoren	4
Psychologie	1983	64	42 Indikatoren	2
Psychologie	1996	307	8 Indikatoren	4
Sozialwissenschaften	(Stichtag 01.01.2014)	2.915	9 Indikatoren, 2 Faktoren	9
Sozialwissenschaften	2003	992	11 Indikatoren	4
Soziologie	2007	54	8 Indikatoren, 2 Faktoren	6
Wirtschaftsinformatik	2012	215	k. A.	14

Tabelle 2-3: Themenfeldübergreifende Übersicht publizierter Clusteranalysen

Verfahren	Algorithmus	Distanzmaß	Datenvorbereitung	Software	Quelle
Hierarchisch agglomerativ	k. A.	Maximum-Likelihood	Hauptkomponentenanalyse	R-MCLUST	[HAL2004]
Hierarchisch agglomerativ	Ward	k. A.	k. A.	SAS-Cluster	[FAU1996]
Hierarchisch agglomerativ	UPGMA	k. A.	k. A.	SAS, NTSYS	[KIM2003]
k. A.	Bayesian	k. A.	k. A.	structure	[PAR2004]
Hierarchisch agglomerativ	Centroid	Euklidische Distanz	k. A.	JMP 4.0.5 (SAS)	[LEM2004]
Two-Step, Regressionsanalyse	k. A.	Log-Likelihood	k. A.	SPSS 18	[HÖP2014]
k-Means, hierarchisch agglom.	Ward	Quadrierte euklidische Distanz	k. A.	k. A.	[FIS2010]
Hierarchisch agglomerativ	Average-Linkage	Quadrierte euklidische Distanz	Faktoranalyse	SPSS 21	[ISE2013], [ISE2014]
k. A.	k. A.	k. A.	k. A.	k. A.	[RON1994]
Hierarchisch agglomerativ	Ward	k. A.	Faktoranalyse	k. A.	[KEE2003]
Hierarchisch agglomerativ, k-Means	Ward	k. A.	k. A.	SPSS 20	[TSE2015]
Hierarchisch agglomerativ	Ward	Maximum-Likelihood.	k. A.	SAS-Cluster	[BUR1994]
k. A.	Ward	Quadrierte euklidische Distanz	Faktoranalyse	k. A.	[HIL1997]
Hierarchisch agglomerativ	Ward	Quadrierte euklidische Distanz	k. A.	k. A.	[HEI1999]
Hierarchisch agglomerativ	Ward	Quadrierte euklidische Distanz	Z-Transformation, Single-Linkage	k. A.	[SCH2004]
Hierarchisch agglomerativ	Ward	Quadrierte euklidische Distanz	k. A.	k. A..	[ALL2004]
Hierarchisch agglomerativ	Ward	k. A.	k. A.	k. A.	[STE2004]
Hierarchisch agglomerativ	Complete Linkage	Euklidische Distanz	k. A.	SAS-Cluster	[FAR1983]
Hierarchisch agglomerativ	Average Linkage	Euklidische Distanz	k. A.	k. A.	[ELL1996]
Hierarchisch agglomerativ, k-Means	Ward	Quadrierte euklidische Distanz	Z-Transformation, Faktoranalyse	k. A.	[BER2014]
k. A.	Aver.- und Compl.-Link.	Phi, Yule's Q	k. A.	k. A.	[JOH2003]
Hierarchisch agglomerativ, k-Means	Ward	Quadrierte euklidische Distanz	Faktoranalyse	k. A.	[STR2007]
k-Means	k. A.	k. A.	k. A.	SAP-BI, APD	[FLA2012]

In Bacher et al. [BAC2010:26] werden einige der in der Tabelle 2-1 aufgeführten Aspekte thematisiert. Auch dort wird darauf hingewiesen, dass eine Vielzahl publizierter Clusteranalysen nicht vollständig beschrieben wird. Eine Auswahl besonders wichtiger Aspekte und Informationen für die Nachvollziehbarkeit einer durchgeführten Clusteranalyse sind nachfolgend aufgeführt [BAC2010: 26]:

- Ausgewählte Objekte
- Ausgewählte Variablen
- Ausgewählte Algorithmen
- Ausgewähltes Clusteranalyseverfahren
- Verwendete Kriterien zur Bestimmung der Clusteranzahl
- Verwendete Software

Es gibt durchaus weitere Aspekte, die für die Nachvollziehbarkeit hilfreich sind, jedoch können die oben Aufgeführten als wichtig angesehen werden. An erster Stelle steht die Darstellung der für die Untersuchung ausgewählten Objekte, die zumindest in den betrachteten Clusteranalysen alle erwähnt wurden. Die Objekte (Städte) für diese Arbeit werden in dem nachfolgenden Unterkapitel 2.5 beschrieben und erläutert.

Es zeigt sich, dass die Struktur der Tabelle 2-3 entsprechend der relevanten Informationen aufgebaut wurde. Aufbauend auf der Gesamtschau der untersuchten Clusteranalysen lässt sich ein passender Anwendungsfall für die Clusterung deutscher kreisfreier Städte ableiten bzw. transferieren. Leider wurde nicht in allen untersuchten Arbeiten die Clusteranalyse in identischer Detailtiefe beschrieben. Die essenziellen Aspekte wurden jedoch erwähnt, wodurch sich eine Methode für die Anwendung auf deutsche kreisfreie Städte ableiten lässt.

Es wurden 23 verschiedene Arbeiten zur Clusteranalyse nach den benannten Gesichtspunkten ausgewertet. Die Tabelle 2-3 zeigt, dass die wesentlichen Anwendungsgebiete der Clusteranalyse derzeit stark in dem psychologisch/medizinischem Bereich (sieben Arbeiten), den Geistes- und Sozialwissenschaften (sechs Arbeiten), den Naturwissenschaften (fünf Arbeiten) sowie der Ökonomie/Wirtschaftsinformatik (drei Arbeiten) liegen. Somit entfallen 21 der 23 Analysen auf diese vier Bereiche. Obwohl der Fokus bei der Literaturrecherche zur Clusteranalyse auf dem ingenieurtechnischen Bereich lag, wurden nur verhältnismäßig wenige Clusteranalysen dazu gefunden. Lediglich zwei der betrachteten Arbeiten entfallen auf diesen Bereich. Das Verhältnis zeigt auf, dass die Clusteranalyse in diesem Wissenschaftsbereich eher unterrepräsentiert ist.

Wie die Auflistung der Clusteranalysen zeigt, findet die hierarchische Clusteranalyse die häufigste Anwendung. Werden auch die Arbeiten hinzugezogen, die aufbauend auf die hierarchische Analyse andere Clusterverfahren anwenden, so wird in 17 der 23 untersuchten Untersuchungen die hierarchische Clusteranalyse angewendet. Auch für die hier vorliegende Arbeit

wurde das hierarchisch agglomerative Verfahren gewählt. Die Aufschlüsselung in der Tabelle 2-3 verdeutlicht ferner weitere Aspekte. Es lassen sich auch andere angewendete Verfahren identifizieren, die aber eher als Sonderfälle angesehen werden, da in diesen Fällen Datensätze mit Umfängen von mehreren Tausend Objekten den Untersuchungen unterzogen werden und somit für die Clusteranalyse deutscher kreisfreier Städte nicht zielführend sind. Der Untersuchungsrahmen deutscher kreisfreier Städte liegt mit 107 Städten weit unter dieser Größenordnung. Auch die geplante Betrachtung deutscher Städte mit Einwohnerzahlen von 10.000 - 20.000 erfordert die Anwendung dieser Verfahren nicht, da in diese Größenordnung lediglich ca. 650 – 1.500 Städte fallen würden. Bereits die Betrachtung der 23 Artikel hat gezeigt, dass die Artikel in der Beschreibung der angewendeten Methode große Unterschiede aufweisen. In vier der betrachteten Arbeiten wurde eine der wichtigsten Informationen nicht gegeben – welches Clusteranalyseverfahren für die Arbeit verwendet wurde. Auch in den weiteren betrachteten Aspekten (Algorithmus, Distanzmaß, Datenvorbereitung, Software) weist die Tabelle 2-3 z.T. viele Lücken auf. Trotz der Lücken lassen sich Abschätzungen treffen. So wurde in 12 Arbeiten der Ward Algorithmus für die Durchführung der Clusteranalyse als Algorithmus gewählt. Diese 12 Untersuchungen betrachten verschiedene Objektanzahlen. Die Spannweite zwischen der Untersuchung mit der geringsten Objektanzahl ist [Fis2010], in der zunächst eine k-Means Clusteranalyse durchgeführt wird und diese Ergebnisse dann in einem weiteren Schritt mit dem Ward-Algorithmus einer hierarchischen Clusteranalyse unterzogen werden. Die Untersuchung mit dem größten Untersuchungsrahmen bildet der Wegweiser Kommune, der einen Umfang von 2.915 Objekten umfasst. Die meisten der Arbeiten, in denen der Ward-Algorithmus mit quadrierter euklidischer Distanz als Distanzmaß angewendet wird, liegt jedoch im Anwendungsbereich zwischen ca. 25 und 337 betrachteter Objekte, sodass sich auch für die Clusterung deutscher kreisfreier Städte (Objektanzahl 107) dieser Algorithmus eignet. Ein ähnlicher Bereich wird auch in [SCH2010] und [BAC2010] als Anwendungsbereich angegeben, wobei ein solcher Bereich heutzutage nur äußerst schwer festgelegt werden kann, da mit dem Fortschritt der Rechenleistung der Computer sich auch der Umfang der rechenintensiven hierarchischen Clusteranalyse nach oben hin öffnet.

Ein weiterer wichtiger Punkt ist die Datenvorbereitung. Eine Datenvorbereitung findet immer statt, da die Daten der betrachteten Problemstellung oftmals nicht in einer Form vorliegen, die einer Clusteranalyse direkt zugeführt werden können. So müssen die Daten auch im Hinblick auf die Aussagekraft vorbereitet werden. Da bei vielen der hierarchischen Clusteranalysealgorithmen als Distanz die euklidische (ED) bzw. quadrierte euklidische Distanz (QED) angewendet wird, würden ohne eine Normierung bzw. Standardisierung der Daten mathematische Fehler eingebaut, da die Variablenwerte durch das Quadrieren bei der Distanzbestimmung ungleich in die Berechnung der Distanz einfließen (siehe Gleichung 2-2 und 2-3). Hohe Vari-

ablenwerte erhalten durch das Quadrieren des Wertes gegenüber sehr geringen Variablenwerten³ eine ungleiche Aussagekraft, wodurch den Variablen im mathematischen Sinn nicht die gleiche Bedeutung zukommt.

$$ED_{(i,j)} = \sqrt{\sum_{k=1}^m (z_{ik} - z_{jk})^2} \quad \text{Gleichung 2-2}$$

$$QED_{(i,j)} = \sum_{k=1}^m (z_{ik} - z_{jk})^2 \quad \text{Gleichung 2-3}$$

Die Distanz berechnet sich im Prinzip für beide Distanzmaße ähnlich. Bei der *ED* werden die Differenzen der Variablenwerte (*i, j*) der betrachteten Objekte (*z*) zwischen allen Objekten berechnet und aufsummiert. Daraufhin wird die Quadratwurzel des Ergebnisses berechnet. Bei der *QED* fehlt dieser letzte Schritt.

Obwohl die Datenvorbereitung einen relevanten Schritt einer validen Clusteranalyse bildet werden in 16 der betrachteten Arbeiten keine Angaben zu diesem Schritt gegeben. Es ist davon auszugehen, dass in mehr als nur den Arbeiten in denen die Datenvorbereitung beschrieben war, diese auch angewendet wurde. Dazu lässt sich jedoch keine Aussage treffen. Aus den Arbeiten in denen zur Datenvorbereitung Angaben gemacht wurden, wird in fünf dieser Arbeiten die Faktoranalyse z. T. mit einer vorgeschalteten Z-Transformation angegeben. Das macht in beiden Fällen Sinn, da bei großen Variablenumfängen die Faktorenanalyse eine Datenreduktion erzeugt, die für die weitere Deutung der Ergebnisse hilfreich sein kann, da nur die essenziellen Indikatoren in Faktoren zusammengefasst werden. Die Z-Transformation sorgt zudem dafür, dass bei Variablenwerten verschiedener Größenordnung diese auf ein einheitliches Skalenniveau „standardisiert“ werden (Bereich ca. zwischen +/-3) wodurch die mathematische Problematik der ungleichen Bedeutung der Indikatorwerte bei verschiedenen Skalengrößen behoben wird. Auch für die Clusteranalyse deutscher kreisfreier Städte wird eine solche Datenvorbereitung und die Faktoranalyse angewendet, um die Ergebnisse zu verbessern und die Clusteranalyse damit regelgerecht anwenden zu können (siehe Kapitel 5).

Als letzter Aspekt wird die verwendete Software betrachtet. In 11 Artikeln wurde die Software benannt. Dabei zeigt sich, dass in den beschriebenen und verwendeten statistischen Softwarelösungen im Wesentlichen die kommerziellen Programme SPSS (3 Artikel) und SAS (5 Artikel) in verschiedenen Versionen benannt werden. Es lässt sich feststellen, dass die SPSS Software besonders in den letzten Jahren angewendet wird. Die Clusteranalyse in Ka-

³ bspw. verfügbares Einkommen in Haushalten –Größenordnung ca. 30.000 €/Jahr und die Bevölkerungsentwicklung in deutschen Städten –Größenordnung im geringen einstelligen Prozentbereich

pitel 3 und Kapitel 5 wird aufgrund der Verfügbarkeit, Aktualität und Anwendbarkeit der hierarchischen Clusteranalyse auf Datensätze, die in ähnlicher Größenordnung zu dem in dieser Arbeit betrachteten Fall liegen, mit der Software SPSS 22.0 von IBM [IBM2016] durchgeführt.

Der Untersuchungsumfang von 23 Clusteranalysen bildet an dieser Stelle nicht alle durchgeführten Clusteranalysen ab. Die Anzahl durchgeführter Clusteranalysen bildet einen viel größeren Umfang, als die hier betrachtete Auswahl, deshalb ist die Auswertung sämtlicher Clusteranalysen nicht möglich. Dennoch zeigt die Darstellung in Tabelle 2-3 eine breite Anwendungsbasis und bestätigt die Aussage, dass die hierarchische Clusteranalyse eins der häufigsten Clusteranalyseverfahren bildet. Dieses Verfahren erlaubt die Betrachtung und Analyse der in ihrem Umfang vielfältiger Untersuchungsrahmen (Objekt- und Indikatoranzahl). Eine Arbeit, die zu dem Thema der Entwicklung der Clusteranalyseanwendung von Kettenring [KET2006] durchgeführt wurde, zeigt ein eindrucksvolles Ergebnis. In dieser Arbeit wurde die Entwicklung der Anzahl der Veröffentlichungen und Artikel in Literaturdatenbanken im zeitlichen Verlauf untersucht. Betrachtet wurden dabei die wissenschaftlichen Artikel in den drei Datenbanken des Web of Science, dem über 7.000 wissenschaftliche Journals angehören. Diese wurden nach Artikeln durchsucht in denen der Ausdruck Clusteranalyse (*cluster analysis*) vorgefunden werden kann und diese Anzahl der gefundenen Artikel dem Jahr der Veröffentlichung zugeordnet. Die dort enthaltenen Analysen entfallen im Wesentlichen auf die Biologie (Pflanzenkunde) sowie die anderen Naturwissenschaften. Die Medizin, Psychologie, Sozialwissenschaften sowie die Ingenieurwissenschaften sind eher unterrepräsentiert. Das liegt jedoch an den Datenbanken, die verstärkt den naturwissenschaftlichen Bereich umfassen. Die Entwicklung in den Zahlen der Artikel zum Thema Clusteranalyse entwickelte sich von 646 Artikeln in 1995 innerhalb von acht Jahren auf 1.156 Artikeln in denselben Datenbanken im Jahr 2003. [KET2006]

Es ist anzunehmen, dass der Trend auch zukünftig eine ähnliche Entwicklung nehmen wird, da besonders jetzt durch die Digitalisierung und die sich stark entwickelnde Computer- und Softwaretechnik neue Möglichkeiten eröffnet werden und die multivariate Clusteranalyse in vielen Wissenschaftsfeldern an Bedeutung gewinnen dürfte.

Ein weiterer Punkt ist, dass besonders viele durchgeführte Clusteranalysen nicht veröffentlicht werden. Dabei sind besonders die wirtschaftlichen Bereiche zu nennen (Ingenieurtechnik, Pharmazie, Ökonomie), wo mit großer Wahrscheinlichkeit die Clusteranalyse angewendet wird, die Clusteranalysen aufgrund von Geheimhaltung von Daten, Unternehmensinteressen sowie sensiblen Kundendaten aber nicht publiziert werden. In der Wirtschaft findet die Clusteranalyse bereits heute sehr häufig Ihre Anwendung, beispielsweise um Kundengruppen oder Märkte zu segmentieren. [TIM2002], [STA2009]

2.5 Untersuchungsrahmen

Der Untersuchungsrahmen bildet einen essenziellen Punkt einer jeden statistischen Untersuchung. Oft sind clusteranalytische Untersuchungen nicht detailliert beschrieben (s. Tabelle 2-3) mit einer Vielzahl an Lücken zu den benannten Bereichen. Daraus lässt sich eine vollständige Nachvollziehbarkeit nicht ohne weiteres erzeugen, da entscheidende Annahmen und das exakte angewendete Vorgehen z.T. nur in Ansätzen beschrieben sind. In diesem Unterkapitel wird der Untersuchungsrahmen detaillierter erläutert. Dazu werden die betrachteten Städte (Objekte) und Sektoren mit der kurzen Vorstellung der gewählten Variablen (Indikatoren), auf die in den Unterkapiteln 3.1 und 4.1 näher eingegangen wird, beschrieben.

2.5.1 Voruntersuchung - Methodenscreening

Der erste Abschnitt dieser Arbeit basiert auf den Daten einer bereits durchgeführten Clusteranalyse. In dieser Clusteranalyse wurden 16 Städte mit jeweils 12 Variablen einer Clusterung unterzogen. Die Ergebnisse dieser ersten Clusteranalyse wurden in [WAL2014] veröffentlicht. Diese Datenbasis wird für die erste Untersuchung als Voruntersuchung gewählt. Zum einen erlauben diese Ergebnisse den ersten Vergleich mit den Ergebnissen aus dem Methodenscreening, bei dem gängige clusteranalytische Algorithmen verwendet werden und zum anderen wird kein zu großer Datenumfang für das Methodenscreening verwendet wodurch die Auswertung überschaubar bleibt. Durch dieses Vorgehen soll ein geeigneter Algorithmus gewählt werden, der einerseits stabile Cluster erzeugt und sich für den Untersuchungsumfang eignet sowie andererseits eine sinnvolle Interpretation der Cluster erlaubt.

In der Voruntersuchung wurde lediglich ein Ausschnitt aller deutschen, kreisfreien Städte betrachtet. Die Zuordnung der Städte zum jeweiligen Stadttyp basiert auf [EU2011] und kann der Tabelle 2-4 entnommen werden. Die Zuordnung ist für die Clusterung nicht notwendig, dennoch stellt das Wissen über die definitionsgemäße Zuordnung einer Stadt zu einem Stadttyp eine Hilfe bei der Interpretation der Stadtcluster dar.

Tabelle 2-4: Verschiedene Städtegrößen und ihre Einwohnerzahlen

Stadt	Einwohnerzahl
Kleinstadt	< 50.000
Mittelstadt	50.000 – 100.000
Großstadt	> 100.000
Millionenstadt	> 1.000.000

Die Tabelle 2-5 beinhaltet in Anlehnung an [WAL2013], [Eic2014] die 16 Städte mit den Einwohnerzahlen, dem Stadttyp sowie dem Bundesland.

Tabelle 2-5: Betrachtete Städte und deren Einordnung - Voruntersuchung

Stadt	Einwohnerzahl 2011 ⁴	Stadttyp	Bundesland
Bochum	373.976	Großstadt	Nordrhein-Westfalen
Braunschweig	250.556	Großstadt	Niedersachsen
Essen	573.468	Großstadt	Nordrhein-Westfalen
Kiel	242.041	Großstadt	Schleswig-Holstein
Koblenz	106.677	Großstadt	Rheinland-Pfalz
Köln	1.017.155	Millionenstadt	Nordrhein-Westfalen
Magdeburg	232.364	Großstadt	Sachsen-Anhalt
Mainz	200.957	Großstadt	Rheinland-Pfalz
München	1.378.176	Millionenstadt	Bayern
Nürnberg	510.602	Großstadt	Bayern
Potsdam	158.902	Großstadt	Brandenburg
Regensburg	136.577	Großstadt	Bayern
Schwerin	95.300	Mittelstadt	Mecklenburg-Vorpommern
Stuttgart	613.392	Großstadt	Baden-Württemberg
Wiesbaden	278.919	Großstadt	Hessen
Wuppertal	349.470	Großstadt	Nordrhein-Westfalen

Bei der Wahl der Städte für die Voruntersuchung wurde der Ansatz verfolgt unterschiedliche Städte in die Betrachtung hinzuzuziehen, um eine möglichst hohe Aussagekraft und ein breites Abbild deutscher, kreisfreier Städte zu generieren. Aus diesem Grund finden sich neben den mit Abstand am meisten vertretenen 13 Großstädten auch zwei Millionenstädte und eine Mittelstadt im Untersuchungsrahmen. Der Querschnitt wird ferner ausgeweitet, indem Landeshauptstädte und normale Städte betrachtet werden. Dadurch kann vermieden werden, dass das Ergebnis verfälscht werden kann, weil die Strukturen von Landeshauptstädten sich von den nicht Landeshauptstädten unterscheiden und der Fehler so mitgetragen werden könnte. Darüber hinaus wurde auch auf eine regionale Diversität geachtet. Durch die Wahl unterschiedlich großer Städte, die aus verschiedenen geographischen, politischen und zum Teil historischen Regionen stammen, wurde dem Anspruch einer möglichst belastbaren Datenbasis für die Clusteranalyse sowie das Methodenscreening entsprochen. [WAL2013] Diese 16 Städte bilden mit ca. 6,5 Millionen Einwohnern knapp 8 % der deutschen Bevölkerung ab.

2.5.2 Untersuchung kreisfreier Städte

Für die Clusterung deutscher kreisfreier Städte wurde eine größere Datenbasis erarbeitet. Die betrachteten Städte sind in dieser Untersuchung sämtliche kreisfreie Städte in Deutschland mit Ausnahme einiger weniger, die im weiteren Verlauf kurz erläutert werden. Die Verwaltung Deutschlands wird durch ca. 400 Kreise und kreisfreie Städte getragen. Das Statistische Bundesamt listet für Deutschland 107 kreisfreie Städte⁵. Eine kreisfreie Stadt ist nicht per se eine Großstadt, obwohl dieser Zusammenhang oft vermutet wird. Eine Großstadt ist hingegen auch nicht immer kreisfrei, sondern zum Teil kreisangehörig (bspw. Hildesheim in Niedersachsen oder Bergisch-Gladbach in Nordrhein Westfalen) [DES2014B].

⁴ Quelle: Statistisches Bundesamt [DES2013A], [DES2014B]

⁵ Stand 2013 [DES2014B]

Eine kreisfreie Stadt zeichnet sich ebenso wie die Landkreise durch deren direkte Unterstellung den Regierungsbezirken der Bundesländer aus. [NDS2010] Dieser Zusammenhang führt zu der Vermutung, dass eine kreisfreie Stadt eine Großstadt sein muss, da in den meisten Fällen erst eine Großstadt die Pflichten einer kreisfreien Stadt alleine verantworten kann.

Die Zahl der kreisfreien Städte in Deutschland verändert sich im zeitlichen Verlauf. Die letzte größere Änderung fand im Jahr 2011 statt. In diesem Jahr wurde vier Städten aus Mecklenburg-Vorpommern der Status einer kreisfreien Stadt entzogen und die Städte den angrenzenden Landkreisen zugeordnet [DES2015B]. Änderungen in der Anzahl kreisfreier Städte finden hin und wieder statt, jedoch halten sich die Veränderungen in der Größenordnung von einer Stadt, die ihren Status verändert. Die aktuelle Anzahl von 107 kreisfreien Städten ist in Tabelle 2-6 aufgeführt. Da die Tabelle umfangreicher ist als Tabelle 2-5 wird im Anschluss an die Tabelle eine quantitative Auswertung der Städte gemacht. Die Daten, die für die Städte angegeben werden sind die Bevölkerungsanzahl, das Bundesland sowie der Stadttyp, dem die Stadt zugeordnet wird.

Tabelle 2-6: Alle deutschen kreisfreien Städte und deren Einordnung

Stadt	Einwohnerzahl 2012 ⁶	Stadttyp	Bundesland
Amberg	41.578	Kleinstadt	Bayern
Ansbach	39.684	Kleinstadt	Bayern
Aschaffenburg	67.681	Mittelstadt	Bayern
Augsburg	272.699	Großstadt	Bayern
Baden-Baden	52.585	Mittelstadt	Baden-Württemberg
Bamberg	70.863	Mittelstadt	Bayern
Bayreuth	71.482	Mittelstadt	Bayern
Berlin	3.375.222	Millionenstadt	Berlin
Bielefeld	328.314	Großstadt	Nordrhein-Westfalen
Bochum	362.213	Großstadt	Nordrhein-Westfalen
Bonn	309.869	Großstadt	Nordrhein-Westfalen
Bottrop	116.498	Großstadt	Nordrhein-Westfalen
Brandenburg an der Havel	71.149	Mittelstadt	Brandenburg
Braunschweig	245.798	Großstadt	Niedersachsen
Bremen	546.451	Großstadt	Bremen
Bremerhaven	108.323	Großstadt	Bremen
Chemnitz	241.210	Großstadt	Sachsen
Coburg	41.022	Kleinstadt	Bayern
Cottbus	99.913	Mittelstadt	Brandenburg
Darmstadt	147.925	Großstadt	Hessen
Delmenhorst	73.588	Mittelstadt	Niedersachsen
Dessau-Roßlau	84.606	Mittelstadt	Sachsen-Anhalt
Dortmund	572.087	Großstadt	Nordrhein-Westfalen
Dresden	525.105	Großstadt	Sachsen
Duisburg	486.816	Großstadt	Nordrhein-Westfalen
Düsseldorf	593.682	Großstadt	Nordrhein-Westfalen
Eisenach	41.744	Kleinstadt	Thüringen
Emden	49.751	Kleinstadt	Niedersachsen
Erfurt	203.485	Großstadt	Thüringen
Erlangen	105.412	Großstadt	Bayern
Essen	566.862	Großstadt	Nordrhein-Westfalen
Flensburg	83.462	Mittelstadt	Schleswig-Holstein
Frankenthal (Pfalz)	47.035	Kleinstadt	Rheinland-Pfalz
Frankfurt (Oder)	58.537	Mittelstadt	Brandenburg
Frankfurt am Main	687.775	Großstadt	Hessen
Freiburg im Breisgau	218.043	Großstadt	Baden-Württemberg
Fürth	118.358	Großstadt	Bayern

⁶ Quelle: Statistisches Bundesamt [DES2013A], [DES2014B]

Stadt	Einwohnerzahl 2012 ⁶	Stadttyp	Bundesland
Gelsenkirchen	257.607	Großstadt	Nordrhein-Westfalen
Gera	95.384	Mittelstadt	Thüringen
Hagen	186.243	Großstadt	Nordrhein-Westfalen
Halle (Saale)	231.440	Großstadt	Sachsen-Anhalt
Hamburg	1.734.272	Millionenstadt	Hamburg
Hamm	176.440	Großstadt	Nordrhein-Westfalen
Heidelberg	150.335	Großstadt	Baden-Württemberg
Heilbronn	117.531	Großstadt	Baden-Württemberg
Herne	154.563	Großstadt	Nordrhein-Westfalen
Hof	44.461	Kleinstadt	Bayern
Ingolstadt	127.886	Großstadt	Bayern
Jena	106.915	Großstadt	Thüringen
Kaiserslautern	97.112	Mittelstadt	Rheinland-Pfalz
Karlsruhe	296.033	Großstadt	Baden-Württemberg
Kassel	192.874	Großstadt	Hessen
Kaufbeuren	41.570	Kleinstadt	Bayern
Kempten (Allgäu)	64.625	Mittelstadt	Bayern
Kiel	239.866	Großstadt	Schleswig-Holstein
Koblenz	109.779	Großstadt	Rheinland-Pfalz
Köln	1.024.373	Millionenstadt	Nordrhein-Westfalen
Krefeld	222.026	Großstadt	Nordrhein-Westfalen
Landau in der Pfalz	43.641	Kleinstadt	Rheinland-Pfalz
Landshut	65.322	Mittelstadt	Bayern
Leipzig	520.838	Großstadt	Sachsen
Leverkusen	159.926	Großstadt	Nordrhein-Westfalen
Lübeck	211.713	Großstadt	Schleswig-Holstein
Ludwigshafen am Rhein	160.179	Großstadt	Rheinland-Pfalz
Magdeburg	229.924	Großstadt	Sachsen-Anhalt
Mainz	202.756	Großstadt	Rheinland-Pfalz
Mannheim	294.627	Großstadt	Baden-Württemberg
Memmingen	41.551	Kleinstadt	Bayern
Mönchengladbach	255.087	Großstadt	Nordrhein-Westfalen
Mülheim an der Ruhr	166.654	Großstadt	Nordrhein-Westfalen
München	1.388.308	Millionenstadt	Bayern
Münster	296.599	Großstadt	Nordrhein-Westfalen
Neumünster	76.951	Mittelstadt	Schleswig-Holstein
Neustadt a. d. W.	52.268	Mittelstadt	Rheinland-Pfalz
Nürnberg	495.121	Großstadt	Bayern
Oberhausen	210.005	Großstadt	Nordrhein-Westfalen
Offenbach am Main	116.945	Großstadt	Hessen
Oldenburg	158.658	Großstadt	Niedersachsen
Osnabrück	155.625	Großstadt	Niedersachsen
Passau	49.038	Kleinstadt	Bayern
Pforzheim	116.425	Großstadt	Baden-Württemberg
Pirmasens	40.267	Kleinstadt	Rheinland-Pfalz
Potsdam	159.456	Großstadt	Brandenburg
Regensburg	138.296	Großstadt	Bayern
Remscheid	109.352	Großstadt	Nordrhein-Westfalen
Rosenheim	59.935	Mittelstadt	Bayern
Rostock	202.887	Großstadt	Mecklenburg
Salzgitter	98.095	Mittelstadt	Niedersachsen
Schwabach	39.137	Kleinstadt	Bayern
Schweinfurt	52.098	Mittelstadt	Bayern
Schwerin	91.264	Mittelstadt	Mecklenburg
Solingen	155.316	Großstadt	Nordrhein-Westfalen
Speyer	49.764	Kleinstadt	Rheinland-Pfalz
Straubing	45.099	Kleinstadt	Bayern
Stuttgart	597.939	Großstadt	Baden-Württemberg
Suhl	35.967	Kleinstadt	Thüringen
Trier	106.544	Großstadt	Rheinland-Pfalz
Ulm	117.977	Großstadt	Baden-Württemberg
Weiden i. d. O.	41.684	Kleinstadt	Bayern
Weimar	63.236	Mittelstadt	Thüringen
Wiesbaden	272.636	Großstadt	Hessen
Wilhelmshaven	76.545	Mittelstadt	Niedersachsen
Wolfsburg	121.758	Großstadt	Niedersachsen
Worms	79.727	Mittelstadt	Rheinland-Pfalz
Wuppertal	342.885	Großstadt	Nordrhein-Westfalen
Würzburg	124.577	Großstadt	Bayern
Zweibrücken	34.064	Kleinstadt	Rheinland-Pfalz

Die aufgeführten 107 Städte sind derzeit alles kreisfreie Städte in Deutschland. Wie in nachfolgender Tabelle 2-8 zu erkennen ist, ist tatsächlich der größte Teil der Städte Großstädte. Neben den vier Millionenstädten in Deutschland gibt es auch 18 Kleinstädte, die den Status einer kreisfreien Stadt tragen. Das Problem wird hieran erkennbar, denn die kleinste Stadt im Betrachtungsrahmen ist die Kleinstadt Zweibrücken in Rheinland-Pfalz und mit Ihren 34.000 Einwohnern nur ca. 1 % der Bevölkerung der ebenfalls betrachteten Stadt Berlin mit Ihren ca. 3.400.000 Einwohnern hat. Verglichen werden können beide Städte, wenn auch nicht in vollem Umfang. Eine komplette Gleichsetzung ist aufgrund der enormen Unterschiede ausgeschlossen. Das zeigt auch, dass die Beschränkung auf kreisfreie Städte eine vertretbare Entscheidung ist, um die Vergleichbarkeit und Belastbarkeit zunächst zu erhöhen. Für weitere Untersuchungen wäre eine Einteilung der betrachteten Städte in zwei Gruppen denkbar und sollte geprüft werden. In der Gegenüberstellung wurden die Großstädte unterteilt nach *großen* und *kleinen* Großstädten. Die Unterteilung kann eine weitere Hilfestellung für die spätere Analyse der Ergebnisse darstellen, da große Großstädte zum Teil schon andere Strukturen aufweisen als sehr kleine Großstädte.

Auch regional ist die Anzahl kreisfreier Städte ungleich verteilt. Die größte Anzahl kreisfreier Städte (25) ist in Bayern vorzufinden. Folgend sind 22 kreisfreie Städte in Nordrhein-Westfalen vorzufinden, wobei dort über die Hälfte der Städte im Ruhrgebiet liegen. Die Anwendung der Clusteranalyse kann besonders gut auf solche Regionen angewendet werden, um zu überprüfen ob die Belastbarkeit der Clusterung hoch ist. Denn Städte aus ähnlichen historischen und wirtschaftlichen Regionen die Vermutung nahelegen, ähnliche Strukturen vorzuweisen. Die geringste Anzahl an kreisfreien Städten weisen die Stadtstaaten Berlin (1), Hamburg (1) und Bremen (2) auf. Bremen ist der einzige Stadtstaat, der aus zwei Städten besteht. Das Saarland ist das einzige Bundesland in dem keine kreisfreie Stadt existiert. Im Saarland gibt es eine Städteregion, die mehrere Städte umfasst [SAA2013].

Tabelle 2-7: Bundesländer nach Anzahl der kreisfreien Städte

Bundesland	Abkürzung	Anzahl kreisfreier Städte	Kumulierte Anzahl der Städte
Bayern	BY	25	25
Nordrhein-Westfalen	NW	22	47
Rheinland-Pfalz	RP	12	59
Baden-Württemberg	BW	9	68
Niedersachsen	NI	8	76
Thüringen	TH	6	82
Hessen	HE	5	87
Schleswig-Holstein	SH	4	91
Brandenburg	BB	4	95
Sachsen	SN	3	98
Sachsen-Anhalt	ST	3	101
Bremen	HB	2	103
Mecklenburg-Vorpommern	MV	2	105
Hamburg	HH	1	106
Berlin	BE	1	107
Saarland	SL	0	107

Die 16 Städte der Voruntersuchung sind der Wohnort von ca. 6,5 Millionen Menschen in Deutschland. Die 6,5 Millionen stehen somit stellvertretend für 8,1 % der Bevölkerung Deutschlands. Werden hingegen alle kreisfreien Städte untersucht, so leben in den 107 Städten knapp 22,5 Millionen Menschen, die mit 31,6 % fast ein Drittel der Gesamtbevölkerung Deutschlands darstellen. Diese Zahlen zeigen die starke Urbanisierung Deutschlands auf. Durch eine Verdopplung der Anzahl betrachteter Städte würden bereits ca. 50 % der Einwohner Deutschlands Teil des Betrachtungsrahmes sein. Dieser Schritt, die Erweiterung des Untersuchungsrahmens über kreisfreie Städte hinaus, ist leider nicht ohne weiteres möglich, da die dafür benötigten Daten in einem solchen Umfang nicht erhoben werden können. Alle Städte, die nicht kreisfrei sind, werden Landkreisen zugerechnet. Dadurch müssen auch die Daten für die Landkreise erhoben werden, diese sind nur in seltensten Fällen stadtsspezifisch verfügbar.

Tabelle 2-8: Gegenüberstellung der beiden Untersuchungen

		Voruntersuchung [WAL2014]	Alle kreisfreien Städte
Kleinstadt		0	18
Mittelstadt		1	23
Großstadt	Klein <500.000	10	54
	Groß >500.000	3	8
Millionenstadt		2	4
Summe		16	107
Einwohner		6.518.532	25.522.828
Anteil an Gesamtbevölkerung in %		8,1	31,6

Durch die Datenlage, die für kreisfreie Städte besser und umfangreicher zu erheben ist, eröffnet sich die Möglichkeit die Städte ganzheitlich, d. h. alle Verbrauchssektoren einer Stadt zu betrachten. Diese werden im Abschnitt 2.5.3 kurz vorgestellt.

2.5.3 Verbrauchssektoren

Die betrachteten Verbrauchssektoren sind die vier in der Wissenschaft verbreiteten Sektoren. Diese sind der Sektor *Private Haushalte*⁷, *Verkehr*, *Industrie und Gewerbe*, *Haushalte und Dienstleistungen (GHD)*. Die vier Sektoren sind besonders geeignet, da es mit ihnen möglich ist die energiebedingten Aktivitäten Bereichen zuzuordnen ohne eine „Ausfransung“ durch zu genaue Unterteilung zu bewirken. Außerdem ist die Betrachtung der vier Sektoren besonders in den kreisfreien Städten sinnvoll, da die drei Sektoren Verkehr, Industrie und GHD in kreisfreien Städten besonders stark ausgeprägt sind [PBB2013].

⁷ Im Folgenden wird von privaten Haushalten ausgegangen auch wenn das Wort Haushalt benutzt wird

Private Haushalte

Die privaten Haushalte bilden definitionsgemäß alle Haushalte ab, die die Organisation des privaten Lebens und Wirtschaftens einschließen. Da das gesamte private Leben in den Haushalten abläuft, findet dementsprechend die Hauswirtschaft auf privater Ebene statt. Es gibt diverse Formen der Haushalte, eine der gängigsten ist die Unterscheidung nach Ein- und Mehrpersonenhaushalten. [WAL2013] Die Lebenssituation der Haushalte stellt häufig den Gradmesser dar [ENG2015]. Damit einher geht auch die Variation der Haushaltstypen in Abhängigkeit von Struktur und Art einer Stadt. So sind z. B. Universitätsstädte durch eine höhere Anzahl an Einpersonenhaushalten geprägt. Andere Städte können viel höhere Anteile an Mehrpersonenhaushalten verfügen, jedoch sind Mehrpersonenhaushalte nicht immer auf Basis der Ehe festzumachen, sondern auch durch eingetragene Lebenspartnerschaften, die zusammen durch ein familiäres, familienähnliches oder emotionales Zusammenleben geprägt werden [WAL2013], [BPB2012], [MEY2013]. Die betrachteten Städte bilden mit insgesamt 25,6 Mio. Einwohnern, die sich auf 13,3 Mio. Haushalte aufteilen, in etwa 33 % aller deutschen Haushalte ab (39,9 Mio. Haushalte in Deutschland, Stand 2013). [DES2015c], [ZEN2015]. In den Haushalten wird auch eine große Energiemenge benötigt, wodurch deren verstärkte Betrachtung notwendig ist. Die Tabelle 2-9 zeigt exemplarisch den Endenergieverbrauch der Haushalte. Der größte Anteil entfällt auf die Bereitstellung von Raumwärme und Warmwasser. Der Bedarf elektrischer Energie kann in etwa mit 20 % des Gesamtverbrauchs angesehen werden, unter der Annahme, dass ein Teil der Warmwasserbereitung durch elektrische Energie erfolgt. [DES2013c] Somit entfällt ein Anteil von ca. 80 % auf die Bereitstellung von Raumwärme und die Bereitstellung von Warmwasser. Diese Posten werden nach wie vor zu einem sehr großen Anteil durch konventionelle Energieträger gedeckt und bewirken somit Emissionen.

Tabelle 2-9: Endenergieverbrauch in Haushalten - 2012

Anwendungsbereich	Endenergie ⁸ [PJ]	Anteile [%]
Raumwärme	1.709	70,3
Warmwasser	308	12,7
Kochen, Trocknen, Bügeln, Haushaltsgeräte, Beleuchtung	414	17,0
Gesamt	2.431	100,0

Es werden viele Ansätze, besonders im Bereich der Wärmebereitung angegangen, da hier auch die größten Potenziale zu realisieren sind. Neben der Verringerung des Energieverbrauchs durch die Verbesserung der eingesetzten Heizungsanlagen und der Gebäudedämmung gibt es auch die Möglichkeit die Energieeffizienz zu steigern. Das kann zum Beispiel durch die effizientere und besser organisierte Nutzung der Endenergie geschehen. Elektronisch programmierbare Heizkörperthermostate zum Beispiel können nach Stiftung Warentest [SW2008] dazu beitragen bis zu 10 % Einsparungen zu erzielen.

⁸ eigene Darstellung nach [DES2013c]

Verkehr

Der Verkehrssektor stellt einen der energieintensivsten Sektoren dar. Dieser Sektor erfordert eine besondere Betrachtung, da er für alle Bereiche des Lebens in einem Land relevant ist. Die Unterteilung des Sektors erfolgt in den Personen- und Güterverkehr, wodurch dieser als Querschnittssektor betrachtet werden kann. Der Verkehr bildet viele Bereiche ab, wozu der Straßen-, Schienen-, Luft- sowie der Schiffsverkehr zählen. [DES2013B] Der Güterverkehr ist besonders in industrialisierten Staaten für die Funktion des Warenaustausches nicht wegzudenken. Da dieser Warenaustausch jedoch fast immer über Stadtgrenzen hinweg erfolgt und die exakte Zuordnung von getätigter Fahrleistung zur jeweiligen Stadt aus diesem Grund nicht möglich ist, wird die Erarbeitung der Indikatoren zur Clusterung der Städte lediglich auf den Personenverkehr fokussiert. Aber auch im Personenverkehr lassen sich nicht alle Verkehrsarten sinnvoll in die Untersuchung von Städten hinzuziehen. So ist der Luft-, Schiffs- und bahnfernverkehr kaum einer einzelnen Stadt zuzuordnen, da der Transport über sehr weite Strecken und Stadtgrenzen hinweg erfolgt. Deswegen werden diese Bereiche nicht in der Untersuchung betrachtet. Der innerstädtische Personenverkehr konzentriert sich somit auf den öffentlichen Personennahverkehr (ÖPNV), den motorisierten Individualverkehr (MIV) sowie die Nutzung des Fahrrads. Letzteres kann zwar Verkehrsleistung substituieren, hat jedoch keine direkte energetische Relevanz. Der ÖPNV, bestehend aus dem Schienen- und Busverkehr, ist in der Diskussion um die Emissionsreduktion im Verkehr ein Hoffnungsträger, da das Potenzial enorm ist. Derzeit entfallen ca. 9 % des Personentransportaufkommens auf den ÖPNV, das Aufkommen des Personentransports am MIV (Pkw und Kraftrad) ist mit ca. 80 % fast um den Faktor 9 größer. [UBA2012], [Eic2014] Die Mobilität hat sich heutzutage verändert. Viele Bereiche und Aktivitäten des Lebens erfordern die Mobilität. So nimmt der Arbeits-Pendelverkehr zu, der Verkehr zum Einkaufen in vielen Fällen ebenfalls. Als weiterer Aspekt ist die Nutzung, insbesondere des MIV, für das Ausleben Freizeitaktivitäten relevant. [MEY2013]

Industrie

Die Industrie ist in Deutschland ein besonders großer Sektor. Deutschland ist ein Industrieland, in dem auf kapitalintensiven Industriesektor ca. 23 % des Bruttoinlandsproduktes (BIP) entfallen (Stand 2014 [DES2015D]). Der Sektor teilt sich auf das produzierende und das verarbeitende Gewerbe auf. Dazu zählt die Herstellung und Verarbeitung von Gütern, die Wartung und Reparatur, die Instandhaltung, das Tätigkeitsfeld der Grundstoffindustrie, der Bergbau sowie das Baugewerbe. [DES2015H] Auf die Industrie entfällt ein Anteil von fast 30 % des gesamten Endenergieverbrauchs in Deutschland. Somit ist dieser Sektor im industrialisierten Deutschland der größte Energieverbraucher [BMW2015]. Das Statistische Bundesamt bietet mit den einzelnen Wirtschaftszweigen eine Möglichkeit Unternehmen, die dem Industriesektor zugeordnet werden, zusammenzufassen. [WAL2013] [DES2003] [DES2008] Es können grob zwei Sparten in der Industrie zusammengefasst werden. Zum einen die energieintensive Industrie,

die z. B. die Bereiche Grundstoff- und Produktionsgütergewerbe, die eisenschaffende Industrie, Bergbau, Energie- und Wassererzeugung sowie die chemische Industrie umfasst. Zum anderen die klassischen, produzierenden oder verarbeitenden Gewerbe (Fahrzeugbau, Maschinenbau, Elektrotechnik), in denen die Grundstoffe benötigt werden, sowie die Leichtindustrie (Textil- und Papierindustrie) [UNG1997], [DES2003], [DES2008].

