



Strukturen und Veränderungen anschaulich machen

Klassifizierung von Daten für Kartendarstellungen in der amtlichen Statistik

Von Kevin Bullenkamp

Karten sind seit vielen Jahren fester Bestandteil der Print- und der Online-Produkte des Statistischen Landesamtes. Mit der wachsenden Zahl an statistischen Indikatoren – insbesondere durch den Ausbau der Geowebdienste – zeigt sich zunehmend, dass aussagekräftige Karten nur über vielfältige Klassifizierungsmethoden zu erreichen sind. Im Vorfeld des anstehenden Relaunchs des Online-Kartenangebots wirft dieser Beitrag einen Blick auf Anforderungen an die Klassifizierung, stellt gängige automatisierte Klassifizierungsmethoden vor und präsentiert eigene Ansätze zur Klassifizierung.

Klassenanzahl und Klassengrößen

Klassenanzahl
variiert je nach
Verwaltungs-
ebene

Für Rheinland-Pfalz mit seiner kleinteiligen Kommunalstruktur ergibt sich die Notwendigkeit, drei Verwaltungsebenen mit aussagekräftigen Karten abzubilden. Die Kreisebene umfasst 36 kreisfreie Städte und Landkreise, die Verbandsgemeindeebene 170 kreisfreie Städte, verbandsfreie Gemeinden und Verbandsgemeinden und die Gemeindeebene 2 302 kreisfreie Städte, kreisangehörige Städte, verbandsfreie Gemeinden und Ortsgemeinden. Rein mathematisch gibt es den Ansatz, dass die Wurzel aus der Gesamtzahl der Elemente jeweils die Zahl der zu bildenden Klassen vorgibt. Das wären z. B. auf der Gemeindeebene 48 Klassen, die sich allerdings nicht mehr sinnvoll in Karten darstellen ließen. Um Aussagekraft und Verständlichkeit zu gewährleisten, werden in der Regel für die Landkreisebene sechs Klassen erzeugt,

für die Verbandsgemeindeebene acht und für die Gemeindeebene zwölf Klassen. Jede Klasse sollte ungefähr die gleiche Anzahl an Elementen enthalten, jedoch möglichst auch die gleiche Breite aufweisen. Es besteht die Möglichkeit die Intervallbreite der obersten und der untersten Klasse zu variieren, da diese beiden Klassen nach oben bzw. unten offen sind.

Datenstruktur und Farben

Das Statistische Landesamt Rheinland-Pfalz verwendet bei seinen Karten in der Regel zwei Grundfarben, die hohe und niedrige Niveaus bzw. positive und negative Veränderungen auf einen Blick unterscheidbar machen. Ausreißer sorgen bei der bisher eingesetzten automatischen Klassenbildung auf der Gemeinde-, teilweise bereits auf der Verbandsgemeindeebene dafür, dass das Farb-

Zweifarbigkeit
macht Unter-
schiede schnell
sichtbar



spektrum nicht zum Tragen kommt; manche Karten sind schlicht monochrom. Dem soll künftig mit dem Einsatz weiterer Verfahren zur Klassenbildung begegnet werden.

Übersicht über Klassifizierungsmethoden

Die Kartenerstellung mit gängigen Geoinformationssystemen enthält verschiedene Klassifizierungsmethoden, deren Anwendung auf unser Datenangebot geprüft wurde. Dazu wurden intervallgesteuerte, elementgesteuerte und andere mathematische Algorithmen verwendet, die im Folgenden näher erläutert werden.

Verzeichnis der Abkürzungen

I_x	= Intervall x
I_G	= Gleiches Intervall
I_D	= Definiertes Intervall
G_x	= Intervallgrenze
Min	= Minimum der Wertemenge
Max	= Maximum der Wertemenge
R	= Range bzw. Spannweite der Wertemenge
k	= Anzahl der Klassen
n	= Anzahl der Elemente in der Wertemenge
mod	= Modulfunktion, gibt den ganzzahligen Rest einer Division zweier ganzzahliger Werte aus
n_{K_x}	= Anzahl an Elementen in Klasse x

Intervallklassifizierung

Geeignet für bestehende Klassifikationen

Bei den intervallgesteuerten Methoden sind manuelle, gleiche und definierte Intervalle zu unterscheiden:

Manuelle Intervalle bieten die Möglichkeit, Intervalle nach Belieben manuell zu setzen. Bei x definierten Klassen ergeben sich folgende Intervalle:

$$I_1 = [Min; G_1]; I_2 = [G_1; G_2]; \dots; I_x = [G_x; Max]$$

Dieser Ansatz ist zwar für vordefinierte Klassifikationen notwendig, führt jedoch angesichts der großen Zahl von Indikatoren in der amtlichen Statistik zu einem unverhältnismäßig hohen manuellen Aufwand.

Bei der Verwendung von **gleichen Intervallen** ist die Klassenanzahl festzulegen. Die Spannweite oder Range der Wertemenge, die sich als Differenz des maximalen und minimalen Werts errechnen lässt, wird durch die festgelegte Klassenanzahl dividiert, wodurch sich die gleichbleibende Intervallgröße ergibt:

$$I_G = \frac{R}{k} = \frac{Max - Min}{k}$$

Gleiche Intervallbreite jeder Klasse, Anzahl der Elemente stark schwankend

Im Gegensatz dazu, spielt bei der **definierten Intervallmethode** die vorgegebene Größe der Intervalle eine Rolle. Die Intervallgrenzen werden bei dieser Methode durch eine vordefinierte Klassenbreite erstellt. Die Grenzen definieren sich anhand aller Vielfachen dieses Werts, die sich zwischen dem Minimum und dem Maximum der Menge befinden. Daraus ergeben sich $x + 1$ Klassen:

$$G_{y-1} < Min < G_y < G_{y+x} < Max < G_{y+x+1}$$

mit G_y = Kleinstes Vielfaches in der Menge und G_x = Größtes Vielfaches in der Menge

Intervallbreite bestimmt die Anzahl der Klassen

Bei dieser Methode ergibt sich keine vordefinierte Klassenanzahl, und sie ist deshalb im weiteren Verlauf nicht relevant.