Gewerbe, Handel und Dienstleistungen - GHD

Die Entwicklung der Wirtschaft in Deutschland hat in den letzten Jahrzehnten einen starken Wandel erlebt. Der Wandel zeigt sich durch die Veränderung der Beschäftigtenstruktur. Noch vor ca. 40 Jahren war ein Großteil der Beschäftigten in Deutschland im Industriesektor tätig. Heutzutage entfällt in Deutschland ein Beschäftigungsanteil von ca. 75 % auf den GHD-Sektor, bei Betrachtung der kreisfreien Städte sogar ein Anteil von fast 80 % [DES2014A] [DES2015E]. Mit ca. 70 % ist auch der Anteil des GHD-Sektors an dem BIP in etwa gleich groß [WiZ2015]. Das zeigt, dass die Entwicklung in Deutschland sich in Richtung des tertiären⁹ Sektors bewegt und dieser Sektor bereits jetzt für die Funktion der Wirtschaft nicht wegzudenken ist. Die Unterteilung der Sektoren Industrie und GHD erfolgt prinzipiell nach dem erzeugten Output. Die Industrie schafft ein materielles Gut, in den Subsektoren des GHD-Sektors wird zumeist ein Output erzeugt der in den meisten Fällen keinen materiellen Wert beinhaltet. Die Subsektoren des GHD werden grob in drei Bereiche zusammengefasst und sind nachfolgend aufgelistet [DES2014A]:

- Handel, Verkehr, Gastgewerbe, Information und Kommunikation, Erziehung, Gesundheit
- Finanz-, Versicherungs-, Unternehmensdienstleistungen, Grundstücks und Wohnungswesen
- Öffentliche und sonstige Dienstleistungen, Erziehung, Gesundheit

⁹ Primärer Sektor = Land- und Forstwirtschaft sowie Fischerei; sekundärer Sektor = produzierendes Gewerbe (Industrie); tertiärer Sektor = Dienstleistungssektor (GHD-Sektor)

3 Voruntersuchung – Anwendung und Analyse diverser Clusteranalysemethoden

Dieses Kapitel umfasst die Anwendung der verschiedenen Fusionsalgorithmen der hierarchisch agglomerativen Clusteranalyse. Zunächst wird kurz die Datenbasis erläutert. In dem darauf folgenden Unterkapitel werden die Algorithmen angewendet und die Unterschiede erläutert. Abgeschlossen wird dieses Kapitel durch die Gegenüberstellung der Ergebnisse, wo die in [WAL2014] ermittelte 7-Clusterlösung als Vergleichskriterium herangezogen wird.

3.1 Datenbasis für die Voruntersuchung

Die Datenbasis für diese Untersuchung, die ein Methodenscreening der in Frage kommenden Clusteranalysemethoden darstellt, wird in diesem Unterkapitel kurz erläutert. Der Datensatz, der hier verwendet wird, existiert bereits und wurde in [WAL2013], [EIC2014] und [WAL2014] ausführlich diskutiert und publiziert, weshalb an dieser Stelle auf diesen nicht erneut im Detail eingegangen wird. Dieser umfasst 16 deutsche Städte (siehe Abschnitt 2.5.1), die mit jeweils 12 Indikatoren beschrieben wurden. Die 12 Indikatoren sind der nachfolgenden Tabelle 3-1 zu entnehmen. Dort sind die Indikatorwerte den Städten gegenübergestellt. Auch für die Untersuchung des geringeren Untersuchungsrahmens von lediglich 12 Indikatoren wurden die vier Verbrauchssektoren (Haushalte, Verkehr, Industrie, GHD) mit je vier Indikatoren (Haushalte, Verkehr) bzw. zwei Indikatoren (Industrie, GHD) betrachtet.

Tabelle 3-1: Für das Methodenscreening verwendete Indikatoren

Private Haushalte	Verkehr	Industrie	GHD
Bevölkerungsentwicklung	Arbeitslosenquote	Anteil der Industrieumsatzsteuer an Gesamtumsatzsteuer	Anteil der GHD-Umsatzsteuer an Gesamtumsatzsteuer
Verfügbares Einkommen der privaten Haushalte	Einwohnerdichte	Energieverwendung je Mio.-€-BWS	Endenergieverbrauch je Mio.-€-BWS
Heiztage	Kraftfahrzeugdichte	---	---
Anteil Einpersonenhaushalte	Fahrten mit ÖPNV	---	---

Die Indikatoren beschreiben die Städte durch deren individuelle Ausprägung stadtspezifisch. Die reine Betrachtung der Indikatoren erlaubt es, sofern einige wenige Indikatoren betrachtet werden, die Städte zu analysieren. Beispielsweise wäre es möglich die Städte auf Basis der Werte eines Indikators zu sortieren. Würden zwei bzw. drei Indikatoren betrachtet werden, so wäre die Darstellung im zwei- bzw. dreidimensionalen Raum möglich. Diese Darstellung würde es erlauben über das Betrachten der so entstandenen Punktwolken eine unvollständige (geometrische) Clusteranalyse durchzuführen. Da zur Klassifizierung der Städte alle 12 Indikatoren betrachtet werden (multivariat) benötigen wir andere Verfahren.

Die Datenvorbereitung für die multivariate Clusteranalyse erfolgte auch im Rahmen dieser Analyse durch die Standardisierung der Variablenwerte mithilfe der Z-Transformation sowie

der Faktorenanalyse. Die Z-Transformation transformiert die Indikatorwerte auf ein einheitliches Wertenniveau, was für belastbare statistische Untersuchungen notwendig ist, um allen Indikatoren im mathematischen Sinn die gleiche Bedeutung zukommen zu lassen [WAL2014]. Die Faktorenanalyse greift die transformierten Indikatorwerte auf und extrahiert daraus statistisch sinnvolle Faktoren. Ein Faktor ist dabei ein statistisches Konstrukt (künstliche Variable, Abschnitt 2.2.3), da die Faktorwerte nicht gemessen werden können, sondern aus den Indikatorwerten, die dieser abbildet, errechnet werden. Die detaillierte Durchführung der Faktoranalyse wurde in [WAL2013] und [WAL2014] beschrieben. Eine tiefergehende Beschreibung der Z-Transformation und der Faktoranalyse erfolgt in Kapitel 5, wo diese auf den Datensatz aller kreisfreien Städte angewendet werden.

Wie im vorherigen Kapitel gezeigt wurde, eignet sich für den Anwendungsfall der Untersuchung die hierarchische agglomerative Clusteranalyse. Da dieses Verfahren einen großen Anteil in der Clusteranalyseanwendung einnimmt, gibt es eine Vielzahl verschiedener Methoden und Fusionsalgorithmen. Die Recherche zeigt, dass einige Algorithmen für diesen Untersuchungsfall nicht geeignet sind. Die Algorithmen, die in Frage kommen, werden in den nachfolgenden Abschnitten angewendet. Um die Ergebnisse besser zu deuten, werden diese den verwendeten Algorithmen gegenübergestellt und ausgewertet, um die Wahl der Methode für die Durchführung der Hauptclusteranalyse aller deutschen kreisfreien Städte zu identifizieren und zu begründen.

3.2 Clusteranalyse mit gängigen, hierarchischen agglomerativen Algorithmen

Die hierarchische agglomerative Clusteranalyse sowie die wesentlichen Fusionsalgorithmen, die in diesem Verfahren ihre Anwendung finden, wurden bereits vorgestellt (Abschnitt 2.3.2). Nähere Informationen zu der Herleitung geben die Arbeiten von [BAC2010], [BAC2016], [STE1977] sowie [EVE2011]. In diesem Unterkapitel werden vier Algorithmen angewendet, um eine Aussage bezüglich der Eignung für die Untersuchung des Datensatzes aller deutschen kreisfreien Städte herzuleiten.

3.2.1 Single-Linkage-Verfahren (nächster Nachbar)

Das Single-Linkage Verfahren ist einer der simpelsten Algorithmen zur Durchführung der clusteranalytischen Betrachtung von Datensätzen. [STE1977] Das Single-Linkage Verfahren wird als „nächstgelegener Nachbar bezeichnet“, da hierbei die beiden Objekte (Cluster, Nachbarn) zu einem Cluster fusioniert werden, die die geringste mathematische Distanz vorweisen. In Abbildung 3-1 ist die Zusammenführung exemplarisch und dimensionslos skizziert.

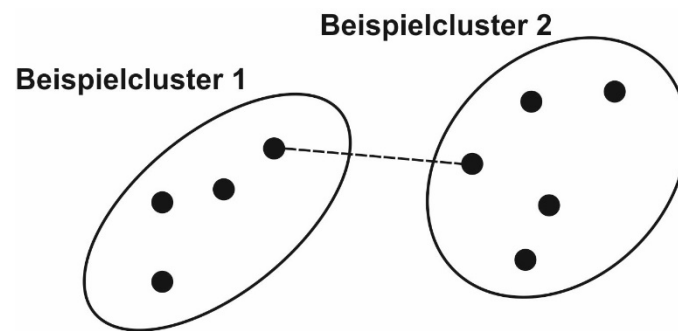


Abbildung 3-1: Single-Linkage (nächster Nachbar) – schematische Darstellung

Durch die Art dieses Fusionsalgorithmus eignet er sich durch das simple Vorgehen bei der Fusion Ausreiser zu entdecken. Die Cluster, die durch das Single Linkage erzeugt werden, neigen zu einer Kettenbildung, da die fusionierten Objekte diejenigen sind, die zumindest zu einem anderen Objekt die geringste Distanz aufweisen. Dieser Umstand führt dazu, dass dieser Algorithmus dazu neigt, kettenförmige Cluster zu erzeugen (Kontraktionseffekt). [STE1977] [BAC2010] Wird dieser Algorithmus auf den Datensatz der 16 Städte und 12 Indikatoren angewendet so ergibt sich ein ungleichmäßiger Pfad der Zusammenführung. In Abbildung 3-2 ist das Dendrogramm (Baumdiagramm) der durchgeführten Clusteranalyse mit dem Single-Linkage dargestellt.

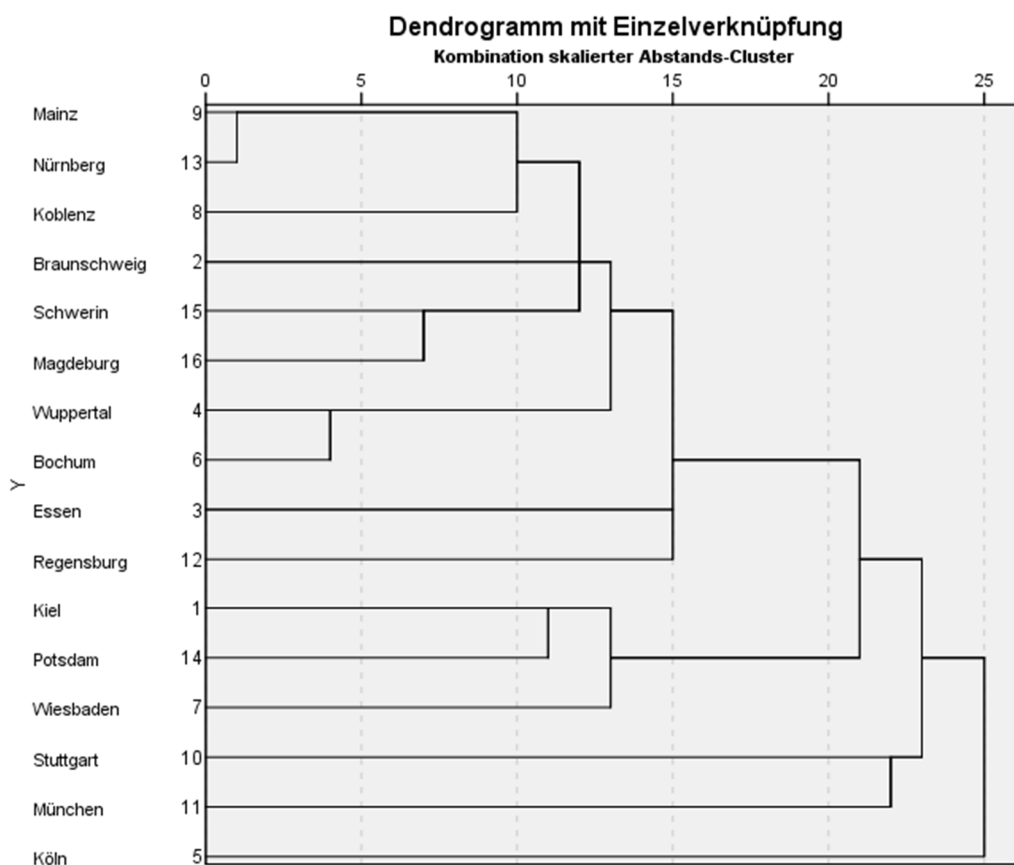


Abbildung 3-2: Dendrogramm der Fusionsschritte – Single-Linkage

In der Form des Dendrogramms des Single-Linkage Algorithmus lässt sich der Verkettungseffekt erkennen. Dies zeigt sich dadurch, dass die Städte, in diesem Fall die Stadt Köln, als

„Ausreiser“ als letzte Stadt fusioniert wird. Da dieser Algorithmus keinen großen Anspruch auf Intraclusterhomogenität legt, werden hierbei Objekte und Cluster zusammengefasst, die durch eine minimale Distanz zu mindestens einem Objekt in bereits bestehenden Clustern die Regel des Algorithmus erfüllen und somit nicht als eigenständiges Cluster identifiziert werden können. Hierdurch werden die Objekte in weiteren Schritten nicht als Cluster betrachtet und deshalb auch weiterhin als einzelne Clusterelemente bzw. eigenständige Objekte den nächsten Fusionsschritten zugeführt. Starke Ausreißer werden bei diesem Algorithmus immer erst am Ende fusioniert, woran die Eignung dieses Algorithmus Ausreißer zu identifizieren deutlich wird. Ebenfalls ist dieser Algorithmus der einzige, der Clusterformen erkennen kann (z. B. langgezogene Cluster). Bei einer 7-Clusterlösung würden wir für diesen Algorithmus die Lösung in Tabelle 3-2 erhalten.

Tabelle 3-2: 7-Clusterlösung mit Single-Linkage

Cluster	Städte
1	Kiel, Wiesbaden, Potsdam
2	Braunschweig, Wuppertal, Bochum, Koblenz, Mainz, Nürnberg, Schwerin, Magdeburg
3	Essen
4	Köln
5	Stuttgart
6	München
7	Regensburg

Das Clusterergebnis ist im Sinne der Clustergrundidee nicht zufriedenstellend, zeigt jedoch, dass besonders die Städte Köln, Stuttgart und München erst spät fusionieren. Dadurch wird angedeutet, dass diese Städte „Ausreißer“ darstellen könnten. Es ist bei der Anwendung des Single-Linkage nicht möglich eine belastbare Clusterlösung zu erhalten, da die Städte Stuttgart und München auf einem hohen Fusionierungsniveau (s. Abbildung 3-2, Dimensionsloses Distanzmaß bei ca. 22) zusammengeführt werden. Somit bilden beide Städte einfach nur zwei Objekte, die, verglichen mit den übrigen Objekten, besonders verschiedene Ausprägungen aufweisen. Lediglich die Stadt Köln wird als letzte Stadt mit dem „Restcluster“, in dem die anderen 15 Städte enthalten sind, vereint (Distanzmaß 25). In diesem Fall würde die Stadt Köln als „Ausreiser“ identifiziert werden. In einer endgültigen Clusteranalyse würde die Elimination dieses „Ausreißers“ die Clusterlösung verbessern. Da hier jedoch nur die Ergebnisse der verschiedenen Algorithmen vorgestellt werden, wird darauf verzichtet. Das Ergebnis zeigt, dass das Single-Linkage als Voruntersuchung angewendet eine Hilfe bei der Elimination von Ausreißern darstellen kann, z. B. in der Analyse in [SCH2004].

Bei einer vollständigen Clusteranalyse mit dem Single-Linkage würden andere Clusterlösungen als die hier gewählte 7-Clusterlösung als sinnvolle Clusterlösung gewählt werden. Auf die Darstellung der vollständigen Single-Linkage Clusteranalyse und -lösung wird an dieser

Stelle¹⁰ jedoch verzichtet, da hierbei die ermittelte 7-Clusterlösung aus der Arbeit [WAL2014] Algorithmen übergreifend verglichen wird.

3.2.2 Complete-Linkage-Verfahren (entferntester Nachbar)

Der zweite Algorithmus der „Nächste-Nachbarn Verfahren“ ist das Complete-Linkage. Hierbei werden die maximalen Distanzen zwischen Objekten und Clustern berechnet. Anders als bei dem Single-Linkage werden hier nur die Objekte zu einem Cluster vereint, die die geringste maximal Distanz zu den Objekten in einem Cluster besitzen (s. Abbildung 3-3). Die Distanz muss dabei die geringste Distanz zu allen Objekten eines Clusters sein. D. h. ein Objekt das fusioniert werden soll, muss zeitgleich nächster Nachbar zu allen bereits im Cluster enthaltenen Objekten sein. Das ist der Grund weshalb dieser Algorithmus sehr homogene Gruppen bildet, jedoch auch bei großen Objektanzahlen dazu neigt sehr viele kleine Gruppen zu generieren (Dilatationseffekt). [STE1977] [BAC2010] [SCH2010]

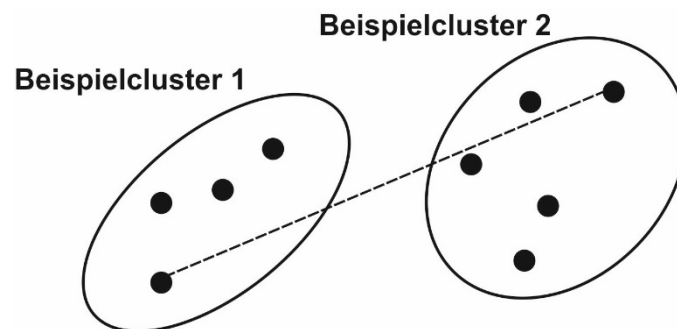


Abbildung 3-3: Complete-Linkage (entferntester Nachbar) – schematische Darstellung

Wie bereits erwähnt neigt der Algorithmus dazu kleine Gruppen zu bilden, wodurch das Problem entstehen kann, dass große Cluster zerstört werden können. Vorteilhaft hierbei ist, dass dieser Algorithmus weniger anfällig gegenüber Ausreißern ist, als das zuvor diskutierte Single-Linkage. [STE1977] [BAC2010] [SCH2010]

In dem Dendrogramm des Complete-Linkage (Abbildung 3-4) ist die Bildung kleiner Gruppen dadurch zu erkennen, dass auf dem Niveau der dimensionslosen Fusionskennzahl von ca. 12 die 7-Clusterlösung (s. Tabelle 3-3) vorzufinden ist. Zeitgleich hat dieser Algorithmus bereits zu sehr frühen Fusionszeitpunkten mehrere kleine Gruppen gebildet, die erst relativ spät zu größeren Gruppen vereint werden. Das zeigt die Fusionsregel dieses Algorithmus. Dieser Algorithmus ist sehr konservativ, liefert gute Cluster, eignet sich jedoch aufgrund der besagten Probleme für große Objektmengen eher nicht.

¹⁰ Bei den weiteren Algorithmen (Complete-, Average-Linkage und Ward) wird ebenfalls auf die Darstellung der Clusterauswahl verzichtet. Die Auswahl der hier betrachteten 7-Clusterlösung erfolgte ausführlich in [WAL2014]. Eine ausführliche Erläuterung des Verfahrens der Wahl der Clusteranzahl erfolgt in Kapitel 5.

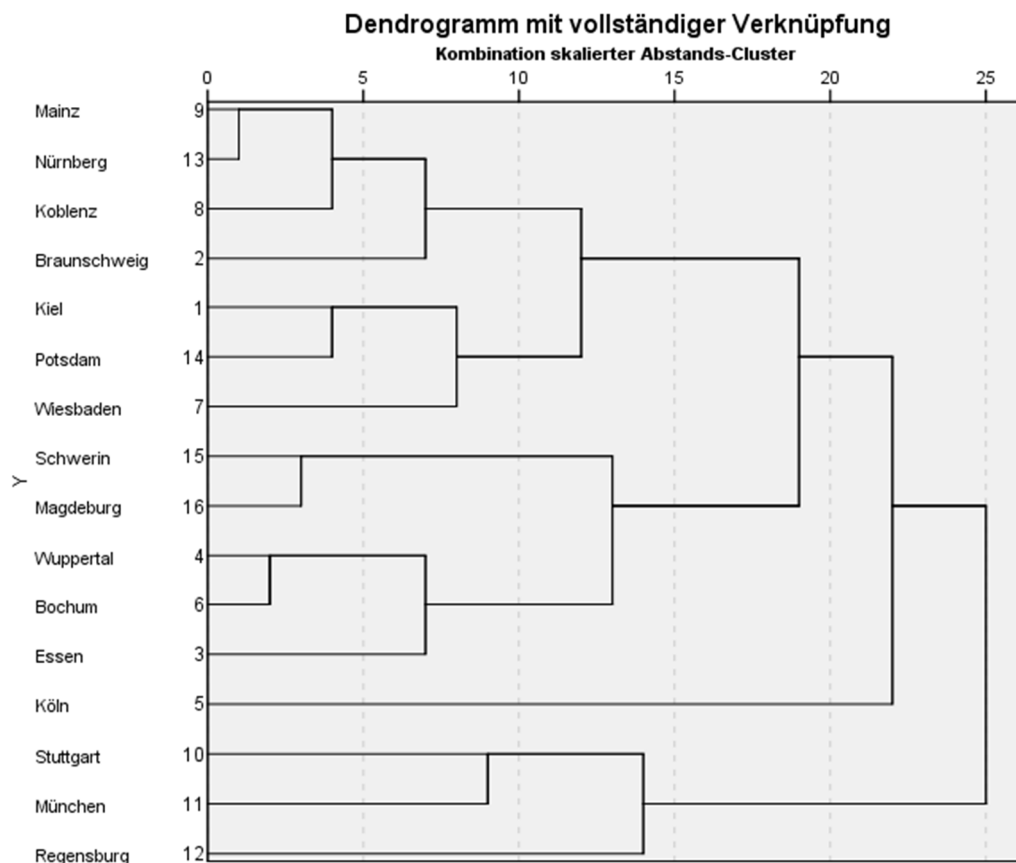


Abbildung 3-4: Dendrogramm der Fusionsschritte – Complete-Linkage

Tabelle 3-3: 7-Clusterlösung mit Complete-Linkage

Cluster	Städte
1	Kiel, Wiesbaden, Potsdam
2	Braunschweig, Koblenz, Mainz, Nürnberg
3	Essen, Wuppertal, Bochum
4	Köln
5	Stuttgart, München
6	Regensburg
7	Schwerin, Magdeburg

Das Dendrogramm sowie die Zuordnung der Städte zeigen, dass die Cluster 1, 2, 3, 5 und 7 bereits bei einer Distanz zwischen 5 und 10 gebildet werden. Auch bei diesem Algorithmus bleibt die Stadt Köln fast bis zuletzt als Objekt ein eigenes Cluster. Für Clusteranalysen, bei denen der verstärkte Fokus auf möglichst homogenen Clustern liegt, eignet sich das Complete-Linkage ideal und wäre somit zu favorisieren [STE2011]. Da die Fusion zu größeren Clustern, jedoch erst auf hohen Niveaus erfolgt, ist dieser Algorithmus zur Clusterung deutscher kreisfreier Städte eher als nicht vorteilhaft zu bewerten, da größere Städtegruppen von diesem Algorithmus nicht bzw. fehlerhaft identifiziert werden.

3.2.3 Average-Linkage-Verfahren (zwischen den Gruppen)

Den dritten hier näher betrachteten Algorithmus stellt das Average-Linkage-Verfahren dar. Hierbei wird anders als bei den vorherigen beiden Algorithmen, die mittlere Distanz aller im Cluster enthaltenen Objekte berechnet. Somit treten bei diesem Algorithmus zeitgleich alle in einem Cluster enthaltenen Objekte in die Berechnung der Distanz- und Ähnlichkeitsmaße ein, wodurch die so ermittelten Cluster dann einer belastbaren Clusteranalyse näher kommen, da nicht nur lediglich ein Objekt (kürzeste bzw. größte Distanz) betrachtet wird. Fusioniert werden in dieser Form der Clusteranalyse diejenigen Objekte oder Cluster, deren mittlere Distanz die geringste aller ermittelten Distanzen darstellt. In Abbildung 3-5 ist schematisch durch Linien die Distanzberechnung dargestellt. Alle Distanzen werden dabei aufaddiert und durch die Anzahl der möglichen Distanzen dividiert.

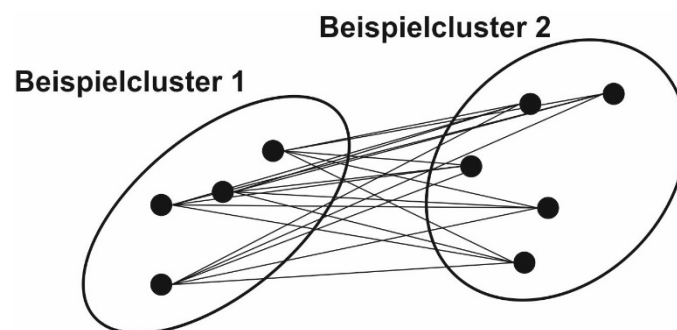


Abbildung 3-5: Average-Linkage (zwischen den Gruppen) – schematische Darstellung

Dieser Algorithmus stellt auch in gewisser Weise eine Fortführung der beiden erstgenannten Algorithmen dar, nur dass hierbei die Schwächen der ersten beiden Algorithmen durch die Mittelwertbildung abgeschwächt werden. [STE2011] [BAC2010]

In der folgenden Abbildung 3-6 ist das Dendrogramm zu diesem Algorithmus dargestellt. Wird dieses Dendrogramm mit dem des vorherigen Complete-Linkage verglichen so fällt auf, dass sich beide ähneln. Auch hier führt der Algorithmus zu derselben Zuordnung der Städte zu den sieben Clustern, wie schon zuvor bei dem Complete-Linkage (siehe dazu Tabelle 3-4). Bei einer größeren Objektanzahl würden die Unterschiede zwischen den beiden Algorithmen jedoch deutlicher ausfallen. Es wird vorgeschlagen das Average-Linkage bzw. die weiteren Algorithmen der Mittelwertverfahren (Weighted- oder Within-Average-Linkage) zu verwenden.

Dieser Algorithmus findet auch in drei der betrachteten 23 Arbeiten Anwendung. Damit wird er deutlich häufiger als das Complete- und Single-Linkage zusammen verwendet. Das Average-Linkage ist die simpelste Form der Algorithmen der Mittelwertverfahren. Es wird in [BAC2010] darauf verwiesen, dass die Average-Algorithmen besonders dann angewendet werden sollen, wenn die Verwendung der quadrierten-euklidischen Distanz nicht sinnvoll bzw. möglich ist. Diese ist für das Ward und k-Means Verfahren notwendig. Bei metrischen Skalenniveaus ist der Ward-Algorithmus anwendbar, wodurch dieser häufig benutzt wird (s. Tabelle 2-3).

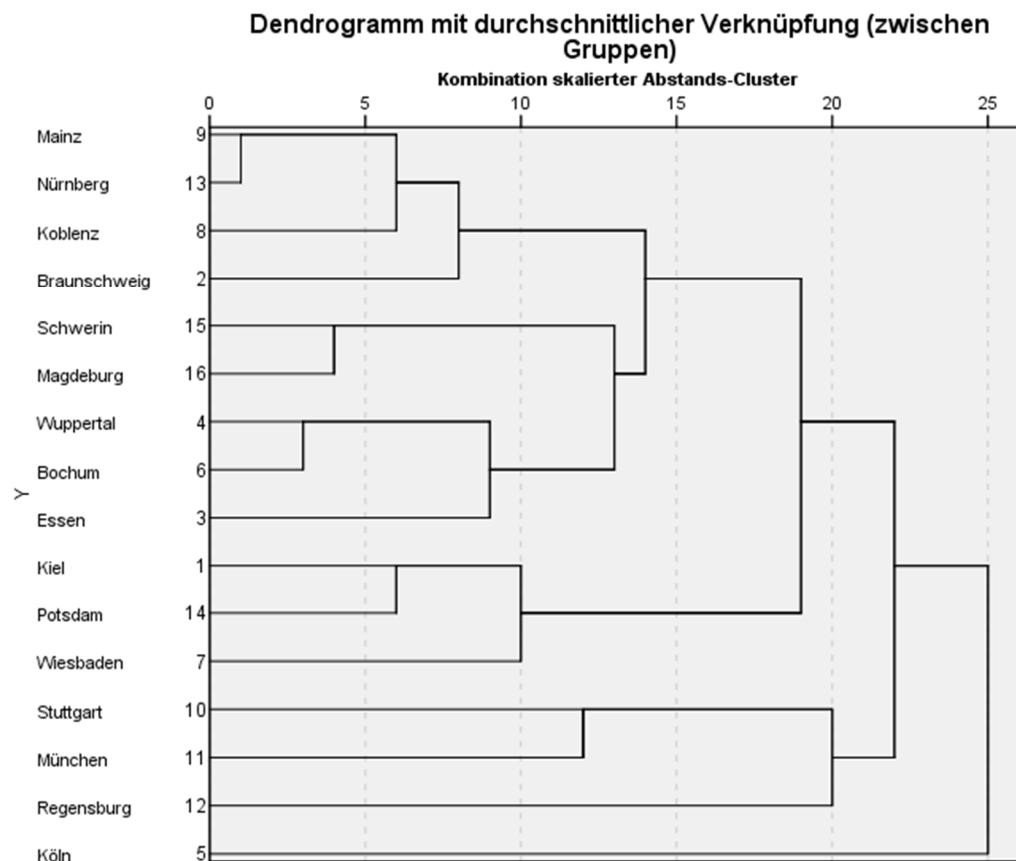


Abbildung 3-6: Dendrogramm der Fusionsschritte – Average-Linkage

Tabelle 3-4: 7-Clusterlösung mit Average-Linkage

Cluster	Städte
1	Kiel, Wiesbaden, Potsdam
2	Braunschweig, Koblenz, Mainz, Nürnberg
3	Essen, Wuppertal, Bochum
4	Köln
5	Stuttgart, München
6	Regensburg
7	Schwerin, Magdeburg

3.2.4 Ward-Verfahren

Der letzte hier betrachtete Algorithmus ist der, der in der Anwendung am häufigsten eingesetzt wird. Der Ward Algorithmus ist anders als die zuvor vorgestellten und angewendeten Algorithmen, bei denen die Distanz für die Clusterfusion herangezogen wird, einer, bei dem die Clusterbildung mithilfe von Clusterzentren erfolgt. Dabei werden die Verfahren Centroid und Median, die ebenfalls Clusterzentren als Repräsentanten verwenden weggelassen, da diese aufgrund von auftretenden Fehlern (Inversionen = Umkehreffekte) vermieden werden sollten und der Vorzug dem Ward-Algorithmus erteilt wird [BAC2010]. Die Fusionierung läuft dabei so ab, dass die Objekte, die zu einem Cluster fusioniert werden, die Varianz (Streuung) in dem Cluster möglichst gering erhöhen. Dadurch liefert dieser Algorithmus sehr homogene Cluster und

wird auch als „*leistungsstärkstes Verfahren unter den hierarchisch-agglomerativen Verfahren* [STE2011]“ bezeichnet. Detaillierte Regeln sowie die Herleitung dieses Algorithmus können [Bac2010], [BAC2016], [STE1977], [SCH2010], [STE2011] entnommen werden.

Das Varianzkriterium (Fehlerquadratsumme), nach dem die Objekte und Cluster gruppiert werden, berechnet sich mit nachfolgender Gleichung 3-1:

$$V_g = \sum_{k=1}^{K_g} \sum_{j=1}^J (x_{kjg} - \bar{x}_{jg})^2 \quad \text{Gleichung 3-1}$$

Die Varianz einer Gruppe (V_g) berechnet sich aus der Summe der quadrierten Differenzen zwischen den zu betrachtenden und zu fusionierenden Objekten (x_{kjg}) sowie den Mittelwerten der jeweiligen Ausprägungen in der betrachteten Gruppe ($\bar{x}_{jg}$). Dabei kennzeichnen die Indices j die jeweilige Variable (Indikator/Faktor) eines Objekts k der betrachteten Gruppe g . Bei den Mittelwerten wird nicht zwischen den Objekten unterschieden, da hierbei die Mittelwerte der Variablen über alle einer Gruppe angehörigen Objekte ermittelt werden.

Die Tabelle 3-5 stellt die Clusterlösung dar, die mit dem Ward Algorithmus ermittelt wurde. Diese Clusterlösung wurde bereits in [WAL2014] veröffentlicht und ausführlich beschrieben.

Tabelle 3-5: Clusterlösung mit Ward

Cluster	Städte
1	Kiel, Wiesbaden, Potsdam
2	Braunschweig, Schwerin, Magdeburg
3	Essen, Wuppertal, Bochum
4	Köln
5	Koblenz, Mainz, Nürnberg
6	Stuttgart, München
7	Regensburg

Eine höhere Aussagekraft als die Darstellung der Clusterlösung in einer Tabelle bietet das Dendrogramm der Clusteranalyse. Hier können die Niveaus, auf denen eine Clusterfusion erfolgt, nachverfolgt werden. Der Ward Algorithmus neigt dazu gleichgroße Cluster zu erzeugen, die jedoch aufgrund der Verwendung der Intracustervarianz sehr homogen sind. Die folgende Abbildung 3-7 zeigt diesen Effekt. Dort tritt die sieben Clusterlösung bereits bei einer dimensionslosen Distanz von ca. 5 auf. Die schnelle Fusion der sieben Clusterlösung zeigt, dass möglichst ideale Cluster erzeugt wurden, deren Varianz gering ist und diese dementsprechend homogen sind. Zudem sind die erzeugten Cluster heterogen zueinander. Das wird in Abbildung 3-7 durch lange waagrechte Linien rechts der dimensionslosen Distanz von 5 dargestellt. Würden geringere Clusteranzahlen als Lösung gewählt, dann würden dementsprechend größere Intracustervarianzen als Folge hervorgehen.

Die Wahl des Ward Algorithmus zur Clusterermittlung ist heutzutage das am weitesten verbreitete Vorgehen bei der hierarchisch agglomerativen Clusteranalyse. Die Tabelle 2-3 in Unterkapitel 2.4 zeigt dies. In 12 der 23 betrachteten Arbeiten verschiedener Forschungsgebiete wird dieser Algorithmus zur Clusterermittlung verwendet. Dabei werden aussagekräftige Ergebnisse für verschiedene Untersuchungsgrößen (Objekt- und Variablenanzahl) ermittelt.

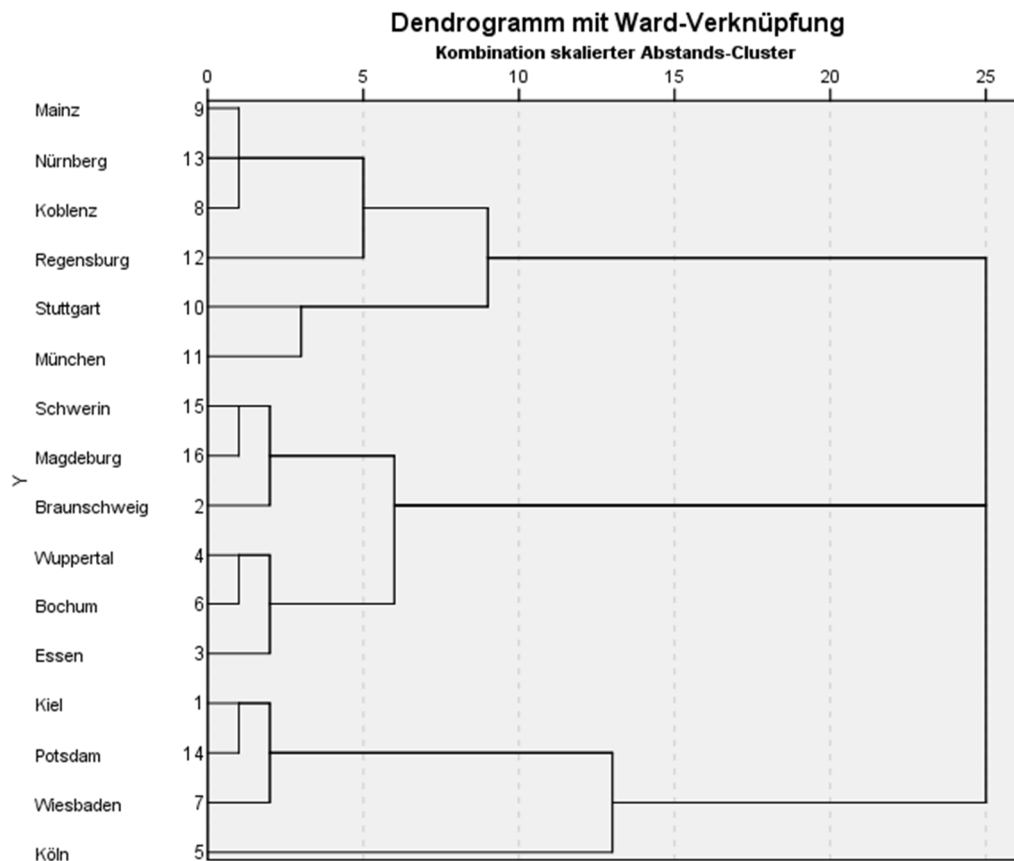


Abbildung 3-7: Dendrogramm der Fusionsschritte – Ward Verknüpfung

3.3 Gegenüberstellung, Analyse und Auswertung der Clusterergebnisse

In diesem Unterkapitel werden die zuvor diskutierten Ergebnisse gegenübergestellt, um direkte Vergleiche zwischen den Ergebnissen der verschiedenen Fusionsalgorithmen aufzeigen und deren Aussagekraft einheitlich quantifizieren zu können. Die Gegenüberstellung erfolgt indem die jeweiligen sieben Clusterlösungen der vier betrachteten Fusionsalgorithmen einheitlich in einer Tabelle aufgetragen werden. Die Dendrogramme hingegen werden nicht gegenübergestellt, da dadurch die Aussagekraft durch deren erneute Darstellung nicht verstärkt werden würde. Es ist zu erkennen, wie sich die Clusterzusammensetzungen in Abhängigkeit des Algorithmus verhält. Die Tabelle 3-6 beinhaltet die zuvor erwähnte Gegenüberstellung.

Tabelle 3-6: Gegenüberstellung der Ergebnisse der untersuchten Fusionsalgorithmen

	Single-Linkage	Complete-Linkage	Average-Linkage	Ward
Distanz	Ca. 15	Ca. 12	Ca. 13	5
Cluster 1	Kiel, Potsdam, Wiesbaden	Kiel, Potsdam, Wiesbaden	Kiel, Potsdam, Wiesbaden	Kiel, Potsdam, Wiesbaden
Cluster 2	Braunschweig , Wuppertal, Mainz, Bochum, Schwerin, Nürnberg, Koblenz, Magdeburg	Braunschweig , Koblenz, Mainz , Nürnberg	Braunschweig , Koblenz, Mainz , Nürnberg	Braunschweig , Schwerin , Magdeburg
Cluster 3	Essen	Essen, Bochum , Wuppertal	Essen, Bochum , Wuppertal	Essen, Bochum , Wuppertal
Cluster 4	Köln	Köln	Köln	Köln
Cluster 5	Stuttgart	Stuttgart , München	Stuttgart , München	Koblenz, Mainz , Nürnberg
Cluster 6	München	Regensburg	Regensburg	Stuttgart , München
Cluster 7	Regensburg	Schwerin , Magdeburg	Schwerin , Magdeburg	Regensburg

In der Gegenüberstellung sind die Städte für die verschiedenen Algorithmen nach den ermittelten sieben Clustern aufgeteilt. Einige Städte sind dabei Fett bzw. Fett und Kursiv. Diese Kennzeichnung wurde gewählt, um bei der Auswertung der Tabelle eine Einheitlichkeit und bessere Nachverfolgbarkeit zu erzeugen. Das einzige Cluster, das sich unabhängig von der Clusterzusammensetzung über die vier Algorithmen nicht unterscheidet, ist das Cluster 1. Dort sind die Städte Kiel, Potsdam und Wiesbaden enthalten und Fett markiert. Fett markiert sind die Städte, da mindestens die Hälfte der Algorithmen diese Clusterzusammensetzung unter Cluster 1 zusammenfügen. Ein ähnliches Bild zeigt sich für das Cluster 4, das lediglich durch die Stadt Köln gebildet wird. Hieran zeigt sich der Vorteil einer vorgeschalteten Single-Linkage Clusteranalyse. Das Single-Linkage hat bereits ermittelt, dass die Stadt Köln einen Ausreißer darstellt. Würde dieser nach dem Single-Linkage eliminiert und dann bspw. das Ward Verfahren angewendet werden, so würde der Untersuchungsrahmen zwar um ein Objekt reduziert, die Clusterlösung hingegen würde jedoch stabiler und aussagekräftiger werden. Ferner zeigt die Gegenüberstellung, dass die Ergebnisse des Single-Linkage die größten Abweichungen zu den Ergebnissen anderer Algorithmen aufweisen (viele Clusterlösungen weder Fett noch Kursiv). Im Cluster 2 stellt die Stadt Braunschweig ein konstantes Element dar. Die weiteren Clusterelemente sind über die Clusterlösungen des Cluster 2 verschieden. Jedoch lassen sich in diesen Teillösungen erkennen, die in anderen Clusterlösungen als einzelne Cluster abgebildet sind, so z. B. die Städte Koblenz, Mainz und Nürnberg bei dem Average- und Complete-Linkage in Cluster 2 sowie in Cluster 5 bei dem Ward Verfahren. Ähnliches lässt sich für die Städte Schwerin und Magdeburg erkennen. Diese sind (Fett und Kursiv) vom Ward Verfahren zusammen mit der Stadt Braunschweig dem Cluster 2 zugeordnet, werden jedoch vom Complete- und Average-Linkage dem Cluster 7 zugewiesen. Ähnlich verhält es sich mit der Stadt

Regensburg, die zwar in den Algorithmen verschiedenen Clustern zugeordnet wird jedoch immer als einzelne Stadt ein eigenes Cluster bildet. Werden die Clusterlösungen des Single-Linkage etwas außen vorgelassen und nur die Ergebnisse der übrigen drei Algorithmen untersucht, so zeigen sich zwei Clusterlösungen (Essen, Bochum, Wuppertal und Stuttgart, München), die zwar in verschiedenen Clustern fusioniert sind, die Clusterzusammensetzung hingegen identisch ist.

Werden die Ergebnisse miteinander verglichen, so fällt auf, dass die Ergebnisse des Complete- und des Average Linkage identisch sind und sogar die Cluster in derselben Reihenfolge hervorgegangen sind. Dies hängt mit der hier verwendeten geringen Objektanzahl zusammen. Wird hingegen die Objektanzahl vergrößert, ist davon auszugehen, dass bei dem Complete Linkage viele kleinere Cluster ermittelt werden würden. Zudem wird dieses Auftreten dadurch gefördert, dass nicht mit jedem Algorithmus eine vollständige Clusteranalyse durchgeführt wurde, sondern die sieben Clusterlösung aus [WAL2014] als Vergleichsbasis gewählt werden. Dennoch zeigt dieser Vergleich, dass bei dem Anwenden der sieben Clusterlösung auf die vier Algorithmen, der Ward-Algorithmus mit die besten Ergebnisse liefert, da die betrachtete sieben Clusterlösung (bei Betrachtung des Dimensionslosen Distanzmaßes) wesentlich früher konstruiert wird, als es bei den Linkage Verfahren der Fall ist.

Nach den Untersuchungen und Auswertungen in den Kapiteln 2 und 3 wird in Kapitel 5 mit der Anwendung des sich als aussagekräftigsten behaupteten Ward-Algorithmus zur Clusterung der im Untersuchungsrahmen befindlichen deutschen kreisfreien Städte verfahren.

4 Datenerhebung für betrachtete kreisfreie Städte

Das Ziel dieses Kapitels ist es eine umfangreiche Datenbasis zu generieren, auf die sich die durchgeführte Clusteranalyse im weiteren Verlauf stützt. Eine Clusteranalyse stützt sich auf Indikatoren, die als Variablen ein betrachtetes Objekt, im Falle dieser Arbeit eine Stadt, beschreiben. Ein Indikator ist dabei für jede Stadt verschieden, wodurch der Kennwert, den der Indikator beschreibt als stadtspezifisch zu bewerten ist. Auf die grundlegenden Definitionen und Ausprägungen von Indikatoren wird hier zunächst kurz eingegangen. Den Hauptteil dieses Kapitels stellen die Indikatoren der vier Verbrauchssektoren dar.

4.1 Definition, Formen, Konzept und Datenbasis von Indikatoren

Der Ursprung des Wortes Indikator findet sich im Lateinischen - *indicare*. Indicare bedeutet „anzeigen“ oder auch „bezeichnen“, wodurch der Nutzen von Indikatoren deutlich wird. Mithilfe erhobener oder beobachteter Kenngrößen zeigen diese einen bestimmten Sachverhalt an und bieten ferner die Möglichkeit durch einen Vergleich diesen zu bewerten. Indikatoren werden zudem in fast allen Fällen durch einen Quotienten gebildet (siehe z. B. Gleichung 4-1).

$$\text{Bevölkerungsdichte} = \frac{\text{Bevölkerung}}{\text{Fläche}} \quad \text{Gleichung 4-1}$$

Im allgemeinen Sprachgebrauch werden Indikatoren daher meist als Kennwerte oder Verhältniszahlen bezeichnet. [SAN2002] Indikatoren sind darüber hinaus in der Lage Sachverhalte auf eine Modellebene zu heben die, reduziert auf Zahlen, andere Möglichkeiten der Beurteilung, Auswertung und des Vergleichs der durch den Indikator beschriebenen Objekte bietet. In dieser Arbeit wurde der Ansatz verfolgt für eine breite Indikatorbasis erhebbare Kennwerte zu wählen, um die betrachteten Städte einer Top-Down Untersuchung zu unterziehen. Der Begriff Top-Down beschreibt ein häufig verwendetes Planungsvorhaben, das aus der Wirtschaft entstammt. Die wörtliche Bedeutung des Top-Down Begriffes ist „von Oben nach Unten“, oder „vom Groben zum Detail“. Dabei wird ein Ziel formuliert, ohne zeitgleich das exakte Vorgehen sowie detaillierte Schritte festzulegen. Der Brückenschlag zur Clusterbildung erfolgt in diesem Zusammenhang, dass das „grobe“ Ziel feststeht (Ermittlung energetisch vergleichbarer Städtecluster) und dieses ausgearbeitet werden soll mit einer breiten erhebbaren Basis an Indikatoren und den durch diese aufgezeigten Zusammenhängen. [PRO2015]

4.1.1 Formen von Indikatoren

Alleinstehende Indikatoren/ Einzelindikatoren

Als alleinstehende Indikatoren werden Indikatoren bezeichnet, die einen gesonderten präzisen Ausdruck einer Größe angeben. Diese Form von Indikatoren ist selten. Als Beispiel könnte die CO₂ Emission eines Unternehmens als alleinstehender Indikator dienen, jedoch wird auch dieser sehr häufig auf eine weitere Größen bezogen wie z. B. ein Jahr oder einen Monat. Dadurch

entsteht ein Einzelindikator, der dann wiederum eine weitere Verwendung, wie z. B. eine Gegenüberstellung oder einen Vergleich mit vorherigen Zeiträumen oder Unternehmen erlaubt.

Indikatoren mit Bezug zueinander/ Zusammengesetzte Indikatoren

Zusammengesetzte Indikatoren erfordern, wie aus dem Namen hervorgeht, mindestens zwei Größen aus denen diese zusammengesetzt werden. Theoretisch können beliebig viele Indikatoren zusammengesetzt werden, jedoch geht mit einer beliebigen Erweiterung der Einzelindikatoranzahl ein Informationsverlust einher, da der Sachverhalt unklar wird. [THE2007]

Nach [DIE2012] finden zusammengesetzte Indikatoren oft ihre Anwendung im politischen Bereich, wo viele komplexe Einzelindikatoren durch ein Zusammenführen zu einem Indikator zusammengefasst werden. Zusammengeführte Indikatoren bilden zumeist die Basis für die Durchführung von Benchmarks für Länder [DIE2012], z. B. der HDI (Human Development Index), bei dem Indikatoren wie die Lebenserwartung, die Ausbildung und auch das Bruttoinlandsprodukt einige der wichtigen Säulen darstellen [BPB2008].

Bereinigte Indikatoren

Um den Einfluss von Unregelmäßigkeiten auf Indikatoren zu verhindern, ist es bei manchen Indikatoren möglich eine Bereinigung der Werte durchzuführen. Die Aussagekraft und Genauigkeit eines Indikators kann erhöht werden, indem bekannte externe Einflüsse aus den Indikatorwerten eliminiert werden. Als Beispiel kann an dieser Stelle die Temperaturbereinigung oder die Inflationsbereinigung aufgeführt werden [PEH2010]. Die Temperaturbereinigung bezieht sich auf Temperaturunterschiede, die anders ausfallen als der Durchschnitt. Ausreißer werden mithilfe zusätzlicher Faktoren aus den Werten eliminiert, wodurch die so bereinigten Indikatorwerte einen Vergleich mit vorherigen Jahren oder zukünftigen Jahren erlauben, wenn diese z. B. auf den Gebäudeenergiebedarf bezogen werden. [PEH2010]

Energieeffizienzindikatoren

Die Energieeffizienz ist ein, nicht zuletzt durch die Diskussion um die Energiewende, sehr großes Thema. Die Energieeffizienz muss überprüfbar und nachvollziehbar sein. An dieser Stelle werden Energieeffizienzindikatoren verwendet. Energieeffizienzindikatoren sind Indikatoren, die so konstruiert werden, dass sie es vermögen, die Energieeffizienz mittelbar oder unmittelbar durch Kenngrößen zu beschreiben. Fragestellungen, wie die Beschreibung einer Stadt, müssen möglichst vielschichtig erfolgen, wodurch eine Vielzahl an Indikatoren notwendig ist. Diese Indikatorsammlung kann unmöglich nur mit direktem Energiebezug erfolgen, sodass auch Indikatoren aus anderen Wissenschaftsfeldern wie z. B. Sozialindikatoren, ökonomische Indikatoren und Umweltindikatoren unbedingt hinzugezogen werden müssen. [DIE2012]

4.1.2 Konzept und Datenbasis

Die Erstellung eines Datensatzes erfordert in erster Linie belastbare Daten. Um den Datensatz möglichst nachvollziehbar und reproduzierbar zu gestalten wurden ausschließlich nationale und frei verfügbare Daten verwendet. Die Beschreibung von Städten erfordert stadspezifische Kenngrößen, die in unterschiedlich gutem Umfang verfügbar sind. Die Kenngrößen mit denen Städte beschrieben werden können, können auf *weichen* und *harten* Daten basieren. Weiche Daten werden durch Beobachtungen oder Befragungen erhoben. Harte Daten sind mathematisch-statistische Daten, die Messungen, Erhebungen oder Fortschreibungen entstammen. Die für diese Arbeit verwendeten Daten sind in fast allen Fällen harte Daten, da es keine Datenbasis für sämtliche deutsche kreisfreie Städte gibt, die beobachtete Werte in zuverlässiger Genauigkeit und gleichem Umfang beinhalten. Lediglich einige Indikatoren aus dem Bereich Verkehr basieren auf weichen Daten, die z. B. von Verkehrsunternehmen durch Befragungen erhoben und vom Verband deutscher Verkehrsunternehmen (VDV) zu einer Statistik zusammengeführt werden [VDV2015].

Die Datenverfügbarkeit, besonders im energetischen Bereich, ist auf kommunaler Ebene limitiert. Energetische Daten sind zumeist in einem größeren Umfang nur auf Landesebene oder Bundesebene verfügbar. Ein Punkt der jedoch alle Ebenen betrifft ist die Aktualität der Daten. Es dauert eine gewisse Zeit bis alle Information in einer Stelle zusammenkommt, sodass z. T. einige Daten erst mit mehrjähriger Verzögerung verfügbar sind. [DIE2012: 11] Wiederum andere Daten werden nicht jährlich erhoben, sodass die bisher verfügbaren Daten in unregelmäßigen Abständen veröffentlicht werden, z. B. Mobilitätsdaten im ÖPNV. Aus diesem Grund kann der erarbeitete Datensatz nicht auf ausschließlich ein Jahr bezogen werden sondern einen Zeitraum von einigen wenigen Jahren, da die Indikatoren z. T. verschiedene Bezugsjahre aufweisen.

Für die Erarbeitung des Indikatorsatzes wurde durch eine breitangelegte Recherche ein Indikatorsatz erarbeitet, der Daten aus diversen Einrichtungen vereint. Ein Großteil der Daten entstammt dem Statistischen Bundesamt, den Statistischen Landesämtern sowie dem Zensus. Weiterhin wurden Daten von weiteren Bundesämtern, wie z. B. dem Kraftfahrtbundesamt, hinzugezogen. Verbände als auch Organisationen sowie einige Unternehmen bilden den letzten Informationsbereich.

In den nachfolgenden drei Unterkapiteln werden die ermittelten Indikatoren, aufgeteilt nach den betrachteten Verbrauchssektoren, vorgestellt. Die Indikatoren werden dafür zunächst kurz beschrieben. Für die Nachvollziehbarkeit der gewählten Indikatoren werden Details zur Datenlage, dem Bezugsjahr, der Information zur Datenrecherche und der Berechnung der jeweiligen Indikatoren erläutert. Diese detaillierte Darstellung ist notwendig, um einem der gesetzten

Ziele, der Reproduzierbarkeit der Ergebnisse, nachkommen zu können und dies sicherzustellen. Insgesamt wurden 41 Indikatoren erhoben, mit denen die 107 kreisfreien Städte dieser Betrachtung beschrieben werden. Somit wird jede Stadt durch 41 Indikatorwerte beschrieben, wodurch ein sehr umfangreicher Datensatz (4.387 Kennwerte) für die breitangelegte, mathematisch-statistische clusteranalytische Betrachtung geschaffen werden konnte. Aufgrund des großen Umfangs des Datensatzes wird darauf verzichtet die Indikatorwerte unmittelbar nach dem Beschreiben der Indikator abzubilden. Die Indikatoren werden gesammelt, jedoch eingeteilt nach den Sektoren, im Anhang dieser Arbeit abgebildet.

4.2 Indikatoren des Verbrauchssektors private Haushalte

In diesem Unterkapitel werden die ermittelten Indikatoren für den Sektor *private Haushalte* vorgestellt. Die Indikatoren werden jeweils kurz beschrieben sowie deren Datenlage und Berechnung erläutert. Dieser Sektor ist für das Erreichen der Klimaschutzziele sehr bedeutend. In ihm steckt ein enormes Potenzial, besonders im Bereich der thermischen Energieverwendung und –versorgung. Nicht selten wird das Potenzial in der Wärme und Warmwasserbereitung für die privaten Haushalte als „*schlafender Riese*“ bezeichnet [IHK2015: 5], da besonders in diesem Sektor die erst in letzter Zeit verstärkt forcierte Wärmewende noch ein enormes Potenzial bereithält. Die für diesen Sektor erhobenen 15 Indikatoren werden nachfolgend aufgelistet und beschrieben. Die Indikatorwerte sind in Tabelle A-1 im Anhang aufgeführt.

- Bevölkerungsentwicklung der letzten elf Jahre
- Wohnfläche pro Person
- Anteil Einpersonenhaushalte
- Durchschnittliche Haushaltsgröße
- Haushaltsdichte
- Pro-Kopf Stromverbrauch
- Erdgasverbrauch
- Anteil Wohnungen in 1-2 Familienhäusern
- Verfügbares Einkommen
- Anteil über 65-Jähriger
- Arbeitslosenquote bezogen auf die abhängigen zivilen Erwerbspersonen
- Anteil der Siedlungs- und Verkehrsfläche an Gesamtfläche
- Heiztage Mittel der Jahre 2009-2013
- Heizgradtage Mittel der Jahre 1970-2013
- Gradtagzahlen Mittel der Jahre 1970-2013

Bevölkerungsentwicklung der letzten 11 Jahre

Der erste Indikator stellt die *Bevölkerungsentwicklung* der untersuchten Stadt dar. Die Feststellung einer Entwicklung erfordert jedoch den Bezug auf einen Zeitraum, um eine Veränderung gegenüber einem festgelegten Ausgangszeitpunkt feststellen zu können. Im vorliegenden Fall wurde die Entwicklung der letzten elf Jahre für den Zeitraum von 2000 bis 2011 ermittelt. Die Tabelle 4-1 stellt die Kenndaten zum Indikator dar.

Tabelle 4-1: Indikatorkenndaten – „Bevölkerungsentwicklung der letzten elf Jahre“

Bevölkerungsentwicklung der letzten elf Jahre	
Aussage	Die Entwicklung der Bevölkerung von 2000-2011
Bezugsjahr	2011
Einheit	Prozent [%]
Quelle	Statistisches Bundesamt [DES2013D]

Die in [DES2013D] bereitgestellten Daten können in verschiedener Detailtiefe abgerufen werden. Die für diese Untersuchung benötigten Daten sind: *Bevölkerung: Kreise, Stichtag*.

Der Indikator ist der Quotient aus Bevölkerungsveränderung zum Bezugsjahr und der Bevölkerung im Ausgangsjahr. Dieser Indikator ist einer von vielen demographischen Indikatoren, die gleichermaßen einen Einfluss auf den Energieverbrauch [DIE1999] und die Umsetzung energieeffizienzsteigernder Maßnahmen haben. Die Beachtung der Bevölkerungsentwicklung ist unabdingbar für die Planung und Entwicklung von Städten, da viele erarbeitete kommunale Konzepte für Städte mit positiver und negativer Bevölkerungsentwicklung differenziert gestaltet werden müssen [EIC2014: 49]. Ein Beispiel für diesen Sachverhalt kann in dem Ausbau der Fernwärme gefunden werden. Ein längerer effektiver Nutzen herrscht dort vor, wo die Auslastung auch auf lange Sicht bestehen bleibt. In einer Stadt mit negativem Wachstum ist dies schwer vorstellbar. [Wis2013], [DIF2011]

Wohnfläche pro Person

Die *Wohnfläche pro Person* stellt den zweiten Indikator dar. Dieser Indikator vermag eine Aussage ableiten zu lassen, wie zum einen die Verhältnisse auf dem Wohnungsmarkt sind und zum anderen wie die Wohnqualität ist [BER2015B]. Über die Jahre entwickelt sich ein Trend in Richtung immer größer werdender Wohnungen. Dieser Trend lässt sich zum Teil auch durch die immer kleiner werdenden Haushalte (s. Indikator: *Einpersonenhaushalte*) sowie die sinkende durchschnittliche Haushaltsgröße der deutschen Haushalte (s. Indikator: *Durchschnittliche Haushaltsgröße*) erklären. Die nachfolgende Tabelle 4-2 beinhaltet einige Kenndaten zu dem Indikator Wohnfläche pro Person.

Tabelle 4-2: Indikatorkenndaten – „Wohnfläche pro Person“

Wohnfläche pro Person	
Aussage	Durchschnittliche Wohnfläche pro Person in m ²
Bezugsjahr	2011
Einheit	Quadratmeter [m ²]
Quelle	Wegweiser Kommune [BER2015A]

Die Daten entstammen dem Wegweiser Kommune der Bertelsmann Stiftung [BER2015A]. Das Bezugsjahr ist das Jahr 2011, das Jahr des letzten erhobenen Zensus. Die Einheit ist der Quadratmeter [m²]. Der Indikator zeigt die durchschnittliche Wohnfläche pro Person. Eine

große Wohnfläche pro Person geht in den meisten Fällen mit einem Anstieg des Energiebedarfs einher, da eine größere Fläche zu beheizen, zu beleuchten etc. ist. Ähnlich sind auch die Auswirkungen bei den beiden nächsten beschriebenen Indikatoren.

Anteil Einpersonenhaushalte und durchschnittliche Haushaltsgröße

Der Energieverbrauch der Privathaushalte wird unter anderem durch demographische Einflussfaktoren bestimmt. Dazu zählt neben dem *Anteil der Einpersonenhaushalte an den Gesamthaushalten*, die die kleinste Form des Haushalts bilden, auch die durchschnittliche Haushaltsgröße, die eine Abschätzung der Haushaltsstruktur ermöglicht.

Tabelle 4-3: Indikatorkenndaten – „Anteil Einpersonenhaushalte“

Anteil Einpersonenhaushalte	
Aussage	Anteil der Einpersonenhaushalte an den Gesamthaushalten in krfr. Städten
Bezugsjahr	2011
Einheit	Prozent [%]
Quelle	Regionalstatistik - Zensusdatenbank [DES2015F]

Die Kenndaten zu den Einpersonenhaushalten sind in Tabelle 4-3 aufgetragen. Die Daten zur Berechnung der Indikatorwerte entstammen der Zensusdatenbank und haben das Bezugsjahr 2011 [BER2015A]. Der Anteil ist der Quotient aus der Anzahl der Einpersonenhaushalte und der Anzahl der Gesamthaushalte multipliziert mit 100.

Der nächste Indikator ist die *Durchschnittliche Haushaltsgröße*. Die Kennwerte dazu sind in Tabelle 4-4 dargestellt. Auch bei diesem Indikator ist das Bezugsjahr das Jahr 2011 und die Datenbasis die Zensusdatenbank [DES2015F]. Die Einheit ist hier eine Anzahl, die durch eine Verhältniszahl aus der Anzahl der beim Zensus erhobenen Einwohnerzahl einer jeden Stadt und der Anzahl der Haushalte derselben Stadt gebildet wird.

Tabelle 4-4: Indikatorkenndaten – „Durchschnittliche Haushaltsgröße“

Durchschnittliche Haushaltsgröße	
Aussage	Durchschnittliche Haushaltsgröße in krfr. Städten
Bezugsjahr	2012
Einheit	Personen/Haushalt [Anzahl]
Quelle	Regionalstatistik - Zensusdatenbank [DES2015F]

Beide Indikatoren weisen einen gewissen Zusammenhang auf. Die Bevölkerungszahl ist seit 2003 auf einen Wert von 80,2 Millionen gesunken. Zeitparallel ist die Zahl der Haushalte in Gesamtdeutschland gestiegen [DES2011]. Dieser Verlauf kann besonders gut verdeutlicht werden, wenn die Entwicklung über viele Jahre hinweg betrachtet wird. In Tabelle 4-5 ist die *Mittlere Haushaltsgröße* in verschiedenen Jahren seit 1900 aufgetragen. Es zeigt sich, dass die Entwicklung stets zu kleineren Haushaltsgrößen hin zu erkennen ist. Derzeit liegt die durchschnittliche Haushaltsgröße in Gesamtdeutschland bei ca. 2,1 Personen je Haushalt, in den kreisfreien Städten sogar nur bei knapp 2 Personen je Haushalt. Diese niedrigen Haushaltsgrößen sind auch den Einpersonenhaushalten geschuldet, die besonders in den letzten

Jahrzehnten enorm an Zuwachs erhalten haben. So stellt die kleinste Haushaltsgruppe, die Einpersonenhaushalte, in den kreisfreien Städten schon heute mit ca. 45,6 % die am häufigsten vertretene Haushaltsgruppe dar.

Tabelle 4-5: Entwicklung der mittleren Haushaltsgröße

Jahr	Mittlere Haushaltsgröße [Personen/Haushalt] aus [MEY2013]
1900	4,5
1918	4
1945	3
2000	2,2

Ein hoher Anteil an Einpersonenhaushalten sowie eine niedrige Haushaltsgröße lässt den Gesamtenergiebedarf der deutschen Haushalte tendenziell steigen. Alleinlebende Personen verbrauchen im Schnitt 2.050 kWh elektrischer Energie im Jahr. Der Pro-Kopf Verbrauch an elektrischer Energie fällt in einem Zwei-Personenhaushalt um ca. 16 %, bei einem Vier-Personenhaushalt sogar um ca. 42 % niedriger aus [BDE2014].

Haushaltsdichte

Die *Haushaltsdichte* ist ein Wesentlicher Indikator für die energetische Beschreibung von Kommunen. Im Zuge der forcierten Dezentralisierung der energetischen Versorgung von Kommunen, sind Kommunen mit verschiedenen Haushaltsdichten differenziert zu betrachten. Eine Kommune mit hoher Haushaltsdichte bietet höhere Potenziale z. B. in der Realisierung von Nahwärmenetzen, die bspw. mit BHKW betrieben werden können und somit neben der KWK auch die benötigte Energie möglichst lokal erzeugen. Die Kenndaten zu diesem Indikator sind in Tabelle 4-6 aufgetragen.

Tabelle 4-6: Indikatorkenndaten – „Haushaltsdichte“

Haushaltsdichte	
Aussage	Anzahl Haushalte pro km ² Gemarkungsfläche
Bezugsjahr	2012
Einheit	Haushalte pro Quadratkilometer [HH/km ²]
Quelle	Statistisches Bundesamt [DES2015F], [DES2015G]

Die Haushaltsdichte gibt einen Einblick in die Haushaltsstruktur einer Kommune. Bei der reinen Betrachtung der Haushaltszahlen oder der Bevölkerungsgröße kann festgestellt werden, wie die Größenverhältnisse in Städten sind. Es kann jedoch nicht abgeschätzt werden, wie groß die Auswirkung einer Maßnahme sein könnte, da ohne diese Verhältniszahl nicht die Anzahl auf einer bestimmten Fläche bestehender Haushalte bekannt ist.

Pro-Kopf Stromverbrauch

Der Sektor private Haushalte trägt mit ca. 27 % einen großen Anteil zum gesamten Stromverbrauch in Deutschland bei [UBA2015]. Trotz der Bestrebungen den Stromverbrauch zu verringern ist der Trend ungebrochen und über die Zeit von 1990 bis zu dem Jahr 2013 nahm der Stromverbrauch der privaten Haushalte um ca. 18 % zu [UBA2015]. In diesem Zuge wurde

von dem Internetvergleichsportal *Preisvergleich.de AG* eine anonyme Auswertung der Eingabe der Stromwechsler im Wesentlichen aus deutschen Großstädten durchgeführt. Dafür wurden stichprobenartig 120.000 Online-Stromvertragsabschlüsse im Zeitraum von 2011-2012 untersucht. Die Daten der durchschnittlichen Haushaltsgrößen stammen von den statistischen Landesämtern und des statistischen Bundesamtes. [WEI2014] [PRE2014] Die Ergebnisse wurden veröffentlicht, sodass es zumindest für die untersuchten Städte relativ genaue Stromverbrauchsgrößen je Person gibt, die für die hier durchgeführte Clusteranalyse ideale Kennwerte darstellen. [PRE2014] Durch diese Werte des *Pro-Kopf-Stromverbrauchs* kann analysiert werden, in welchen der betrachteten Städte der Stromverbrauch besonders hoch ist und Einsparungen erreicht werden, bzw. in welchen Städten bereits wenig Strom verbraucht wird und diese Städte u. U. als Vorbilder dienen können.

Tabelle 4-7: Indikatorkennndaten – „Pro-Kopf Stromverbrauch“

Pro-Kopf-Stromverbrauch	
Aussage	Verbrauchte Menge elektrischen Stroms pro Person
Bezugsjahr	2013
Einheit	Kilowattstunde [kWh]
Quelle	Preisvergleich.de [PRE2014]

Die Maßzahl des Kennwertes stellt eine durchschnittliche Größe dar und kann in der veröffentlichten Form bereits verwendet werden. Lediglich einige der hier betrachteten kreisfreien Städte wurden nicht in den von [PRE2014] veröffentlichten Städten abgedeckt. Für diese wurden, soweit möglich, über Anfragen der Werte bei den Kommunen die Datenlücken geschlossen. Bei den Städten, bei denen die Anfragen keine Daten hervorbrachten wurden die Lücken mit dem Durchschnitt der gesamten Stromverbrauchswerte geschlossen, um die Städte weiterhin in der Betrachtung berücksichtigen zu können.

Erdgasverbrauch

Der *Erdgasverbrauch* ist der zweite direkte energetische Verbrauchsindikator dieser Untersuchung. Dieser sagt den Erdgasverbrauch je Quadratmeter Wohnfläche und Jahr aus. Das Bezugsjahr ist das Jahr 2012 mit dem Stichtag 31.12.2012. Die Werte entstammen der Auswertung in der Veröffentlichung der Techem Energie Services GmbH [TEC2013]. Die Kennwerte werden bereits temperaturbereinigt angegeben, sodass eine Vergleichbarkeit der Werte untereinander sichergestellt werden kann. Die Tabelle 4-8 beinhaltet die Kennndaten zum Indikator Erdgasverbrauch.

Tabelle 4-8: Indikatorkennndaten – „Erdgasverbrauch“

Erdgasverbrauch	
Aussage	Erdgasverbrauch pro Quadratmeter und Jahr
Bezugsjahr	2012
Einheit	Kilowattstunde pro Quadratmeter und Jahr [kWh/m ² *a]
Quelle	Techem Energie Services GmbH [TEC2013]

Ein negativer Aspekt ist bei diesem Indikator die Datenlage. Auf der einen Seite sind die Daten sehr belastbar und haben eine sehr hohe Aussagekraft, da sie tatsächlich aus erhobenen Heizenergieverbräuchen der Bewohner der Städte bestehen. Auf der anderen Seite sind diese Daten im Wesentlichen vom Umfang auf kreisfreie Großstädte limitiert, wodurch nicht für alle kreisfreien Städte Indikatorwerte zur Verfügung stehen. Es besteht eine große Anzahl an Lücken (ca. 20 %) der Daten, die ebenfalls durch Mittelwerte geschlossen wurden, um auch diesen Indikator vorerst für die clusteranalytische Untersuchung verwenden zu können.

Anteil Wohnungen in 1 und 2-Familienhäusern

Der *Anteil der Wohnungen in 1 und 2-Familienhäusern* wurde aufgrund seiner Aussagekraft in Bezug auf die Anzahl der Wohngebäude, die zu großen Teilen Privatpersonen gehören und von diesen selbst bewohnt werden als Indikator gewählt. 1 und 2-Familienhäuser werden in den meisten Fällen vom Eigentümer bewohnt. Der Sanierungsstatus ist in 1 und 2-Familienhäusern oft besser als in Mehrfamilienhäusern, da kein Mieter-Vermieter Dilemma auftritt [INW2011]. Das Mieter-Vermieter Dilemma entsteht, wenn der Eigentümer Investitionen (z. B. Energetische Sanierung) durchführt, die Vorteile in Form von Einsparungen jedoch bei dem Mieter bleiben [INW2011]. Zudem vermag dieser Indikator einen Überblick über die bauliche Struktur einer Stadt geben [BER2015B].

Tabelle 4-9: Indikatorckenndaten – „Anteil Wohnungen in 1 und 2-Familienhäusern“

Anteil Wohnungen in 1 und 2-Familienhäusern	
Aussage	Anteil der Wohnungen an den gesamten Wohnungen einer Stadt, die in 1 und 2-Familienhäusern liegen [BER2015B]
Bezugsjahr	2010
Einheit	Prozent [%]
Quelle	Wegweiser Kommune [BER2015A]

Die Tabelle 4-9 beinhaltet die Kennwerte dieses Indikators. Die Differenzen in den Indikatorwerten sind zum Teil sehr verschieden. So weisen einige Städte Anteile von über 50 % auf. Andere wiederum liegen im niedrigen zweistelligen Bereich. Die Werte dieses Indikators sind einheitlich und für alle Städte des Betrachtungsrahmens abrufbar.

Verfügbares Einkommen

Das *verfügbare Einkommen der Haushalte* ist ein ökonomischer Indikator, der einen Überblick über die Einkommensstruktur einer Stadt zu geben vermag. Dieser Indikator steht häufig im Zusammenhang mit dem Indikator „*Anteil Wohnungen in 1 und 2-Familienhäusern*“. Der Zusammenhang entsteht dadurch, dass Menschen, die in Wohnungen in 1 und 2-Familienhäusern Wohnen, oftmals ein höheres verfügbares Einkommen je Haushalt zur Verfügung haben. Die Bevölkerung muss zum Erreichen der Klimaschutzziele mit einbezogen werden und das verfügbare Einkommen hängt kausal mit der Investitionsbereitschaft und der Fähigkeit energieeffizienzsteigernde Maßnahmen umzusetzen zusammen. Ein höheres verfügbares Ein-

kommen senkt die ökonomische Hürde Energieeffizienzmaßnahmen besonders im Wohngebäudebereich umzusetzen. [WAL2013] Die nachfolgende Tabelle 4-10 beinhaltet die Kenndaten des Indikators „Verfügbares Einkommen“.

Tabelle 4-10: Indiktorkennndaten – „Verfügbares Einkommen“

Verfügbares Einkommen	
Aussage	Das Verfügbare Jahreseinkommens eines Haushalts
Bezugsjahr	2011
Einheit	Euro [€]
Quelle	Statistisches Bundesamt [DES2014c]

Dieser Indikator stellt die Einkommensverhältnisse im Bundesgebiet dar. So sind die Differenzen ebenfalls z. T. sehr hoch. Das höchste verfügbare Einkommen je Haushalt findet sich in Heilbronn (ca. 71.000 Euro) wieder. Das niedrigste Einkommen mit ca. 28.000 Euro weisen die Haushalte in Rostock auf. Somit liegt auch hier die Spannweite der Indikatorwerte in etwa in derselben Größenordnung wie bei dem vorherigen Indikator.

Anteil über 65-Jähriger

Der *Anteil über 65-Jähriger* ist besonders in Deutschland aktuell, denn dieser hängt direkt mit dem demografischen Wandel zusammen. Sinkende Geburtenraten und das „Älterwerden“ der geburtenstarken Jahre verstärken diese Entwicklung besonders.

Der hohe Anteil der Bevölkerung im Alter von über 65-Jahren ist bereits heute bei der Planung der Energieversorgung zu beachten. Auch in dem Bereich der Gebäudesanierung erfordert dies ein anderes Herangehen, da die Beweggründe für z. B. eine energetische Gebäudesanierung hier anders betrachtet werden müssen. Ein weiterer Finanzkredit für eine energetische Sanierung wird in einem solchen Alter tendenziell vermieden bzw. eher in einen altersgerechten Umbau investiert. Bereits jetzt erfolgt ein immer stärkerer Fokus auf das „Älterwerden“ der deutschen Bevölkerung. In den sozialen Bereichen wird schon heute die Forschung angegangen, inwiefern die Sanierungsoptionen (energetisch oder altersgerecht) im Konflikt zueinander stehen. [EHF2015]

Vor diesem Hintergrund erfordert eine allumfassende Betrachtung sozio-energetischer Indikatoren auch einen Indikator, der diesen demografischen Aspekt abdeckt. Die nachfolgende Tabelle 4-11 beinhaltet Kenndaten zu dem Indikator Anteil 65-Jähriger.

Tabelle 4-11: Indiktorkennndaten – „Anteil über 65-Jähriger“

Anteil über 65-Jähriger	
Aussage	Anteil der über 65-Jährigen an der Gesamtbevölkerung der Stadt
Bezugsjahr	2011
Einheit	Prozent [%]
Quelle	Zensus [ZEN2015]

Im Durchschnitt liegt der Anteil der über 65-Jährigen in den betrachteten kreisfreien Städten bei ca. 20 %. In den nächsten Jahren kann von einer sukzessiven Zunahme dieses Anteils

ausgegangen werden, wodurch besonders in Deutschland der demografische Wandel auch in Zukunft eine aktuelle Herausforderung darstellen wird und auch den Bereich der Energieverwendung und –versorgung betreffen wird.

Arbeitslosenquote bezogen auf die abhängigen zivilen Erwerbspersonen

Die *Arbeitslosenquote bezogen auf die abhängigen zivilen Erwerbspersonen* ist ein weiterer Indikator des Verbrauchssektors Haushalte. Die Arbeitslosenquote hat dabei auch eine Aussagekraft im Verbrauchssektor Verkehr, wo angenommen wird, dass in Städten mit hohen Arbeitslosenquoten ein geringeres Verkehrsaufkommen in den modalen Mobilitätsbereichen ÖPNV und MIV auftritt [Eic2014]. Im Haushaltsbereich zeigt sich eine erhöhte Arbeitslosigkeit in höheren Anwesenheitszeiten in der Wohnung, woraus beispielsweise längere Heizzeiten sowie erhöhte elektrische Energieverbräuche (z. B. Unterhaltungselektronik) resultieren. Die nachfolgende Tabelle 4-12 stellt die Kenndaten zu diesem Indikator dar.

Tabelle 4-12: Indikatorkenndaten – „Arbeitslosenquote“

Arbeitslosenquote bezogen auf die abhängigen zivilen Erwerbspersonen¹¹	
Aussage	Arbeitslose bez. auf zivile Personen in erwerbstätigen Alter ohne Selbstständige
Bezugsjahr	2013
Einheit	Prozent [%]
Quelle	Statistisches Bundesamt [DES2014D]

Die saisonbereinigte Arbeitslosenquote liegt in Deutschland derzeit bei ca. 5 % (Stand: Mai 2015) [EUR2015]. Die Betrachtung der deutschen kreisfreien Städte zeigt hingegen eine Spanne zwischen 4,0 und 16,3 %. Die niedrigste Arbeitslosigkeit ist in Ingolstadt (4,0 %) festzustellen. Die höchste Arbeitslosigkeit des Betrachtungsrahmens weist die Stadt Bremerhaven (16,3 %) auf. In den Daten für diesen Indikator sind nur die abhängigen zivilen Erwerbspersonen berücksichtigt worden. Diese Form der Datenberechnung und Darstellung der Arbeitslosenquote hat in Deutschland eine erhöhte Verbreitung, da diese traditionell in der Erhebung und Berichterstattung verwendet wird [BFA2015].

Anteil der Siedlung- und Verkehrsfläche an Gesamtfläche

Die Fläche stellt in den ökonomisch entwickelten Ländern ein knapp werdendes Gut dar [GOT2008]. Durch den Ausbau von Verkehrsinfrastruktur und von Wohn- und Arbeitsraum wird immer mehr der Gesamtfläche zerschnitten und versiegelt. So kommt es bereits jetzt schon in manchen deutschen kreisfreien Städten zu sehr hohen Anteilen in diesem Indikator. Die Stadt Herne weist mit 77,8 % den höchsten Anteil an bebauter Fläche an der Gesamtfläche auf. Die Tendenz zeigt, dass je größer eine Stadt ist, umso höher ist der Anteil der bebauten und somit versiegelten Fläche. Die Tabelle 4-13 zeigt die Kenndaten des Indikators.

¹¹ Abhängige zivile Erwerbspersonen sind Personen im Erwerbstätigenalter, die von einem Arbeitgeber oder dem Staat abhängig sind. D. h. Selbstständige sind in dieser Statistik nicht enthalten. [BFA2015]

Tabelle 4-13: Indikatorkenndaten – „Anteil Siedlungs- und Verkehrsfl. an Gesamtfläche“

Anteil der Siedlungs- und Verkehrsfläche an Gesamtfläche	
Aussage	Anteil der Siedlungs- und Verkehrsfläche an der Gesamtfläche
Bezugsjahr	2012
Einheit	Prozent [%]
Quelle	Statistisches Bundesamt [DES2014E]

Die Flächendaten können für viele Bereiche nützlich sein. Städte mit hohen bebauten Flächenanteilen haben beispielsweise ein geringeres Potenzial für den stadtinternen Ausbau von z. B. der Windenergie. Städte mit geringen bebauten Flächenanteilen hingegen haben z. T. – sofern sich die Standorte aufgrund der Windhöflichkeit sowie den Naturschutzbelangen eignen – die Möglichkeit Windenergiekonverter auf dem Stadtgebiet zu installieren und den Anteil der Erneuerbaren Energien „rechnerisch“ stark anzuheben. Als gutes Beispiel kann die Stadt Wolfhagen aufgeführt werden, die dies durch den Bau von vier Windenergiekonvertern auf dem verhältnismäßig großen Stadtgebiet verdeutlicht. [WAL2016A]

Heiztage, Heizgradtage und Gradtagzahlen

Die letzten drei Indikatoren des Verbrauchssektors Haushalte sind die *Heiztage* im Mittel der Jahre 2009-2013, die *Heizgradtage* im Mittel der Jahre 1970-2013 sowie die *Gradtagzahlen* im Mittel der Jahre 1970-2013. Die drei Begriffe Heiztage, Heizgradtage sowie Gradtagzahlen sind Begriffe aus dem Hauswärmebereich.

Den ersten Indikator bilden die Heiztage. Der Heiztag (HT) ist definiert als Tag, an dem die mittlere Tagesaußenlufttemperatur unter 15 °C¹² [VDI2013], d. h. unter der als Heizgrenztemperatur definierten Größe, liegt [IWU2015] [Eic2014]. Heiztage weisen eine Aussagekraft über die regionale Witterung sowie die klimatischen Bedingungen einer Stadt auf, was sich besonders aus den Gebäudewärmeenergiebedarf auswirkt [Eic2014]. Auf die Heiztage kann kaum Einfluss genommen werden, da dieser natürliche Indikator klimatisch bedingt ist, dennoch in großem Maße Einfluss auf das Heizverhalten in den privaten Haushalten nimmt.

Die folgende Tabelle 4-14 stellt die Kenndaten des Indikators Heiztage dar. Bei diesem Indikator erfolgt der Bezug auf eine Zeitspanne von fünf Jahren. Die beiden folgenden Indikatoren basieren auf längeren Zeitreihen. Der Gedanke hinter der Wahl der verschiedenen Zeitreihen war bei der Clusteranalyse zu prüfen, ob eine Abhängigkeit der Aussage der drei Indikatoren von der Zeitreihe besteht.

Tabelle 4-14: Indikatorkenndaten – „Heiztage Mittel der Jahre 2009-2013“

Heiztage Mittel der Jahre 2009-2013	
Aussage	Anzahl der Tage mit mittlerer Außenlufttemperatur unter 15 °C
Bezugsjahr	Mittelwert 2009-2013
Einheit	Anzahl HT [-]
Quelle	Institut Wohnen und Umwelt [IWU2015]

¹² 15 °C bei normalem Gebäudebestand, 12 °C bei einem Neubau oder saniertem Altbau, 10 °C bei einem Passivhaus [IWU2015], [VDI2013]

Die Heizgradtage (HGT) und die Gradtagzahlen (GTZ) werden als Maße für Gebäudewärmebedarfe während der Heizperiode verwendet [ALT2015], [IWU2015]. Diese beiden Maße stellen den Zusammenhang zwischen der Raum- und der Außenlufttemperatur dar und werden aus diesem Grund als Hilfsmittel zur Bestimmung der Heizkosten und des Heizstoffbedarfs herangezogen [ALT2015], [IWU2015]. Die Gradtagzahlen sind aufsummierte Differenzen zwischen der Raumtemperatur, die in Deutschland zumeist mit 20 °C angesetzt werden und der Luftaußentemperatur unter der eingangs genannten Heiztag-Grenze vom 15 °C eines jeden Tages [VDI2012], [IWU2015], [WAL2013: 35], [DIE1999].

Die Indikatorwerte der beiden Indikatoren (HGT und GTZ) wurden mithilfe eines Softwaretools [IWU2015] ermittelt, dem die Berechnungsgleichungen HGT (Gleichung 4-2) [VDI2012] und GTZ (Gleichung 4-3) [VDI2013] hinterlegt sind. Die Summe der HGT entspricht den aufsummierten ermittelten Differenzen zwischen der Heizgrenztemperatur (t_{hg}), hier 15 °C (HGT_{15}), und der tatsächlichen Außenlufttemperatur (t_a). Es werden alle Differenzen aufsummiert, die für Heiztage in der Heizperiode ermittelt werden können (z). Die GTZ hingegen summieren ebenfalls Differenzen für Heiztage auf, jedoch wird die Differenz zwischen der mittleren Innentemperatur (t_i), die mit 20 °C angesetzt ist und der mittleren Außenlufttemperatur (t_a), die mit 15 °C angesetzt ist, ermittelt. Auch hier werden die Differenzen für alle Heiztage aufsummiert.

$$HGT_{15} = \sum_1^z (t_{hg} - t_a) \quad \text{Gleichung 4-2}$$

$$GTZ_{t_{20/15}} = \sum_1^z (t_i - t_a) \quad \text{Gleichung 4-3}$$

Die Kenndaten zu den beiden Indikatoren werden in Tabelle 4-15 und Tabelle 4-16 dargestellt. Die Indikatorwerte werden als Mittelwerte des Betrachtungszeitraums von 1970 bis 2013 ermittelt. Die Quelle ist auch hier das Institut Wohnen und Umwelt [IWU2015], das die verwendeten Daten auf Basis der Langfristwetterdaten vom Deutschen Wetterdienst (DWD) erfasst.

Tabelle 4-15: Indikatorkenndaten – „Heizgradtage Mittel der Jahre 1970-2013“

Heizgradtage Mittel der Jahre 1970-2013	
Aussage	Anzahl der Heizgradtage im Mittel der Jahre 1970-2013
Bezugsjahr	1970-2013
Einheit	HGT ₁₅ [Kd]
Quelle	Institut Wohnen und Umwelt [IWU2015]

Tabelle 4-16: Indikatorkenndaten – „Gradtagzahlen Mittel der Jahre 1970-2013“

Gradtagzahlen Mittel der Jahre 1970-2013	
Aussage	Anzahl der Gradtagzahlen im Mittel der Jahre 1970-2013
Bezugsjahr	1970-2013
Einheit	GTZ _{20/15} [Kd]
Quelle	Institut Wohnung und Umwelt [IWU2015]

Es gibt teilweise große Unterschiede der betrachteten Städte. Diese Indikatoren führen zu einer Aussage, ob eine Stadt in einer kalten oder eher warmen Region liegt. Die größte Anzahl an Heiztagen weisen mit 297 die Städte München und Rosenheim auf. Die „wärmste“ Stadt des Untersuchungsrahmens ist mit 231 Heiztagen die Stadt Frankfurt am Main.

4.3 Indikatoren des Verbrauchssektors Verkehr

Dieses Unterkapitel stellt die Indikatoren des zweiten betrachteten Verbrauchssektors Verkehr dar. Dieser Verbrauchssektor wird auch durch eine Reihe an Indikatoren beschrieben. Es lässt sich hier aber feststellen, dass verstärkt Daten auf Landes- und Bundesebene zur Verfügung stehen, wodurch nicht jeder gewünschte Indikator aus verfügbaren Daten zu ermitteln ist. Für die Beschreibung dieses Verbrauchssektors durch Indikatoren wurden nicht ausschließlich energetische Indikatoren gewählt, sondern auch sozio-ökonomische.

Die Indikatoren werden in diesem Unterkapitel äquivalent zum Unterkapitel 4.2 jeweils kurz beschrieben sowie deren Datenlage und Berechnung erläutert. Der Verkehrssektor basiert derzeit fast ausschließlich auf fossilen Treibstoffen, wodurch hier das Potenzial groß und der Sektor zum Erreichen der Klimaschutzziele sehr bedeutend ist.

Der Verkehrssektor wird durch elf sozio-energetische Indikatoren abgebildet. Die ersten fünf Indikatoren decken den gegenwärtigen Zustand im bereits beschriebenen Bereich des motorisierten Individualverkehrs (MIV) ab. Die darauf folgenden drei Indikatoren bilden den sozio-strukturellen Bereich dieses Verbrauchssektors. Die letzten drei Indikatoren stellen die gegenwärtige Situation im ÖPNV dar. Die für diesen Sektor erhobenen Indikatoren sind nachfolgend aufgelistet und werden im Folgenden kurz erläutert. Die Indikatorwerte des Sektors Verkehr können im Anhang (Tabelle A-2) eingesehen werden.