Quantil-Methode

Anders als bei den Intervallberechnungen sind bei der Quantil-Methode nicht das Minimum und das Maximum der Wertemenge entscheidend, sondern die Gesamtzahl der enthaltenen Elemente. Sie wird durch die vorgegebene Klassenanzahl geteilt

Gleiche Anzahl an Elementen je Klasse, Klassenbreiten stark variierend



und ergibt somit die Anzahl der Elemente, die einer Klasse angehören sollen:

$$\frac{n}{k} = x \text{ (Anzahl der Elemente je Klasse)}$$

Daraus folgt, dass nach jedem x-ten Element eine Klassengrenze gezogen wird. Falls das Ergebnis dieser Berechnung nicht ganzzahlig ist, wird die Anzahl der überschüssigen Werte als Rest definiert:

$$\begin{aligned} n \bmod k \\ = r \text{ (Anzahl überschüssiger Elemente)} \end{aligned}$$

Den ersten r-Klassen wird daraufhin jeweils ein Element hinzugefügt.

Beispiel: Bei einer Gesamtzahl von 170 Elementen (Anzahl der Einheiten auf der Verbandsgemeindeebene) mit einer Vorgabe von acht Klassen ergibt sich ein Rest von zwei, was zu folgender Verteilung führt:

$$\begin{aligned} n_{K_1} &= 22; n_{K_2} = 22; n_{K_3} = 21; \\ n_{K_4} &= 21; n_{K_5} = 21; n_{K_6} = 21; \\ n_{K_7} &= 21; n_{K_8} = 21 \end{aligned}$$

Natürliche Unterbrechungen nach dem Jenks-Algorithmus

Clusterbildung von ähnlichen Elementen, Klassenbreite und Anzahl der Elemente unterschiedlich

Diese Klassifikation soll möglichst harmonische Klassen in der Wertemenge erzeugen. Das heißt, dass einzelne Cluster entstehen, deren Ausreißer innerhalb einer Klasse minimiert werden. Dieser iterative Algorithmus vergleicht benachbarte Klassen miteinander und verschiebt je nach Resultat das größte Element der unteren Klasse in die obere Klasse oder das kleinste Element der oberen Klasse in die untere Klasse. Dieser Prozess wird so lange wiederholt, bis der Algorithmus keine Verschiebung mehr als sinnvoll erachtet. Dadurch entstehen Cluster, die die Daten klassifizieren.

Ablauf des Jenks-Algorithmus¹

1. Die Anzahl der Klassen wird definiert
2. Die Klassen werden durch die Methode der gleichen Intervalle gebildet und dadurch die vorläufigen Klassengrenzen bestimmt.
3. In jeder Klasse wird der Mittelwert der Elemente gebildet.
4. Der Abstand des größten Elements der ersten Klasse zu deren Mittelwert wird mit dem Abstand zum Mittelwert der zweiten Klasse verglichen.
5. Sollte der Abstand zu dem Mittelwert der zweiten Klasse geringer sein als zum Mittelwert der ersten Klasse, wird das Element in die zweite Klasse verschoben und die Klassengrenze angepasst.
6. Falls dies nicht der Fall ist, werden Schritt 4 und 5 auf das kleinste Element der zweiten Klasse angewandt.
7. Die Überprüfung (Schritte 4 bis 6) findet zwischen allen benachbarten Klassen statt.
8. Sofern innerhalb der Schritte 3 bis 7 Elemente in eine andere Klasse verschoben wurden, wird der Prozess von Schritt 3 an wiederholt. Sollte dies nicht der Fall sein, ist die Clusterung abgeschlossen.

Geometrische Intervalle

Die Berechnung der geometrischen Intervalle ist ein Kompromiss zwischen den Klassifizierungsmethoden „Quantil“, „Gleiche Intervalle“ und „Natürliche Unterbrechungen“. Grundlegend für diese Klassifizierung

Klassenbildung anhand von geometrischen Reihen

¹ Jenks, G. F./Caspall, F. C.: „Error on Choroplethic Maps. Definition, Measurement, Reduction“. In: Annals of the Association of American Geographers, Vol. 61 (1971), S. 217–244.



ist die Bildung einer geometrischen Reihe, anhand derer die Klassenintervalle definiert werden. Der Algorithmus hat die Aufgabe, eine geometrische Verteilung zu finden, mit der die Elemente möglichst gleichmäßig auf die Klassen verteilt werden. Dies wird durch eine Minimierung der Summe der Quadrate der Elemente je Klasse erreicht.²

Neben den genannten Methoden zur Klassifizierung von Mengen gibt es noch zahlreiche weitere, die bei der Kartenproduktion normalerweise nicht verwendet werden.

Vergleich der Klassifizierungsmethoden – Aufbau der Daten

Um die Vor- und Nachteile bzw. die Aussagekraft der Klassifizierungsverfahren aufzuzeigen, wurden auf gleicher Datenbasis mit verschiedenen Verfahren Karten erstellt.