- Motorisierungsgrad
- Personenkraftwagen (Pkw) pro 1.000 Einwohner (EW)
- Dieselfahrzeuge pro 1.000 Einwohner
- Benzinfahrzeuge pro 1.000 Einwohner
- Elektrofahrzeuge (E-Fahrzeuge) pro 1.000 Einwohner
- Bevölkerungsdichte
- Anteil 18 – 65 Jähriger
- Anteil Auspendler an sozialversicherungspflichtig Beschäftigter am Wohnort
- ÖPNV Fahrten pro Einwohner
- Personenkilometer (P.km) pro Einwohner
- Mittlere Reiseweite

Motorisierungsgrad und Personenkraftwagen pro 1.000 Einwohner

Der Energieverbrauch sowie die CO₂-Emissionen fallen im Verkehrssektor zu einem großen Anteil im MIV an. Zeitgleich gibt es keine zufriedenstellenden und flächendeckenden, auf kommunaler Ebene verfügbaren Daten bezüglich der tatsächlichen Nutzung der Verkehrsmittel des Mobilitätsbereichs. Das zeigt, dass der MIV durch Indikatoren nur schwer erfasst werden

kann [DIE1999], [EIC2014]. Die ersten fünf Indikatoren beschreiben den Motorisierten Individualverkehr und dienen dabei als „Umweg“, um diesen Bereich der Mobilität durch Indikatoren vergleichbar darstellen zu können. Das geschieht, indem zu den klassischen Kenngrößen (*Motorisierungsgrad* und *Pkw pro 1.000 Einwohner*) noch spezifische Kenngrößen erhoben werden (Benzin-, Diesel- sowie Elektrofahrzeuge pro 1.000 Einwohner), um auch eine spezifischere Aufteilung und Auswertung der Städte zu ermöglichen. Da der reine Kfz-Bestand nicht in Zahlen die tatsächliche Verkehrsleistung, die Nutzungsintensität sowie das Verkehrsaufkommen beziffern kann, wird für diese Indikatoren die Nutzung der vorhandenen Fahrzeuge vorausgesetzt. [WAL2013] Somit finden auch für den Verbrauchssektor Verkehr mittelbare und unmittelbare Indikatoren Anwendung.

Der Motorisierungsgrad wird durch die Verhältniszahl zwischen der Kraftfahrzeuganzahl eines definierten Gebietes, in dieser Untersuchung einer Stadt, und der Einwohnerzahl dieses Gebietes definiert. Der Motorisierungsgrad beinhaltet somit neben den Pkw auch die Motorräder. Hier werden die Kfz auf jeweils 1.000 Einwohner bezogen. Der zweite Indikator stellt das Verhältnis zwischen Pkw und der Einwohnerzahl dar, somit fallen hier die Krafträder aus der Betrachtung heraus. Die Tabelle 4-17 beinhaltet die Kennwerte zum Motorisierungsgrad. Es lassen sich unter den betrachteten 107 kreisfreien Städten zum Teil erhebliche Unterschiede in den Indikatorwerten der beiden Indikatoren erkennen. So weisen meist große Großstädte nicht die höchsten Motorisierungsgrade auf, da dort die ÖPNV Infrastruktur besser ausgebaut ist.

Tabelle 4-17: Indikatorkenndaten – „Motorisierungsgrad“

Motorisierungsgrad	
Aussage	Anzahl der Kfz pro 1.000 Einwohner in kreisfreien Städten
Bezugsjahr	2014
Einheit	Anzahl [-]
Quelle	Kraftfahrtbundesamt [KBA2014], Statistisches Bundesamt [DES2013D]

In Deutschland sind ca. 47,9 Millionen Pkw und Krafträder zugelassen. Auf die hier betrachteten kreisfreien Städte entfallen mit ca. 13 Millionen Kfz ca. 27 %. Der durchschnittliche Motorisierungsgrad liegt in Deutschland bei 595 Kfz je 1.000 Einwohner. Der Motorisierungsgrad in den kreisfreien Städten liegt mit 533 Kfz je 1.000 Einwohner im Mittel unter dem bundesdeutschen Durchschnitt. [KBA2014] [WEI2014] Den höchsten Motorisierungsgrad sowie die höchste Anzahl an Pkw je 1.000 Einwohner weist die Stadt Wolfsburg auf (1.183 Kfz/1.000 EW). An zweiter Stelle folgt Ingolstadt (762 Kfz/1.000 EW). Die beiden Städte stellen zwei Ausreißer dar, denn dort liegen die Unternehmenssitze zweier Kfz-Großkonzerne (Volkswagen, Audi), wodurch neben den Flottenfahrzeugen der ansässigen Kfz-Hersteller auch die Haushalte der Mitarbeiter als gut ausgestattet angenommen werden können.

Tabelle 4-18: Indikatorkenndaten – „Personenkraftwagen pro 1.000 Einwohner“

Personenkraftwagen pro 1.000 Einwohner	
Aussage	Anzahl der Pkw pro 1.000 Einwohner in kreisfreien Städten
Bezugsjahr	2014
Einheit	Anzahl [-]
Quelle	Kraftfahrtbundesamt [KBA2014], Statistisches Bundesamt [DES2013D]

Die Tabelle 4-18 stellt die Kenndaten des Indikators Pkw pro 1.000 Einwohner dar. Hier ist die Verteilung identisch zu den Zahlen des Motorisierungsgrades. Auch hier ist das Bezugsjahr das Jahr 2014 mit Stichtag am ersten Januar (01.01.2014). Die Daten werden auf kreisebene vom Kraftfahrtbundesamt erhoben und jährlich veröffentlicht.

Diesel-, Benzin- und Elektrofahrzeuge pro 1.000 Einwohner

In diesem Abschnitt werden drei weitere Indikatoren des MIV kurz erläutert. Die Indikatorwerte beziehen sich wie bei dem bereits erläuterten Indikator (Motorisierungsgrad) auf 1.000 Einwohner der jeweiligen Stadt. Die Datenbasis entstammt den Statistiken des Kraftfahrtbundesamtes [KBA2014]. Das Bezugsjahr ist 2014. Zum Stichtag des 01.01.2014 konnte die größte Anzahl der Pkw mit Dieselmotor je 1.000 Einwohner in Wolfsburg ermittelt werden. Die geringste Anzahl an Dieselfahrzeugen je 1.000 Einwohner (79) weist die Stadt Halle an der Saale auf. Interessanter als diese Zahlen ist jedoch, dass die Zahlen der Dieselmotorbetriebenen Pkw in den Städten verschiedene Anteile an allen Kfz ausweisen. Die Anteile der Diesel-Pkw liegen zwischen ca. 20 und 42 %. Die Kenndaten stellt die Tabelle 4-19 zusammen.

Tabelle 4-19: Indikatorkenndaten – „Dieselfahrzeuge pro 1.000 Einwohner“

Dieselfahrzeuge pro 1.000 Einwohner	
Aussage	Anzahl der Pkw mit Dieselmotor pro 1.000 Einwohner in kreisfreien Städten
Bezugsjahr	2014
Einheit	Anzahl [-]
Quelle	Kraftfahrtbundesamt [KBA2014], Statistisches Bundesamt [DES2013D]

Die Kenndaten zum Indikator *Benzinfahrzeuge pro 1.000 Einwohner* sind in Tabelle 4-20 abgebildet. Die Verteilung ist hier ähnlich zum Indikator Personenkraftwagen pro 1.000 Einwohner, da die größeren Anteile der einzelnen Fahrzeuge klar auf der Seite der Benzinfahrzeuge liegen. Die Diesel- und Benzin-Pkw bilden mit knapp 99 % den Hauptanteil der Kraftfahrzeuge des MIV ab [KBA2014].

Tabelle 4-20: Indikatorkenndaten – „Benzinfahrzeuge pro 1.000 Einwohner“

Benzinfahrzeuge pro 1.000 Einwohner	
Aussage	Anzahl der Pkw mit Benzinmotor pro 1.000 Einwohner in kreisfreien Städten
Bezugsjahr	2014
Einheit	Anzahl [-]
Quelle	Kraftfahrtbundesamt [KBA2014], Statistisches Bundesamt [DES2013D]

Neben der Diskussion über die Effizienz der konventionellen Pkw erfolgt in letzter Zeit eine präsenste Diskussion um Pkw mit alternativen Antrieben. Den E-Fahrzeugen wird ein enormes

Potenzial zugesprochen, wodurch die Bundesregierung sogar das Ziel formuliert hat bis zu dem Jahr 2020 den Bestand an E-Fahrzeugen auf eine Million zu erhöhen [BUN2015].

Allen voran sind es Städte, in denen die Elektrofahrzeuge verstärkt ihre Anwendung finden können, denn die Infrastruktur ist in Kommunen und besonders in kreisfreien Städten besser ausgebaut und die Fahrzyklen der E-Fahrzeuge, die derzeit stark durch die Kapazität der Batterie und der daraus resultierenden geringeren Reichweite eher dem innerstädtischen individuellen Verkehr entsprechen. Hier ist das Potenzial der E-Fahrzeuge erkennbar, die während der Zeit, in der sie nicht genutzt werden, geladen werden können. Darüber hinaus fallen lokal keine CO₂-Emissionen an, da diese am Erzeugungsort entstehen, was der innerstädtischen Luftqualität entgegenkommt. Aufbauend auf der aktuellen Diskussion um die E-Fahrzeuge wurde ein Indikator zum Abdecken dieser Thematik erarbeitet, der aus der Anzahl der *E-Fahrzeuge pro 1.000 Einwohner* besteht (Tabelle 4-21).

Tabelle 4-21: Indikatorkennndaten – „Elektrofahrzeuge pro 1.000 Einwohner“

Elektrofahrzeuge pro 1.000 Einwohner	
Aussage	Anzahl der Pkw mit Elektromotor pro 1.000 Einwohner in kreisfreien Städten
Bezugsjahr	2014
Einheit	Anzahl [-]
Quelle	Kraftfahrtbundesamt [KBA2014], Statistisches Bundesamt [DES2013D]

Die Ziele der Bundesregierung, bis 2020 die Anzahl der E-Kfz auf eine Million zu erhöhen, sind ambitioniert. Derzeit sind die Zulassungszahlen der Elektrofahrzeuge mit 12.156 Elektrofahrzeugen (01.01.2014) [KBA2014] weit davon entfernt. Auf die kreisfreien Städte entfallen mit 6.823 Elektrofahrzeugen mehr als die Hälfte aller zugelassenen Elektrofahrzeuge in Deutschland [KBA2014]. Die mit Abstand meisten zugelassenen E-Fahrzeuge je 1.000 Einwohner (5,66) entfallen auf Wolfsburg. Im Wesentlichen liegt die Anzahl der zugelassenen E-Fahrzeuge pro 1.000 jedoch bei 0,2. Das entspricht 2 Fahrzeugen je 10.000 Einwohner.

Bevölkerungsdichte

Die *Bevölkerungsdichte* ist ein allgemeingebräuchlicher Indikator zum Vergleich von z. B. Ländern und Städten, die aussagekräftig ist [WAL2013], [EIC2014]. Die Bevölkerungsdichte hat Einfluss auf viele Bereiche, besonders die Mobilität. Eine geringe Einwohnerdichte bewirkt zum Beispiel größere mittlere Weglängen des innerörtlichen Verkehrs [MEY2013], [WAL2013]. Die Abhängigkeit vom Individualverkehr wächst somit antiproportional mit der Einwohnerdichte [WAL2013: 42]. Die nachfolgende Tabelle beinhaltet nähere Information zu diesem Indikator.

Tabelle 4-22: Indikatorkennndaten – „Bevölkerungsdichte“

Bevölkerungsdichte	
Aussage	Einwohner pro km ² Gebietsfläche einer Stadt
Bezugsjahr	2012, Stichtag: 31.12.2012
Einheit	Einwohner/km ² , EW/km ²
Quelle	Statistisches Bundesamt [DES2014D]

Die geringste Einwohnerdichte weist die Stadt Brandenburg an der Havel (311 EW/km²) auf. Die höchsten Einwohnerdichten weisen die großen Großstädte sowie die Millionenstädte, allen voran die Stadt München mit 4.468 EW/km², auf. Somit übersteigen auch hier die Differenzen in den Indikatorwerten zwischen dem größten und dem kleinsten den Faktor 10.

Anteil 18 – 65 Jähriger

Der *Anteil der Bevölkerung zwischen 18 und 65 Jahren* wird hier als Indikator aufgeführt, da die Personen in diesem Alter, abzüglich der Arbeitslosen, in erwerbstätigen Alter sind und somit auf die Mobilität angewiesen sind, um zu den Arbeitsplätzen zu gelangen. Die Wege zur Arbeit, zur Fortbildung sowie dienstliche Wege stellen mit knapp 30 % einen großen Teil der Mobilität dar, was die Untersuchung „Mobilität in Deutschland 2008“ [INF2010: 116] zeigt. Darüber hinaus nimmt die Mobilität mit dem Anstieg des Alters ab. Es entfallen mit dem Alter z. T. Freizeitaktivitäten, die Mobilität erfordern. Zeitgleich entfällt mit dem Rentenalter auch der regelmäßige Weg zur Arbeit. [INF2010] Die Tabelle 4-23 beinhaltet Informationen zum Indikator.

Tabelle 4-23: Indikatorckenndaten – „Anteil 18 – 65 Jähriger“

Anteil 18 – 65 Jähriger	
Aussage	Anteil der Einwohner einer Stadt im Alter zwischen 18 und 65 Jahren
Bezugsjahr	2014
Einheit	Prozent [%]
Quelle	Statistisches Bundesamt [DES2014F]

Im Durchschnitt deutscher kreisfreier Städte liegt der Anteil der Personen im hier betrachteten Alter bei ca. 64 %. Der größte Anteil lässt sich mit 69,6 % in der Stadt Heidelberg identifizieren. Den niedrigsten Anteil in diesem Altersschnitt weist die Stadt Baden-Baden aus, dieser beträgt 58,8 % und ist unter allen betrachteten Städten der einzige Wert der unter 60 % liegt.

Anteil Auspendler an sozialversicherungspflichtig Beschäftigter am Wohnort

Die *Auspendler* einer Stadt haben einen nicht zu verachtenden Einfluss auf das individuelle Verkehrsaufkommen der Einwohner. Es gibt Städte, die eine geringe Arbeitsplatzzentralität aufweisen und somit eher „Schlafstädte“ sind. Andere Städte haben hingegen eine erhöhte Arbeitsplatzzentralität wodurch die Leute morgens zur Arbeit in die Stadt einpendeln und nach der Arbeit auspendeln müssen. In beiden Fällen erhöht sich das Verkehrsaufkommen einer Stadt. Die Anzahl der Auspendler ist somit für eine Stadt relevant. Für die Vergleichbarkeit der Ergebnisse werden die jeweiligen Auspendlerzahlen auf die sozialversicherungspflichtigen Beschäftigten am Wohnort bezogen. Kenndaten dazu sind Tabelle 4-24 zu entnehmen.

Tabelle 4-24: Indikatorckenndaten – „Anteil Auspendler an Beschäftigten am Wohnort“

Anteil Auspendler an sozialversicherungspflichtig Beschäftigten am Wohnort	
Aussage	Auspendleranteil an den sozialversicherungspflichtig Beschäftigten am Wohnort
Bezugsjahr	(Stichtag: 30.06.2012)
Einheit	Prozent [%]
Quelle	Statistisches Bundesamt [DES2014G]

Der geringste Anteil an Auspendlern und somit die größte Arbeitsplatzzentralität lässt sich mit lediglich 14,04 % in der Bundeshauptstadt Berlin feststellen. Der höchste Auspendleranteil im Untersuchungsrahmen konnte in Fürth ermittelt werden. Dort verlassen 70,21 % die Stadtgrenze zwecks Arbeit und kehren nach dem Ende der Arbeit zurück. [DES2014G] Somit kann diese Stadt als Beispiel für eine Schlafstadt angeführt werden.

ÖPNV Fahrten pro Einwohner

Der ÖPNV bildet ca. 9 % des Personentransportaufkommens in Deutschland ab (s. Abschnitt 2.4.3). Dennoch ist dieser für die Städte essenziell, denn dort liegt dieser Wert um einiges höher als im Bundesdurchschnitt, der viele großflächige Landkreise umfasst, wo bekanntermaßen der ÖPNV eher unterrepräsentiert ist. Grund dafür ist oftmals, dass dieser dort nur schwer wirtschaftlich betrieben werden kann [MEY2013]. Die Inanspruchnahme des ÖPNV hängt von vielen Faktoren ab. Zum einen hat die Struktur und die Entwicklung einer Stadt darauf Einfluss. Städte unterscheiden sich zudem oft durch Aspekte wie z. B. „Lifestyle“, die nur schwer durch Indikatoren erfasst werden können, jedoch einen Einfluss auf das Verhalten der Einwohner einer Stadt nehmen. Ein Beispiel hierfür ist die Stadt Freiburg, die deutschlandweit den höchsten Anteil an Fahrradfahrern am Modal Split der Personenmobilität einnimmt. [SCH2013] Andererseits nehmen auch andere Aspekte, z. B. das Einkommen der Einwohner Einfluss auf das Mobilitätsverhalten [MEY2013].

Messbar, zumindest für die oftmals größeren kreisfreien Städte sind die beförderten Personen durch die regional aktiven Verkehrsunternehmen, was die Intensität der ÖPNV Nutzung wiedergibt. So wurde der Indikator – *ÖPNV Fahrten pro Einwohner* - auf Basis der vom Verband Deutscher Verkehrsunternehmen e. V. (VDV) veröffentlichten Fahrgastzahlen der deutschen Verkehrsunternehmen errechnet. Einige Kenndaten sind in Tabelle 4-25 enthalten.

Tabelle 4-25: Indikatorkenndaten – „ÖPNV Fahrten pro Einwohner“

ÖPNV Fahrten pro Einwohner	
Aussage	Fahrten pro Einwohner und Jahr
Bezugsjahr	2012
Einheit	Anzahl [-]
Quelle	Verband Deutscher Verkehrsunternehmen e. V. [VDV2013]

Die häufigste Nutzung des ÖPNV fällt in der Stadt Bottrop an. Dort fährt jeder Einwohner pro Jahr im Schnitt 541-mal mit öffentlichen Verkehrsmitteln. Es ist jedoch davon auszugehen, dass diese hohe Zahl überschätzt ist, da die vom VDV erhobenen Werte zwar die Fahrgäste der Verkehrsunternehmen einer Stadt beinhalten, die tatsächlichen Fahrgäste jedoch, besonders im dichtbesiedelten Ruhrgebiet, oft nur auf Durchreise durch die Stadtfläche einer Stadt, in diesem Fall Bottrop, sind. Die geringste Nutzung des ÖPNV lässt sich für Schwabach feststellen. Schwabach als eine der kleinsten Städte des Betrachtungsrahmens, weist einen ho-

hen Motorisierungsgrad auf und bildet mit lediglich drei Fahrten mit dem öffentlichen Nahverkehr pro Person und Jahr die Stadt mit der geringsten Nutzung des ÖPNV ab. Der Durchschnitt liegt deutschlandweit bei 120,8 Fahrten je Einwohner und Jahr [VDV2013], [WEI2014]. Der Durchschnitt der kreisfreien Städte liegt mit ca. 197 Fahrten je Einwohner und Jahr ca. um die Hälfte höher als im Bundesdurchschnitt. Diese Zahlen verdeutlichen den wichtigen Stellenwert des ÖPNV für die innerstädtische Mobilität.

Da nicht alle Verkehrsunternehmen in Deutschland Mitglieder im VDV sind, gab es Datenlücken für 17 Städte. Die Daten wurden dann aus den regionalen Daten der dortigen Verkehrsunternehmen errechnet. Dieser Indikator kann somit nicht vollständig aus einer Quelle abgebildet werden. Die übrigen Datenlücken wurden durch Mittelwertbildung aller Städte geschlossen. Bei den beiden folgenden Indikatoren erfolgt ein ähnliches Vorgehen für das Schließen der Datenlücken. Die Datenbasis ist auch dort die Statistik des VDV.

Personenkilometer pro Einwohner

Der zweite Indikator des ÖPNV stellt die *Anzahl gefahrener Kilometer im ÖPNV je Einwohner* einer Stadt dar. Auch hier ist die Datenquelle die Erhebung des VDV. Die Anzahl der Personenkilometer pro Jahr wurde berechnet indem alle erbrachten Personenkilometer des lokalen Verkehrsunternehmens durch die Einwohnerzahl der jeweiligen Stadt dividiert wurden. Der Personenkilometer ist eine Größe aus dem Bereich der Personenmobilität und beschreibt die zurückgelegte Strecke einer Person in einem Verkehrsmittel. Ist beispielsweise ein Bus mit zehn Personen besetzt und fährt eine Strecke von einem Kilometer so hat der Bus zehn Personenkilometer zurückgelegt. Die Tabelle 4-26 beinhaltet die Indikatorkenndaten.

Tabelle 4-26: Indikatorkenndaten – „Personenkilometer pro Einwohner“

Personenkilometer pro Einwohner	
Aussage	Zurückgelegte Kilometer im ÖPNV pro Einwohner und Jahr
Bezugsjahr	2012
Einheit	Kilometer [km]
Quelle	Verband Deutscher Verkehrsunternehmen e. V. [VDV2013]

Die Spannweite der Indikatorwerte liegt zwischen einigen wenigen Kilometern im Jahr und ca. 3.000 Personenkilometern im Jahr. Die Differenzen sind hier nicht zu vermeiden, da die Mobilität im ÖPNV besonders in Ballungsräumen nur schwer einer Stadt zugeordnet werden kann. Ein weiterer Punkt ist, dass trotz der ausführlichen Statistik des VDV einige Datenlücken manuell geschlossen werden müssen. Zudem können Verzerrungen auftreten, wenn ein großes Verkehrsunternehmen in einer kleineren kreisfreien Stadt liegt. In solchen Fällen wurde versucht über die Bevölkerungsdichten der Städte des Gebietes, das von diesem Verkehrsunternehmen bedient wird Anteile zu errechnen, um die Ergebnisse zu plausibilisieren. Ein Beispiel, wo dieses Vorgehen angewendet wurde, sind die Daten des Verkehrsunternehmens *BOGE-*

STRA, das die Städte Bochum und Gelsenkirchen bedient sowie einige Strecken in den umliegenden Städten anbietet. Der Bundesdurchschnitt liegt bei 1.580 Personenkilometer und Jahr [WEI2014], [VDV2013]. In kreisfreien Städten liegt dieser Durchschnitt mit 1.230 Personenkilometern unter dem Bundesdurchschnitt.

Mittlere Reiseweite

Der letzte Indikator des ÖPNV und des Verbrauchssektors Verkehr stellt die *mittlere Reiseweite einer ÖPNV Fahrt* dar. Dieser Indikator basiert auf den Statistiken des VDV. In nachfolgender Tabelle sind die Kenndaten zu diesem Indikator enthalten.

Tabelle 4-27: Indikatorkenndaten – „Mittlere Reiseweite“

Mittlere Reiseweite	
Aussage	Mittlere Reiseweite einer ÖPNV Fahrt
Bezugsjahr	2012
Einheit	Kilometer [km]
Quelle	Verband Deutscher Verkehrsunternehmen e. V. [VDV2013]

Im Mittel aller ÖPNV Fahrten in kreisfreien Städten werden ca. 5,6 Kilometer zurückgelegt, wobei auch bei diesem Indikator die Spannweite der Indikatorwerte zwischen 27 und ca. 0,3 Kilometern liegt. Auch hier können die zum Teil großen Differenzen so begründet werden, wie bei dem vorherigen Indikator. Der Bundesdurchschnitt für diesen Indikator liegt bei 13 Kilometern. Auch hier liegt das Mittel der kreisfreien Städte unter dem Bundesdurchschnitt. [WEI2014], [VDV2013] Die Wege in Städten sind kürzer als im ländlichen Raum. Das Verkehrsaufkommen in kreisfreien Städten ist oft viel höher als im ländlichen Raum, da in dichtbesiedelten Regionen der ÖPNV trotz einer höheren Taktung und einer besseren Flächenabdeckung wirtschaftlicher betrieben werden kann.

4.4 Indikatoren der Verbrauchssektoren Industrie und GHD

In diesem Unterkapitel werden die Indikatoren für die Verbrauchssektoren Industrie und GHD zeitgleich beschrieben. Diese beiden Sektoren sind zusammen mit einem erheblichen Anteil am deutschen Energiebedarf beteiligt, wobei davon der größere Teil auf die energieintensivere Industrie entfällt [Eic2013]. Die Datenlage für diese beiden Sektoren ist auf kommunaler Ebene eher limitiert. Ein Grund dafür sind zum einen die Geheimhaltungsinteressen der wirtschaftenden Unternehmen und zum anderen gibt es wenige flächendeckende Erhebungen der Energie und Betriebsmittelverbräuche, die zentral von zum Beispiel einer Behörde oder einem Amt zusammengetragen und dokumentiert werden. [WAL2013] Es ist denkbar, dass besonders für diese beiden Sektoren, in Zukunft die Datenlage durch das forcieren des Energiemanagements nach DIN EN ISO 50001 [ISO2011] besser werden könnte.

Die beiden Verbrauchssektoren wurden durch sieben identische Indikatoren beschrieben. Für den Sektor Industrie wurde ein Indikator mehr erhoben. Nachfolgend sind die Indikatoren aufgelistet. Nach jedem Indikator ist gekennzeichnet für welchen der beiden Sektoren dieser erhoben wurde. Die berechneten Indikatorwerte sind im Anhang in Tabelle A-3 dargestellt.

- Beschäftigtenanteil im jeweiligen Sektor (Industrie, GHD)
- Energieproduktivität im jeweiligen Sektor (Industrie, GHD)
- Energieintensität im jeweiligen Sektor (Industrie, GHD)
- Produktivitätsniveau im jeweiligen Sektor (Industrie, GHD)
- Spezifischer Energieverbrauch im jeweiligen Sektor (Industrie, GHD)
- Umsatzsteuer des jeweiligen Sektors an der Gesamtumsatzsteuer (Industrie, GHD)
- Entwicklung des Beschäftigtenanteile des jeweiligen Sektors (Industrie, GHD)
- Veränderung der Energieintensität zwischen 2007 und 2011 (Industrie)

Beschäftigtenanteil im jeweiligen Sektor (Industrie, GHD)

Der erste für die beiden Sektoren Industrie und GHD ermittelte Indikator stellt den Anteil der Erwerbstätigen im jeweiligen Sektor dar. Die Erwerbstätigenzahlen des jeweiligen Sektors werden auf die gesamten Erwerbstätigenzahlen einer Stadt bezogen. Eine weitere Möglichkeit wäre es die Erwerbstätigenzahlen auf die gesamte Bevölkerung zu beziehen [BUD2014], jedoch kann ein Teil des Informationsgehaltes des Indikators verloren gehen, da das Verhältnis zwischen den Anteilen der Beschäftigten der Industrie und GHD verzerrt wird. Dies erfolgt durch den Einbezug von nicht erwerbstätigen Personen, Rentnern sowie Personen, die noch nicht im erwerbstätigen Alter sind. Die Größen werden in anderen Indikatoren berücksichtigt. Die Kennwerte dieses Indikators erlauben es eine Differenzierung der Erwerbsstrukturen sowie der Art (Dienstleistungs- oder Industriestadt) einer Stadt abschätzen. Die Indikatorwerte schaffen einen Überblick über die verschiedenen Stadtarten. So wird vermieden, dass Industriestädte den Dienstleistungsstädten gegenübergestellt werden. Der anzunehmende Unterschied des Wirtschaftsaufbaus zwischen den verschiedenen Stadtarten macht die Erhebung dieses Indikators sinnvoll. Eine Industriestadt weist höhere Energieverbräuche als eine GHD Stadt auf, die zwar einen hohen Wert schafft, jedoch weniger energieintensiv ist als z. B. die Grundstoffproduktion, die hoch energieintensiv ist. [BUD2014], [Eic2013], [WAL2013]

In Tabelle 4-28 sind die Kenndaten zu diesem Indikator aufgetragen. Die Daten zur Indikatorberechnung entstammen der Regionaldatenbank des Statistischen Bundesamtes.

Tabelle 4-28: Indikatorkenndaten – „Beschäftigtenanteil im jeweiligen Sektor“

Beschäftigtenanteil im jeweiligen Sektor	
Aussage	Beschäftigtenanteil der Beschäftigten im jeweiligen Sektor
Bezugsjahr	2011
Sektor	Industrie, GHD
Einheit	Anteil [%]
Quelle	Statistisches Bundesamt [DES2014H]

Der höchste Beschäftigtenanteil im Industriesektor lässt sich für die Stadt Wolfsburg ermitteln. Die Stadt Wolfsburg hat mit einem Anteil von 49,9 % den höchsten unter den kreisfreien Städten. Den höchsten Beschäftigtenanteil im GHD Sektor weist die Stadt Potsdam (93,6 %) auf. Dementsprechend gering fällt in Potsdam auch der Anteil der Beschäftigten im Industriesektor aus. Dieser beträgt lediglich 6,3 %.

Energieproduktivität/Energieintensität im jeweiligen Sektor (Industrie, GHD)

Die *Energieproduktivität* sowie die *Energieintensität* im jeweiligen Sektor bilden zwei Indikatoren, für die sich die zeitgleiche Betrachtung beider Kennwerte anbietet. Beide Kennwerte bilden in der Diskussion um die Energieeffizienz in den Sektoren Industrie und GHD mit die wichtigsten Kennziffern. Diese werden z. B. herangezogen, um Vergleiche zwischen Wirtschaftszweigen oder zwischen Ländern zu ermöglichen. Es werden verschiedene Bezugsgrößen verwendet. Diese können z. B. die Bevölkerung, das Bruttoinlandsprodukt (BIP), der Produktionswert sowie die Bruttowertschöpfung (BWS) sein. Zur Berechnung der Indikatorwerte wird, bedingt durch die flächendeckende Verfügbarkeit der BIP Zahlen für Kreise und kreisfreie Städte, als stadt spezifische Bezugsgrößen das BIP gewählt. Dieses wird auf den Energieverbrauch bezogen. Der Energieverbrauch wird, aufgeschlüsselt nach Kreisen und kreisfreien Städten, vom Statistischen Bundesamt statistisch zur Verfügung gestellt. Die Umrechnung der Anteile der Sektoren Industrie und GHD am Energieverbrauch erfolgte auf Basis der Beschäftigungszahlen im jeweiligen Sektor [DES2014H] der betrachteten Stadt, da die Energieverbräuche [DES2014I] lediglich für die Stadt veröffentlicht werden nicht aber aufgeteilt nach dem jeweiligen Sektor.

Die Energieproduktivität ist definiert als erzeugtes BIP pro eingesetzte Energiemenge in der jeweils betrachteten Stadt. Der höchste Wert, also die höchste Energieproduktivität (ca. 5600 €/GJ), lässt sich im Sektor Industrie für die Stadt Potsdam feststellen. Städte, in denen viel Kfz-Industrie ansässig ist, liegen ebenfalls im hohen Energieproduktivitätsbereich. Den niedrigsten Wert für diesen Indikator im Sektor Industrie lässt sich für Köln feststellen. Dort werden lediglich 4,1 Euro je eingesetzter Energiemenge produziert. Köln stellt an dieser Stelle einen Sonderfall dar, da der Kölner Raum einen der größten Chemiestandorte Europas bildet und somit auch der große nichtenergetische Verbrauch mit unter den Energieverbrauch fällt, der wiederum nicht energetisch genutzt wird sondern als Produkt die Industrieunternehmen verlässt. Im Sektor GHD fallen die Produktivitätswerte im Schnitt höher aus als im Sektor Industrie. Grund dafür ist, dass die Erzeugung eines materiellen Guts oftmals einen höheren Energieeinsatz erfordert als bei der Erzeugung eines nichtmateriellen Outputs, wie es im GHD Sektor meist der Fall ist [Eic2014]. Einige Kenndaten zur Energieproduktivität sind in Tabelle 4-29 aufgeführt.

Tabelle 4-29: Indikatorkenndaten – „Energieproduktivität im jeweiligen Sektor“

Energieproduktivität im jeweiligen Sektor	
Aussage	Erzeugtes BIP pro eingesetzte Energiemenge
Bezugsjahr	2011
Sektor	Industrie, GHD
Einheit	Euro/Gigajoule Energieeinsatz [€/GJ]
Quelle	Statistisches Bundesamt [DES2014H], [DES2014I], [DES2014J]

Tabelle 4-30: Indikatorkenndaten – „Energieintensität im jeweiligen Sektor“

Energieintensität im jeweiligen Sektor	
Aussage	Notwendiger Energieeinsatz zur Erzeugung des BIP
Bezugsjahr	2011
Sektor	Industrie, GHD
Einheit	Megajoule Energieeinsatz/Euro [MJ/€]
Quelle	Statistisches Bundesamt [DES2014H], [DES2014I], [DES2014J]

Die Energieintensität errechnet sich aus dem Kehrwert der Energieproduktivität. Diese Kenngröße wird in der Berichterstattung oft stellvertretend für den Ausdruck der Energieeffizienz verwendet, da angenommen werden kann, dass je geringer der Energieeinsatz zur Erzeugung derselben Menge an Wertschöpfung ist (BIP, BWS, Produktionswerte, etc.), die Nutzung der Energie effizienter erfolgt. In Energieprognosen wird dieser Indikator ebenfalls verwendet, wobei die Betrachtung einzelner Städte in solchen Prognosen kaum erfolgt, sondern vielmehr die Ebene der Länder bzw. Staaten betrachtet wird [Exx2015]. In Tabelle 4-30 sind einige Kenndaten zur Energieintensität aufgetragen. Die verwendeten Quellen sind dieselben, wie bei der Berechnung der Indikatorwerte der Energieproduktivität.

Produktivitätsniveau/Spez. Energieverbrauch im jeweiligen Sektor (Industrie, GHD)

Für das Beschreiben der beiden Indikatoren in diesem Absatz werden, äquivalent zu den beiden vorherigen Indikatoren, beide zusammengefasst und zeitgleich erläutert. Das Produktivitätsniveau ist ein dem Indikator Energieproduktivität ähnlicher Kennwert. Das Produktivitätsniveau wird ausgedrückt durch den Quotienten aus der Bruttowertschöpfung (BWS) je eingesetzter, zur Erzeugung dieser BWS notwendigen, Energiemenge. Der Kehrwert des Indikators Produktivitätsniveau ist der spezifische Energieverbrauch. Der spezifische Energieverbrauch drückt den aufzuwenden Energiebedarf zum Erzeugen der Bruttowertschöpfung aus. Die Kenndaten zu diesen Indikatoren sind in Tabelle 4-31 und Tabelle 4-32 dargestellt.

Tabelle 4-31: Indikatorkenndaten – „Produktivitätsniveau im jeweiligen Sektor“

Produktivitätsniveau im jeweiligen Sektor	
Aussage	Erzeugtes BWS pro eingesetzte Energiemenge
Bezugsjahr	2011
Sektor	Industrie, GHD
Einheit	Euro/Gigajoule Energieeinsatz [€/GJ]
Quelle	Statistisches Bundesamt [DES2014H], [DES2014I], [DES2014K]

Tabelle 4-32: Indikatorkenndaten – „Spezifischer Energieverbrauch im jeweiligen Sektor“

Spezifischer Energieverbrauch im jeweiligen Sektor	
Aussage	Notwendiger Energieeinsatz zur Erzeugung der BWS
Bezugsjahr	2011
Sektor	Industrie, GHD
Einheit	Megajoule Energieeinsatz/Euro [MJ/€]
Quelle	Statistisches Bundesamt [DES2014H], [DES2014I], [DES2014K]

Auch zur Berechnung dieses Indikators muss der Umweg über die Beschäftigtenzahlen genommen werden, da die Energieverbräuche lediglich auf Kreisebene und nicht gesondert aufgeteilt nach den Verbrauchssektoren erhoben und veröffentlicht werden. Diese Umrechnung erfolgt mit erhobenen Kennzahlen derselben Institution (Statistisches Bundesamt), wodurch davon auszugehen ist, dass derselbe Detaillierungsgrad sowie dieselbe Sorgfalt zum Erheben der Daten aufgewendet wurden. Die niedrigsten Energieproduktivitätswerte in der Industrie können für Köln und für Ludwigshafen am Rhein ermittelt werden. Grund ist die äußerst prä-sente chemische Industrie, die einen hohen energetischen Aufwand generiert, wenngleich ein Großteil der aufgewendeten Energieträger nicht energetische genutzt werden.

Umsatzsteuer des jeweiligen Sektors an der Gesamtumsatzsteuer (Industrie, GHD)

Die Wertschöpfungsstruktur einer Stadt lässt sich durch den Indikator *Umsatzsteuer an der Gesamtumsatzsteuer des jeweiligen Sektors* bewerten. Hohe Umsatzsteuerwerte eines Sektors an der gesamten Umsatzsteuer deuten auf eine hohe Industrialisierung, sofern die Werte der Industrieumsatzsteuer an der Gesamtumsatzsteuer hoch sind. Äquivalent dazu ist die GHD Orientierung prägend, wenn die GHD-Umsatzsteuer einen hohen Anteil aufweist. Die größten Anteile wurden im Industriesektor für die Städte, in denen eine große chemische Industrie (Leverkusen, Ludwigshafen) sowie große Produzenten von Kraftfahrzeugen (Stuttgart, München) ansässig sind, ermittelt. Im GHD-Sektor fallen hohe Anteile der GHD-Umsatzsteuer auf die Städte Mülheim an der Ruhr sowie Frankfurt am Main. Der hohe Anteil in Mülheim deutet z. T. die Transformation, die im Ruhrgebiet nach und nach vorangeht (Industrie → Dienstleistung), an. In Frankfurt ist der hohe Anteil eher wenig verwunderlich, da Frankfurt als Finanzzentrum Deutschlands einen erheblichen Anteil der Umsatzsteuer mit Finanzdienstleistungen erwirtschaftet. Die niedrigsten Anteile im Sektor lassen sich für die Städte ermitteln, in denen die Anteile der umgekehrten Sektoren besonders hoch sind. Die Tabelle 4-33 zeigt die Indikatorkenndaten.

Tabelle 4-33: Indikatorkenndaten – „Umsatzsteuer des jeweiligen Sektors an der Gesamtumsatzsteuer“

Umsatzsteuer des jeweiligen Sektors an der Gesamtumsatzsteuer	
Aussage	Anteil der Umsatzsteuer an der Gesamtumsatzsteuer der Stadt
Bezugsjahr	2012
Sektor	Industrie, GHD
Einheit	Prozent [%]
Quelle	Statistisches Bundesamt [DES2008], [DES2015A]

Tabelle 4-34: Zuordnung der WZ08 zu den Sektoren (eigene Darstellung [DES2008])

Industrie	GHD
Bergbau und Gewinnung von Steinen und Erden (B)	Handel, Instandhaltung und Reparatur von Kfz (G)
Verarbeitendes Gewerbe (C)	Verkehr und Lagerei (H)
Energieversorgung (D)	Gastgewerbe (I)
Wasserversorgung, Entsorgung, Beseitigung von Umweltverschmutzung (E)	Information und Kommunikation (J)
Baugewerbe (F)	Erbringung von Finanz- und Versicherungsleistungen (K)
---	Grundstücks- und Wohnungswesen (L)
---	Freiberufliche, wissenschaftl. und techn. Dienstleistungen (M)
---	Sonstige wirtschaftliche Dienstleistungen (N)
---	Öffentliche Verwaltung, Verteidigung, Sozialversicherung
---	Erziehung und Unterricht (P)
---	Gesundheits- und Sozialwesen (Q)
---	Kunst, Unterhaltung und Erholung (R)
---	Erbringung von sonstigen Dienstleistungen (S)

Die Berechnung der Indikatorwerte erfolgt mithilfe der Einteilung der einzelnen Wirtschaftszweige zu den Sektoren. So wurden die Werte der Umsatzsteuer [DES2015A] den Sektoren zugeordnet. Die Zuordnung erfolgt unter Zuhilfenahme der vom Statistischen Bundesamt definierten Wirtschaftszweige (Tabelle 4-34). Lediglich der Wirtschaftszweig (A) ist nicht in der Tabelle enthalten, da dieser die Tätigkeitsfelder beinhaltet, die dem Primärsektor zugeordnet werden (Fischerei, Forst- und Landwirtschaft).

Entwicklung der Beschäftigtenanteile des jeweiligen Sektors (Industrie, GHD)

Um den zeitlichen Verlauf in der Entwicklung der Städte in den gewerblichen Sektoren zu quantifizieren, wurde der Indikator *Entwicklung der Beschäftigtenanteile im jeweiligen Sektor* ermittelt. In Tabelle 4-35 sind die wichtigsten Indikatorckenndaten aufgeführt.

Tabelle 4-35: Indikatorckenndaten – „Entwicklung der Beschäftigtenanteile des jeweiligen Sektors“

Entwicklung der Beschäftigtenanteile des jeweiligen Sektors	
Aussage	Veränderung der Beschäftigtenzahlen in dem jeweiligen Sektor von 2005-2011
Bezugsjahr	2011
Sektor	Industrie, GHD
Einheit	Prozent [%]
Quelle	Statistisches Bundesamt [DES2014H]

Fast über alle Städte hinweg lässt sich in der Industrie ein zum Teil enormer Rückgang in den Beschäftigtenzahlen feststellen. So verzeichnen die Städte Flensburg und Rosenheim (Rückgänge jeweils über 25 %) die größten Rückgänge in diesem Sektor. Gesamtheitlich betrachtet haben die Beschäftigtenanteile in dem Bereich große Anteile, dennoch sind die absoluten Veränderungen im Vergleich zu anderen Städten eher gering. Im Schnitt lässt sich für diesen Zeitraum ein Rückgang von ca. 7 % für die Industrie verzeichnen. Im GHD-Sektor hingegen

ist ein mittlerer Zuwachs von ca. 2,5 % festzustellen. Am stärksten wachsen die Beschäftigtenanteile in der Stadt Wolfsburg. Es ist davon auszugehen, dass in diesem Zeitraum verstärkt Dienstleistungen vom dort ansässigen Kfz-Konzern ausgebaut wurden.

Veränderung der Energieintensität zwischen 2007 und 2011 (Industrie)

Die *Veränderung der Energieintensität* wurde als Indikator gewählt, da diese Kenngröße eine Aussage bezüglich der Energieeffizienz erlaubt. Wenn bei gleichbleibendem auf die Industrie entfallendem BIP weniger Energie eingesetzt wird so lässt sich eine effizientere Nutzung der Energie feststellen. Das ist einer der wenigen Indikatoren, die im zeitlichen Verlauf anzeigen können, ob Veränderungen zu verzeichnen sind. In der Tabelle 4-36 sind die notwendigen Kenndaten aufgetragen mit denen der Indikator konstruiert wurde.

Tabelle 4-36: Indikatorkenndaten – „Veränderung Energieintensität 2007-2011 Industrie“

Veränderung der Energieintensität zwischen 2007 und 2011 im Sektor Industrie	
Aussage	Veränderung der Energieintensität in der Industrie - Effizienzsteigerung
Bezugsjahr	2011
Sektor	Industrie
Einheit	Prozent [%]
Quelle	Statistisches Bundesamt [DES2014H], [DES2014I], [DES2014J]

Im Durchschnitt wurde eine Verringerung der Energieintensität um ca. 2,8 % errechnet. Da das BIP nicht inflationsbereinigt ist, ist die Intensität im Mittel in etwa gleich geblieben ist. In den einzelnen Städten hingegen zeigt dieser Indikator z. T. große Veränderungen.

5 Clusteranalyse der kreisfreien Städte

In diesem Kapitel ist die Clusteranalyse deutscher kreisfreier Städte der Gegenstand. Dazu wurde dieses Kapitel in drei Unterkapitel eingeteilt, die die für eine Clusteranalyse notwendigen Schritte nachvollziehbar darstellen. Die Datenvorbereitung bildet den ersten Schritt der Clusteranalyse. Darauf baut die Faktoranalyse der im vorherigen Kapitel vorgestellten Indikatoren auf. Abschließend wird mithilfe der Faktoren die Clusteranalyse durchgeführt.