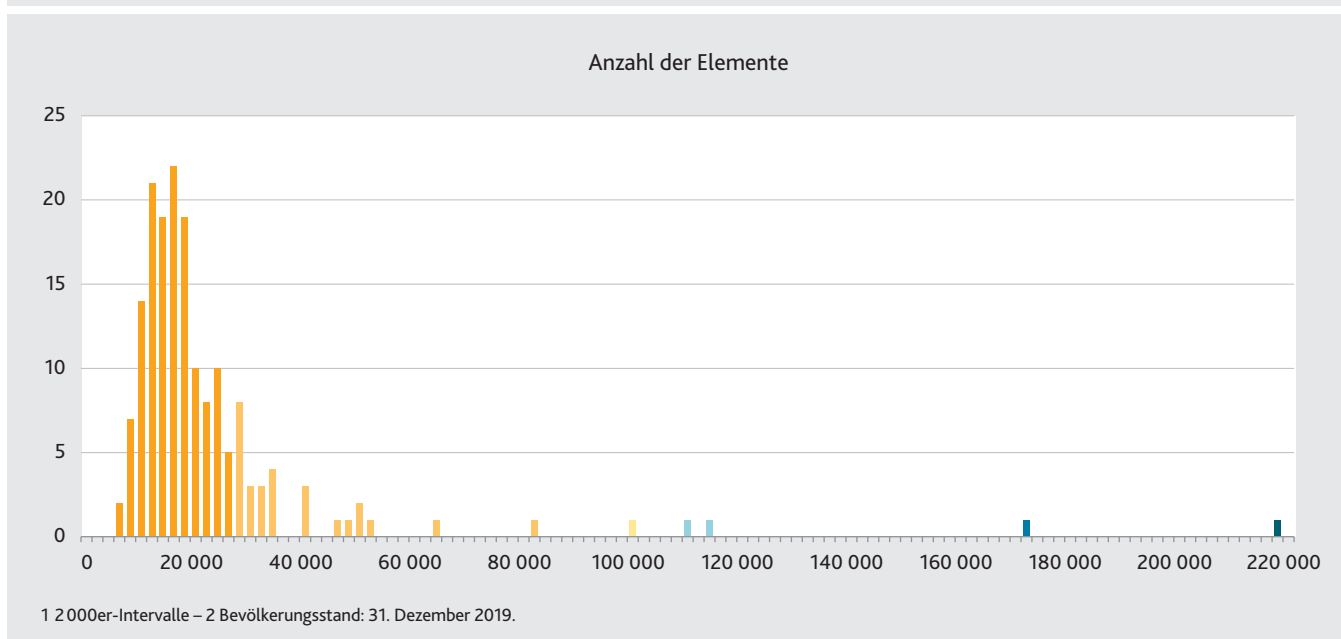
² ESRI: Datenklassifizierungsmethoden: Geometrische Intervalle; Quelle: <https://pro.arcgis.com/de/pro-app/help/mapping/layer-properties/data-classification-methods.htm> (abgerufen am 15.07.2020)

Grafik 1 zeigt die Bevölkerungszahl auf Verbandsgemeindeebene am 31. Dezember 2019. Grafik 2 stellt dar, wie sich die Bevölkerungszahl am 31. Dezember 2019 gegenüber dem 31. Dezember 2011 prozentual verändert hat. Beide Datengrundlagen sind unterschiedlich aufgebaut. Bei der Bevölkerungszahl handelt es sich um eine exponentielle Verteilung, während die Bevölkerungsveränderung annähernd normalverteilt ist. Diese beiden Datenstrukturen repräsentieren die Strukturen der meisten amtlichen Statistiken. Werte ballen sich oft um einen Mittelwert oder in einem anderen Bereich; eine stetige Gleichverteilung der Daten kommt eher selten vor.

Die Karten zeigen das Ergebnis der automatischen Klassifizierung unter Verwendung der Methoden „Gleiche Intervalle“, „Natürliche Grenzen“, „Quantil“ und „Geometrische Intervalle“.

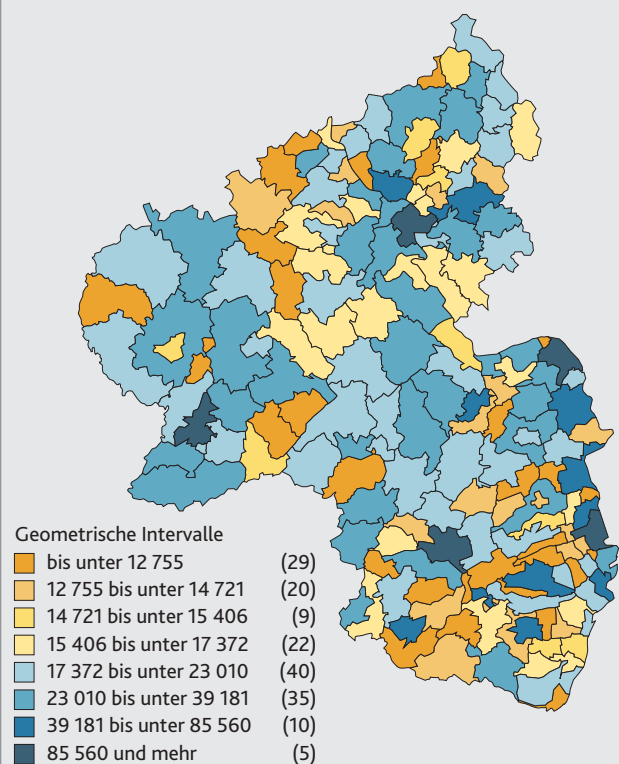
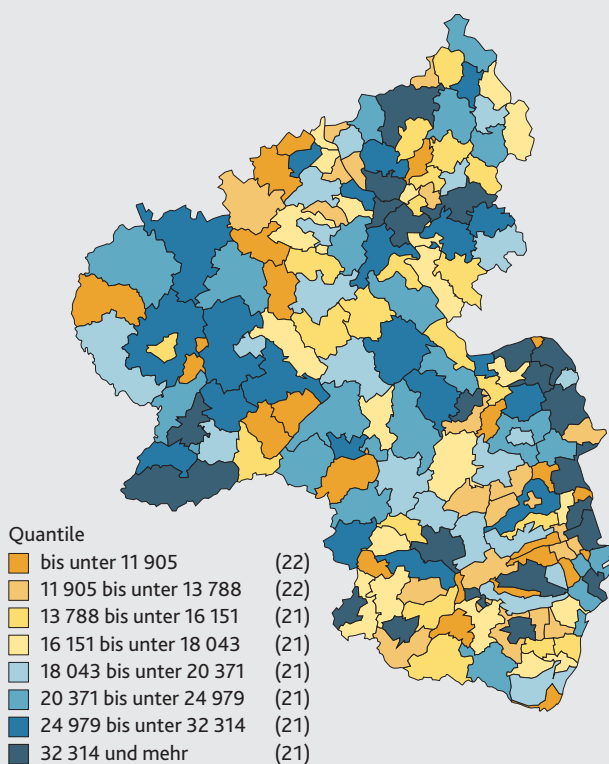
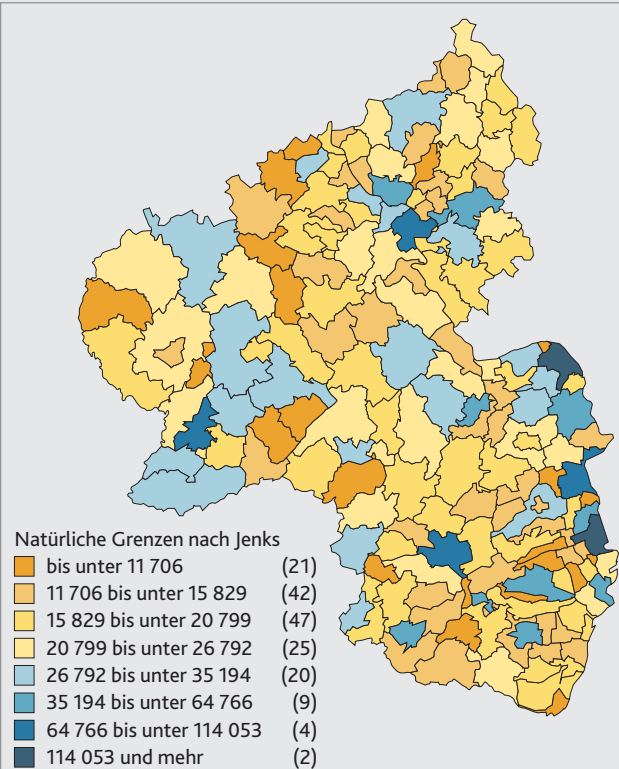
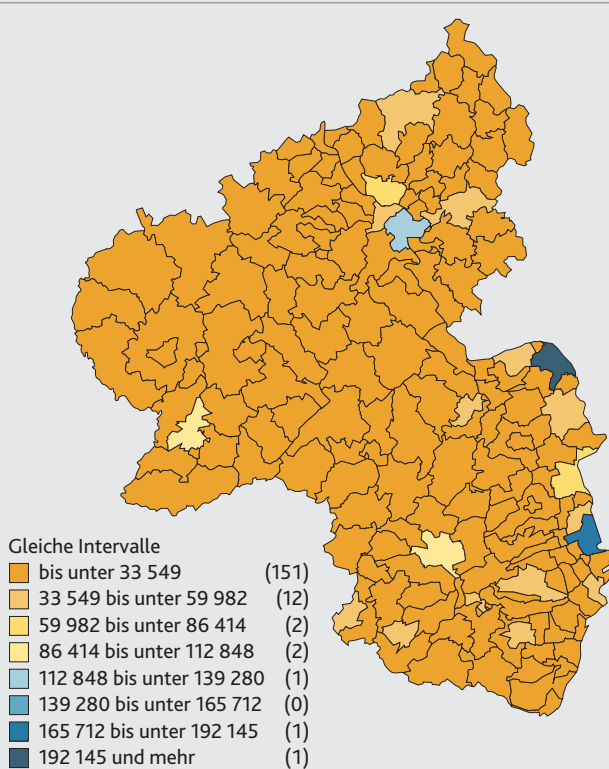
Daten meist exponentiell- oder normalverteilt

G1 Intervallaufteilung¹ der rheinland-pfälzischen Verbandsgemeindeebene nach der absoluten Bevölkerung²