5.1 Datenvorbereitung

Die Datenvorbereitung stellt einen wichtigen Schritt für die korrekte Durchführung der Clusteranalyse dar. Daten werden benötigt, um die Clusterung mathematisch valide durchführen zu können. In diesem Unterkapitel wird die durchgeführte Datenvorbereitung erläutert. Zunächst wird kurz auf das Schließen der Datenlücken eingegangen. Abschließend wird die Standardisierung der Indikatorwerte mithilfe der Z-Transformation behandelt.

5.1.1 Datenlücken

Für die Anwendung der Clusteranalyse müssen die Daten einheitlich und ohne Datenlücken vorliegen. Wie in dem Kapitel zuvor bereits für die einzelnen Indikatoren erklärt wurde, gibt es aufgrund der Datenlage auf kommunaler Ebene z. T. Probleme bei der Datenverfügbarkeit. So treten bei manchen Indikatoren Datenlücken auf. Dabei wird darauf geachtet, dass die Anzahl an Datenlücken in einem Indikator vertretbar ist. Die Datenlücken werden dabei durch die Bildung der Mittelwerte des Indikators geschlossen. Dieses Vorgehen erlaubt es, den Indikator in der Untersuchung zu behalten. Dabei wird in Kauf genommen, dass durch die Mittelwertbildung ein geringer Fehler in die Analyse eingetragen wird, der aber aufgrund der großen Anzahl an Indikatoren die Analyse unwesentlich beeinflusst. Die Anwendung der Mittelwertmethode zum Schließen der Datenlücken wurde lediglich bei den Indikatoren und Werten durchgeführt, die nicht durch eine gesonderte Recherche ermittelt werden konnten.

5.1.2 Z-Transformation der Indikatorwerte

Die Z-Transformation ist eine in SPSS implementierte Methode, die die statistisch zu betrachtenden Variablenwerte standardisiert [IBM2016]. Dieser erste Schritt knüpft an die zuvor erwähnte primäre Datenvorbereitung, bestehend aus der Ermittlung der Indikatoren und Indikatorwerte sowie dem Schließen der Datenlücken als auch der Elimination von ungeeigneten Objekten an und ist für eine belastbare Analyse notwendig. Durch die Z-Transformation werden Indikatoren verschiedener Skalenniveaus und Spannweiten auf eine einheitliche Größenordnung skaliert. [BUE2012] [SCH2004] [BER2014] Hierdurch werden in der Analyse Fehler vermieden. Die so standardisierten Werte ähneln sich in Ihrer Größe und erfahren dabei keinen Informationsverlust. Die Ergebnisse der Clusteranalyse sind durch die vorgeschaltete Z-Transformation belastbarer, da die verwendeten Distanzmaße (quadrierte euklidische Distanz,

Unterkapitel 2.4) durch die Größenordnung der Indikatorwerte beeinflusst werden. Standardisierte Werte hingegen eliminieren diesen Fehler und den Objekten kommt im mathematischen Sinn dieselbe Bedeutung zu. [WAL2014]

Die Z-Transformation homogenisiert die Werte der erhobenen Indikatoren unter der Zuhilfenahme mathematischer Werte wie der Maxima, der Minima, der Mittelwerte sowie dem Standardfehler der Indikatorwerte eines Indikators. Diese Werte fließen in die Berechnung der Z-transformierten Indikatorwerte (Z-Werte) ein. Die Z-Werte sind nach der Transformation auf einen Wertebereich zwischen +/- 3 standardisiert. [BUE2012] Es gibt dennoch vereinzelt transformierte Werte, die ober- bzw. unterhalb dieses Werts liegen. Dabei handelt es sich dann z. B. um Ausreißer, die in einem Indikator sehr starke Ausprägungen aufweisen, jedoch mathematisch auf die Clusteranalyse bezogen keine Verfälschung des Ergebnisses bewirken. Die Berechnung der Z-Werte erfolgt mit nachfolgender Gleichung 5-1. In Gleichung 5-1 stellt z die neue, transformierte Variable (Z-Wert), x den ursprünglichen Indikatorwert, m den arithmetischen Mittelwert und s die Standardabweichung dar. Der arithmetische Mittelwert wird über die Indikatorwerte eines jeden Indikators gebildet. [BUE2012]

$$z = \frac{x - m}{s} \quad \text{Gleichung 5-1}$$

Wird die Clusteranalyse in einer grafischen Skizze dargestellt (s. Abbildung 5-1), so bilden die Z-transformierten Werte die Basis einer belastbaren Analyse auf der die Faktoranalyse und schließlich die eigentliche Clusteranalyse aufbauen. Trotzdem gibt es einige wenige Z-Werte, die als Ausreiser zu bezeichnen sind. Diese liegen dann außerhalb des benannten Bereichs. Sie sind zumeist Sonderfälle, z. B. die Stadt Köln mit dem durch die starke chemische Industrie einhergehenden hohen Energieverbrauch (auch nichtenergetisch) oder die Stadt Wolfsburg mit dem höchsten Anteil an zugelassenen E-Kraftfahrzeugen. Diese Ausreißer haben jedoch keinen negativen Einfluss auf die Aussagekraft der Ergebnisse.

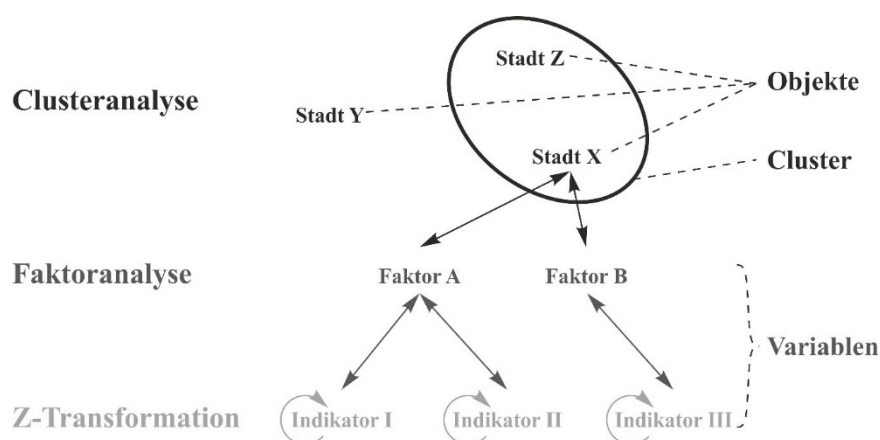


Abbildung 5-1: Schematische Struktur des gewählten Vorgehens (aus [WAL2014])

5.2 Faktorenanalyse

In der multivariaten Statistik gibt es zwei Faktoranalyseformen, die explorative und die konfirmatorische Faktorenanalyse. Die explorative Faktorenanalyse wird auch als Sinn entdeckendes Verfahren bezeichnet [SCH2010], da diese in der Lage ist, mithilfe der Korrelation Variablen (Indikatoren) zu Gruppen zu „bündeln“. Die konfirmatorische Faktorenanalyse hingegen erlaubt es, vermutete Zusammenhänge und Faktoren aufgestellten Hypothesen gegenüberzustellen und diese zu überprüfen. Die konfirmatorische Faktorenanalyse fällt aus diesem Grund tendenziell eher in die Markt- und Meinungsforschung als in den Bereich, in dem lediglich eine Datenreduktion der Variablenanzahl das Ziel darstellt. Fortan wird die explorative Faktorenanalyse angewendet, die den zweiten Schritt der durchgeführten Untersuchung bildet. Diese baut auf den im vorherigen Schritt ermittelten Z-Werten der Indikatoren auf. Die Faktoranalyse wird auch als sinn entdeckendes Verfahren bezeichnet [SCH2010]. Dieses Verfahren stellt einen wichtigen Schritt in der statistischen Betrachtung dar, da hier gleich mehrere wesentliche Vorteile für verlässliche Clusterergebnisse vereint werden. Die Faktoranalyse vereint die nachfolgend aufgelisteten Aspekte, auf die in dem nachfolgenden Abschnitt eingegangen wird.

- Statistisch sinnvolle Datenreduktion
- Variablen werden auf Abhängigkeit geprüft
- Gleiche Gewichtung der Indikatoren

5.2.1 Vorgehen der Faktorenanalyse

Die Faktorenanalyse ist ein statistisches multivariates Analyseverfahren, das sich dazu eignet Dimensionen in dem zu untersuchenden Datensatz zu reduzieren. Die Reduktion bezieht sich bei diesem Analyseverfahren, anders als bei der Clusteranalyse bei der die Objektanzahl reduziert wird, darauf die Anzahl der betrachteten Variablen (Indikatoren) auf ein sinnvolles Maß zu verringern. Diese Datenreduktion erfolgt idealerweise unter Aufrechterhaltung der statistischen Aussagekraft. Durch die Faktorenanalyse werden aus den betrachteten Indikatoren latente Variablen, die auch als „künstliche Variablen“ bezeichnet werden, extrahiert. Diese künstlichen Variablen bilden die dem extrahierten Faktor angehörenden Indikatoren hinreichend genau ab, wodurch die Deutung dieses Faktors ermöglicht wird. [MON2008] Ein Faktor wird als künstliche Variable bezeichnet, da dieser nicht erhoben werden kann sondern über den Umweg der Faktorenanalyse konstruiert wird. Hierfür wurde die Varimax-Rotation verwendet. Dieses Vorgehen ist ein gängiges Verfahren in der Statistik, bei dem die Matrizen der zu untersuchenden Daten so lange mathematisch rotiert werden bis die Ausrichtung der betrachteten Indikatoren mit den Faktoren möglichst identisch ist und die so ermittelten Faktoren voneinander zu trennen sind. Ein auf diese Weise ermittelter Faktor hat zunächst keine Bedeu-

tung, diese erhält er erst nach einer Deutung der Faktoren mithilfe der mit diesem Faktor korrelierenden Indikatoren. Die auf diese Weise extrahierten Faktoren sollen gemeinsame, inhaltlich interpretierbare künstliche Variablen darstellen [BAC2010].

Die Qualität mit der ein Indikator in einem Faktor abgebildet wird, lässt sich über die Faktorladungen nachvollziehen. Faktorladungen werden für jeden Indikator und jeden Faktor berechnet. Diese fallen unterschiedlich groß aus, sollten jedoch betragsmäßig mindestens über einen Wert von 0,5 verfügen [BUE2012], [BER2014], [BAC2016], da erst ab dieser Größe angenommen werden kann, dass die Indikatoren vom extrahierten Faktor verlässlich abgebildet werden. Eine höhere Faktorladung ist gewünscht. Faktorladungen $<0,5$ sollten aus der Faktorenanalyse eliminiert werden, da geringere Faktorladungen darauf hindeuten, dass es sinnvoller ist, einen Indikator aus der Untersuchung zu eliminieren als diesen mit wenig Aussagekraft einem Faktor zuzuweisen [BAC2016]. Diese Grenze wird auch als Kaiser-Meyer-Olkin-Kriterium bezeichnet und stellt die gebräuchlichste Prüfgröße für Faktorenanalysen dar [BAC2016].

Die Faktorenladungen können positiv und negativ sein. Das Vorzeichen erlaubt die Deutung eines Faktors und ist demnach für die weiteren Schritte nützlich. Positive Faktorladungen deuten darauf hin, dass der jeweilige Indikator positiv auf den Faktor lädt, und bei negativen umgekehrt. Das bedeutet, dass wenn ein Indikator stark positive Werte aufweist (z. B. stark positive Bevölkerungsentwicklung) und dieser Indikator eine hohe positive Faktorladung vorweist, so wird auch der Faktorwert größer. Bei negativen Faktorladungen und negativen Indikatorwerten verstärkt der negative Indikatorwert den Faktorwert positiv (s. Abschnitt 5.2.2).

Neben der Datenreduktion hilft die Faktorenanalyse dabei die betrachteten Indikatoren gleichmäßig in die spätere Clusteranalyse einfließen zu lassen. [HIL1997] Mithilfe der Faktorenladungen fließen die Indikatoren zwar unterschiedlich stark in einen Faktor ein, der Faktor fließt hingegen gleichmäßig in die folgende Clusteranalyse ein. Weiterhin sind die ermittelten Faktoren voneinander unabhängig. Das ist eine der wichtigsten Forderungen einer statistischen Untersuchung. Extrahierte Faktoren sind immer unabhängig voneinander, wodurch die Faktoren eine solide Datenbasis für den letzten Schritt der Clusteranalyse bilden.

5.2.2 Ergebnisse der Faktorenanalyse

Die durchgeführte Faktorenanalyse der betrachteten 41 Indikatoren hat im letzten Schritt insgesamt 11 Faktoren extrahiert, auf die auf die folgend näher eingegangen wird. Die Faktorenanalyse wurde zunächst mehrfach auf den betrachteten Indikatorsatz angewendet, um valide und belastbare Faktoren für die weitere Untersuchung zu extrahieren. Die so erhobenen Faktoren haben zunächst keine Bedeutung, die sich jedoch mit dem Zuordnen der Indikatoren, die einem Faktor zugehören, einstellt. Die Extraktion erfolgt nach mathematischen Gesichtspunkten, die eine objektive Analyse der Indikatoren erlaubt. Durch die Faktorenanalyse wird eine Zusammenfassung aus der Vielzahl an Indikatoren ermöglicht, die sich statistisch ähneln

und somit gebündelt in einem Faktor interpretiert werden können. Zudem korrelieren die Faktoren nicht miteinander und sind deshalb unabhängig.

In einem ersten Schritt wurden die Indikatoren einer ersten Faktorenanalyse unterzogen. Diese wurde mithilfe der Hauptkomponentenmethode¹³ sowie der Berechnung der Korrelationsmatrix durchgeführt. Ebenfalls wurde für die Analyse keine Anzahl an zu extrahierenden Faktoren vorgegeben, da aus dem Indikatorsatz eine statistisch sinnvolle Faktorenanzahl identifiziert werden soll, die die betrachteten Indikatoren verlässlich erklärt. Es wurde die z. B. in [BUE2012], [BAC2016] benannte Methode der Zuhilfenahme der Eigenwerte zur Wahl einer geeigneten Faktorenanzahl verwendet. Dabei werden die Gesamtwerte der anfänglichen Eigenwerte hinzugezogen, die größer als 1 sind. Diese Faktoren erklären mehr Varianz als ein einzelner Indikator. In nachfolgender Tabelle 5-1 ist die erklärte Gesamtvarianz der untersuchten 41 Indikatoren aufgeführt. In der ersten durchgeführten Faktorenanalyse zeigt die Gesamtsumme der anfänglichen Eigenwerte (>1), dass eine Faktorenanzahl von 13 gewählt werden müsste, um unter statistischen Gesichtspunkten ein belastbares Ergebnis zu erhalten.

Tabelle 5-1: Faktorenanalyse – Erklärte Varianz der untersuchten 41 Indikatoren

Komponente	Anfängliche Eigenwerte			Rotierte Summen von quadrierten Ladungen		
	Gesamtsumme	% der Varianz	Kumulativ %	Gesamtsumme	% der Varianz	Kumulativ %
1	7,642	18,639	18,639	5,852	14,272	14,272
2	5,897	14,382	33,022	4,431	10,807	25,079
3	4,291	10,465	43,487	4,076	9,942	35,021
4	2,510	6,123	49,610	2,910	7,098	42,119
5	2,112	5,150	54,760	2,849	6,948	49,067
6	1,973	4,811	59,572	2,556	6,235	55,302
7	1,733	4,227	63,798	2,447	5,969	61,271
8	1,659	4,045	67,844	2,229	5,436	66,706
9	1,525	3,719	71,563	1,768	4,311	71,018
10	1,300	3,171	74,734	1,523	3,716	74,734
11	1,200	2,927	77,661			
12	1,077	2,626	80,287			
13	1,004	2,449	82,736			
14	,945	2,304	85,040			
15	,861	2,100	87,140			
...	...	...	...			
40	,001	,002	99,999			
41	,000	,001	100,000			

Diese 13 Faktoren erklären bereits 82,736 % der betrachteten Indikatoren. Es ist hierdurch eine Datenreduktion erfolgt. Würde ein vierzehnter Faktor betrachtet werden, dann würden lediglich 2,3 % der Gesamtvarianz mehr erklärt werden, wodurch sich aber ein erhöhter Interpretationsbedarf einstellen würde. Erst bei der Extraktion von 41 Faktoren würde die gesamte Varianz erklärt. Das bedeutet aber auch, dass keine Datenreduktion erfolgen würde. Es lässt

¹³ Bei der Hauptkomponentenmethode wird versucht möglichst viele der betrachteten Indikatoren in den ermittelten Faktoren abzubilden

sich erkennen, dass mit jeder Datenreduktion auch ein gewisser Informationsverlust einhergeht, in dem hier aufgezeigten Fall ca. 17,3 %. Der Informationsverlust von 17,3 % ist vertretbar, da durch die Extraktion einer geringeren Faktorenanzahl ein Mehrwert generiert wird, der die folgende Clusteranalyse qualitativ (z. B. durch unabhängige Variablen) aufwertet.

Tabelle 5-2: Faktorenanalyse – Übersicht der Zuordnung der Indikatoren (Durchlauf 1)

Faktor	1	2	3	4	5	6	7	8	9	10	11	12	13
Motorisierungsgrad 2014	0,909												
Pkw pro 1000 EW 2014	0,924												
Dieselfahrzeuge pro Tsd. EW 2014	0,907												
Benzinfahrzeuge pro Tsd. EW 2014	0,727												
Elektrofahrzeuge pro Tsd. EW 2014	0,872												
Energieproduktivität GHD	0,553												
HH-Dichte		0,892											
Ant. Siedlungs- und Verk.-fl. an Ges.-fl.		0,803											
Bevölkerungsdichte		0,897											
Energieintensität GHD		-0,57											
Bev.-entw. der letzten 11 Jahre			0,803										
Anteil über 65 Jähriger			-0,84										
Arb.-losenq. bez. abh. Ziv. Erw.			-0,67										
Anteil zwischen 25 und 65 Jähriger			0,767										
Heiztage Mittel (2009-2013)				0,815									
Heizgradtage Mittel (1970-2013)				0,877									
Gradtagzahlen Mittel (1970-2013)				0,849									
Energieproduktivität Industrie					0,938								
Produktivitätsniveau Industrie					0,939								
Beschäftigtenanteil Industrie 2011						-0,70							
Beschäftigtenanteil GHD 2011						0,69							
ÖPNV Fahrten pro EW 2012						0,636							
Durchschnittliche HH Größe							0,737						
Verfügbares Einkommen							0,734						
Ant. Whg. in 1-2 Fam. Häusern							0,589						
Energieintensität Industrie								0,916					
Spez. Energieverbrauch Industrie								0,782					
Produktivitätsniveau GHD								0,59					
Mittlere Reiseweite									0,89				
Personenkilometer pro EW 2012									0,785				
GHD Umsatzst. an Ges.-umsatzst.										0,841			
Ind. Umsatzst. an Ges.-umsatzst										-0,70			
Entw. Besch.-ant. Ind. 2005-2011											-0,85		
Entw. Besch.-ant. GHD 2005-2011											0,701		
Pro-Kopf Stromverbrauch												0,817	
Wohnfläche pro Person												0,587	
Spez. Energieverbrauch GHD													-0,87
Anteil Einpersonenhaushalte 2012	-0,07	0,423	0,423	0,054	-0,15	0,284	-0,48	-0,10	0,139	0,131	-0,04	0,05	0,127
Erdgasverbrauch	-0,04	0,238	0,011	0,329	0,048	0,01	0,34	-0,28	-0,15	0,409	0,088	0,069	0,313
Ant. Ausp. an soz. Vers.-pfl. am Wohnort	-0,05	-0,17	0,13	-0,03	0,405	0,238	-0,05	0,101	-0,32	-0,15	-0,14	-0,34	-0,04
Veränderung Energieintensität 2007-2011	0,065	-0,06	0,032	0,069	-0,37	0,436	0,306	0,125	-0,07	-0,04	-0,20	-0,10	-0,35

Die Tabelle 5-2 stellt die Zuordnungsübersicht der Indikatoren zu den Faktoren im ersten Durchlauf dar. Es fällt auf, dass die Anwendung der Faktorenanalyse auf den Gesamtdatensatz von 41 Indikatoren 13 Faktoren extrahiert. Diese sind zu erkennen, indem die Spalten 1-13 den jeweils grau/weiß hinterlegten Zeilen gegenübergestellt werden. Der größte Faktor ist in dem Fall der Faktor 1 mit dem 6 Indikatoren korrelieren. Die Korrelation der Indikatoren mit dem Faktor wird durch die Faktorladungen angezeigt. Diese sind in dieser Tabelle nur aufgetragen, sofern ein Indikator auf einen Faktor mit einer Faktorladung von mindestens 0,5 lädt. Geringere Faktorladungen wurden für die Indikatoren, die eindeutig einem Faktor zugeordnet werden können, der Übersichtlichkeit halber nicht aufgeführt. Es fällt auf, dass die meisten Indikatoren gute Faktorladungen, in den meisten Fällen $>0,7$, aufweisen. Dennoch zeigt die tabellarische Darstellung, dass vier der 41 Indikatoren auf keinen Faktor ausreichend hoch laden (untere vier Zeilen). Um durch diese Tatsache die spätere Clusteranalyse nicht zu destabilisieren, wurden vier Indikatoren aus dem Gesamtdatensatz eliminiert, um die Belastbarkeit der weiteren Analyse nicht zu gefährden. Die nachfolgende Tabelle 5-3 führt die aus der Untersuchung eliminierten Indikatoren auf. Dabei ist jeweils auch der Durchlauf der Faktorenanalyse mit aufgetragen. Nachdem ein oder mehrere Indikatoren eliminiert wurden, erfolgte eine erneute Faktorenanalyse, wodurch die Stabilität der extrahierten Faktoren überprüft werden soll. Auf die Darstellung der Zuordnungsübersichten eines jeden Durchlaufs wurde hierbei ebenfalls verzichtet. Lediglich die Zuordnungsübersicht des fünften Durchlaufs der Faktorenanalyse wird aufgeführt. Durch das beschriebene Verfahren wurde aus dem anfänglichen Datensatz von 41 Indikatoren in fünf Durchläufen die Indikatoranzahl auf 34 reduziert. Dabei wurden, anders als in der ersten Zuordnungsübersicht zu erkennen ist, 11 Faktoren ermittelt. Das plausible Ergebnis zeigt sich darin, dass trotz einer verringerten Anzahl an Faktoren eine höhere Gesamtvarianz (84,7 %) des Indikatorsatzes beschrieben werden kann, als es mit den 13 Faktoren des ersten Durchlaufs der Fall war (82,7 %). Als Beispiel kann der Indikator Erdgasverbrauch der privaten Haushalte aufgeführt werden. Dieser Indikator war derjenige, bei dem aufgrund der Datenlage die meisten Lücken bestehen und geschlossen werden mussten. Dadurch ist die Linearität des Indikators z. T. verändert und korreliert mit keinem Faktor ausreichend hoch. Es wird deutlich, dass die Faktorenanalyse auch bei der Elimination von Indikatoren mit schlechterer Datenqualität helfen kann.

Tabelle 5-3: Faktorenanalyse – Übersicht der fünf Durchläufe

Durchlauf	Betrachtete Indikatoren	Faktoranzahl	Eliminierte Indikatoren
1	41	13	Anteil Einpersonenhaushalte
			Erdgasverbrauch der priv. HH
			Ant. Absp. an soz. Vers.-pfl. am Wohnort
			Veränderung Energieintensität 2007-2011
2	37	12	Energieintensität GHD
3	36	12	Durchschnittliche HH-Größe
4	35	11	Spez. Energieverbrauch GHD
5	34	11	---

Obwohl sich der Datensatz nun von 41 auf 34 Indikatoren reduziert hat, konnte durch die Elimination von sieben Indikatoren eine stabilere Datenbasis geschaffen werden, wodurch eine belastbare Beschreibung der untersuchten Städte erreicht werden kann.

Tabelle 5-4: Faktorenanalyse – Übersicht der Zuordnung der Indikatoren (Durchlauf 5)

Faktor	1	2	3	4	5	6	7	8	9	10	11
Pkw pro 1000 EW 2014	0,912										
Dieselfahrzeuge pro Tsd. EW 2014	0,9										
Motorisierungsgrad 2014	0,895										
Elektrofahrzeuge pro Tsd. EW 2014	0,88										
Benzinfahrzeuge pro Tsd. EW 2014	0,715										
Energieproduktivität GHD	0,578										
HH-Dichte		0,906									
Bevölkerungsdichte		0,901									
Ant. Siedl.- und Verk.-fl. an Ges.-fl.		0,852									
Ant. Whg. in 1-2 Fam. Häusern		-0,56						0,521			
Anteil über 65 Jähriger			-0,87								
Bev.-entw. der letzten 11 Jahre			0,804								
Anteil zwischen 25 und 65 Jähriger			0,776								
Arb.-losenq. bez. abh. Ziv. Erw.			-0,65								
Heizgradtage Mittel (1970-2013)				0,914							
Gradtagzahlen Mittel (1970-2013)				0,851							
Heiztage Mittel (2009-2013)				0,817							
Energieintensität Industrie					0,963						
Spez. Energieverbrauch Industrie					0,748						
Produktivitätsniveau GHD					0,721						
Beschäftigtenanteil Industrie 2011						-0,70					
Beschäftigtenanteil GHD 2011						0,71					
ÖPNV Fahrten pro EW 2012						0,684					
Energieproduktivität Industrie							0,965				
Produktivitätsniveau Industrie							0,964				
Wohnfläche pro Person								0,693			
Pro-Kopf Stromverbrauch								0,648			
Verfügbares Einkommen								0,635			
Ind. Umsatzst. an Ges.-umsatzst									0,835		
GHD Umsatzst. an Ges.-umsatzst									-0,82		
Mittlere Reiseweite										0,902	
Personenkilometer pro EW 2012										0,784	
Entw. Besch.-ant. Ind. 2005-2011											-0,89
Entw. Besch.-ant. GHD 2005-2011	0,566										0,718

Die elf ermittelten Faktoren sind der Tabelle 5-4 zu entnehmen. Dabei sind die jeweils zu einem Faktor gehörenden Indikatoren gekennzeichnet. Im Folgenden werden die Faktoren kurz erläutert und deren mögliche Aussage mithilfe der Indikatoren identifiziert. Dabei werden die Faktorenladungen zu Hilfe genommen. Hier zeigen betragsmäßig hohe Faktorenladung an, dass eine sehr gute Übereinstimmung des Indikators mit dem Faktor vorliegt. Eine negative bzw. positive Übereinstimmung zeigt, dass ein Indikatorwert besonders gering (negativ) bzw. hoch (positiv) ausfällt.

Der **Faktor 1** besteht aus sechs Indikatoren und korreliert mit sämtlichen Indikatoren des motorisierten Individualverkehrs (MIV) sowie der *Energieproduktivität des GHD-Sektors*, der mit der geringsten Faktorladung nur wenig in den Faktor einfließt. Dennoch laden alle Indikatoren positiv, was bedeutet, dass je höher die Indikatorwerte ausfallen, desto höher fallen auch die Faktorwerte aus, die auf den einzelnen Indikatorwerten basieren. Dieser Faktor kann aufgrund des Schwerpunkts im Bereich des MIV als **Individualverkehrsfaktor** bezeichnet werden.

Faktor 2 besteht aus vier Indikatoren, der die **Gemarkungsfläche der betrachteten Städte in Bezug zu weiteren Größen wie den Haushalten, der Bevölkerung oder der Siedlungs- und Verkehrsfläche setzt**. Diese drei Größen laden stark positiv auf den Faktor (Faktorladungen $>0,8$). Ebenfalls lädt der Indikator *Anteil Wohnungen in 1- und 2-Familienhäusern* auf den Faktor 2, jedoch negativ. Diese Konstellation zeigt ein logisches Bild, das sich in den betrachteten Städten zeigt. Städte mit einer hohen Bevölkerungsdichte weisen auch eine hohe Haushaltsdichte sowie einen hohen *Anteil an Siedlungs- und Verkehrsflächen an der Gesamtfläche* der Stadt auf. Zeitgleich haben diese Städte tendenziell einen niedrigen *Anteil an Wohnungen in 1- und 2-Familienhäusern* (negative Faktorladung).

Ebenfalls setzt sich der **Faktor 3** aus vier Indikatoren zusammen. Diese Indikatoren bilden den **sozio-demographischen Faktor** ab, da das Alter der Bevölkerung (*Anteil über 65-Jähriger, Anteil Personen zwischen 25 und 65 Jahren*), als auch die *Bevölkerungsentwicklung* und die *Arbeitslosenquote* den Faktor abbilden. Der *Anteil über 65-Jähriger* als auch die *Arbeitslosenquote* weisen negative Faktorenladungen auf. Die beiden anderen Indikatoren laden positiv, sodass sich der Zusammenhang ableiten lässt, dass in Städten mit einer Bevölkerungszunahme und einem großen Anteil an Einwohnern im Altersschnitt zwischen 25 und 65 Jahren, wenig Arbeitslosigkeit herrscht und der Anteil der über 65 Jährigen tendenziell eher gering ist.

Der **Faktor 4** wird durch klimatische Indikatoren gebildet, wodurch dieser als **klimatischer Faktor** bezeichnet werden kann. Die gemittelten Heiztage, Heizgradtage und Gradtagzahlen laden alle positiv auf diesen Faktor, wodurch sich ableiten lässt, welche Stadt in einer klimatisch eher kälteren oder wärmeren Region des Bundesgebietes liegt.

Faktor 5 vereint in drei Indikatoren zwei **Wertschöpfungsindikatoren**, die den Energieverbrauch ins Verhältnis zur geleisteten Wertschöpfung setzen. Der dritte Indikator ist das *Produktivitätsniveau des GHD Sektors*. Dieser Zusammenhang korreliert statistisch mit dem Faktor 5. Wird das Ergebnis gedeutet, so lässt sich feststellen, dass die Indikatoren allesamt positive Faktorenladungen aufweisen. Das *Produktivitätsniveau im GHD-Sektor* scheint besonders in den Städten gut zu sein, in denen die Industrie hohe Energieaufwendungen zur Erzeugung der anfallenden Bruttowertschöpfung aufwendet. In diesen Städten generiert der GHD-Sektor eine hohe Bruttowertschöpfung bei geringem Energieeinsatz.

Die **Beschäftigtenanteile der Sektoren GHD und Industrie** laden betragsmäßig in etwa gleichhoch auf den **Faktor 6**. Zudem lädt der Indikator *ÖPNV Fahrten je Einwohner und Jahr* ebenfalls positiv auf diesen Faktor. Dieser Zusammenhang zeigt, dass in Städten mit einem hohen Beschäftigtenanteil im GHD-Sektor (oder geringem Anteil an beschäftigten im Industriesektor) die Einwohner den ÖPNV deutlich häufiger nutzen als umgekehrt.

Im **Faktor 7** wird die **Industrieproduktivität** einer Stadt abgebildet. Die beiden Indikatoren *Energieproduktivität* und *Produktivitätsniveau* laden beide stark positiv auf diesen Faktor. Der Zusammenhang beider Indikatoren, der bereits bei der Erarbeitung des Datensatzes vermutet wurde, wird deutlich herausgestellt. Ein ähnliches Bild, das bereits bei der Datensatzerarbeitung vermutet wurde bestätigt sich auch für den **Faktor 9** und **Faktor 11**. Faktor 9 besteht aus den beiden Indikatoren *Industrieumsatzsteuer an der Gesamtumsatzsteuer* und *GHD-Umsatzsteuer an der Gesamtumsatzsteuer*, die betragsmäßig gleichhoch laden. Ein identisches Bild zeigt sich bei Faktor 11, bei dem die Indikatoren **Entwicklung der Beschäftigtenzahlen** der beiden Sektoren GHD und Industrie betragsmäßig ähnlich laden.

Der **Faktor 8** stellt, anders als die soeben erläuterten Faktoren, in gewisser Weise den **Wohlstand, bzw. die finanzielle Struktur** einer Stadt bezogen auf die privaten Haushalte dar. Die Indikatoren verfügbares Einkommen, Wohnfläche pro Person und der *Pro-Kopf Stromverbrauch* korrelieren allesamt positiv und sind demnach gleichgerichtet. Das bedeutet, dass in Städten, in denen das verfügbare Einkommen höher liegt, sich die Leute auch größere Wohnungen leisten können. Zudem ist so auch der höhere Stromverbrauch erklärbar, da dann eine bessere Haushaltsausstattung mit elektrischen Geräten vorzufinden ist.

Der **Faktor 10** wird als letzter Faktor betrachtet. Dieser wurde aus den Indikatoren *Mittlere Reiseweite im ÖPNV* und der *Anzahl an Personenkilometern im ÖPNV pro Person und Jahr* extrahiert. Dieser Faktor kann auch der **ÖPNV Faktor** bezeichnet werden, da beide Indikatoren den ÖPNV beschreiben und die Fahrleistung auf eine vergleichbare Größe in zwei Indikatoren transformieren. Beide Indikatoren laden positiv.

Die gesamte Übersicht der Faktorenwerte ist den betrachteten Städten im Anhang gegenübergestellt. Diese erhobenen Faktorenwerte werden nun für die im nachfolgenden Unterkapitel thematisierten Clusteranalyse verwendet.

5.3 Clusteranalyse und Auswahl der Clusteranzahl

Das Ziel dieser Untersuchung stellt die Clusteranalyse der betrachteten Städte dar. Aufbauend auf den Ergebnissen der durchgeführten Z-Transformation sowie der Extraktion von 11 Faktoren in der Faktorenanalyse in den vorherigen Unterkapiteln, wird nun die hierarchische agglomerative Clusteranalyse unter Verwendung des Ward-Algorithmus durchgeführt (siehe Ka-

pitel 2 und 3). Teile der Clusteranalyse wurden bereits in [WAL2016B] publiziert. Die hier durchgeführte Analyse führt einen Schritt weiter. Die ermittelten Städtegruppen weisen vereinzelt Cluster auf, die die Clustergrundidee nicht wirklich erfüllen, da auch einzelne Städte Cluster bilden. Diese Cluster werden in weiterführenden Schritten, äquivalent zur Faktorenanalyse, bei den Indikatoren, die mit keinem Faktor korrelieren eliminiert werden, ebenfalls eliminiert. Dadurch soll eine finale Clusterlösung generiert werden, die im Kapitel 6 näher diskutiert werden.

5.3.1 Durchführung des zweiten Clusteranalysedurchlaufs

Die hierarchisch agglomerative Clusteranalyse als auch deren wesentlichen Fusionsalgorithmen wurden bereits in Abschnitt 2.3.5 sowie Kapitel 3 vorgestellt und auf einen reduzierten Datensatz angewendet. Es hat sich herausgestellt, dass die Clusterung mit einem Untersuchungsrahmen von 107 Objekten und unter der Berücksichtigung weiterer Aspekte wie z. B. der Ergebnisse sowie der Schwächen der einzelnen Fusionsalgorithmen, dass der Ward Algorithmus als beste Wahl identifiziert werden kann.

Die ermittelten 11 Faktoren beschreiben die deutschen kreisfreien Städte und erlauben die gesammelte Interpretation der Städtegruppen. So wurden auch die Faktoren für die Clusterung der Städte verwendet. Zu Beginn einer Clusterung weist der Untersuchungsrahmen gleichviele Cluster wie betrachtete Objekte auf. In hierarchisch agglomerativen Verfahren wachsen die Cluster mit dem Fortlaufen der Clusteranalyse zu immer größeren Clustern an und die Clusteranzahl verringert sich, bis letztlich alle betrachteten Städte in einem einzigen Cluster vereint sind. Beispielhaft ist dieses hierarchische Vorgehen der Clusteranalyse in Abbildung 5-2 aufgeführt.

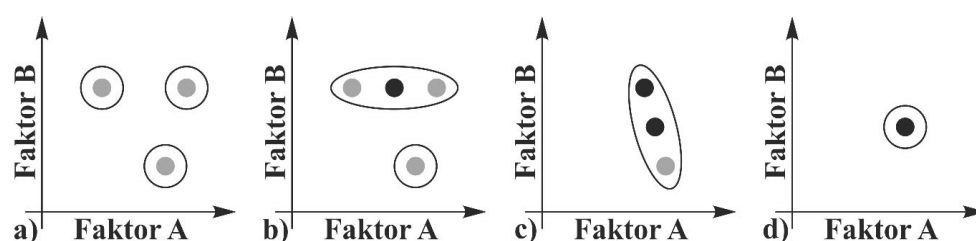


Abbildung 5-2: Clusteranalyse – Schematische Clusterzusammenführung (aus [WAL2014])

Es werden dabei stets die Städte und im späteren Clusteranalyseverlauf auch Städtecluster zu weiteren Clustern vereint, die sich mathematisch am meisten ähneln. Diese „Ähnlichkeit“ wird mithilfe des Varianzkriteriums (Abschnitt 3.2.4) gemessen. Diese Maßzahl wird als Koeffizient für die Zuordnung der Städte zu Clustern herangezogen. Zudem erlaubt die Entwicklung dieses Koeffizienten mit jedem Fusionsschritt eine sinnvolle Wahl der Clusteranzahl. Dieses Verfahren wurde mehrfach auf die kreisfreien Städte angewendet. Durch die mehrmalige An-

wendung wurden aussagekräftige Städtecluster erzeugt. Im ersten Durchgang der Clusteranalyse der 107 kreisfreien deutschen Städte konnten 12 Städtecluster erzeugt werden. Die Ergebnisse des ersten Durchlaufs der Clusteranalyse wurden bereits in [WAL2016B] publiziert.

Dabei kam heraus, dass vier Städte Einzelcluster bilden. Cluster, die aus nur einer Stadt bestehen, erfüllen die Clusteridee nicht, da diese oftmals Extremfälle darstellen oder evtl. Fehler in der Datenbasis aufweisen können. Deshalb wurden für weitere Durchläufe diese „ein Stadt Cluster“ eliminiert. In der ersten Durchführung der Clusteranalyse aller 107 Städte konnten die in der Tabelle 5-5 aufgeführten Städte als „ein Stadt Cluster“ identifiziert werden. Diese wurden vor der Durchführung des zweiten Durchlaufs aus dem Untersuchungsrahmen entfernt.

Tabelle 5-5: Clusteranalyse – Ein Stadt Cluster des ersten Durchlaufs (107 Städte)

Cluster/Stadt	Extremfaktor	Ausprägung
Wolfsburg	Faktor 1	Dominanz des MIV
Bonn	Faktor 8	Hohe Wohnfläche pro Kopf, hoher Stromverbrauch, hohes verfügbares Einkommen
Erlangen	Faktor 10	Starke ÖPNV Nutzung
Potsdam	Faktor 5	Industrie Energieintensität und Produktivität GHD hoch

Der zweite Clusteranalysedurchlauf erfolgte mit 103 Städten im Untersuchungsrahmen, die weiteren Parameter blieben unverändert. Die nachfolgende Tabelle 5-6 stellt die Zuordnungsübersicht der betrachteten Städte im zweiten Durchlauf dar. In dieser Tabelle sind ausschließlich die Fusionskoeffizienten der Clusteranalyseschritte aufgetragen, die einen relevanten Bereich im Hinblick auf die Wahl einer geeigneten Clusteranzahl darstellen. Die Tabelle wurde um die übrigen Abschnitte gekürzt, um die Clusteranzahl des zweiten Durchlaufs fokussiert ermitteln zu können.

Tabelle 5-6: Clusteranalyse – Auswahl aus der Zuordnungsübersicht der Fusionsschritte (103 Städte)

Fusionsschritt	Fusionskoeffizienten	Steigung	Differenz zwischen den Schritten	Clusteranzahl
1	0,364	0,364	0,623	102
2	0,949	0,584	0,731	101
3	1,748	0,800	0,991	100
...	...	...	...	...
83	311,640	12,407	0,984	20
84	324,255	12,615	0,862	19
85	338,897	14,642	0,980	18
86	353,834	14,937	0,970	17
87	369,232	15,398	0,832	16
88	387,735	18,503	0,980	15
...	...	...	...	...
94	522,557	25,853	0,953	9
95	549,694	27,136	0,846	8
96	581,751	32,057	0,929	7
97	616,254	34,503	0,788	6
98	660,032	43,778	0,888	5
99	709,318	49,286	0,835	4

Auch im zweiten Clusteranalysedurchlauf wurde mithilfe des Fusionskoeffizienten mehr als nur eine mögliche Clusteranzahl als mathematisch sinnvoll ermittelt. Die Clusteranzahl wird unter Zuhilfenahme der Fusionskoeffizienten sowie der daraus errechneten Steigung der Werte errechnet. Die dabei entstehende Kurve wird auch als Scree-Kurve oder Scree-Diagramm bezeichnet und erlaubt die Anwendung des „Ellenbogenkriteriums“. Dieses Kriterium stellt die Steigungsanstiege über der Clusteranzahl dar. [BAC2010] Das bedeutet, dass bei der Fusionierung eines Clusters die Kurve einen Steigungsanstieg erfährt, da die Fusionskoeffizienten ansteigen. Wenn durch eine Zusammenführung ein heterogenes Cluster entsteht, kann der Anstieg der Kurve wie ein „Ellenbogen“ aussehen. Auf die Darstellung der Kurve wurde an dieser Stelle verzichtet, da bei großen Datenumfängen die Kurve stark vergrößert betrachtet werden muss, um das „Ellenbogenkriterium“ visuell anwenden zu können. Um die Clusteranzahl bestimmen zu können, wird die mathematische Berechnung der Differenzen in der Steigung der Fusionsschritte angewendet. Die Fusionskoeffizienten, die Steigung und die Differenzen sind in Tabelle 5-6 abgebildet. Mithilfe dieser Werte kann in einem letzten Schritt die Clusteranzahl bestimmt werden. Dafür ist es notwendig die Anzahl der der Clusteranalyse zugeführten Objekte zu kennen. In diesem zweiten Analysedurchlauf waren es 103 Städte. Zudem benötigen wir den Fusionsschritt in dem die Steigung verstärkt ansteigt. In der Tabelle 5-6 geschieht dies bei den Fusionsschritten 1, 2, 84, 87, 95, 97, 98 und 99. Diese Schritte sind in der Tabelle als grau hinterlegte Zeilen dargestellt. Mit der Gleichung 5-2 kann nun die mathematisch sinnvolle Clusteranzahl bestimmt werden. [BUE2012]

$$\text{Clusteranzahl} = \text{Anzahl Städte} - \text{Schritt mit Steigungsanstieg} \quad \text{Gleichung 5-2}$$

Mathematisch betrachtet entsprechen somit die 102, 101, 19, 16, 8, 6, 5 und die 4 Clusterlösung einer sinnvollen Lösung. Auch hier sind die beiden ersten Clusterlösungen 102 und 101 zu ignorieren, da an dieser Stelle, äquivalent zum ersten Clusteranalysedurchlauf, die Clustergrundidee nicht erfüllt wird und die Clusterlösung demnach mathematisch sowie statistisch keinen Sinn ergibt. Als aussagekräftigste Clusterlösungen wurden die 8, 6, 5 und 4 Clusterlösung identifiziert. Im ersten Durchlauf [WAL2016B] wurde eine Clusteranzahl von 12 Clustern als sinnvollste Lösung identifiziert und folgend vier „ein Stadt Cluster“ eliminiert. Damit kann von einer Clusteranzahl unter zehn ausgegangen werden. Aus diesem Grund werden die beiden Clusterlösungen 19 und 16 nicht weiter betrachtet. Die Fusionskoeffizienten zeigen, dass sich aufgrund der nun um vier Städte geringeren Objektanzahl im Untersuchungsrahmen in den ersten 96 Fusionsschritten keine Veränderungen in der Clustergröße und Clusterzusammensetzung aufgrund der neu zu berechnenden Varianzkriterien einstellen. Da jedoch die verwendeten Variablen unverändert bleiben und lediglich die Z-transformierten Variablensätze der eliminierten Städte entfernt wurden, sind in den ersten 96 Fusionsschritten keine wesentlichen Veränderungen aufgetreten. Veränderungen in den Fusionskoeffizienten treten ab dem

97 Schritt auf, da auch im ersten Fusionsdurchlauf die eliminierten vier Städten erst ab diesem Schritt fusioniert wurden.

So wurde mithilfe des Fusionskoeffizienten die 8 Clusterlösung als sinnvoll erachtet. Diese Clusterlösung zeigt erneut, dass durch die Elimination der vier Städte Wolfsburg, Bonn, Erlangen und Potsdam die Clusterzusammensetzung nicht verändert wurde. Die so ermittelten Cluster sind stabil, da trotz der Veränderung der Objektanzahl keine Veränderung in der Clustergröße und Clusterzusammensetzung aufgetreten ist. Die Abbildung 5-3 stellt das Dendrogramm der Fusionsschritte des zweiten Clusteranalysedurchlaufs mit 103 Städten dar. Die 8 Clusterlösung ist dort durch eine Strich-Punkt Linie eingetragen.

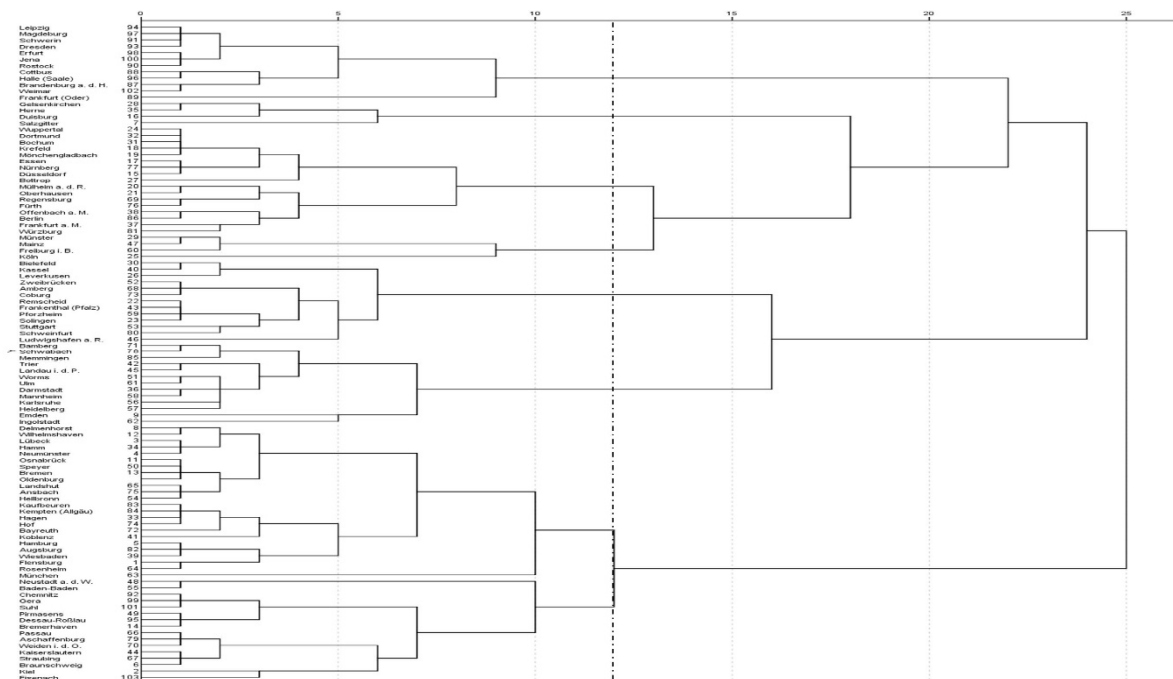


Abbildung 5-3: Clusteranalyse – Dendrogramm der Fusionsschritte (103 Städte)

Das Dendrogramm ist eine grafische Darstellung der Fusionsschritte einer Clusteranalyse. Die einzelnen Schritte lassen sich hier nachvollziehen, wobei das Distanzmaß dimensionslos ist. Waagrechte Linien stehen hierbei für Objekte oder Cluster. Senkrechte Linien stellen Clusterfusionen dar. Die Strichpunktlinie zeigt das Niveau an, ab dem die hier gewählte acht Clusterlösung zu einer sieben Clusterlösung werden würde. Die Linie stellt demnach die acht Clusterlösung dar (acht waagrechte Linien links der Strichpunktlinie). Je länger eine waagrechte Linie bis zu einer Fusion (senkrechte Linie) ist, desto inhomogener wird das entstehende Cluster. Eine zu hohe Clusteranzahl erfüllt die Clustergrundidee nicht. Eine zu geringe Clusteranzahl führt zu inhomogenen Clustern, da Objekte zu Clustern vereint werden, die in ihren Ausprägungen verschieden sind, auf Städte bezogen würden sich somit äußerst ungleiche Städte zu Clustern vereinen.

5.3.2 Analyse und Auswertung der Clusterergebnisse der kreisfreien Städte

In diesem Unterkapitel werden die ermittelten Städtecluster näher betrachtet und deren Zusammensetzung durch Bezüge zu den Faktoren sowie Indikatoren analysiert und deren mögliche Aussage gedeutet. Mithilfe der Clusteranalyse konnten 8 Clustergruppen ermittelt werden, die gemessen an den Indikatoren des Untersuchungsrahmens statistisch als ähnlich bezeichnet werden können. Abschließend werden die Ergebnisse gedeutet sowie die durchgeführten Schritte auf deren positiven als auch negativen Nutzen reflektiert.

Die Clusteranalyse hat nach der Elimination von vier „ein Stadt Clustern“ acht finale Cluster als Clusterlösung ergeben. Dabei wurden die Clusterlösungen in der Tabelle 5-7 mit den dem Cluster angehörenden Städten aufgetragen. Das größte der acht Cluster umfasst 24 Städte. Fünf weitere Cluster umfassen jeweils 12 bis 17 Städte. Die kleinsten identifizierten Cluster bestehen aus jeweils vier Städten und umfassen Städte, die in kleineren Clustern sinnvoller als Gruppe abgebildet werden können, als durch eine Fusion mit den größeren Clustern.

Tabelle 5-7: Clusteranalyse – Die acht ermittelten Cluster sowie die zugeordneten 103 Städte

Cluster	Zugeordnete Städte
Cluster 1	Flensburg, Lübeck, Neumünster, Hamburg, Delmenhorst, Oldenburg, Osnabrück, Wilhelmshaven, Bremen, Hagen, Hamm, Wiesbaden, Koblenz, Speyer, Heilbronn, München, Rosenheim, Landshut, Bayreuth, Hof, Ansbach, Augsburg, Kaufbeuren, Kempten (Allgäu) (24 Städte)
Cluster 2	Kiel, Braunschweig, Bremerhaven, Kaiserslautern, Neustadt a. d. W., Pirmasens, Suhl, Baden-Baden, Passau, Straubing, Weiden i. d. O., Aschaffenburg, Chemnitz, Gera, Dessau-Roßlau, Eisenach (16 Städte)
Cluster 3	Salzgitter, Duisburg, Gelsenkirchen, Herne (4 Städte)
Cluster 4	Emden, Darmstadt, Trier, Landau i. d. P., Worms, Karlsruhe, Heidelberg, Mannheim, Ulm, Ingolstadt, Bamberg, Schwabach, Memmingen (13 Städte)
Cluster 5	Düsseldorf, Essen, Krefeld, Mönchengladbach, Mülheim a. d. R., Oberhausen, Bottrop, Wuppertal, Bochum, Dortmund, Frankfurt a. M., Offenbach a. M., Regensburg, Fürth, Nürnberg, Würzburg, Berlin (17 Städte)
Cluster 6	Remscheid, Solingen, Leverkusen, Bielefeld, Kassel, Frankenthal (Pfalz), Pforzheim, Amberg, Ludwigshafen a. R., Zweibrücken, Stuttgart, Coburg, Schweinfurt (13 Städte)
Cluster 7	Köln, Münster, Mainz, Freiburg i. B. (4 Städte)
Cluster 8	Brandenburg a. d. H., Cottbus, Frankfurt, Rostock, Schwerin, Dresden, Leipzig, Halle (Saale), Magdeburg, Erfurt, Jena, Weimar (12 Städte)

Bei der Betrachtung der Städtegruppen in obiger Tabelle neigt der Betrachter dazu die Clusterzusammensetzung zu interpretieren. Dabei wird oftmals die eigene Erfahrung hinzugezogen, wodurch Fragen auftreten können, weshalb die eine oder andere Stadt sich in einer Gruppe mit einer Stadt wiederfindet, die basierend auf der eigenen Erfahrung nicht in derselben Gruppe vermutet wird. Daran werden die Problematik und auch der Vorteil der Clusteranalyse deutlich. Die Ergebnisse müssen gedeutet werden, jedoch anders als bei dem aufgeführten Beispiel nicht subjektiv sondern Objektiv. Die Faktoren sowie die Indikatoren müssen

bei der Interpretation hinzugezogen werden. Das verdeutlicht die multivariate Herangehensweise der Clusteranalyse bei der zeitgleich sämtliche betrachteten Variablen in die Analyse einfließen. Die eigene Erfahrung, die oftmals Fragen bei der Betrachtung der Städtegruppen aufwirft, basiert lediglich auf dem Vergleich einer oder einiger wenigen Variablen wie z. B. die Stadtgröße und z. B. die bekannte lokale Industrie einer Stadt. Da jedoch viele weitere Indikatoren in dieser Untersuchung dazwischenliegen, ist eine solide Clusterdeutung nur mithilfe der Ausprägungen einer Clusterlösung möglich. Die Ausprägungen bauen auf den Variablen auf. D. h. die Faktoren, die eine Stadt mathematisch beschreiben werden weiter rückverfolgt bis zu den einzelnen Indikatoren, die durch den Faktor abgebildet werden. In Tabelle 5-8 sind die Mittelwerte der Faktoren der acht Cluster dargestellt. Die jeweils stärkste Ausprägung eines Faktors ist dabei grau hinterlegt. Bereits diese Tabelle zeigt die verschiedenen starken Faktorwerte der einzelnen Cluster.

Tabelle 5-8: Clusteranalyse – Clustermittelwerte der Faktoren

Cluster	1	2	3	4	5	6	7	8
Faktor 1	-0,139	0,033	-0,244	0,070	-0,036	0,156	-0,187	-0,503
Faktor 2	0,112	-0,495	0,768	-0,672	1,052	0,092	-0,359	-0,540
Faktor 3	-0,061	-0,799	-0,930	1,207	0,048	-0,075	1,099	-0,335
Faktor 4	0,931	0,347	-1,084	-0,237	-0,841	-0,306	-1,302	0,180
Faktor 5	-0,353	0,214	-0,341	-0,261	-0,075	0,126	-0,982	0,304
Faktor 6	0,070	0,064	-1,067	-0,860	0,270	-0,374	1,832	0,375
Faktor 7	0,563	0,332	-0,517	0,006	0,220	0,298	-0,607	-1,770
Faktor 8	-0,164	-0,247	2,961	-0,204	-0,138	-0,181	-0,606	-0,152
Faktor 9	-0,189	-0,636	-0,251	0,089	-0,438	1,921	-0,311	-0,067
Faktor 10	-0,368	0,362	0,227	-0,556	-0,147	0,260	1,132	-0,142
Faktor 11	0,604	-0,365	-0,073	-0,137	-0,281	-0,127	0,946	-0,519

Die betrachteten Städte weisen stadtspezifische Ausprägungen in den elf Faktoren auf. Diese fließen in die Faktorwerte der ermittelten Cluster ein. Um diese Werte in Tabelle 5-8 einfacher betrachten zu können, sind die clusterspezifischen Faktorwertediese in Abbildung 5-4 grafisch aufgetragen. Die grafische Darstellung beinhaltet viel Information. Ziel der Darstellung ist es nicht die jedoch nicht die Tabelle 5-8 zu ersetzen sondern vielmehr eine grafische Hilfe für die Deutung der Cluster bilden. Die Abbildung zeigt grafisch, dass die Clusterausprägungen auf Basis der Faktormittelwerte verschieden sind. Die exakten Werte können in der bereits aufgeführten Tabelle genauer eingesehen werden.

Bei der grafischen Darstellung fällt auf, dass die ersten beiden Cluster (24 und 16 Städte) die geringsten Ausschläge in den Faktorwerten aufweisen. Besonders bei dem ersten Cluster bestätigt das Ergebnis die Vermutung, dass sich mehr Städte ähneln je näher die Faktorausprägungen der Städte am Mittelwert liegen.

Das andere Extremum bildet das Cluster 7. Dort fallen fünf der betrachteten elf Faktoren am höchsten aus. Dieses „Extremcluster“ mit den starken Ausprägungen in den Faktorwerten beschreibt nur ein kleines Cluster (4 Städte). Die höchste Faktorausprägung eines Clusters kann für das Cluster 3 festgestellt werden. Es ist ebenfalls ein Cluster bestehend aus vier Städten. Dieser Zusammenhang verdeutlicht die theoriegeleitete Vermutung, dass je mehr eine Stadt einen Sonderfall darstellt, umso geringer ist die Wahrscheinlichkeit, dass in einem zufälligen Betrachtungsrahmen weitere Städte mit ähnlichen Extremwerten enthalten sind. Die Cluster 4, 5, 6 und 8 weisen in jeweils einem Faktor den größten Faktorwert auf. Diese vier Cluster sind kleiner als das Cluster 1 und deutlich größer als die beiden Extremcluster 3 und 7.

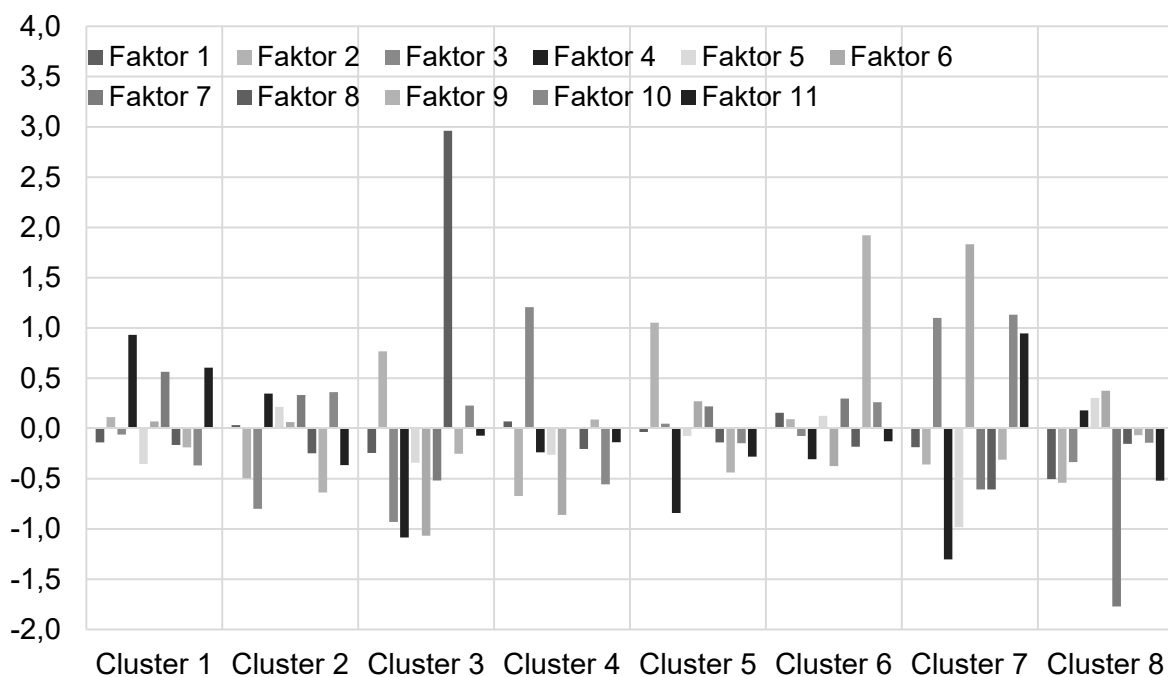


Abbildung 5-4: Clusteranalyse – Mittelwerte der 8 Cluster bezogen auf die 11 Faktoren

Ergänzend zu den Clustermittelwerten der einzelnen Faktoren können die Mittelwerte der Einzelnen Indikatoren zur Interpretation der ermittelten Cluster herangezogen werden. Hier ist die in der Faktoranalyse bereits erläuterte Faktorenladung ein hilfreiches Instrument. Die nachfolgende Tabelle 5-9 beinhaltet alle Faktoren, aufgeschlüsselt nach den dazugehörigen Indikatoren sowie den jeweiligen Faktorladungen. So kann zu jedem der acht ermittelten Cluster der jeweilige Indikatormittelwert mit den der anderen Cluster verglichen und so Schlüsse bezüglich der Ausprägungen einzelner Städtegruppen gezogen werden. Die arithmetischen Indikatormittelwerte wurden aus den Indikatorwerten der Städte, die einem Cluster angehören, berechnet. So kommt es, dass ähnlich wie bei den Faktorwerten, die gemittelten Indikatorwerte verschieden groß ausfallen. Diese Differenzen erklären somit die verschiedenen Faktorwerte, da mithilfe der Faktorenladungen auf diese geschlossen werden kann.

Tabelle 5-9: Clusteranalyse – Clustermittelwerte der Ausgangsindikatoren und jeweiligen Faktorladungen

Faktor	Indikator	Faktorladung	Cluster							
			1	2	3	4	5	6	7	8
1	Pkw pro 1000 EW	0,912	502	523	468	514	478	513	439	434
	Dieselfz./Tsd. EW	0,9	154	152	111	168	138	152	146	111
	Motorisierungsgrad	0,895	543	562	507	559	518	556	473	459
	Elektrofz./Tsd. EW	0,88	0,17	0,20	0,07	0,23	0,13	0,23	0,27	0,13
	Benzinfz./Tsd. EW	0,715	340	364	347	340	332	354	286	332
	Energieprod. GHD	0,578	577	510	649	619	627	630	659	465
2	HH-Dichte	0,906	674	408	969	523	1131	712	878	530
	Bevölkerungsdichte	0,901	1.339	781	1.998	1.057	2.213	1.472	1.752	982
	Siedl.- u. Verk.-fl. an Ges.-fl.	0,852	43,5	33,8	60,5	35,0	57,1	45,6	44,1	30,1
	Whg. in 1-2Fam.-H.	-0,56	30,5	29,1	23,8	33,6	22,3	31,2	24,2	16,8
3	Ant. ü. 65 Jähriger	-0,87	21,1	23,5	21,9	19,2	20,2	21,2	18,3	22,8
	Bev.-ent. 11 Jahre	0,804	1,12	-3,88	-7,11	4,58	0,72	-1,04	9,27	-0,23
	Ant.zw. 25 u. 65 J.	0,776	63,2	62,5	61,7	64,9	64,3	62,8	66,3	64,3
	Arbeitslosenquote	-0,65	8,8	10,0	14,0	6,3	10,7	8,8	7,7	12,0
4	HGT (1970-2013)	0,914	2.602	2.504	2.084	2.356	2.151	2.348	2.103	2.515
	GTZ (1970-2013)	0,851	3.905	3.876	3.385	3.678	3.446	3.692	3.378	3.874
	HT (2009-2013)	0,817	279	266	251	264	251	261	260	260
5	Energieint. Ind.	0,963	4,29	1,89	61,96	4,36	4,81	5,32	3,95	2,33
	Spez. Energieverbr. Industrie	0,748	3,45	1,80	49,25	3,29	4,33	3,79	5,04	1,72
	Prod.-niv. GHD	0,721	577,0	509,8	649,3	619,2	627,2	629,9	659,2	464,9
6	Besch.-ant. Ind.	-0,7	17,1	19,9	29,3	24,5	19,2	29,2	11,4	14,5
	Besch.-ant. GHD	0,71	82,5	79,6	70,5	75,1	80,5	70,5	88,3	85,2
	ÖPNV Fahrten p. EW	0,684	142	134	154	87	244	157	336	170
7	Energieprod. Ind.	0,965	406,3	802,8	18,4	536,2	439,4	676,3	367,3	739,0
	Prod.-niveau Ind.	0,964	500,9	815,3	22,2	651,2	516,1	821,6	385,0	917,4
8	Wohnfl. pro Person	0,693	41,34	43,39	38,88	40,65	39,01	41,11	37,83	37,08
	Pro-Kopf Stromver.	0,648	1.876	1.832	1.834	1.775	1.905	1.832	1.794	1.580
	Verfügb. Einkommen	0,635	41.672	38.550	35.485	42.426	39.128	42.238	39.447	30.188
9	Ind. Ums. an Ges.-umsatzsteuer	0,835	36,76	30,88	28,21	41,39	30,54	63,30	31,40	40,93
	GHD Ums. an Ges.-umsatzsteuer	-0,82	53,10	61,32	48,86	53,04	58,77	30,53	66,12	53,42
10	Mittlere Reiseweite	0,902	4,76	6,75	5,18	4,48	4,47	5,24	6,30	4,93
	P.km. pro EW	0,784	781	845	793	559	1.019	1.008	2.196	859
11	Entwicklung Beschäftigtenanteil Industrie	-0,89	-12,9	-4,2	-4,1	-4,9	-7,8	-5,8	-11,2	-0,5
	Entwicklung Beschäftigtenanteil GHD	0,718	3,2	1,2	3,3	2,5	1,9	2,6	1,7	0,4

Die so individuellen Ausprägungen der Städte eines einzelnen Clusters erlaubt eine Deutung. An dieser Stelle werden die betragsmäßig stärksten Faktoren zu jedem Cluster unter Zuhilfenahme der Indikatoren fokussiert, um so die Typisierung deutscher kreisfreier Städte auf die aussagekräftigsten Ausprägungen zu stützen und diese nachvollziehbar erklären zu können.

Das erste Cluster ist das größte Städtecluster dieser Untersuchung und vereint 24 Städte (s. Tabelle 5-10). Alle Städte sind in keinem der erhobenen Faktoren extrem ausgeprägt. Vielmehr liegen die Faktoren- als auch die Indikatorwerte nah am Mittelwert, was die besonders große Anzahl an Städten in diesem Cluster erklärt. Die stärkste Faktorausprägung dieses

Clusters lässt sich für Faktor 4 identifizieren. Der Faktor 4 vereint dabei die klimatischen Indikatoren *Heiztage*, *Heizgradtage* sowie *Gradtagzahlen*, die allesamt stark miteinander korrelieren und dementsprechend hohe Faktorenladungen aufweisen. Die Städte des ersten Clusters liegen demnach in Regionen, die allesamt die kältesten des Untersuchungsrahmens sind. Ebenfalls lässt sich für die Städte dieses Clusters eine besonders starke Verringerung des Beschäftigtenanteils im Industriesektor feststellen. Zeitgleich erfolgte jedoch der zweitstärkste Zuwachs aller ermittelten Cluster in dem GHD Sektor, was auf eine verstärkte Transformation der Städte hin zu Dienstleistungsstädten deutet. Diese Städte des ersten Clusters sind im Wesentlichen durchschnittliche Städte in kälteren Regionen mit stärkeren Veränderungen in der Erwerbsstruktur von der Industrie hin zur Dienstleistung.

Tabelle 5-10: Clusteranalyse – Städte des ersten Clusters

Cluster 1	24 Städte	Flensburg, Lübeck, Neumünster, Hamburg, Delmenhorst, Oldenburg, Osnabrück, Wilhelmshaven, Bremen, Hagen, Hamm, Wiesbaden, Koblenz, Speyer, Heilbronn, München, Rosenheim, Landshut, Bayreuth, Hof, Ansbach, Augsburg, Kaufbeuren, Kempten (Allgäu)
	Beschreibung der Städtecluster	Durchschnittliche Städte in kalten Regionen Deutschlands mit starken Veränderungen in der Erwerbsstruktur.

Das zweite ermittelte Cluster (Tabelle 5-11) ist ähnlich wie das erste Cluster in keinem der betrachteten Indikatoren am stärksten ausgeprägt. Der Großteil der Faktoren und demnach auch der Indikatoren liegt nah am Mittelwert. Drei Faktoren (Faktor 2, 3 und 9) weisen bei diesem Cluster höhere Ausprägungen auf. Der zweite Faktor (-0,495) vereint Indikatoren mit starkem Flächenbezug. Die Städte dieses Clusters sind flächenmäßig groß und haben demnach eine geringe Haushalts- und Bevölkerungsdichte sowie einen niedrigen Anteil an Siedlungs- und Verkehrsfläche an der Gesamtfläche. Zudem weisen die Städte dieses Clusters den dritthöchsten Anteil an Wohnungen in 1 und 2-Familienhäusern auf, was mit einer dünnen Besiedlung einhergeht. Der Faktor 3 weist mit einem Faktorwert von (-0,799) den betragsmäßig vierthöchsten Wert des Faktor 3 auf. Unter den negativen Faktorwerten ist es nach dem dritten Cluster der zweithöchste. Gebildet wird dieser Faktor durch die Indikatoren *Anteil über 65-Jähriger*, *die Bevölkerungsentwicklung der letzten 11 Jahre*, *den Anteil der Bevölkerung zwischen 25 und 65 Jahren* sowie *der Arbeitslosenquote* und stellt in gewisser Weise den sozio-demographischen Faktor dieser Untersuchung dar. Auch hier weisen die Städte höhere Werte auf. Der *Anteil über 65-Jähriger* ist in diesen Städten der höchste des Betrachtungsrahmens. Die zweithöchste Indikatoreausprägung unter allen Clustern weisen diese Städte in den Indikatoren *Bevölkerungsentwicklung* (zweitstärkster Rückgang) sowie *Anteil Bewohner zwischen 25 und 65 Jahren* (zweitniedrigster Wert) auf. Auch die Arbeitslosenquote ist mit 10 % für diese Städte der vierthöchste Wert unter den betrachteten Städtegruppen. Im Faktor 9 weisen die Städte dieses Clusters den zweithöchsten Wert aller Cluster auf. Hier entfallen

ca. 61 % der Gesamtumsatzsteuer auf den GHD-Sektor. Somit sind die Städte dieser Städtegruppe flächenmäßig groß und weisen sozio-demografisch im Vergleich mit den anderen Gruppen Schwächen auf. Zudem sind diese Städte regional in eher kalten Regionen gelegen. Die Industrie trägt in diesen Städten nur zu ca. einem Drittel zu der Gesamtumsatzsteuer bei, wodurch diese Städte stärker in dem GHD-Sektor vertreten sind. Energetisch wirkt sich das auch auf den Energieverbrauch aus, da die Energieaufwendungen im GHD-Sektor weniger energieintensiv sind als bei der Industrie.

Tabelle 5-11: Clusteranalyse – Städte des zweiten Clusters

Cluster 2	16 Städte	Kiel, Braunschweig, Bremerhaven, Kaiserslautern, Neustadt a. d. W., Pirmasens, Suhl, Baden-Baden, Passau, Straubing, Weiden i. d. O., Aschaffenburg, Chemnitz, Gera, Dessau-Roßlau, Eisenach
	Beschreibung der Städtecluster	Großflächige Städte in kälteren Regionen Deutschlands hohem GHD-Anteil und Schwächen im sozio-demografischen Bereich

Die Städte Salzgitter, Duisburg, Gelsenkirchen und Herne bilden das dritte Cluster (siehe Tabelle 5-12). Alle vier Städte bilden Industriestandorte, die, obwohl auch dort eine Transformation weg von der Schwerindustrie zu verzeichnen ist, immer noch den höchsten *Beschäftigtenanteil* der betrachteten Städte im Industriesektor und den niedrigsten im GHD Sektor aufweisen (Faktor 6). Zeitgleich lässt sich mit ca. 18 € je MJ Energieeinsatz die geringste *Energieproduktivität der Industrie* in diesen Städten wiederfinden, was stark auf eine Grundstoffindustrie hindeutet (Faktor 7). Zum Vergleich weisen die Städte des Cluster 2 eine Energieproduktivität der Industrie von ca. 802 € je MJ Energieeinsatz auf.

Tabelle 5-12: Clusteranalyse – Städte des dritten Clusters

Cluster 3	4 Städte	Salzgitter, Duisburg, Gelsenkirchen, Herne
	Beschreibung der Städtecluster	Kleinflächige Städte mit hoher Haushalts- und Bevölkerungsdichte in warmen Regionen Deutschlands mit großem Industrieanteil und sehr schwacher sozio-demografischen Dynamik sowie geringem Wohlstand

Neben der starken Industrie besitzen die Städte des Cluster 3 im dritten Faktor mit -0,930 die niedrigste Faktorausprägung unter allen Clustern. Die Indikatoren, die auf diesen Faktor laden sind die *Arbeitslosenquote*, die mit 14 % die höchste aller Cluster darstellt. Ebenfalls ist der *Anteil über 65-Jähriger* der dritthöchste aller betrachteten Städte. Der *Bevölkerungsrückgang* wie auch der *Anteil der Bevölkerung im erwerbsfähigen Alter (25-65)* bleiben von den anderen Städtegruppen im Betrachtungsrahmen unerreicht. Diese Städte müssen somit sozio-demografischen Problematiken begegnen, die im Gegensatz zu den übrigen Städten, anders angegangen werden müssen. Zu den sozio-demografischen Aspekten und der Transformation in der Industrie kommt historisch in diesen Städten die ehemalige Montan- und Schwerindustriegeschichte, die sich in einer hohen Flächenversiegelung, ca. 60 % der *Gesamtfläche werden von Siedlungs- und Verkehrsflächen* bedeckt, sowie hohen *Bevölkerungs- und Haushaltsdichten* widerspiegelt, hinzu. Die Region kann mit durchschnittlich 251 Heiztagen pro Jahr als

warme Region bezeichnet werden. Obwohl die Haushaltsdichte als auch die Bevölkerungsdichte gering sind, weisen diese Städte die niedrigste Wohnfläche pro Person auf. Zeitgleich verfügen die Haushalte über ein niedriges Einkommen, das dennoch zu dem zweithöchsten Stromverbrauch aller betrachteten Objekte führt. Somit sind die Städte dieses Cluster eher heterogen zu den anderen, da viele spezielle Ausprägungen (Arbeitslosigkeit, Energieverbrauch, Industrieschwerpunkt etc.) zusammenkommen. Diese Umstände erklären auch die geringe Stadtanzahl in diesem Stadttyp.

Tabelle 5-13: Clusteranalyse – Städte des vierten Clusters

Cluster 4	13 Städte	Emden, Darmstadt, Trier, Landau i. d. P., Worms, Karlsruhe, Heidelberg, Mannheim, Ulm, Ingolstadt, Bamberg, Schwabach, Memmingen
	Beschreibung der Städtecluster	Flächengroße Städte mit niedriger Haushalts- und Bevölkerungsdichte in durchschnittlichen klimatischen Regionen Deutschlands mit ausgeglichener Erwerbsstruktur, starker sozio-demografischen Dynamik, hohem Einkommen und einer verstärkten Pkw Orientierung

Im vierten Cluster (siehe Tabelle 5-13) fallen drei Faktoren durch deren Werte auf. Der Faktor 2 bildet den niedrigsten aller Cluster ab. Dort fließt vor allem ein, dass die Städte des vierten Clusters große Stadtflächen aufweisen. Das führt zu der geringsten Bevölkerungs- sowie Haushaltsdichte und zudem zu dem geringsten Anteil an Siedlungs- und Verkehrsfläche an der kommunalen Gesamtfläche. Dieser Faktor stellt dar, dass auf das Gesamtgebiet der Stadt bezogen bspw. ein Fernwärmeausbau nicht sinnvoll erscheint. Im dritten Faktor sind die Städte mit dem höchsten Faktorwert beaufschlagt. Die sozio-demographische Struktur dieser Städte ist stark positiv. *Der stärkste Bevölkerungszuwachs, der geringste Anteil der Bevölkerung über 65 Jahren sowie die niedrigste Arbeitsquote aller betrachteten Städte sind für dieses Cluster charakteristisch.* Die Erwerbsstruktur ist ausgeglichen, wobei ein höherer Anteil der Beschäftigten im GHD Sektor beschäftigt ist. Die bereits erwähnte große Gemarkungsfläche der Kommunen ist womöglich auch ein Grund dafür, dass in den Städten dieser Städtegruppe die geringste Nutzung des ÖPNV stattfindet. Zudem sind die Strecken (P.km./EW und Mittlere Reiseweite), die im ÖPNV zurückgelegt werden, verglichen mit den Kommunen des Untersuchungsrahmens, mit am geringsten. So legen die Einwohner der Städte dieses Clusters im Mittel pro Einwohner und Jahr ca. 559 P.km. im ÖPNV zurück. Einen geringeren Wert weist kein anderes Cluster auf. Zum Vergleich werden in den Kommunen des siebten Clusters 2.196 P.km. pro Person und Jahr im ÖPNV zurückgelegt. Dieser Zusammenhang zeigt die Abhängigkeit der Nutzung des ÖPNV und der Fläche einer Kommune. Es erklärt auch, dass in diesem Cluster mit 514 Kfz je Tsd.-Einwohner der zweithöchste Motorisierungsgrad aller betrachteten Kommunen erreicht wird. Diese Kommunen müssen im Hinblick auf eine kommunale Energieeinsparung differenziert betrachtet werden, als bspw. die Kommunen des vorherigen Clusters, wo allesamt sehr hohe Bevölkerungsdichten aufgewiesen haben.

Das Cluster 5 ist mit 17 Städten (Tabelle 5-14) das zweitgrößte des Untersuchungsrahmens. Die Städte dieses Clusters zeichnen sich durch drei wesentliche Aspekte von den anderen ab, während acht Faktoren in etwa dem Durchschnitt des Untersuchungsrahmens entsprechen. Zu nennen ist zunächst der charakterisierende Faktor 9, der den niedrigsten Wert aller betrachteten Städtegruppen vorweist. Die Indikatoren der Industrie- und GHD-Umsatzsteuer an der Gesamtumsatzsteuer sind dabei ausschlaggebend. Die Städte dieses Clusters bilden mit deren Indikatorwerten den dritthöchsten Anteil des GHD an der Gesamtumsatzsteuer ab, wohingegen der Anteil der Industrie mit einem Anteil von 30,54 % den zweithöchsten Wert abbildet. Alle Städte liegen klimatisch in wärmeren Regionen, was z. T. durch große Flüsse begünstigt wird (z. B. Rhein, Ruhr, Main). Alle Städte haben zudem gemeinsam, dass die Flächenausnutzung intensiv ist. Das schlägt sich in den höchsten Bevölkerungs- und Haushaltsdichten wieder. Zudem korrelierten diese Indikatoren mit einem niedrigem Anteil an Wohnungen in 1- und 2 –Familienhäusern, was auf einen höheren Anteil an Mehrfamilienhäusern hindeutet. Ebenfalls ist der Anteil der Siedlungs- und Verkehrsfläche an der kommunalen Gesamtfläche mit 57,1 % der zweithöchste aller Werte dieses Indikators. Kommunen dieses Clusters sind demnach, verglichen mit den anderen Clustern, verschieden. Ein differenziertes Herangehen an beispielsweise infrastrukturelle Umbaumaßnahmen (z. B. Fernwärme) kann andere Potenziale aufweisen, als in Kommunen mit niedrigen Haushaltsdichten. Zudem weisen diese Kommunen in den Bereichen Industrie und GHD differenzierte Potenziale auf, als beispielsweise Industriestädte wie im nachfolgenden Cluster sechs.

Tabelle 5-14: Clusteranalyse – Städte des fünften Clusters

Cluster 5	17 Städte	Düsseldorf, Essen, Krefeld, Mönchengladbach, Mülheim a. d. R., Oberhausen, Bottrop, Wuppertal, Bochum, Dortmund, Frankfurt a. M., Offenbach a. M., Regensburg, Fürth, Nürnberg, Würzburg, Berlin
	Beschreibung der Städtecluster	Dichtbesiedelte Städte in warmen Regionen Deutschlands mit großem GHD-Anteil und ausgeglichener sozio-demografischen Struktur

Die sechste Stadtgruppe bilden 13 Städte, die sich besonders stark dadurch abzeichnen, dass diese mit einem Anteil der Industrieumsatzsteuer an der Gesamtumsatzsteuer von fast Zweidritteln den höchsten Wert aller Cluster für diesen Indikator aufweisen. Die Erwerbsstruktur in diesen Städten ist demnach stark verschieden als beispielsweise im Cluster 5. Trotz des hohen Industrieanteils fällt auch die Energieproduktivität im Vergleich aller Cluster relativ hoch aus. Die Energieintensität der Industrie markiert zwar den zweithöchsten Wert nach den Städten des Clusters 3, ist jedoch um ca. den Faktor 12 geringer als die Werte des Clusters 3. Das zeigt, dass die Industrie in den Städten des Clusters 6 (z. B. Werkzeugbau, Maschinenbau, Kfz-Herstellung), eher auf der verarbeitenden Industrie aufbaut, wo eine höhere Wertschöpfung erfolgt als bei der Grund- und Schwerindustrie der Städte des Clusters 3. Zu dem großen Industrieanteil weisen die Städte dieses Clusters auch den höchsten Wert im Faktor 1 auf. Die Indikatoren dieses Clusters beziehen sich auf die Kfz-Verfügbarkeit in den Städten bezogen

auf die Einwohnerzahl. Diese Zahl wird als verstärkte Pkw Nutzung interpretiert, da vorausgesetzt werden kann, dass ein Vorhandensein eines Kfz dessen Nutzung auch einschließt. Deshalb ist auch der ÖPNV eher schwächer ausgeprägt und wird auch weniger genutzt.

Tabelle 5-15: Clusteranalyse – Städte des sechsten Clusters

Cluster 6	13 Städte	Remscheid, Solingen, Leverkusen, Bielefeld, Kassel, Frankenthal (Pfalz), Pforzheim, Amberg, Ludwigshafen a R., Zweibrücken, Stuttgart, Coburg, Schweinfurt
	Beschreibung der Städtecluster	Industriestädte mit geringer Energieintensität in der Industrie in durchschnittlichen klimatischen Regionen Deutschlands mit hoher Individualverkehrsorientierung und hohem Einkommen

Das zweite vier-Städte Cluster, bestehend aus den Städten Köln, Münster, Mainz und Freiburg im Breisgau stellt das Cluster 7 dar. Dieses Cluster weist in sechs der elf Faktoren die höchsten positiven bzw. negativen Faktorwerte auf. Dies erklärt auch die kleine Clustergröße, da anders als z. B. bei dem Cluster 1, bei dem viele Städte enthalten sind, die in fast allen Faktoren Faktorwerte vorweisen, die nah am Mittelwert liegen, sind in diesem Cluster kaum Faktorwerte, die nah am Mittelwert liegen, festzustellen. Die vier Städte grenzen sich von den anderen durch starke Ausprägungen in den Faktoren 4, 5, 6, 8, 10 und 11 ab. Faktor 4 ist der am stärksten ausgeprägte Faktor und deutet darauf hin, dass die Städte in warmen Regionen liegen. Der fünfte Faktor ist stark negativ, da die diesen Faktor bildenden Indikatoren die Energieintensität der Industrie als auch den spezifischen Energieverbrauch der Industrie sowie das Produktivitätsniveau des GHD Sektors sind. Das Produktivitätsniveau des GHD Sektors weist hier den höchsten Wert aller Cluster auf. Die Faktoren 6 und 11 werden hier zusammengefasst betrachtet. Die beiden Faktoren setzen die Industrie sowie den GHD-Sektor ins Verhältnis. Die vier Städte weisen den mit Abstand höchsten Beschäftigtenanteil von 88 % im GHD-Sektor und den entsprechend niedrigsten in der Industrie auf. Genauso ist auch das Ergebnis bei der Betrachtung der Beschäftigtenentwicklung (Faktor 11). Hier können die beiden Erwerbssektoren im Vergleich betrachtet werden. Der hohe GHD Anteil lässt sich durch die stark rückläufigen Beschäftigtenzahlen im Industriebereich erklären. Der GHD-Sektor erfährt in demselben Zeitraum einen geringen Zuwachs. Jedoch ist auch ein geringer Zuwachs von einem sehr hohen Niveau eine beachtliche Entwicklung. Die starke GHD-Orientierung wird auch hier ein weiteres Mal unterstrichen. Zudem ist die sozio-strukturelle Entwicklung stark positiv. In dem Betrachtungszeitraum der Bevölkerungsentwicklung erfuhren die Städte dieses Clusters einen Bevölkerungszuwachs von fast 10 %, was auch die niedrigen Werte des Faktors 8 erklärt (z. B. geringe Wohnfläche pro Einwohner). Die Indikatoren des ÖPNV, die die im ÖPNV zurückgelegte Strecke pro Jahr und Einwohner als auch pro Fahrt vereinen übersteigen alle anderen. Der ÖPNV wird in diesen Städten am intensivsten nachgefragt. Wenig überraschend ist auch, dass in diesen Städten der MIV den niedrigsten Anteil einnimmt. In nachfolgender Tabelle 5-16 sind die Städte als auch die Beschreibung der siebten Städtegruppe (Cluster) enthalten.

Tabelle 5-16: Clusteranalyse – Städte des siebten Clusters

Cluster 7	4 Städte	Köln, Münster, Mainz, Freiburg i. B.
	Beschreibung der Städtecluster	Dienstleistungsstädte in klimatisch warmen Regionen mit hohem Produktivitätsniveau des GHD-Sektors, hoher ÖPNV und geringer MIV Orientierung und stark positiver sozio-demografischer Dynamik

Das letzte ermittelte Cluster 8 zeichnet sich, dadurch aus, dass alle enthaltenen Städte in den neuen Bundesländern liegen. Vier der elf Faktoren weisen Maximalwerte auf, wovon drei die niedrigsten Werte des Untersuchungsrahmens sind. Die MIV Orientierung in den Städten dieses Clusters ist die niedrigste, was mit einem niedrigen verfügbaren Einkommen sowie einer hohen Arbeitslosigkeit in Zusammenhang gebracht werden kann. Solche Voraussetzungen müssen bei der Planung und Umsetzung energieeffizienzsteigernder Maßnahmen betrachtet werden, da in solchen Kommunen beispielsweise die Energiearmut eher eintreten kann. Wird der Erwerbssektor dieser Städte fokussiert, so sticht mit einem Beschäftigtenanteil von ca. 85 % die starke GHD Orientierung hervor. Mit einem geringen Industrieanteil in diesem Cluster und einem hohen GHD Anteil fällt der Anteil der GHD-Umsatzsteuer an der Gesamtumsatzsteuer mit lediglich ca. 50 % deutlich geringer aus, als bei den Städten des Clusters 7. Die Städte des Clusters 7 sind in der Erwerbsstruktur ähnlich stark auf den GHD-Sektor ausgerichtet, wobei der GHD-Anteil an der Gesamtumsatzsteuer jedoch deutlich höher liegt als im Cluster 8. Der siebte Faktor setzt sich aus der Energieproduktivität sowie dem Produktivitätsniveau der Industrie zusammen. Beide Indikatoren bewirken den mit Abstand höchsten Faktorwert, der zum einen den höheren Industrieumsatzsteueranteil und zum anderen die hohe *Energieproduktivität* als auch das hohe *Produktivitätsniveau* der Industrie erklärt. So kommt es, dass die Industrie in diesen Städte eher höherwertige Produkte erzeugt, als beispielsweise die Montan- bzw. Schwerindustrie (z. T. in den Städten des Ruhrgebiets). Ferner sind im Bereich der Haushalte auch deutliche Unterschiede zu den anderen Stadttypen festzustellen. Der Pro-Kopf-Stromverbrauch und die pro Einwohner zur Verfügung stehende Wohnfläche sind deutlich geringer als in den zuvor betrachteten Städten. Aus energetischer Sicht kann von einem niedrigeren Ausstattungsgrad der Haushalte ausgegangen werden, woraus auch geringere Energieverbräuche in den Städten dieses Clusters vermutet werden können. Die nachfolgende Tabelle 5-17 beinhaltet die 12 Städte des Clusters und deren Beschreibung.

Tabelle 5-17: Clusteranalyse – Städte des achten Clusters

Cluster 8	12 Städte	Brandenburg a. d. H., Cottbus, Frankfurt, Rostock, Schwerin, Dresden, Leipzig, Halle (Saale), Magdeburg, Erfurt, Jena, Weimar
	Beschreibung der Städtecluster	Dienstleistungsstädte mit hohem Flächenpotenzial in klimatisch kälteren Regionen mit geringer MIV- und verstärkter ÖPNV-Nutzung, geringem verfügbaren Einkommen und niedrigem Energieverbrauch in Haushalten.