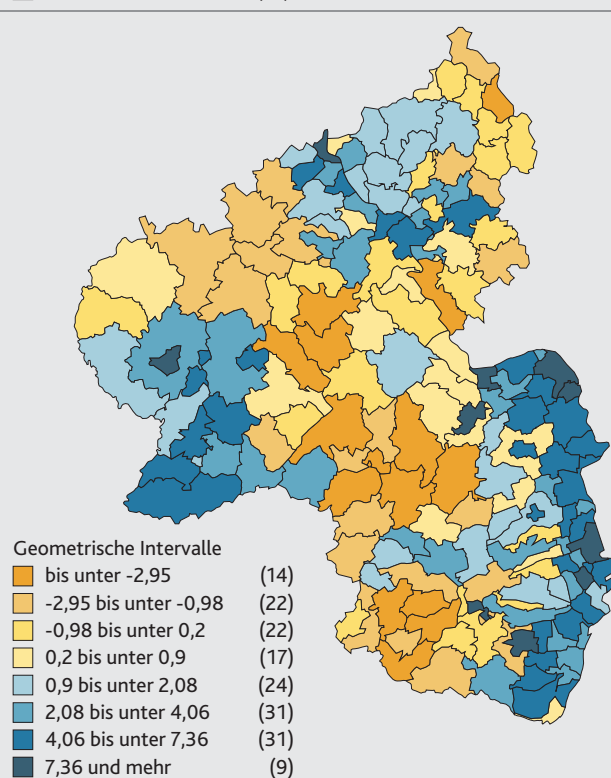
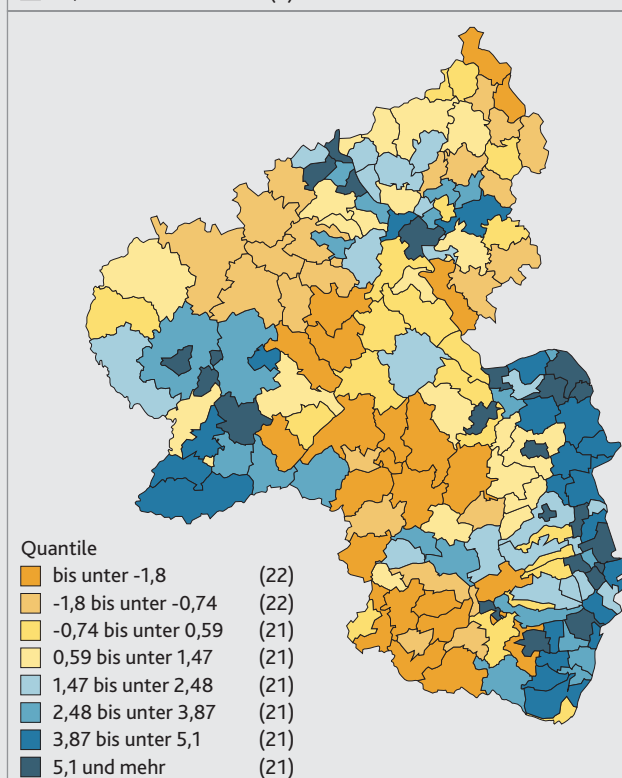
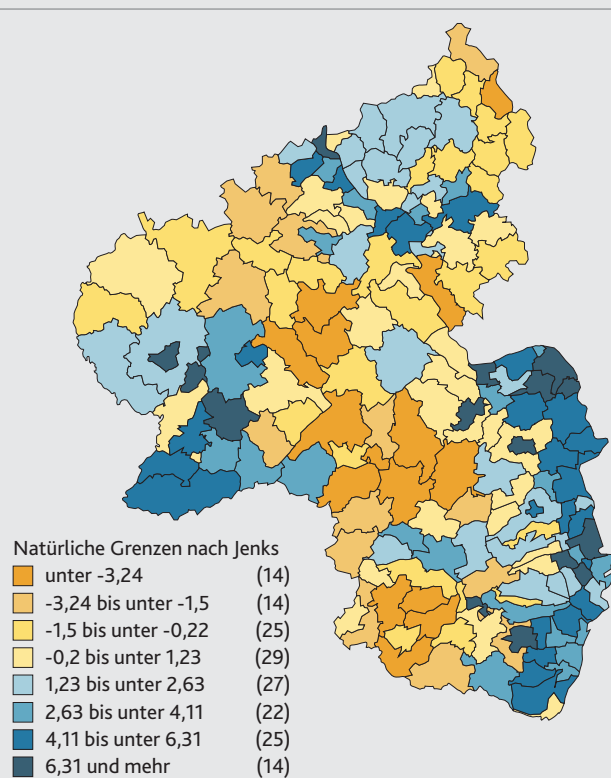
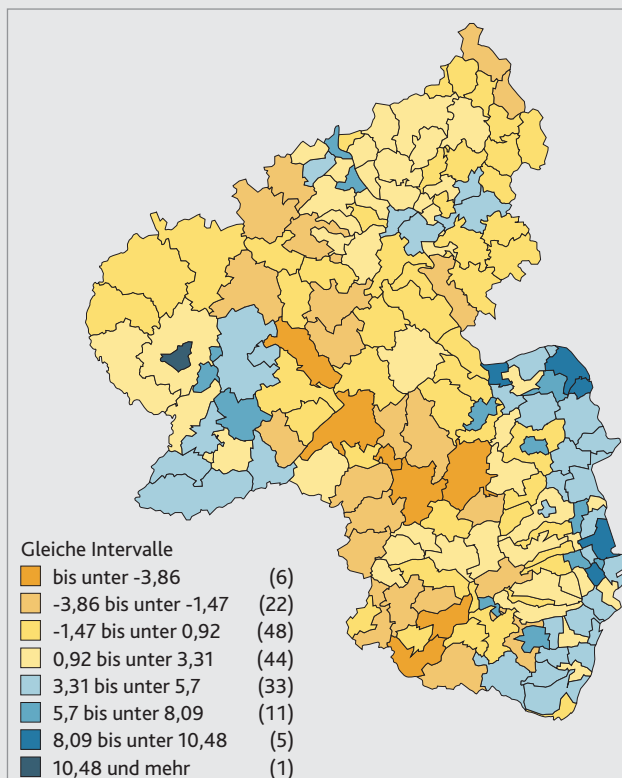




K1 Anwendung von Standardklassifikationen auf einen exponentiell verteilten Datensatz