6 Zusammenfassung und Ausblick

Das Ziel dieser Arbeit beinhaltet einen ersten Schritt in der Untersuchung der Eignung der statistischen Clusteranalyse als Methode im energetischen Bereich sowie einer Grundlagen-generierung für weitere Untersuchungen. Zunächst wurde die in Ihrem Einsatzbereich umfangreiche Methode der Clusteranalyse untersucht, eingegrenzt sowie mögliche Clusteranalysetypen für eine Betrachtung von Kommunen unter sozio-energetischen Gesichtspunkten identifiziert. Als Analysemethode ist die Clusteranalyse in dem ingenieurtechnischen Bereich unterrepräsentiert, während in den Sozial-, Natur- und Wirtschaftswissenschaften sowie der Medizin diese Methode umfangreich und bereits seit langem angewandt wird. Aus einer Auswahl wissenschaftlicher Veröffentlichungen konnte ein Überblick über die Anwendung der Clusteranalyse erstellt werden. So konnte die hierarchische Clusteranalysemethode als die am weitesten verbreitete Variante identifiziert werden. Das hierarchische Herangehen bietet viele Vorteile, wobei ein großer Vorteil die häufige Anwendung darstellt. Zudem erlaubt die hierarchische Clusteranalyse die statistische Betrachtung kleiner sowie großer Untersuchungsumfänge, wodurch diese auf die kreisfreien Städte und auch weitere Untersuchungen mit größeren Untersuchungsumfängen angewendet werden kann.

In der hier durchgeführten Untersuchung erfolgt die Charakterisierung deutscher kreisfreier Städte mit Indikatoren, die aus der Literatur als auch der Berechnung aus Daten Statistischer Landes- und Bundesämter, Unternehmensverbände, Vereine als auch weiterer Organisationen, die Daten auf kommunaler Ebene zur Verfügung stellen, hervorgehen. Dieser Top-Down Ansatz wurde auf den Untersuchungsrahmen, bestehend aus allen 107 deutschen kreisfreien Städten in denen ca. 31,6 % der Gesamtbevölkerung Deutschlands leben, angewendet. Für diese Städte wurde ein Indikatorsatz von 41 Indikatoren erarbeitet, der gleichermaßen Indikatoren zu den vier Hauptverbrauchssektoren private Haushalte, Verkehr, Industrie und GHD beinhaltet. Im Rahmen einer Faktorenanalyse der erhobenen Indikatoren konnten diese zu elf Faktoren zusammengefasst werden. Bei der Faktorenanalyse wurde der Indikatorsatz um 7 Indikatoren reduziert, da diese in keinem Analysedurchlauf mit weiteren Indikatoren korrelierten. Durch die Ermittlung von Faktoren kommt jedem Indikator mathematisch die gleiche Bedeutung zu. Zudem erleichtert eine geringere Anzahl an zu betrachtenden Variablen die Interpretation der Cluster. In einem letzten Schritt wurden die Städte in mehreren Durchläufen typisiert und acht Cluster ermittelt, die letztendlich nur 103 deutsche kreisfreie Städte gruppieren. Die einzelnen Städtegruppen (Cluster) erfordern bei deren Interpretation lediglich die Betrachtung der Städtegruppen und nicht die einzelnen 103 Städte, wodurch der Aufwand reduziert werden kann. Weiterhin wird eine Möglichkeit dargestellt Städte zu identifizieren, auf die bereits in anderen Städten des selben Clusters umgesetzte Maßnahmen übertragen werden könnten.

Die mithilfe der Clusteranalyse ermittelten Cluster reduzieren die Anzahl zu betrachtender Städte auf Städtegruppen, die sich statistisch ähneln. Sofern die Städte durch belastbare Daten beschrieben werden können, ist deren Charakterisierung belastbar. Die Interpretation der Cluster wird durch die Clusteranalyse erleichtert und auch die Anzahl der Gruppen wird verringert. Dennoch ist bei einem großen Datensatz (mehr bezogen auf die Anzahl an Variablen (Indikatoren, Faktoren) als die Anzahl der Städte) die Interpretation mitunter umfangreich und nicht immer eindeutig, sodass der Indikatorsatz deutlich reduziert werden sollte. Zudem ist für die Untersuchung einer möglichen Übertragbarkeit spezifischer Maßnahmen eher eine detaillierte Untersuchung der für diese Maßnahme notwendigen Kennwerte und deren anschließende Clusterung zu favorisieren. Nützliche Ideen und Zusammenhänge können so identifiziert werden, pauschale Übertragbarkeiten können jedoch basierend auf den Ergebnissen einer Clusteranalyse nicht abgeleitet werden. Vielmehr sollten die Ergebnisse als Hilfestellung gedeutet werden, um die Vielzahl der heutzutage erhobenen Daten mit in Entscheidungsprozesse einfließen zu lassen. Für die Identifikation von Gruppen, die beispielsweise ähnlichen Herausforderungen begegnen müssen, die durch Indikatoren abbildbar sind, eignet sich eine Clusteranalyse hingegen schon, da die ermittelten Gruppen statistisch Ähnlichkeiten aufweisen und bei der Identifikation in Frage kommender Kommunen dienen können. [WAL2016B]

Für weitere Arbeiten kann das hier erarbeitete und angewendete Clusteranalysevorgehen verwendet werden. Um hingegen die Clusteranalyse auf eine konkrete Maßnahme anzuwenden, sollte lediglich ein Verbrauchssektor betrachtet werden. Dazu ist ein Bottom-Up Ansatz denkbar, da in diesem die betrachtete Maßnahme durch darauf angepasste und der Maßnahme dienliche Indikatoren beschrieben wird. Zudem führt die Fokussierung auf geringere Variablensätze zu einer leichteren Deutung sowie Gruppierung der betrachteten Städte. Aufbauend auf solche Ergebnisse ist eine Untersuchung der Clusteranalyse im Hinblick auf differenzierte Anwendungsbereiche, z. B. die überschlägige Potenzialabschätzung von beispielsweise Energieeinsparungen, möglich. Die Möglichkeiten der Clusteranalyse sollten in dieser Hinsicht betrachtet werden, da größere Datenmengen statistisch dazu beitragen könnten, Potenziale von betrachteten Maßnahmen zu beziffern ohne die Potenziale für alle betrachteten Objekte einzeln berechnen zu müssen, wodurch auf effiziente Weise Potenziale abgeschätzt werden könnten.

Quellenverzeichnis

- [ABO2007] ABONYI, J.; FEIL, B.: *Cluster Analysis for Data Mining and System Identification*. Basel: Birkhäuser Verlag AG, 2007. ISBN: 9783764379872.
- [ALL2004] ALLIK, J.; MCCRAE, R. R.: *Toward a Geography of Personality Traits. Patterns of Profiles across 36 Cultures*. IN: *Journal of Cross-Cultural Psychology* 35 Nr. 1, S. 13–28.
- [ALT2015] ALT, H.: *Ausgewählte Kapitel der Energiewirtschaft. Gradtagzahl und Heizgradtage*. FACHHOCHSCHULE AACHEN (HRSG.). Aachen. www.alt.fh-aachen.de/downloads/.../Hilfsb%2020%20Gradzahltag.pdf. Abruf: 01.12.2015.
- [BAC2006] BACKHAUS, K.; ERICHSON, B.; PLINKE, W. ET AL.: *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung*. 11. Aufl. Berlin: Springer, 2006. (Springer-Lehrbuch). - ISBN: 978-3540278702.
- [BAC2010] BACHER, J.; PÖGE, A.; WENZIG, K. ET AL.: *Clusteranalyse. Anwendungsorientierte Einführung in Klassifikationsverfahren*. 3., erg., vollst. überarb. und neu gestaltete Aufl. München: Oldenbourg, 2010. - ISBN: 9783486584578.
- [BAC2013] BACKHAUS, K.; ERICHSON, B.; WEIBER, R. ET AL.: *Fortgeschrittene multivariate Analysemethoden. Eine anwendungsorientierte Einführung ; [mit Extras im Web]*. 2., überarb. und erw. Aufl. Berlin: Springer Gabler, 2013. (Springer-Lehrbuch). - ISBN: 9783642373633.
- [BAC2016] BACKHAUS, K.; ERICHSON, B.; PLINKE, W. ET AL.: *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung*. 14. Auflage. Berlin, Heidelberg, s.l.: Springer Berlin Heidelberg, 2016. ISBN: 978-3-662-46076-4.
- [BDE2014] BUNDESVERBAND DER ENERGIE- UND WASSERWIRTSCHAFT E.V.: *Energie-Info, Stromverbrauch im Haushalt*. BDEW (HRSG.). Berlin. [https://www.bdew.de/internet.nsf/id/6966C7CB65D8D8FAC1257D5E0043D565/\\$file/705_BDEW_Stromverbrauch%20im%20Haushalt_Stand_September%202014.pdf](https://www.bdew.de/internet.nsf/id/6966C7CB65D8D8FAC1257D5E0043D565/$file/705_BDEW_Stromverbrauch%20im%20Haushalt_Stand_September%202014.pdf), 2014. Abruf: 01.11.2015.
- [BER2014] BERTELSMANN STIFTUNG: *Gemeindetypisierung – Wegweiser Kommune. Methodisches Vorgehen und empirische Befunde*. BERTELSMANN STIFTUNG (HRSG.), 2014. <http://www.wegweiser-kommune.de/methodik>. Abruf: 03.03.2016.
- [BER2015A] BERTELSMANN STIFTUNG: *Wegweiser Kommune, Statistik*, 2015. <https://www.wegweiser-kommune.de/statistik>. Abruf: 01.11.2015.
- [BER2015B] BERTELSMANN STIFTUNG: *Indikatorerläuterung*, 2015. https://www.wegweiser-kommune.de/documents/10184/21656/Indikatorenerl%C3%A4uterung_neu/0b49cb56-c883-4e0b-a3de-363b5fa59b02, zuletzt aktualisiert am März 2015. Abruf: 01.11.2015.
- [BFA2015] BUNDESAGENTUR FÜR ARBEIT: *Grundlagen, Arbeitslosenquote*. BUNDESAGENTUR FÜR ARBEIT (HRSG.), 2015. http://statistik.arbeitsagentur.de/nn_4236/Statistischer-Content/Grundlagen/Berechnung-Arbeitslosenquote/Berechnung-Arbeitslosenquote.html, zuletzt aktualisiert am 01.06.2015. Abruf: 31.11.2015.
- [BMB2015] BUNDESMINISTERIUM FÜR BILDUNG UND FORSCHUNG (Hrsg.): *Deutschlands Spitzencluster. Broschüre*. Berlin, 2015.
- [BMU2014] BUNDESMINISTERIUM FÜR UMWELT, NATURSCHUTZ, BAU UND REAKTORSICHERHEIT: *Nationale Klimapolitik*. Hg. v. BMUB, 2014. www.bmub.bund.de/P215/. Abruf: 31.10.2015.
- [BMW2015] BUNDESMINISTERIUM FÜR WIRTSCHAFT UND ENERGIE: *Zahlen und Fakten. Energiedaten. Nationale und internationale Entwicklung*. Hg. v. BMWi, 2015. <http://www.bmwi.de/DE/Themen/Energie/Energiedaten-und-analysen/Energiedaten/gesamtausgabe>, zuletzt aktualisiert am 16.03.2015. Abruf: 31.10.2015.
- [BP2014] BP P.L.C.: *Statistical Review of World Energy June 2014*. Hg. v. BP, 2014. <http://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html>. Abruf: 30.05.2015.
- [BPB2008] BUNDESZENTRALE FÜR POLITISCHE BILDUNG - BPB: *Human Development Index (HDI)*. BPB (HRSG.), 2008. http://www.bpb.de/themen/26g2cn,0,0,human_development_index_. Abruf: 31.10.2015.
- [BPB2010] BUNDESZENTRALE FÜR POLITISCHE BILDUNG - BPB: *Zahlen und Fakten - Globalisierung*. BPB (HRSG.), 2010. <http://www.bpb.de/nachschlagen/zahlen-und-fakten/globalisierung/52705/verstaedterung>. Abruf: 31.10.2015.

- [BPB2012] BUNDESZENTRALE FÜR POLITISCHE BILDUNG - BPB: *Zahlen und Fakten, soziale Situation in Deutschland, Haushalte nach Zahl der Personen*. BPB (HRSG.), 2012. <http://www.bpb.de/nachschlagen/zahlen-und-fakten/soziale-situation-in-deutschland/61587/haushalte-nach-zahl-der-personen>. Abruf: 31.10.2015.
- [BUD2014] BUDA, A.: *Untersuchung energetischer Indikatoren für die Sektoren Industrie und Gewerbe, Handel, Dienstleistungen in Städten und Kommunen*, Bochum. RUHR-UNIVERSITÄT BOCHUM (HRSG.). ET, LEE, 2014.
- [BUE2012] BÜHL, A.: *SPSS 20. Einführung in die moderne Datenanalyse*. 13., aktualisierte Aufl. München: Pearson, 2012. ISBN: 9783868941500.
- [BUN2015] DIE BUNDESREGIERUNG: *Energiewende, Mobilität, Elektromobilität. Leitmarkt und Leitanbieter für Elektromobilität*. Hg. v. Die Bundesregierung, 2015. http://www.bundesregierung.de/Webs/Breg/DE/Themen/Energiewende/Mobilitaet/podcast/_node.html. Abruf: 04.12.2015.
- [BUR1994] BURK, R. O. W.; KÖNIG, J.; ROHRSCHEIDER, K. ET AL.: *Dreidimensionale Papillentopographie mittels LaserScanning Tomographie: Klinisches Korrelat einer Cluster-Analyse*. IN: *Klinische Monatsblätter für Augenheilkunde* Nr. 204, S. 504–512, 1994.
- [DES2003] STATISTISCHES BUNDESAMT: *Klassifikation der Wirtschaftszweige (WZ2003)*. DESTATIS (HRSG.), 2003. <https://www.destatis.de/DE/Methoden/Klassifikationen/GueterWirtschaftsklassifikationen/Content75/KlassifikationWZ2003.html>. Abruf: 31.10.2015.
- [DES2008] STATISTISCHES BUNDESAMT: *Klassifikation der Wirtschaftszweige – Mit Erläuterungen*. DESTATIS (HRSG.), 2008. <https://www.destatis.de/DE/Methoden/Klassifikationen/GueterWirtschaftsklassifikationen/Content75/KlassifikationWZ08.html>. Abruf: 31.10.2015.
- [DES2011] STATISTISCHES BUNDESAMT: *Zensusergebnisse*, 2015. https://ergebnisse.zensus2011.de/#StaticContent:00,GWZ_4_3_0,m,table. Abruf: 01.11.2015.
- [DES2013A] STATISTISCHES BUNDESAMT: *Zahlen und Fakten, Datenbanken, Regionaldatenbank, Bevölkerungsstand*. DESTATIS (HRSG.), 2013. www.destatis.de. Abruf: 11.06.2013.
- [DES2013B] *Wirtschaftsbereiche, Transport und Verkehr*. DESTATIS (HRSG.), 2013. <https://www.destatis.de/DE/ZahlenFakten/Wirtschaftsbereiche/TransportVerkehr/TransportVerkehr.html>. Abruf: 25.07.2015.
- [DES2013C] STATISTISCHES BUNDESAMT: *Energie, Rohstoffe, Emissionen. Energieverbrauch der privaten Haushalte für Wohnen*. DESTATIS (HRSG.), 2013. <https://www.destatis.de/DE/ZahlenFakten/GesamtwirtschaftUmwelt/Umwelt/UmweltoekonomischeGesamtrechnungen/EnergieRohstoffeEmissionen/Tabellen/EnergieverbrauchHaushalte.html>. Abruf: 24.07.2014.
- [DES2013D] STATISTISCHES BUNDESAMT: *Zahlen und Fakten, Datenbanken, Genesis, Bevölkerung: Kreise, Stichtag*. DESTATIS (HRSG.), 2013. www.destatis.de. Abruf: 31.10.2015.
- [DES2014A] STATISTISCHES BUNDESAMT: *Regionaldatenbank, Erwerbstätige nach Wirtschaftsbereichen 638-61-4*. DESTATIS (HRSG.), 2014. www.destatis.de. Abruf: 31.10.2015.
- [DES2014B] STATISTISCHES BUNDESAMT: *Kreisfreie Städte und Landkreise nach Fläche, Bevölkerung und Bevölkerungsdichte am 31.12.2013*. DESTATIS (HRSG.), 2014. <https://www.destatis.de/DE/ZahlenFakten/LaenderRegionen/Regionales/Gemeindeverzeichnis/Administrativ/Aktuell/04Kreise.html>. Abruf: 31.10.2015.
- [DES2014C] STATISTISCHES BUNDESAMT: *Verfügbares Einkommen der privaten Haushalte einschließlich privater Organisationen ohne Erwerbszweck - Jahressumme - regionale Tiefe: Kreise und kreisfr. Städte*. DESTATIS (HRSG.), 2014. www.destatis.de. Abruf: 28.11.2015.
- [DES2014D] STATISTISCHES BUNDESAMT: *Arbeitslose nach ausgewählten Personengruppen sowie Arbeitslosenquoten - Jahresdurchschnitt -regionale Tiefe: Kreise und kreisfr. Städte*. Hg. v. DESTATIS, 2014. www.destatis.de. Abruf: 30.11.2015.
- [DES2014E] STATISTISCHES BUNDESAMT: *Bodenfläche nach Art der tatsächlichen Nutzung - Stichtag 31.12. - Kreise und kreisfr. Städte*. DESTATIS (HRSG.), 2014. www.destatis.de. Abruf: 30.11.2015.

- [DES2014F] STATISTISCHES BUNDESAMT: *Zensus 2011. Personen nach Alter (5 Altersklassen) für Flensburg, Stadt (Kreisfreie Stadt), Kiel, Landeshauptstadt (Kreisfreie Stadt) und weitere Orte*. Hg. v. Statistische Ämter des Bundes und der Länder, 2014. <https://ergebnisse.zensus2011.de/#dynTable>. Abruf: 04.12.2015.
- [DES2014G] STATISTISCHES BUNDESAMT: *Sozialversicherungspflichtig Beschäftigte am Arbeits- und Wohnort sowie Ein- und Auspendler über Gemeindegrenzen nach Geschlecht - Stichtag 30.06. - Kreise und krfr. Städte*. DESTATIS (HRSG.), 2014. www.destatis.de. Abruf: 04.12.2015.
- [DES2014H] STATISTISCHES BUNDESAMT: *Erwerbstätige nach Wirtschaftsbereichen - Jahresdurchschnitt - (WZ2008) regionale Tiefe: Kreise und krfr. Städte Erwerbstätigenrechnung des Bundes und der Länder. Erwerbstätige nach Wirtschaftsbereichen 638-61-4*. DESTATIS (HRSG.), 2014. www.destatis.de. Abruf: 03.01.2016.
- [DES2014I] STATISTISCHES BUNDESAMT: *Energieverwendung der Betriebe des Verarbeitenden Gewerbes sowie des Bergbaus und der Gewinnung von Steinen und Erden - Jahressumme -regionale Tiefe: Kreise und kreisfreie Städte*. DESTATIS (HRSG.), 2014. www.destatis.de. Abruf: 10.01.2016.
- [DES2014J] STATISTISCHES BUNDESAMT: *Regionalatlas Deutschland Indikatoren des Themenbereichs 'Bruttoinlandsprodukt' Regionalatlas Deutschland*. DESTATIS (HRSG.), 2014. www.destatis.de. Abruf: 10.01.2016.
- [DES2014K] STATISTISCHES BUNDESAMT: *Bruttoinlandsprodukt/Bruttowertschöpfung (WZ 2008)- Jahressumme - regionale Ebenen*. DESTATIS (HRSG.), 2014. www.destatis.de. Abruf: 10.01.2016.
- [DES2015A] *Umsatzsteuerpflichtige, steuerbarer Umsatz für Lieferungen u. Leistungen - Wirtschaftsabschnitt (WZ08) - Jahr - regionale Tiefe: Kreise und krfr. Städte*. DESTATIS (HRSG.), 2015. www.destatis.de. Abruf: 22.01.2016.
- [DES2015B] STATISTISCHES BUNDESAMT: *Bevölkerung: Kreise, Stichtag*. DESTATIS (HRSG.), 2015. <https://www-genesis.destatis.de/genesis/online/data>. Abruf: 31.10.2015.
- [DES2015C] STATISTISCHES BUNDESAMT: *Zahlen und Fakten, Haushalte 2014: rund 40 Millionen Privathaushalte in Deutschland*. DESTATIS (HRSG.), 2015. <https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Bevoelkerung/HaushalteFamilien/Aktuell.html;jsessionid=F5ED0DD96A19978E9A3BF992D2E52C3D.cae1>. Abruf: 31.10.2014.
- [DES2015D] STATISTISCHES BUNDESAMT: *Pressemitteilung Nr. 124, Anteil der Industrie am BIP*. DESTATIS (HRSG.), 2015. https://www.destatis.de/DE/PresseService/Presse/Pressemitteilungen/2015/04/PD15_124_811.html. Abruf: 31.10.2015.
- [DES2015E] STATISTISCHES BUNDESAMT: *Zahlen und Fakten, Wirtschaftsbereiche, Dienstleistungen*. DESTATIS (HRSG.), 2015. <https://www.destatis.de/DE/ZahlenFakten/Wirtschaftsbereiche/Dienstleistungen/Methoden/InfoDienstleistungen.html>. Abruf: 31.10.2015.
- [DES2015F] STATISTISCHES BUNDESAMT: *Regionaldatenbank Deutschland, Zensusdatenbank*. DESTATIS (HRSG.), 2015. <https://www.regionalstatistik.de/genesis/online/>. Abruf: 01.11.2015.
- [DES2015G] STATISTISCHES BUNDESAMT: *Zahlen und Fakten, Datenbanken, Regionaldatenbank Deutschland, Gebietsstand: Gebietsfläche in qkm - Stichtag 31.12. - regionale Ebenen*. DESTATIS (HRSG.), 2015. www.destatis.de. Abruf: 31.10.2015.
- [DES2015H] STATISTISCHES BUNDESAMT: *Zahlen und Fakten, Wirtschaftsbereiche, Industrie – Verarbeitendes Gewerbe*. DESTATIS (HRSG.), 2015. <https://www.destatis.de/DE/ZahlenFakten/Wirtschaftsbereiche/IndustrieVerarbeitendesGewerbe/IndustrieVerarbeitendesGewerbe.html>. Abruf: 31.10.2015.
- [DFG2013] DEUTSCHE FORSCHUNGSGEMEINSCHAFT (Hrsg.): *Exzellenzinitiative auf einen Blick. Der Wettbewerb des Bundes und der Länder zur Stärkung der universitären Spitzenforschung*. 5. Aufl. Bonn, 2013.
- [DIE1999] DIEKMANN, J.; EICHHAMMER, W.; NEUBERT, A. ET AL.: *Energie-Effizienz-Indikatoren. Statistische Grundlagen, theoretische Fundierung und Orientierungsbasis für die politische Praxis*. Heidelberg: Physica-Verlag HD, 1999. (Umwelt und Ökonomie, 32). - ISBN: 978-3-7908-1243-5.

- [DIE2012] DIEKMANN, J.; GROBA, F.; VOGEL-SPERL, A. ET AL.: *Vergleich der Bundesländer: Analyse der Erfolgsfaktoren für den Ausbau der Erneuerbaren Energien 2012 - Indikatoren und Ranking. Endbericht ; Forschungsprojekt des DIW Berlin und des ZSW Stuttgart im Auftrag und in Kooperation mit der Agentur für Erneuerbare Energien e.V. Berlin.* Berlin: DIW, 2012. ISBN: 978-3-938762-61-5.
- [DIF2011] RÖSLER, C. (Hrsg.): *Klimaschutz in Kommunen. Praxisleitfaden.* DEUTSCHES INSTITUT FÜR URBANISTIK (HRSG.). Loseblattaussg. Berlin: Difu, 2011.
- [DPC2016] DPC SOFTWARE: *Stata.* DPC SOFTWARE (HRSG.), 2016. <http://www.dpc-software.de/?id=357>. Abruf: 28.02.2016.
- [EHF2015] EVANGELISCHE HOCHSCHULE FREIBURG - EHF: *Haussanierung belastet ältere Menschen*, 2015. <http://www.eh-freiburg.de/news-detail/haussanierung-belastet-aeltere-menschen/447>, zuletzt aktualisiert am 19.03.2015. Abruf: 30.11.2015.
- [EIC2013] EICKELKAMP, T.: *Input-Output Ansatz zur Berücksichtigung von Anlagegütern in Ökobilanzen - angewendet für einen Offshore-Windpark.* Univ., Dissertation, Bochum, 2013. Berlin: LIT-Verl., 2013. (Energie und Nachhaltigkeit, 13). ISBN: 9783643122704.
- [EIC2014] EICKELKAMP, T.; WALL, W.; KOCH, M. K. ET AL.: *Indikatoren zur energetischen Charakterisierung und Clusterung von Städten.* In: H.-J. Wagner und J. Görres (Hg.): *Wettbewerb Energieeffiziente Stadt, Band 2 - Energieversorgung, Energiebilanzierung und Monitoring.* Berlin: LIT-Verlag, 2014., S. 43–52, ISBN: 9783643127112.
- [ELL1996] ELLIS, T. E.; RUDD, M. D.; RAJAB, M. H. ET AL.: *Cluster Analysis of MCMI Scores of suicidal psychiatric Patients: Four personality Profiles.* IN: *Journal of Clinical Psychology* 52 Nr. 4, S. 411–422, 1996.
- [ENG2015] ENGELHARDT, C.: *Familienstadt mit Zukunft – Was haben wir uns Vorgenommen.* STADT FRANKENBERG/EDER (HRSG.), 2015. http://frankenberger.active-city.net/city_info/anzeige/redaktionssystem/main/show.cfm?region_id=84&id=19473&modul_id=5&print=1, Abruf: 12.06.2015.
- [EU2011] EUROPÄISCHE KOMMISSION: *Städte von morgen. Herausforderungen, Visionen, Wege nach vorn.* Luxemburg: Amt für Veröffentlichungen der Europäischen Union, 2011. ISBN: 9789279231537.
- [EUR2015] EUROPÄISCHE KOMMISSION, E.: *Beschäftigung und Arbeitslosigkeit (LFS/AKE), Statistiken Illustriert.* Hg. v. Eurostat. EUROPÄISCHE KOMMISSION (HRSG.), 2015. <http://ec.europa.eu/eurostat/de/web/lfs/statistics-illustrated>. Abruf: 30.11.2015.
- [EVE2011] EVERITT, B. S.: *Cluster analysis.* 5. ed. Chichester: Wiley, 2011. (Wiley series in probability and statistics). ISBN: 978-0-470-74991-3.
- [EXX2015] EXXONMOBIL CENTRAL EUROPE HOLDING GMBH: *Energieprognose Deutschland 2015-2040. Broschüre.* Hg. v. ExxonMobil Central Europe Holding GmbH. Hamburg, 2015.
- [EXZ2014] MINISTERIUM FÜR WIRTSCHAFT, ENERGIE, INDUSTRIE, MITTELSTAND UND HANDWERK (Hrsg.): *Exzellenz NRW. Cluster Nordrhein-Westfalens*, 2014. <http://www.exzellenz.nrw.de/clusterinfo/landescluster/>. Abruf: 20.12.2015.
- [FAU1996] FAUNDEZ-ABANS, M.; ORMENO, M. I.; OLIVEIRA-ABANS, M. de et al.: *Classification of planetary nebulae by cluster analysis and artificial neural networks.* IN: *Astronomy & Astrophysics* Nr. 116, S. 395–402, 1996.
- [FIS2010] FISCHER, H. K.: *Soziale Ungleichheit im Spiegel des Konsums. Eine clusteranalytische Untersuchung zum differentiellen Konsum im Kaiserreich. Bericht.* Köln, 2010.
- [FLA2012] FLATH, C.; NICOLAY, D.; CONTE, T. ET AL.: *Clusteranalyse von Smart-Meter-Daten.* IN: *Wirtschaftsinformatik* 2012 Nr. Volume 54, Issue 1, S. 33–42.
- [GOT2008] GOTTHARD MEINEL, ULRICH SCHUMACHER (Hrsg.): *Auszug aus: Flächennutzungsmonitoring, Grundlagen, Statistik, Indikatoren, Konzepte*, 2008, S. 1–14. https://www.destatis.de/DE/Publikationen/Thematisch/LandForstwirtschaft/Flaechennutzung/FlaechenstatistikDL.pdf?__blob=publicationFile. Abruf: 30.11.2015.
- [GRI1961] GRIMM, J.; GRIMM, W.: *Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm*, 1854-1961. <http://www.woerterbuchnetz.de/DWB?lemma=kluster>. Abruf: 24.02.2016.

- [GRI1999] GRIMM, J.; GRIMM, W.; HEYNE, M. ET AL.: *Deutsches Wörterbuch // Deutsches Wörterbuch von Jacob und Wilhelm Grimm. Viertes Band I. Abteilung 4. Teil: Gewöhnlich - Gleeve*. Fotomechan. Nachdr. d. Erstausg. 1949-1984. München: Deutscher Taschenbuch Verlag, 1999 // 1991. (Dtv, 59045). - ISBN: 9783423059459.
- [HAL2004] HALL, M. E.: *Pottery production during the Late Jomon period. Insights from the chemical analyses of Kasori B pottery*. IN: *Journal of Archaeological Science* 31 Nr. 10, S. 1439–1450. 2004
- [HEI1999] HEIDENREICH, M.: *Das europäische Städte- und Regionensystem im Wandel*. In: H. Voelzkow H.-G. Brose (Hg.): *Institutioneller Kontext wirtschaftlichen Handelns und Globalisierung*. Marburg: Metropolis, 1999., S. 291–317.
- [HOL2015] HOLM, K.: *Almo Statistiksystem 15*, 2015. <http://www.almo-statistik.de/>. Abruf: 28.02.2016.
- [HÖP1997] HÖPPNER, F.; KLAUWONN, F.; KRUSE, R.: *Fuzzy-Clusteranalyse. Verfahren für die Bilderkennung, Klassifizierung und Datenanalyse*. Wiesbaden: Vieweg+Teubner Verlag, 1997. (Computational Intelligence). - ISBN: 3-528-05543-X.
- [IBM2016] IBM: *SPSS Software*. IBM (HRSG.), 2016. <http://www-01.ibm.com/software/de/analytics/spss/>. Abruf: 28.02.2016.
- [IHK2015] INDUSTRIE- UND HANDELSKAMMER HESSEN: *Nutzung von Wärmetechnologien, Ratgeber Wärme in Hessen*. IHK HESSEN (HRSG.). http://www.ihk-hessen.de/pdf/umwelt_energie/waermestudie-online_10.02.2015.pdf, 2015. Abruf: 31.10.2015.
- [INF2010] INFAS UND DLR: *Mobilität in Deutschland 2008. Ergebnisbericht*. Struktur - Aufkommen - Emissionen - Trends. Hg. v. infas Institut für angewandte Sozialwissenschaft GmbH und Deutsches Zentrum für Luft- und Raumfahrt e.V. Bonn und Berlin, 2010. Abruf: 04.12.2015.
- [INS2016] INSTITUTE OF MEDICAL SCIENCE UNIVERSITY OF TOKYO: *Open Source clustering Software*, 2016. <http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm>. Abruf: 28.02.2016.
- [INW2011] IMMOBILIENWIRTSCHAFT, STADT- UND REGIONALENTWICKLUNG GMBH: *Wege aus dem Vermieter-Mieter Dilemma, Konzeptstudie*. INWIS (HRSG.), 2011. <http://web.gdw.de/uploads/pdf/InWIS-Vermieter-Mieter-Dilemma.pdf>. Abruf: 28.11.2015.
- [ISE2013] ISERMANN, J.; ULRICH, S.; BRUNS, R. ET AL.: *Generierung von Konstruktionsmerkmalen zur Entwicklung eines bio-inspirierten Kommissionier- Greifers für die Intralogistik mittels Clusteranalyse*. IN: *Logistics Journal: Proceedings*, 2013. S. 1–12.
- [ISE2014] ISERMANN, J.; ULRICH, S.; BRUNS, R. ET AL.: *Greifergestützte Automatisierung in der Intralogistik. Die Clusteranalyse hilft bei der Konstruktion*. IN: *Hebezeuge Fördermittel Nr. 1-2*, S. 36–39, 2014.
- [ISO2011] DIN EN ISO 50001:2011, 12.2011: *Energiemanagementsysteme - Anforderungen mit Anleitung zur Anwendung (ISO 50001:2011); Deutsche Fassung EN ISO 50001:2011*.
- [IWU2015] IWU, 2015.: *Gradtagzahlen_Deutschland.xls*: Institut Wohnen und Umwelt. www.iwu.de/fileadmin/user_upload/.../Gradtagzahlen_Deutschland.xls. Abruf: 02.12.2015.
- [JOH2003] JOHN, R.; KERBY, D. S.; HENNESSY, C. H. ET AL.: *Patterns and Impact of Comorbidity and Multimorbidity Among Community-Resident American Indian Elders*. IN: *The Gerontologist* 43 Nr. 5, S. 649–660, 2003.
- [KBA2014] KRAFTFAHRTBUNDESAMT: *Fahrzeugzulassungen (FZ). Bestand an Kraftfahrzeugen und Kraftfahrzeuganhängern nach Zulassungsbezirken 1. Januar 2014*. Statistik. Hg. v. KBA, 2014. Abruf: 03.12.2015.
- [KEE2003] KEES, V. R.; KOEN, V. E.: *Media repertoires of selective audiences. The impact of status, gender, and age on media use*. IN: *Poetics* 31 Nr. 5-6, S. 465–490.
- [KET2006] KETTENRING, J. R.: *The Practice of Cluster Analysis*. IN: *Journal of Classification* 23 Nr. 1, S. 3–30. 2006.
- [KIM2003] KIM, C.; SHIN, H.; CHOI, H.-K. ET AL.: *A phenetic analysis of Typha in Korea and far east Russia*. IN: *Aquatic Botany* Nr. 75, S. 33–43, 2003.
- [KOP2014] KOPATZ, M.: *Suffizienz als Teil der Energiewende*. IN: *Energiewirtschaftliche Tagesfragen* 64 Nr. 10, S. 8–12, 2014.

- [LEM2004] LE MAPIHAN, K.; VIAL, J.; JARDY, A. ET AL.: *Testing of "special base" columns in reversed-phase liquid chromatography*. IN: *Journal of Chromatography A* Nr. 1-2, S. 135–147.
- [MEY2013] MEYER, J.: *Nachhaltige Stadt- und Verkehrsplanung. Grundlagen und Lösungsvorschläge*. Wiesbaden: Vieweg+Teubner Verlag, 2013. - ISBN: 978-3-8348-2411-0.
- [MON2008] MONKA, M.; SCHÖNECK-VOß, N. M.; VOß, W. ET AL.: *Statistik am PC. Lösungen mit Excel*. 5., aktualisierte und erw. Aufl. München: Hanser, 2008. - ISBN: 978-3446415553.
- [NAT2016] NATIONAL AERONAUTICS AND SPACE ADMINISTRATION: *AutoClass*. NASA (HRSG.). <http://ti.arc.nasa.gov/tech/rse/synthesis-projects-applications/autoclass/autoclass-c/>. Abruf: 28.02.2016.
- [PAR2004] PARKER, H. G.; ET AL.: *Genetic Structure of the Purebred Domestic Dog*. IN: *Science* Nr. 5674, S. 1160–1164. 2004.
- [PBB2013] POLITISCHE BILDUNG BRANDENBURG: *Kommunalpolitik. Stadt, Kreis, Land*. BRANDENBURGISCHE LANDESZENTRALE FÜR POLITISCHE BILDUNG (HRSG.), 2013. <http://www.politische-bildung-brandenburg.de/node/2111>. Abruf: 31.10.2015.
- [PEH2010] PEHNT, M.: *Energieeffizienz: Ein Lehr- und Handbuch*. Springer Berlin Heidelberg, 2010. ISBN: 978-3-642-14250-5.
- [POP2015] POPPENBORG, R.: *Literaturrecherche zum Anwendungsfeld der Clusteranalyse*, Bochum. RUHR-UNIVERSITÄT BOCHUM (HRSG.). ET, LEE, 2015.
- [PRE2014] PREISVERGLEICH.DE: *120-Städte-Studie: Stromverschwender in Deutschland*, 2014. <http://strom.preisvergleich.de/info/8972/120-staedte-studie-stromverschwender/>. Abruf: 01.11.2015.
- [PRO2015] ANGERMEIER, G.: *Glossar, Deutsche Begriffe, Top-Down*. PROJEKTMAGAZIN (HRSG.), 2015. <https://www.projektmagazin.de/glossarterm/top-down>. Abruf: 01.11.2015.
- [RON1994] RONG, G. H.; SLATER, K.; FEI, R. C. ET AL.: *The Use of Cluster Analysis for Grading Textile Yarns*. IN: *The Journal of the Textile Institute* 1994 Nr. 85, S. 389–396.
- [RPR2016] R PROJECT: *The R Project for Statistical Computing*. PROJECT R (HRSG.), 2016. <https://www.r-project.org/>. Abruf: 28.02.2016.
- [SAA2013] SAARLAND: *Fläche, Bevölkerung in den Gemeinden am 31.12.2013*. SAARLAND (HRSG.), 2013. http://www.saarland.de/dokumente/thema_statistik/FB311213_nZPROECK2.pdf. Abruf: 31.10.2015.
- [SAN2002] SANK, C.: *Untersuchung des Nachhaltigkeitsindikatorensatzes "Indikatoren für eine nachhaltige Entwicklung in NRW" und dessen Anwendung in der kommunalen Planungspraxis. Dargestellt am Beispiel der Stadt Herdecke*, Dortmund. UNIVERSITÄT DORTMUND (HRSG.), 2002.
- [SAS2016] SAS: *SAS/STAT Software*. SAS (HRSG.), 2016. http://www.sas.com/en_us/software/analytics/stat.html. Abruf: 28.02.2016.
- [SCH2004] SCHÜNZEL, J.-O.: *Clusteranalyse zur Typenbildung von Unternehmen*. München: GRIN Verlag GmbH, 2007. ISBN: 9783638708500.
- [SCH2010] SCHENDERA, C. FG: *Clusteranalyse mit SPSS. Mit Faktorenanalyse*. München: Oldenbourg, 2010. (Wirtschaftsmathematik- und Statistik 8-2011). ISBN: 9783486586916.
- [SCH2013] SCHMIDT, J. A.: *ÖPNV und Lifestyle in verschiedenen Städten*. Bochum. mündlich an W. Wall, 2013., 23.09.2013.
- [SPR2016] SPRINGER GABLER VERLAG (Hrsg.), Gabler Wirtschaftslexikon, Stichwort: Skalenniveau, 2016. <http://wirtschaftslexikon.gabler.de/Archiv/11233/skalenniveau-v11.html>. Abruf 05.02.2016
- [STA2009] MICHAEL STASIUS: *Clusteranalyse. Arten und Anwendungen*, Aachen. RWTH AACHEN (HRSG.), 2009. https://www.google.de/search?q=stasius+2009+cluster&ie=utf-8&oe=utf-8&gws_rd=cr&ei=9p7IVsaKMMra6AS6ybjYCw. Abruf: 13.03.2016.
- [STE1977] STEINHAUSEN, D.; LANGER, K.: *Clusteranalyse. Einführung in Methoden und Verfahren der automatischen Klassifikation ; mit zahlreichen Algorithmen, FORTRAN-Programmen, Anwendungsbeispielen und einer Kurzdarstellung der multivariaten statistischen Verfahren*. Berlin: De Gruyter, 1977. (De Gruyter Lehrbuch). ISBN: 3110070545.

- [STE2004] STEFURAK, T.; CALHOUN, G. B.; GLASER, B. A. ET AL.: *Personality Typologies of Male Juvenile Offenders Using a Cluster Analysis of the Millon Adolescent Clinical Inventory Introduction*. IN: *International Journal of Offender Therapy and Comparative Criminology* 48 Nr. 1, S. 96–110.
- [STE2007] STEINLEY, D.; BRUSCO, M. J.: *Initializing K-means Batch Clustering. A Critical Evaluation of Several Techniques*. In: *Journal of Classification*, Bd. 24, S. 99–121. DOI: 10.1007/s00357-007-0003-0
- [STE2011] STEIN, P.; VOLLNHALS, S.: *Grundlagen clusteranalytischer Verfahren*, Duisburg, Essen. UNIVERSITÄT DUISBURG-ESSEN (HRSG.). Institut für Soziologie, 2011. https://www.uni-due.de/imperia/md/content/soziologie/stein/skript_clusteranalyse_sose2011.pdf. Abruf: 08.02.2016.
- [STR2007] STROHMEIER, K.; SCHULTZ, A.; BARDEHLE, D. ET AL.: *Sozialräumliche Clusteranalyse der Kreise und kreisfreien Städte und Gesundheitsindikatoren in NRW*. IN: *Gesundheitswesen* 69 Nr. 1, S. 26–33.
- [SW2008] STIFTUNG WARENTEST: *Heizkörperthermostate: Auf Sparen programmiert*. IN: *Stiftung Warentest - Test Nr. 5*, S. 68–71, 2008. <https://www.test.de/Heizkoerperthermostate-Auf-Sparen-programmiert-1672635-0/>. Abruf: 31.10.2015.
- [SYS2016] SYSTATSOFTWARE: *Systat*. SYSTATSOFTWARE (HRSG.), 2016. <https://systatsoftware.com/>. Abruf: 28.02.2016.
- [TEC2013] TECHEM ENERGIESERVICES GMBH: *Energiekennwerte 2013, Hilfen für den Wohnungswirt, Eine Studie von Techem*. 14. Auflage. Eschborn, 2013.
- [THE2007] THEES, N.: *Statistische Qualität zusammengesetzter Indikatoren am Beispiel der Wissensökonomie*. In: Statistisches Bundesamt (Hg.): *Wissenschaftliches Kolloquium 2007*, S. 1–10, 2007.
- [TIM2002] TIMM, H.: *Fuzzy-Clusteranalyse: Methoden zur Exploration von Daten mit fehlenden Werten sowie klassifizierten Daten*. Dissertation. Magdeburg, 2002. [iks.cs.ovgu.de/~htimm/data/dissertation.pdf](https://www.iks.cs.ovgu.de/~htimm/data/dissertation.pdf). Abruf: 26.02.2016.
- [TSE2015] TSENG, S.-F.; LEE, T.-S.; DENG, C.-Y. ET AL.: *Cluster analysis of medical service resources at district hospitals in Taiwan, 2007–2011*. IN: *Journal of the Chinese Medical Association* 78 Nr. 12, S. 732–745.
- [UBA2012] UMWELTBUNDESAMT: *Daten zum Verkehr. Ausgabe 2012*. UBA (HRSG.), 2012. <https://www.umweltbundesamt.de/publikationen/daten-verkehr>. Abruf: 31.10.2015.
- [UBA2015] UMWELTBUNDESAMT: *Energieverbrauch der privaten Haushalte*. UBA (HRSG.), 2015. <https://www.umweltbundesamt.de/daten/private-haushalte-konsum/energieverbrauch-der-privaten-haushalte>. Abruf: 01.11.2015.
- [UNG1997] UNGER, H.; GERNHARDT, D.; MOHR, M. ET AL.: *Kategorisierung der Städte und Gemeinden in Nordrhein-Westfalen unter energetischen und strukturellen Aspekten*. Hg. v. Ruhr-Universität Bochum, Lehrstuhl für Nukleare und Neue EnergieSysteme. Bochum (Technischer Fachbericht), 1993.
- [UZH2015] SCHWARZ, J.: *Methodenberatung, Datenanalyse, Deskriptive, univariate Analyse*. UNIVERSITÄT ZÜRICH - UZH (HRSG.), 2014. <http://www.methodenberatung.uzh.ch/datenanalyse/deskuniv.html>, zuletzt aktualisiert am 15.10.2014. Abruf: 31.10.2015.
- [VDI2012] VEREIN DEUTSCHER INGENIEURE E. V.: *VDI 2067 Blatt 1. Wirtschaftlichkeit gebäudetechnischer Anlagen, Grundlagen und Kostenberechnung*. VDI E.V. (HRSG.), 2012. Abruf: 02.12.2015.
- [VDI2013] VEREIN DEUTSCHER INGENIEURE E. V.: *VDI 3807 Blatt 1. Verbrauchskennwerte für Gebäude, Grundlagen*. VDI E.V. (HRSG.), 2013. Abruf: 02.12.2015.
- [VDI2015] MATZER, M.: *Statistiksprache R beliebt für vorausschauende Analysen*. IN: *VDI Nachrichten Nr. 20*, 2015.
- [VDV2013] VERBAND DEUTSCHER VERKEHRSUNTERNEHMEN E.V.: *Statistik 2012*. VDV (HRSG.). Köln. www.vdv.de, 2013. Abruf: 05.12.2015.
- [VDV2015] VERBAND DEUTSCHER VERKEHRSUNTERNEHMEN E.V.: *Homepage*. VDV (HRSG.), 2015. <https://www.vdv.de/>. Abruf: 31.10.2015.