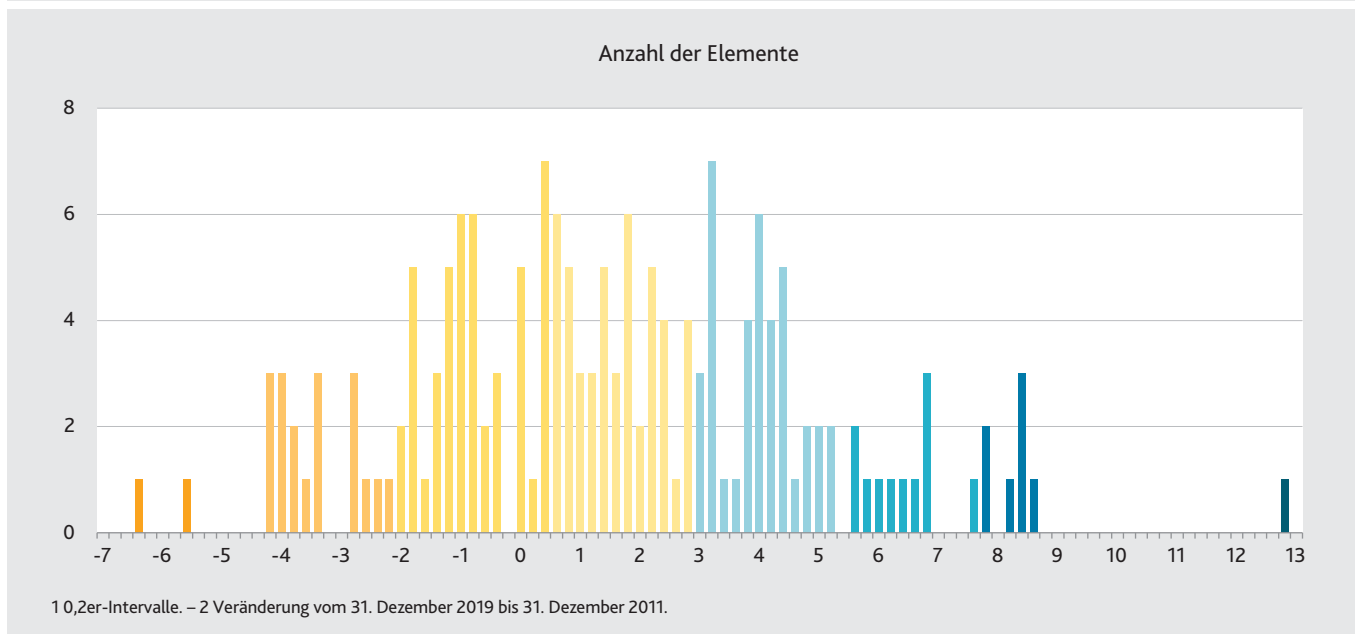


K2 Anwendung von Standardklassifikationen auf einen normalverteilten Datensatz





G2 Intervallaufteilung¹ der rheinland-pfälzischen Verbandsgemeindeebene für die prozentuale Veränderung der Bevölkerung²



Vor- und Nachteile gleicher Intervalle

Extremwerte
werden her-
vorgehoben

Die Karte 1 zur Bevölkerung der kreisfreien Städte, verbandsfreien und Verbandsgemeinden (Verbandsgemeindeebene) zeigt sehr gut, welche Auswirkung die Anwendung dieser Klassifizierungsmethode auf exponentiell verteilte Daten hat. Ausreißer (wie in diesem Fall die Stadt Mainz) können zu einer inhomogenen Klassifizierung und sogar zu leeren Klassen führen. Die unterste Klasse enthält circa 90 Prozent aller Elemente der Verbandsgemeindeebene. Die Karte stellt zwar sehr gut die bevölkerungsreichsten Kommunen in Rheinland-Pfalz dar; eine Differenzierung der bevölkerungsmäßig kleinen verbandsfreien und Verbandsgemeinden ist jedoch nicht möglich.

Die Verwendung gleicher Intervalle bei der Veränderung der Bevölkerungszahl (Karte 2, links oben) zeigt, dass die Elementanzahl in den mittleren Klassen wesentlich höher ist als in den äußeren. Die Verwendung von

gleichen Intervallen für die Klassifizierung rückt die Extremwerte in den Fokus.

Vor- und Nachteile der Methode „Natürliche Grenzen“

Der Algorithmus nach Jenks, der eine automatische Clusterung der Daten erzeugt, zeigt im gelb-blau-Verhältnis der Klassen kaum einen Unterschied zu den Ergebnissen der gleichen Intervallklassifizierung, aber innerhalb der gelben und der blauen Klassen gibt es eine größere Varianz. Gleiche Klassengrößen ergeben sich durch diese Methode jedoch nicht mehr. So ist in Karte 1 sogar die Spannweite der obersten Klasse nahezu genauso groß wie die Gesamtspannweite der übrigen Klassen. Die Normalverteilung der Elemente in der zweiten Karte ist auch bei der Clusterung klar erkennbar. Wie bei der ersten Methode können durch diese Methode größere Kommunen direkt identifiziert werden; die einzelnen Cluster beinhalten jedoch mehr Elemente.

Eine klare
Clusterung ist
erkennbar



Anders verhält es sich in Karte 2; hier wirkt die Elementanzahl der einzelnen Klassen homogener. Soll eine Karte mithilfe der Klassifizierung Informationen über prägnante Cluster liefern, ist diese Methode sicherlich keine schlechte Herangehensweise. Zu beachten ist jedoch, dass die Intervallgrößen sehr stark variieren und diese vom Betrachter nur bedingt nachvollziehbar sind. Problematisch ist auch der Vergleich von Karten zum gleichen Thema in verschiedenen Jahren, da die Klassifizierungen sich mit hoher Wahrscheinlichkeit stark voneinander unterscheiden werden und somit nicht vergleichbar sind.