- [WAL2013] WALL, W.: *Entwicklung von Energieclustern zur Einteilung deutscher Städte*. Ruhr-Universität Bochum, Lehrstuhl Energiesysteme und Energiewirtschaft, Bochum (Masterarbeit), 2013.
- [WAL2014] WALL, W.; EICKELKAMP, T.; KOCH, M. K.: *Clusterung deutscher Städte anhand energetischer Kennwerte*. In: Marco K. Koch und Russell McKenna (Hg.): *Wettbewerb Energieeffiziente Stadt*, Band 3: Methoden und Modelle. Berlin: LIT-Verlag, 2014., S. 61–70-978-3-643-12828-7.
- [WAL2016A] WALL, W.; SCHAACK, F.; KOCH, M. K.: *Analyse ausgewählter Energieeffizienzmaßnahmen der Gewinnerstädte des Wettbewerbs Energieeffiziente Stadt*. In: H.-J. Wagner und P. von Both (Hg.): *Wettbewerb Energieeffiziente Stadt*, Band 8: Gute Beispiele der Umsetzungsphase. Berlin: LIT-Verlag, 2016.
- [WAL2016B] WALL, W.; WAGNER, H.-J.; KOCH, M. K.: *Energiecluster deutscher Städte – Clusteranalyse deutscher Städte anhand sozio-energetischer Indikatoren*. In: Institut für Elektrizitätswirtschaft und Energieinnovation (IEE), Technische Universität Graz (TUG) (Hg.): *EnInnov2016. 14. Symposium Energieinnovation. Energie für unser Europa*, Bd. 14. Graz: Verlag der Technischen Universität Graz, 2016., S. 1–12. https://www.tu-graz.at/fileadmin/user_upload/Events/Eninnov2016/.../LF_Wall.pdf. Abruf: 16.05.2016- 9783851254488.
- [WEI2014] WEISCHE, L.: *Untersuchung energetischer Indikatoren für die Sektoren Private Haushalte und Verkehr in Städten und Kommunen*, Bochum. RUHR-UNIVERSITÄT BOCHUM (HRSG.). ET, LEE, 2014.
- [Wis2013] WISCHERMANN, S.: *Zukünftiger regionaler Endenergiebedarf deutscher Haushalte unter Berücksichtigung des demografischen Wandels*. Promotionsvortrag. Bochum, Mai 2013, 2013.
- [WIZ2015] WELT IN ZAHLEN: *Welt in Zahlen*. Unter Mitarbeit von M. Kania. Welt-in-Zahlen.de (Hrsg.), 2015. <http://www.welt-in-zahlen.de/laendervergleich.phtml?indicator=68>. Abruf: 31.10.2015.
- [ZEN2015] ZENSUS 2011: *Zensusdatenbank*. ZENSUS2011 (HRSG.), 2015. https://www.zensus2011.de/DE/Home/home_node.html. Abruf: 31.10.2015.

Anhang

Abkürzungen zum Anhang 1:

Bev = Bevölkerungsentwicklung
 WF = Wohnfläche pro Person
 1.2Fam = Anteil Wohnungen in 1 und 2 Familienhäusern
 HHDi = Haushaltsdichte
 Eink = verfügbares Einkommen der Haushalte
 HAT/HGT/GTZ = Heiztage/Heizgradtage/Gradtagzahlen
 PKS = Pro-Kopf-Stromverbrauch
 A65 = Anteil über 65-Jähriger
 ASV = Anteil der Siedlungs- und Verkehrsfläche an Gesamtfläche
 ARB = Arbeitslosenquote

Tabelle A-1: Indikatoren des Verbrauchssektor private Haushalte

	Bev [%]	WF [qm]	1.2Fam [%]	HHDi [HH/km²]	Eink [€]	HT [-]	HGT [-]	GTZ [-]	PKS [kWh]	A65 [%]	ASV [-]	ARB. [%]
Flensburg	6,02	40,2	25,4	789	35.155	288	2.548	4.042	1.876	20,6	51,8	12,3
Kiel	4,05	37,5	20,7	1.078	32.243	288	2.548	4.042	1.655	18,3	54,0	11,4
Lübeck	-1,33	38,9	32,3	509	35.606	273	2.501	3.911	1.797	23,0	36,7	11,6
Neumünster	-3,62	41,6	41,5	529	35.817	273	2.501	3.911	k. A.	21,9	49,9	12,5
Hamburg	4,86	35,7	21,4	1.160	43.581	273	2.376	3.782	1.814	19,0	59,8	8,6
Braunschweig	1,93	41,7	23,3	671	37.504	252	2.407	3.754	1.832	21,1	46,7	7,8
Salzgitter	-9,40	42,4	35,7	210	37.980	252	2.407	3.754	1.903	23,5	28,5	10,4
Wolfsburg	0,64	42,9	36,2	291	39.661	252	2.407	3.754	2.020	23,1	31,1	5,4
Delmenhorst	-2,80	44,2	56,9	554	38.768	268	2.322	3.720	k. A.	21,6	48,2	11,9
Emden	1,11	40,3	53,2	213	36.048	281	2.140	3.525	1.166	20,3	27,2	10,0
Oldenburg	4,94	44,4	48,0	787	37.054	281	2.140	3.525	2.116	18,4	60,0	9,1
Osnabrück	0,56	41,4	32,9	666	40.020	262	2.328	3.697	2.180	19,9	49,7	8,6
Wilhelmshaven	-5,00	45,0	28,2	382	34.307	281	2.140	3.525	1.920	25,0	44,8	13,7
Bremen	1,65	40,4	36,3	843	41.771	268	2.322	3.720	1.815	20,9	58,9	11,5
Bremerhaven	-6,49	43,5	27,4	604	34.626	268	2.322	3.720	2.101	22,3	46,4	16,3
Düsseldorf	4,04	39,0	13,3	1.464	42.414	251	1.976	3.262	1.832	20,1	59,9	9,8
Duisburg	-5,23	37,5	23,1	1.007	34.128	251	1.976	3.262	1.915	21,1	61,2	14,3
Essen	-3,66	39,8	19,0	1.371	38.794	251	1.976	3.262	1.964	22,0	67,6	13,6
Krefeld	-2,30	41,3	31,8	790	41.108	251	1.976	3.262	1.984	21,5	55,6	12,5
Mönchengl.	-2,21	39,6	39,2	718	41.190	251	1.976	3.262	1.885	20,5	49,1	12,7
Mülheim a.d.R.	-3,30	43,2	27,2	909	45.714	251	1.976	3.262	1.854	23,8	55,6	8,3
Oberhausen	-4,31	36,8	27,7	1.324	35.945	251	1.976	3.262	2.000	20,9	75,6	13,6
Remscheid	-8,12	41,7	29,8	705	47.361	251	1.976	3.262	1.901	21,8	44,0	10,3
Solingen	-3,20	37,4	32,8	821	46.630	251	1.976	3.262	1.823	21,7	48,4	9,9
Wuppertal	-4,63	40,3	20,2	1.029	40.981	251	1.976	3.262	1.828	21,3	49,1	13,1
Bonn	8,49	38,7	28,7	1.103	42.553	291	2.606	4.074	1.837	17,7	50,5	7,8
Köln	5,64	37,4	18,7	1.272	39.720	251	1.976	3.262	1.832	22,1	60,8	10,6
Leverkusen	0,09	38,7	32,1	968	39.947	291	2.606	4.074	1.832	18,4	59,5	9,3
Bottrop	-3,52	37,1	39,4	544	39.233	251	1.976	3.262	1.874	20,9	43,9	9,9
Gelsenkirchen	-7,91	38,4	16,2	1.193	33.384	251	1.976	3.262	1.826	20,9	74,6	16,0
Münster	9,84	40,2	31,4	489	40.465	262	2.328	3.697	1.782	17,0	33,3	6,5
Bielefeld	0,51	38,2	35,1	609	40.855	272	2.316	3.650	1.777	20,2	42,2	10,4
Bochum	-4,39	37,3	22,0	1.279	37.138	251	1.976	3.262	1.852	21,7	70,7	11,4
Dortmund	-1,36	39,4	23,3	1.032	35.227	251	1.976	3.262	1.830	20,5	59,2	14,6
Hagen	-7,73	42,0	21,6	570	40.631	291	2.606	4.074	1.940	22,6	35,0	11,8
Hamm	-0,17	38,5	44,7	349	38.235	291	2.606	4.074	k. A.	20,1	32,6	11,5
Herne	-5,89	37,2	20,1	1.468	36.449	251	1.976	3.262	1.692	22,0	77,8	15,2
Darmstadt	7,82	39,8	25,2	593	43.071	251	2.372	3.688	2.004	18,0	34,9	7,4
Frankfurt a.M.	6,96	36,1	13,6	1.429	38.537	231	2.142	3.396	1.750	16,3	57,9	8,3
Offenbach a.M.	4,40	34,5	17,3	1.243	35.904	251	2.372	3.688	1.970	18,0	49,2	12,5
Wiesbaden	3,26	38,8	21,4	650	44.014	291	2.606	4.074	1.621	19,5	38,9	8,3
Kassel	0,90	39,5	24,5	932	33.504	272	2.316	3.650	1.962	19,9	60,0	11,3
Koblenz	-1,18	45,3	31,1	528	38.276	291	2.606	4.074	1.888	21,8	36,5	7,2
Trier	6,30	41,1	34,6	476	36.660	255	2.279	3.630	1.974	18,1	31,3	6,2
Frankenth. i.d.P.	-1,59	43,0	36,7	492	41.341	262	2.371	3.736	k. A.	22,6	36,1	7,8

Kaiserslautern	-0,04	44,3	34,8	364	34.387	262	2.371	3.736	1.953	20,0	28,0	11,8
Landau i.d.P.	6,89	44,7	47,5	250	42.824	-	2.054	3.277	k. A.	19,1	23,8	5,5
Ludwigsh. a.R.	2,05	39,2	30,3	958	38.855	262	2.371	3.736	2.077	19,9	61,1	10,2
Mainz	9,89	39,1	26,9	1.049	39.159	-	2.054	3.277	1.923	17,7	50,5	6,8
Neustadt a.d.W.	-1,81	47,8	51,4	210	48.869	262	2.371	3.736	k. A.	23,0	19,5	6,6
Pirmasens	-11,51	51,3	41,4	334	38.175	262	2.371	3.736	k. A.	25,0	32,5	14,7
Speyer	0,52	42,4	38,0	545	50.323	262	2.371	3.736	k. A.	20,8	44,4	7,1
Worms	2,00	41,0	45,9	341	41.199	262	2.371	3.736	1.607	19,8	29,6	9,4
Zweibrücken	-5,35	49,6	58,5	226	40.951	262	2.371	3.736	k. A.	22,2	28,1	8,8
Stuttgart	5,06	35,9	16,9	1.400	48.718	251	2.426	3.761	1.705	18,7	51,5	6,5
Heilbronn	4,15	38,5	35,0	539	71.611	251	2.426	3.761	2.065	20,6	35,6	7,4
Baden-Baden	3,28	51,6	33,4	187	55.810	-	2.054	3.277	k. A.	26,5	14,7	6,3
Karlsruhe	6,80	39,0	22,5	855	41.107	-	2.054	3.277	1.976	18,9	46,5	6,2
Heidelberg	6,68	37,1	22,1	670	45.186	251	2.372	3.688	1.720	16,3	30,1	5,8
Mannheim	2,67	38,9	18,9	1.042	38.427	-	2.054	3.277	1.779	19,1	58,1	6,7
Pforzheim	3,03	39,2	26,4	563	45.090	251	2.426	3.761	1.951	21,3	30,8	8,9
Freiburg i.B.	11,72	34,6	19,6	704	38.443	-	2.054	3.277	1.637	16,5	31,8	6,7
Ulm	5,49	37,7	31,9	476	49.367	277	2.656	4.041	1.587	18,5	32,1	5,6
Ingolstadt	9,51	40,0	36,6	447	43.236	262	2.621	3.951	1.559	18,9	36,9	4,0
München	13,88	37,0	10,5	2.359	47.634	297	3.084	4.606	1.779	18,0	75,6	5,9
Rosenheim	4,42	38,9	23,1	807	45.142	297	3.084	4.606	k. A.	19,7	40,0	5,7
Landshut	9,38	43,7	30,7	488	43.787	262	2.621	3.951	k. A.	21,5	32,2	5,5
Passau	0,02	43,1	28,3	400	35.410	262	2.621	3.951	2.040	21,8	31,9	6,3
Straubing	1,61	44,3	34,0	323	43.282	262	2.621	3.951	1.959	21,3	29,4	6,4
Amberg	-0,61	43,3	31,2	410	43.851	258	2.484	3.815	k. A.	21,9	34,4	6,9
Regensburg	8,67	42,4	17,2	995	34.626	262	2.621	3.951	2.183	18,2	54,0	5,3
Weiden i.d.O.	-2,50	46,4	29,3	304	38.445	258	2.484	3.815	k. A.	22,3	28,6	8,0
Bamberg	1,52	42,8	21,2	714	35.492	265	2.519	3.869	2.235	20,6	46,2	5,8
Bayreuth	-1,41	41,4	24,1	570	39.349	293	3.114	3.114	2.077	20,9	42,0	7,1
Coburg	-4,31	47,2	31,2	446	44.060	265	2.519	3.869	1.826	22,9	41,5	6,9
Hof	-9,53	46,8	25,0	406	37.661	293	3.114	3.114	1.808	24,0	35,7	7,9
Ansbach	0,33	46,9	34,3	195	41.087	258	2.484	3.815	k. A.	21,6	24,0	5,5
Erlangen	5,51	40,3	24,20	754	39.632	258	2.484	3.815	2.135	18,3	40,8	4,7
Fürth	5,29	40,4	21,4	906	46.435	258	2.484	3.815	2.079	18,6	45,6	7,9
Nürnberg	4,55	38,4	17,6	1.368	40.606	258	2.484	3.815	1.842	20,7	59,4	8,8
Schwabach	2,35	44,4	41,20	436	49.601	258	2.484	3.815	k. A.	21,0	33,0	5,2
Aschaffenburg	1,80	42,6	23,2	530	46.647	251	2.372	3.688	2.025	19,8	35,8	7,3
Schweinfurt	-1,98	41,5	20,7	726	37.931	251	2.372	3.688	1.496	24,0	55,7	7,1
Würzburg	4,57	39,5	18,7	813	38.506	251	2.372	3.688	1.911	20,0	47,7	5,4
Augsburg	4,57	38,7	17,0	946	35.340	277	2.656	4.041	1.940	20,5	43,4	7,3
Kaufbeuren	-0,38	41,5	32,3	502	42.729	291	2.948	4.425	k. A.	22,5	37,4	7,0
Kempten (Allgäu)	1,39	39,9	21,2	510	42.229	291	2.948	4.425	k. A.	22,2	32,0	5,5
Memmingen	0,44	41,7	35,6	283	49.323	277	2.656	4.041	k. A.	21,4	24,7	4,4
Berlin	3,54	38,0	10,5	2.013	32.812	245	2.326	3.606	1.751	19,3	70,1	13,9
Brandenb. a.d.H.	-7,72	40,2	21,2	168	29.866	256	2.479	3.814	1.654	26,3	22,1	14,5
Cottbus	-5,86	37,9	21,3	327	31.650	252	2.522	3.854	1.628	23,0	30,2	13,2
Frankfurt a.d.O.	-16,82	37,2	18,7	215	30.448	252	2.522	3.854	1.651	22,8	23,8	16,0
Potsdam	22,87	35,8	16,3	441	32.650	256	2.479	3.814	1.845	19,6	30,0	8,4
Rostock	1,87	34,7	10,5	626	28.120	268	2.415	3.812	1.551	22,9	39,7	12,6
Schwerin	-5,89	36,9	14,3	393	30.128	252	2.522	3.854	1.350	24,3	30,8	12,8
Chemnitz	-6,20	39,5	14,4	595	33.042	273	2.671	4.086	1.677	26,9	43,7	11,5
Dresden	10,88	35,2	127,0	840	32.284	256	2.466	3.791	1.463	22,2	11,6	9,9
Leipzig	7,83	39,3	10,6	954	29.539	257	2.417	3.745	1.670	22,7	40,7	12,3
Dessau-Roßlau	-9,27	43,9	29,9	184	33.366	257	2.417	3.745	1.683	28,2	52,0	12,9
Halle (Saale)	-2,67	38,9	12,5	929	29.169	257	2.417	3.745	1.635	23,5	20,8	13,2
Magdeburg	2,12	36,6	17,9	619	29.879	253	2.407	3.754	1.515	23,6	52,9	12,5
Erfurt	2,90	36,1	18,5	396	31.634	273	2.671	4.086	1.394	21,1	46,2	9,5
Gera	-12,47	39,9	17,1	347	32.212	273	2.671	4.086	1.638	26,4	27,0	13,2
Jena	5,58	35,5	20,5	503	29.529	273	2.671	4.086	1.622	20,3	23,5	7,7
Suhl	-20,42	38,8	32,4	191	39.394	282	2.881	4.343	1.644	27,0	23,9	8,3
Weimar	4,99	36,5	22,6	384	30.009	273	2.671	4.086	k. A.	21,3	18,9	9,5
Eisenach	-4,01	38,1	24,7	209	33.395	282	2.881	4.343	k. A.	25,2	26,3	10,5

Abkürzungen zum Anhang 2:

Kfz/TEW – Pkw/TEW = Motorisierungsgrad – Pkw pro tausend Einwohner

D-Kfz, B-Kfz, E-Kfz/TEW = Diesel-, Benzin-, Elektrofahrzeuge pro tausend Einwohner

A18.65J = Anteil der Bevölkerung zwischen 18 und 65 Jahren

BeD = Bevölkerungsdichte

ÖPNV/EW = Anzahl Fahrten im ÖPNV pro Einwohner und Jahr

P.km./EW = Anzahl gefahrener Personenkilometer im ÖPNV pro Einwohner und Jahr

MRW = Mittlere Reiseweite im ÖPNV

Tabelle A-2: Indikatoren des Verbrauchssektors Verkehr

	Kfz/TEW [-]	Pkw/TEW [-]	D-Kfz/ TEW [-]	B-Kfz/ TEW [-]	E-Kfz/ TEW [-]	A18.65J [%]	BeD [EW/km ²]	ÖPNV/ EW [-]	P.km./EW [km]	MRW [km]
Flensburg	516	481	166	309	0,20	64,59	1.471	110	279	2,53
Kiel	459	426	129	292	0,09	67,35	2.022	115	1.695	14,73
Lübeck	468	434	119	309	0,05	62,03	988	110	469	4,28
Neumünster	546	505	150	347	0,10	61,33	1.074	35	k.A.	k.A.
Hamburg	458	428	136	287	0,25	65,32	2.296	200	980	4,90
Braunschweig	554	515	168	337	1,07	64,21	1.279	149	1.480	13,21
Salzgitter	558	515	136	366	0,15	59,97	438	142	1.126	7,91
Wolfsburg	1.183	1.139	485	638	5,66	61,70	597	113	657	5,83
Delmenhorst	557	517	122	384	0,07	61,57	1.180	38	210	5,50
Emden	491	449	122	316	0,08	62,49	443	30	145	4,83
Oldenburg	546	509	165	330	0,18	66,01	1.541	112	k.A.	k.A.
Osnabrück	529	495	155	331	0,33	65,24	1.299	229	967	4,21
Wilhelmshaven	536	494	118	364	0,03	60,93	716	76	339	4,46
Bremen	452	421	121	293	0,32	63,63	1.679	189	943	5,00
Bremerhaven	456	426	111	306	0,07	61,43	1.155	125	558	4,45
Düsseldorf	527	491	182	302	0,12	65,00	2.731	393	1.915	4,87
Duisburg	484	446	108	327	0,02	62,34	2.091	130	560	4,32
Essen	516	479	124	345	0,10	62,71	2.695	221	1.105	5,00
Krefeld	531	493	133	347	0,09	61,92	1.612	187	777	4,15
Mönchengl.	535	497	138	346	0,04	63,01	1.497	185	986	5,32
Mülheim a.d.R.	572	530	135	386	0,07	60,96	1.826	163	796	4,89
Oberhausen	535	493	111	371	0,02	63,39	2.724	161	729	4,52
Remscheid	572	525	145	372	0,05	61,70	1.467	157	690	4,39
Solingen	588	537	139	388	0,13	61,07	1.734	161	644	4,00
Wuppertal	509	462	124	331	0,10	62,19	2.036	258	973	3,78
Bonn	588	555	234	314	0,07	66,59	2.196	286	1.585	5,55
Köln	473	439	142	289	0,53	61,30	2.528	290	1.584	5,47
Leverkusen	559	516	143	364	0,06	67,21	2.028	204	1.488	7,30
Bottrop	594	542	115	414	0,03	62,83	1.158	541	163	0,30
Gelsenkirchen	482	447	100	337	0,07	62,65	2.455	216	969	4,48
Münster	487	453	151	292	0,13	68,03	979	353	2.787	7,90
Bielefeld	506	473	140	325	0,12	62,67	1.269	296	2.399	8,11
Bochum	544	503	121	374	0,12	63,99	2.487	161	723	4,48
Dortmund	500	461	122	331	0,08	63,74	2.038	236	1.108	4,70
Hagen	533	488	125	356	0,09	61,05	1.161	185	1.063	5,76
Hamm	541	504	134	361	0,06	61,94	779	74	289	3,91
Herne	504	465	99	356	0,03	61,84	3.006	128	517	4,03
Darmstadt	505	470	155	307	0,26	66,07	1.212	181	908	5,01
Frankfurt a.M.	487	458	184	268	0,33	67,96	2.770	231	1.408	4,48
Offenbach a.M.	467	436	133	297	0,26	64,06	2.605	92	k.A.	k.A.
Wiesbaden	545	508	164	338	0,19	63,46	1.337	185	929	5,03
Kassel	462	430	128	295	0,18	64,67	1.806	310	1.982	6,40
Koblenz	556	517	167	343	0,07	63,26	1.044	367	1.691	4,60
Trier	486	446	158	285	0,19	67,91	910	155	772	4,98
Frankenthal i.d.P.	571	530	146	378	0,09	60,99	1.075	20	k.A.	k.A.
Kaiserslautern	531	497	135	355	0,08	65,36	695	189	1.450	7,68
Landau i.d.P.	611	564	160	398	0,02	65,50	526	26	k.A.	k.A.
Ludwigsh. a.R.	517	479	130	343	0,14	62,67	2.065	63	242	3,84
Mainz	495	461	150	305	0,15	67,99	2.074	319	2.655	7,27
Neustadt a.d.W.	636	588	173	409	0,11	61,29	446	k.A.	k.A.	k.A.

Pirmasens	586	546	148	392	0,00	60,08	656	52	283	5,49
Speyer	595	547	157	381	0,12	62,42	1.168	9	k.A.	k.A.
Worms	573	526	141	378	0,15	63,00	733	11	k.A.	k.A.
Zweibrücken	637	586	159	419	0,09	62,35	482	10	k.A.	k.A.
Stuttgart	511	475	166	302	1,50	65,87	2.884	201	1.053	5,23
Heilbronn	553	518	159	353	0,15	62,13	1.177	146	737	5,06
Baden-Baden	662	611	204	402	0,51	58,79	375	180	874	4,85
Karlsruhe	493	455	149	300	0,23	66,34	1.707	48	238	4,94
Heidelberg	410	380	124	252	0,09	69,60	1.381	63	242	3,84
Mannheim	520	483	146	331	0,10	65,27	2.032	187	719	3,84
Pforzheim	514	480	144	330	0,08	61,47	1.188	198	k.A.	k.A.
Freiburg i.B.	436	402	139	257	0,28	67,69	1.425	384	1.759	4,58
Ulm	559	517	193	318	0,29	65,25	994	181	1.162	6,43
Ingolstadt	762	704	283	412	1,26	64,17	959	110	k.A.	k.A.
München	535	493	193	295	0,56	67,54	4.468	277	1.413	5,09
Rosenheim	589	535	188	342	0,18	63,92	1.610	k.A.	k.A.	k.A.
Landshut	558	511	177	329	0,06	63,31	993	104	k.A.	k.A.
Passau	576	521	177	339	0,18	64,59	705	146	584	4,01
Straubing	596	548	175	368	0,27	64,53	667	k.A.	k.A.	k.A.
Amberg	607	557	175	376	0,07	62,67	829	k.A.	k.A.	k.A.
Regensburg	556	508	183	317	0,20	67,43	1.714	133	1.194	8,96
Weiden i.d.O.	640	588	185	397	0,07	62,00	591	k.A.	k.A.	k.A.
Bamberg	573	518	174	338	0,13	64,84	1.297	125	311	2,49
Bayreuth	553	511	163	341	0,10	65,35	1.068	117	385	3,28
Coburg	631	578	195	377	0,20	62,42	849	107	389	3,65
Hof	569	522	138	379	0,20	61,18	766	117	563	4,83
Ansbach	624	572	192	372	0,03	62,44	397	42	162	3,91
Erlangen	537	490	163	322	0,19	66,18	1.370	114	3.064	26,89
Fürth	528	482	143	334	0,12	65,37	1.868	256	758	2,96
Nürnberg	507	465	152	308	0,13	64,37	2.657	306	1.313	4,29
Schwabach	668	602	179	417	0,05	61,79	959	3	14	5,59
Aschaffenburg	644	597	205	384	0,22	64,06	1.083	123	566	4,60
Schweinfurt	548	509	171	330	0,25	60,42	1.459	145	543	3,75
Würzburg	527	482	161	313	0,14	67,93	1.422	343	1.154	3,37
Augsburg	497	459	140	312	0,08	64,32	1.857	204	897	4,40
Kaufbeuren	576	525	164	354	0,22	60,85	1.038	k.A.	k.A.	k.A.
Kempten(Allgäu)	602	547	190	351	0,37	62,25	1.021	k.A.	k.A.	k.A.
Memmingen	622	572	194	371	0,10	61,66	592	15	4	0,30
Berlin	372	342	78	257	0,19	65,72	3.785	281	1.302	4,63
Brandenb. a.d.H.	491	463	99	357	0,06	62,11	311	103	445	4,31
Cottbus	502	475	101	368	0,38	65,10	608	105	421	4,02
Frankfurt a.d.O.	507	478	89	380	0,09	64,85	396	160	443	2,76
Potsdam	451	424	109	308	0,06	65,32	850	184	753	4,09
Rostock	415	394	96	293	0,06	65,16	1.119	177	683	3,87
Schwerin	477	457	119	330	0,09	62,49	699	166	830	5,00
Chemnitz	529	499	k.A.	k.A.	0,10	60,71	1.092	163	838	5,15
Dresden	436	411	k.A.	k.A.	0,15	63,81	1.599	286	1.375	4,81
Leipzig	420	398	k.A.	k.A.	0,25	63,97	1.751	176	1.273	7,23
Dessau-Roßlau	526	495	100	386	0,02	60,34	346	54	307	5,65
Halle(Saale)	412	391	79	305	0,03	63,55	1.714	183	1.137	6,21
Magdeburg	480	457	104	348	0,07	63,95	1.144	196	1.332	6,79
Erfurt	480	450	119	323	0,10	65,29	756	204	1.172	5,74
Gera	497	471	112	351	0,03	62,12	628	167	836	5,01
Jena	427	401	111	284	0,05	66,96	934	186	824	4,42
Suhl	577	541	145	387	0,08	62,37	350	143	719	5,03
Weimar	464	435	113	314	0,00	64,30	751	94	378	4,02
Eisenach	529	494	118	369	0,05	61,47	402	47	573	12,10

Abkürzungen zum Anhang 3:

BA = Beschäftigtenanteil

EP = Energieproduktivität

EI = Energieintensität

PN = Produktivitätsniveau

SEV = Spezifischer Energieverbrauch

IU/GU – GHDU/GU= Industrie/GHD Umsatzsteuer an Gesamtumsatzsteuer

EBA = Entwicklung Beschäftigtenanteil

Tabelle A-3: Indikatoren der Verbrauchssektoren Industrie und GHD

	Industrie							GHD				
	BA [%]	EP [€/GJ]	EI [MJ/€]	PN [€/GJ]	SEV [MJ/€]	IU/GU [%]	EBA [%]	BA [%]	EP [€/GJ]	PN [€/GJ]	GHDU/GU [%]	EBA [%]
Flensburg	12,8	312,8	3,20	432,9	2,31	37,768	-28,5	87,0	548,9	452,0	62,12	6,2
Kiel	12,1	945,7	1,06	1165,9	0,86	18,45	-8,8	87,9	555,0	470,5	81,53	1,3
Lübeck	17,5	689,4	1,45	850,8	1,18	37,86	-5,8	82,2	516,0	424,5	57,45	1,4
Neumünster	19,1	629,0	1,59	737,4	1,36	19,17	-13,8	79,8	531,5	442,6	75,75	2,9
Hamburg	12,8	175,1	5,71	183,3	5,46	44,62	-12,5	87,0	789,4	688,8	55,36	2,2
Braunschweig	17,9	605,9	1,65	840,1	1,19	31,32	-8,3	81,8	564,1	444,2	60,52	2,1
Salzgitter	44,4	16,6	60,23	20,2	49,43	1,98	-10,5	55,1	733,0	467,0	12,51	10,4
Wolfsburg	49,9	k.A.	k.A.	k.A.	k.A.	k.A.	-14,7	49,9	1096,0	509,1	k.A.	21,3
Delmenhorst	15,2	223,4	4,48	225,9	4,43	37,33	-19,4	84,1	427,0	373,3	56,23	4,6
Emden	35,6	426,7	2,34	552,2	1,81	31,13	-9,3	64,1	657,3	442,0	64,26	6,6
Oldenburg	11,1	658,7	1,52	767,5	1,30	12,13	-12,8	88,7	527,1	453,7	38,48	2,3
Osnabrück	17,0	184,5	5,42	198,7	5,03	36,97	-17,0	82,7	531,9	456,3	60,17	4,5
Wilhelmshaven	14,6	130,8	7,65	226,6	4,41	53,68	-14,7	85,4	616,7	462,5	41,94	3,0
Bremen	19,2	58,7	17,05	76,0	13,15	37,17	-11,3	80,6	652,7	522,0	61,94	3,1
Bremerhaven	18,0	469,3	2,13	377,1	2,65	44,90	-0,2	81,7	550,0	503,4	51,35	-0,1
Düsseldorf	11,6	256,5	3,90	291,3	3,43	41,59	-15,2	88,2	797,2	688,2	58,38	2,4
Duisburg	24,7	10,2	98,26	13,6	73,35	27,45	-3,4	75,1	713,2	533,9	66,92	1,2
Essen	14,8	237,2	4,22	457,7	2,18	37,36	-13,8	85,0	722,5	516,1	30,62	2,9
Krefeld	26,8	96,2	10,39	126,3	7,92	46,05	-4,5	72,8	633,5	471,0	46,53	1,6
Mönchengl.	20,7	654,3	1,53	668,5	1,50	25,44	-14,0	78,9	553,2	477,2	61,25	4,7
Mülheim a.d.R.	26,4	307,6	3,25	275,3	3,63	10,77	3,1	73,3	743,0	665,3	89,06	-1,2
Oberhausen	23,4	80,5	12,43	105,8	9,45	36,73	11,4	76,6	563,2	430,8	59,46	-2,9
Remscheid	37,8	375,8	2,66	391,7	2,55	64,22	-7,4	62,0	573,2	461,7	33,26	5,1
Solingen	29,9	487,3	2,05	514,8	1,94	57,71	-16,6	70,0	563,6	465,6	42,17	9,3
Wuppertal	23,8	383,9	2,60	511,9	1,95	39,65	-12,6	76,1	626,0	474,2	52,69	4,7
Bonn	6,7	4,1	246,22	27,7	36,14	6,70	-25,5	93,1	661,3	1537,6	22,38	2,4
Köln	13,6	229,1	4,36	86,6	11,55	28,96	-9,1	86,3	791,0	48,5	67,28	1,6
Leverkusen	27,5	k.A.	k.A.	k.A.	k.A.	72,93	-11,7	72,4	k.A.	0,0	19,71	5,3
Boitrop	25,9	67,5	14,82	54,4	18,37	15,89	-7,3	73,0	455,1	423,8	33,99	2,6
Gelsenkirchen	20,8	18,4	54,33	30,9	32,36	40,52	2,0	79,1	654,6	449,7	59,25	-0,5
Münster	10,2	805,6	1,24	806,5	1,24	28,90	-10,1	89,1	685,3	607,5	70,55	1,3
Bielefeld	18,3	425,0	2,35	504,1	1,98	54,73	-16,4	81,3	554,9	460,9	40,10	4,7
Bochum	18,3	135,3	7,39	128,5	7,78	38,83	-14,0	81,6	575,1	507,1	61,12	3,7
Dortmund	14,7	103,6	9,65	124,6	8,03	39,01	-7,5	85,2	569,7	479,3	60,92	1,5
Hagen	24,0	125,2	7,99	152,2	6,57	28,56	-4,9	75,8	599,9	476,4	62,83	1,8
Hamm	21,0	173,2	5,77	242,5	4,12	39,46	-8,6	78,5	592,7	450,9	50,92	2,6
Herne	27,2	28,6	35,01	23,9	41,88	42,87	-4,4	72,8	496,3	454,3	56,76	2,0
Darmstadt	17,9	522,3	1,91	680,8	1,47	58,87	-9,3	82,0	604,7	486,6	41,05	2,4
Frankfurt a.M.	10,8	256,9	3,89	344,7	2,90	24,31	-13,6	89,1	824,3	692,4	75,66	2,0
Offenbach a.M.	18,7	1175,9	0,85	1081,1	0,92	27,09	-8,0	81,3	588,3	522,3	57,62	2,2
Wiesbaden	13,6	257,4	3,88	266,5	3,75	45,47	-7,3	86,1	769,2	672,0	54,46	1,3
Kassel	15,9	924,9	1,08	1136,0	0,88	55,86	-6,9	83,9	563,2	467,9	35,23	1,5
Koblenz	12,1	249,8	4,00	303,8	3,29	40,38	-2,2	87,5	580,4	494,2	41,93	0,2
Trier	15,7	189,8	5,27	247,8	4,04	33,43	1,8	84,0	491,8	402,3	63,29	-0,1
Frankenthal i.d.P.	33,6	567,9	1,76	711,0	1,41	67,28	-1,9	65,5	597,0	419,3	25,39	1,8
Kaiserslautern	19,5	307,0	3,26	368,0	2,72	35,68	-10,4	80,5	507,5	415,7	57,61	2,9
Landau i.d.P.	15,3	316,2	3,16	338,5	2,95	34,24	9,8	83,7	533,8	461,1	58,06	-0,7
Ludwigsh. a.R.	42,1	23,0	43,50	34,3	29,19	83,09	-7,0	57,6	953,3	438,7	9,27	5,9
Mainz	10,1	149,3	6,70	193,9	5,16	27,45	-14,8	89,5	606,3	515,0	67,06	2,0

Neustadt a.d.W.	16,2	2930,6	0,34	2433,9	0,41	25,05	-4,0	81,3	471,2	429,1	68,11	0,3
Pirmasens	27,4	592,0	1,69	533,5	1,87	40,74	-0,5	72,2	505,4	452,4	52,29	0,2
Speyer	19,7	89,9	11,12	127,2	7,86	40,86	-6,8	80,3	537,9	412,0	52,85	1,8
Worms	25,8	100,5	9,95	139,6	7,17	51,04	-4,9	73,3	575,7	412,6	35,38	2,3
Zweibrücken	28,4	939,2	1,06	1231,5	0,81	65,75	7,2	71,1	533,3	387,3	29,71	-2,6
Stuttgart	18,4	1052,3	0,95	1584,6	0,63	71,05	-12,3	81,4	764,1	578,7	28,81	3,3
Heilbronn	20,1	539,4	1,85	708,0	1,41	40,24	-14,7	79,2	517,4	410,0	56,86	4,8
Baden-Baden	18,3	1837,4	0,54	1772,6	0,56	19,57	-12,6	81,2	616,7	541,4	76,58	3,4
Karlsruhe	16,3	45,1	22,17	63,4	15,78	24,46	-5,1	83,5	630,7	501,2	41,47	1,1
Heidelberg	12,9	1056,0	0,95	1235,9	0,81	50,72	-7,5	86,8	605,3	516,7	46,19	1,1
Mannheim	25,5	233,7	4,28	309,7	3,23	53,08	-5,5	74,4	662,4	495,4	46,81	1,9
Pforzheim	27,2	1086,3	0,92	1261,4	0,79	43,11	-10,1	72,6	593,2	471,6	37,25	4,4
Freiburg i.B.	11,6	285,2	3,51	453,0	2,21	40,30	-10,8	88,2	554,1	445,0	59,60	1,7
Ulm	20,5	753,0	1,33	1011,4	0,99	48,40	-9,7	79,3	608,7	474,6	51,53	2,7
Ingolstadt	41,8	922,7	1,08	1356,6	0,74	16,80	-8,4	57,9	1045,7	502,2	83,10	7,6
München	13,7	1335,4	0,75	2036,3	0,49	56,25	-19,4	86,2	785,2	623,6	43,70	4,0
Rosenheim	16,7	1038,8	0,96	1095,1	0,91	36,12	-26,1	83,1	518,5	446,4	60,45	7,7
Landshut	14,9	247,3	4,04	415,1	2,41	40,80	-23,1	84,9	547,4	413,6	55,73	5,3
Passau	20,5	911,4	1,10	1188,9	0,84	18,94	-5,6	79,5	613,9	483,5	77,50	2,1
Straubing	20,8	328,9	3,04	354,1	2,82	37,75	-4,8	78,7	524,6	444,8	58,02	1,4
Amberg	32,6	512,6	1,95	567,2	1,76	58,79	2,6	67,1	543,7	429,4	37,94	-1,2
Regensburg	25,2	639,0	1,56	1102,0	0,91	17,80	-3,6	74,6	708,1	434,2	45,02	1,7
Weiden i.d.O.	19,9	485,3	2,06	424,3	2,36	18,70	-1,1	79,5	475,7	427,5	80,75	0,6
Bamberg	27,4	528,3	1,89	640,3	1,56	42,01	-8,5	72,3	541,9	420,3	57,81	3,5
Bayreuth	18,3	540,4	1,85	685,7	1,46	14,65	0,1	81,4	553,3	448,8	19,06	-0,2
Coburg	25,2	1216,8	0,82	1292,3	0,77	73,63	-0,9	74,6	662,6	555,6	19,84	0,7
Hof	17,4	356,3	2,81	321,0	3,12	34,90	-12,9	82,0	508,9	455,9	58,61	2,9
Ansbach	20,2	616,9	1,62	775,9	1,29	31,53	-18,4	78,9	533,8	430,2	62,20	6,2
Erlangen	32,2	2070,8	0,48	2663,1	0,38	26,86	-14,1	67,6	662,3	468,7	58,23	8,6
Fürth	23,8	806,0	1,24	794,4	1,26	20,74	4,9	75,6	609,4	530,0	76,28	-1,7
Nürnberg	18,2	603,3	1,66	714,3	1,40	42,19	-13,6	81,5	603,2	501,1	57,64	3,7
Schwabach	31,4	1177,9	0,85	1105,4	0,90	37,87	-2,8	68,6	555,5	485,3	58,51	2,1
Aschaffenburg	22,2	218,3	4,58	247,3	4,04	29,36	-8,0	77,5	665,1	549,6	68,31	2,3
Schweinfurt	42,7	577,7	1,73	743,4	1,35	54,76	6,3	57,3	701,8	422,0	38,18	-4,5
Würzburg	11,3	940,5	1,06	1087,2	0,92	22,47	-14,0	88,2	520,4	448,5	66,11	2,2
Augsburg	21,6	315,5	3,17	393,7	2,54	45,46	-10,2	78,2	606,0	482,4	52,98	3,4
Kaufbeuren	20,5	395,6	2,53	355,3	2,81	38,86	-5,6	78,7	495,7	446,5	54,93	1,0
Kempten(Allgäu)	17,0	409,0	2,45	444,1	2,25	31,99	-13,1	82,4	560,5	481,2	37,52	3,6
Memmingen	32,2	698,8	1,43	784,7	1,27	56,03	-4,4	66,7	536,6	426,2	42,09	1,6
Berlin	12,5	725,7	1,38	906,0	1,10	33,31	-9,5	87,4	569,9	480,6	66,67	1,6
Brandenb. a.d.H.	22,0	k.A.	k.A.	k.A.	k.A.	23,79	-3,3	77,7	483,5	395,1	64,94	0,6
Cottbus	9,7	2028,7	0,49	2429,0	0,41	43,81	-15,1	89,9	428,8	369,9	55,97	1,9
Frankfurt a.d.O.	13,4	305,1	3,28	465,7	2,15	36,73	53,3	85,9	478,6	382,4	51,27	-5,2
Potsdam	6,3	5620,6	0,18	6805,9	0,15	40,80	-20,2	93,6	468,0	408,3	57,90	1,9
Rostock	12,7	452,6	2,21	524,0	1,91	41,70	-4,3	87,2	557,4	468,7	53,46	0,5
Schwerin	12,6	827,5	1,21	942,3	1,06	58,37	1,1	87,2	436,7	372,3	41,52	0,0
Chemnitz	20,0	641,8	1,56	744,5	1,34	24,37	-1,9	79,7	450,6	373,2	38,14	0,4
Dresden	17,2	392,3	2,55	480,1	2,08	40,92	-7,4	82,7	473,6	391,3	52,92	1,7
Leipzig	13,9	583,2	1,71	759,5	1,32	54,17	-6,9	86,0	470,2	389,5	45,65	1,2
Dessau-Roßlau	22,5	469,8	2,13	501,7	1,99	38,58	-0,2	76,6	456,6	387,0	49,46	0,1
Halle(Saale)	10,8	1442,8	0,69	1943,1	0,51	40,33	-4,0	89,1	441,4	370,5	59,58	0,4
Magdeburg	14,3	489,1	2,04	598,5	1,67	47,01	-5,3	85,5	472,5	396,4	51,94	0,9
Erfurt	14,5	484,4	2,06	593,8	1,68	42,88	0,3	85,1	415,7	348,3	45,47	-0,1
Gera	15,9	907,4	1,10	971,5	1,03	33,18	7,7	83,7	404,1	347,6	53,52	-1,6
Jena	19,9	457,8	2,18	664,2	1,51	38,54	-5,5	79,8	495,8	375,4	44,90	1,3
Suhl	17,7	827,5	1,21	756,0	1,32	31,55	-0,3	81,9	416,9	371,8	56,31	-0,5
Weimar	13,2	801,1	1,25	899,9	1,11	22,89	-9,4	86,5	424,1	364,5	73,42	1,6
Eisenach	29,4	367,1	2,72	365,5	2,74	45,86	-7,9	70,3	380,1	323,8	51,11	3,8

LEE Lehrstuhl Energiesysteme und Energiewirtschaft

Eingebunden in das Institut für Energietechnik gehört der Lehrstuhl zur Fakultät für Maschinenbau der Ruhr-Universität Bochum. Die Arbeitsgebiete lassen sich durch die Begriffe ressourcenschonende Energiewirtschaft sowie Anlagensimulation und -sicherheit charakterisieren. Der Lehrstuhl unterhält vielfältige internationale Kooperationen und gute Beziehungen zu einer Vielzahl in- und ausländischer Organisationen.

Ansprechpartner

Prof. Dr.-Ing. H.-J. Wagner

Ruhr-Universität Bochum

Lehrstuhl Energiesysteme und Energiewirtschaft

Geb. IC 2-181

Universitätsstr. 150

44780 Bochum

Tel.: +49 (0) 234 - 32 - 28044

Fax: +49 (0) 234 - 32 - 14158

Internet: <http://www.lee.ruhr-uni-bochum.de>

E-Mail: lee@lee.ruhr-uni-bochum.de