Vor- und Nachteile der Quantil-Methode

Starke Varianz der Klassenbreite führt zu geringer Aussagekraft

Durch den elementgesteuerten Ansatz erzeugt diese Klassifizierung immer eine gleiche Anzahl von Elementen in den einzelnen Klassen. Das kann dazu führen, dass es in einer Karte zum einen sehr verdichtete Klassen gibt, zum anderen aber auch Klassen mit weit verstreuten Werten. Klare fachliche Aussagen lassen sich aus solchen Karten nicht ziehen. Allenfalls ist die Erkenntnis zu gewinnen, in welchem Bereich der Wertemenge sich ein Wert befindet. So bildet die obere Klasse in Karte 1 etwa 90 Prozent der gesamten Spannweite ab. Durch diese Klassifizierung wird ein mehr oder weniger harmonisches gleichverteiltes Produkt erzeugt, das jedoch wenig Aussagekraft besitzt.

Vor- und Nachteile der Methode „Geometrische Intervalle“

Dass diese Methode eine Kombination der drei vorangegangenen Methoden darstellt, zeigt sich durch die Verteilung der Objekte in den Karten. Eine klare Struktur ist dadurch nicht erkennbar und die Klassenbildung ist

nicht unbedingt nachvollziehbar. Wie bei dem Algorithmus zu den natürlichen Grenzen lassen sich verschiedene Cluster gut erkennen. Dabei handelt es sich jedoch hauptsächlich um die Ausreißer in der obersten Klasse.

Zusammenfassung

In geografischen Informationssystemen gibt es vorgefertigte Algorithmen, die Klassen bilden, die je nach Aufbau des Datensatzes eine bestimmte Aussagekraft haben. Die automatisierte Klassenbildung ist aus mehreren Gründen problematisch. So ist z. B. der Wert 0 bzw. der Mittelwert in den Berechnungen nahezu irrelevant, und Ausreißer führen zu größeren Verzerrungen in der Klassenbildung. Die exakte Klassenberechnung der Algorithmen bewirkt außerdem, dass keine „schönen“ gerundeten Zahlen die Grenzen bilden, sondern immer exakte Werte. Mit Ausnahme der Methode der gleichen Intervalle ist keine klare Intervallstruktur erkennbar.

Das Ziel des Statistischen Landesamtes, in den kartenbasierten Regionaldatenangeboten aussagekräftige und leicht nachvollziehbare Karten darzustellen, ist mit keiner der beschriebenen automatisierten Klassifizierungsmethoden erreichbar. Daher soll künftig ein auf verschiedene Datenarten abgestimmter Mix an Klassenbildungsverfahren eingesetzt werden.

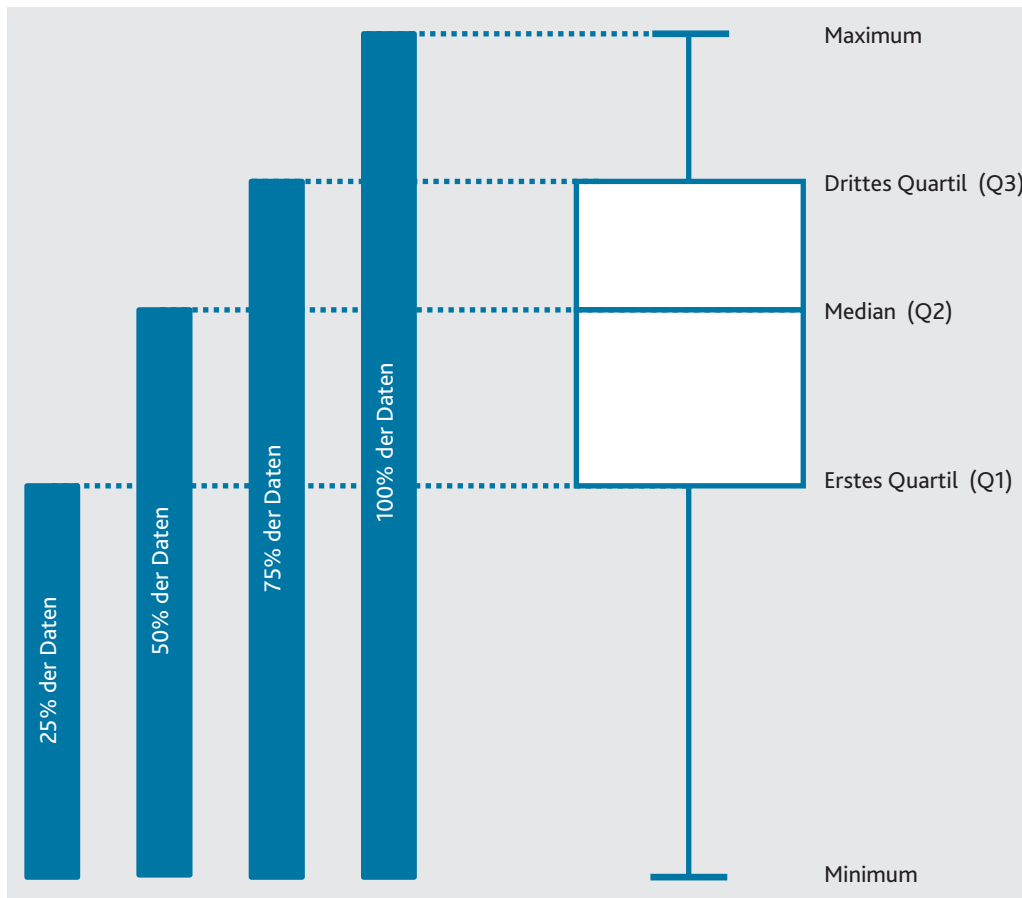
Zu differenzieren sind drei verschiedene Datenarten, die unterschiedliche Methoden erfordern. Für Daten, die schon klassifiziert sind, wie z. B. die Einwohnerinnen und Einwohner nach Größenklassen, muss eine manuelle Klassifizierung vorgenommen werden. Für Daten, die ausschließlich positive

Eine perfekte Klassifizierung gibt es nicht

Symmetrische und positive/negative Intervalle als Ansatz für die Kartenklassifizierung



G3 Definition von Quartilen



oder ausschließlich negative Werte umfassen (z. B. Veränderungen), wird anhand des Mittelwerts eine symmetrische Intervallklassifizierung angestrebt. Und schließlich soll bei Daten, die sowohl positive als auch negative Werte aufweisen, ein Algorithmus verwendet werden, der zu einer klaren Trennung der Datenmenge bei 0 führt. Da Ausreißer auch bei dieser Klassifizierung zu Verzerrungen führen können, muss der Algorithmus die Möglichkeit bieten, Ausreißer zu identifizieren und herauszufiltern. Gerundete Klassengrenzen sind zwar ungenauer als exakte, führen jedoch zu einem harmonischeren Bild. Somit muss ein Automatismus gefunden werden, der definiert, wann welche Rundung anzuwenden ist.

Behandlung von Ausreißern

Als Ausreißer werden in der Statistik z. B. Werte definiert, die um mehr als den 1,5-fachen Interquartilsabstand unter dem Wert des unteren Quartils (Q1) oder über dem Wert des oberen Quartils (Q3) liegen. Der Interquartilsabstand (IQA) ist die Differenz zwischen dem Wert von Q1 und Q3. Das mittlere Element ist der Median (Q2) der Wertemenge.

Der Wert, der nach 25 Prozent aller Elemente über bzw. unter dem Median liegt, definiert Q1 bzw. Q3. Daraus folgt, dass Q1 und Q3 in der geordneten Wertemenge x an folgenden Stellen liegen:

Ein Ausreißer wird durch den 1,5-fachen Interquartilsabstand definiert



Wertemenge: $x = \{x_1, x_2, \dots, x_n\}$
 Daraus folgt: $Q_1 = x_{\frac{1}{4} \cdot n}$; $Q_3 = x_{\frac{3}{4} \cdot n}$

Falls $\frac{1}{4} \cdot n$ und $\frac{3}{4} \cdot n$ nicht ganzzahlig sind,
 müssen diese aufgerundet werden

Die so definierten Ausreißer werden für die weitere Berechnung nicht verwendet, landen bei der Klassifizierung jedoch in der untersten bzw. obersten Klasse. Dies bedeutet, dass diese beiden Intervalle nach unten bzw. oben offen sind. Ausreißer können auch anhand anderer Methoden bestimmt werden, z. B. mittels der Standardabweichung. Die Festlegung auf den Interquartilsabstand erfolgt, weil dieser auch bei schiefen Verteilungen verlässliche Ergebnisse liefert.

Runden der Klassengrenzen

Die benötigte
Rundung wird
durch die
Spannweite
bestimmt

Anhand der Wertemenge soll definiert werden, ob eine Rundung der Klassengrenzen bzw. der Klassenbreiten stattfinden soll oder nicht. Für diese Berechnung ist die Spannweite zwischen Minimum und Maximum der Menge relevant. Der Wert a als Ergebnis der Division von Spannweite und 2,5 ist ausschlaggebend dafür, welche Rundung stattfinden soll:

$0 \leq a < 10 \Rightarrow 1 \text{ Dezimalstelle}$
 $10 \leq a < 100 \Rightarrow \text{Ganzzahl}$
 $100 \leq a < 1\,000 \Rightarrow 10\text{er} - \text{Rundung}$
 $1000 \leq a < 10\,000 \Rightarrow 100\text{er} - \text{Rundung}$

Symmetrische Intervallklassifikation

Mittelwert und
Extremwerte
sind für die
Bildung der
Klassen ver-
antwortlich

Bei dieser Klassifizierungsmethode ist der Mittelwert die Klassengrenze und die Grenze zwischen den beiden verwendeten Farben. Es wird zunächst überprüft, ob der Mittelwert näher am Minimum oder näher am Maximum liegt. Der kleinere Abstand wird zur Weiterberechnung verwendet,

indem er durch die Hälfte der geforderten Anzahl an Klassen geteilt wird. Dieser Wert gibt die Klassenbreite an, die bei Bedarf je nach Größe noch auf- bzw. abgerundet wird. Mittels der Intervallbreite werden nun die Klassen von der Farbgrenze aus aufgespannt. Dadurch entstehen, ausgehend vom Mittelwert, auf beiden Seiten gleichviele Klassen, wobei die beiden letzten Klassen nach außen geöffnet sein können.

Beispiel:

$Min = 17; Max = 113;$
 $Mittelwert = 41; 6 \text{ Klassen}$
 Daraus folgt:
 $R_{Min} = 41 - 17 = 24 < 72 = 113 - 41$
 $= R_{Max}$

Klassenbreite:

$$b = \frac{R_{Min}}{6} \cdot 2 = \frac{24}{6} \cdot 2 = 8$$

Dadurch ergeben sich folgende Intervalle

$I_1 = [17; 25]; I_2 = [25; 33];$
 $I_3 = [33; 41]; I_4 = [41; 49];$
 $I_5 = [49; 57]; I_6 = [57; 103]$

Diese Methode hat den Vorteil, dass einseitige Ausreißer für die Klassenberechnung keine Rolle spielen.

Karte 3 zeigt durch symmetrische Klassifizierung entstandene Klassen für die Gesamtbevölkerung. Hierbei wurden sowohl der Mittelwert als auch die Klassengrenzen auf Tausenderstellen gerundet.

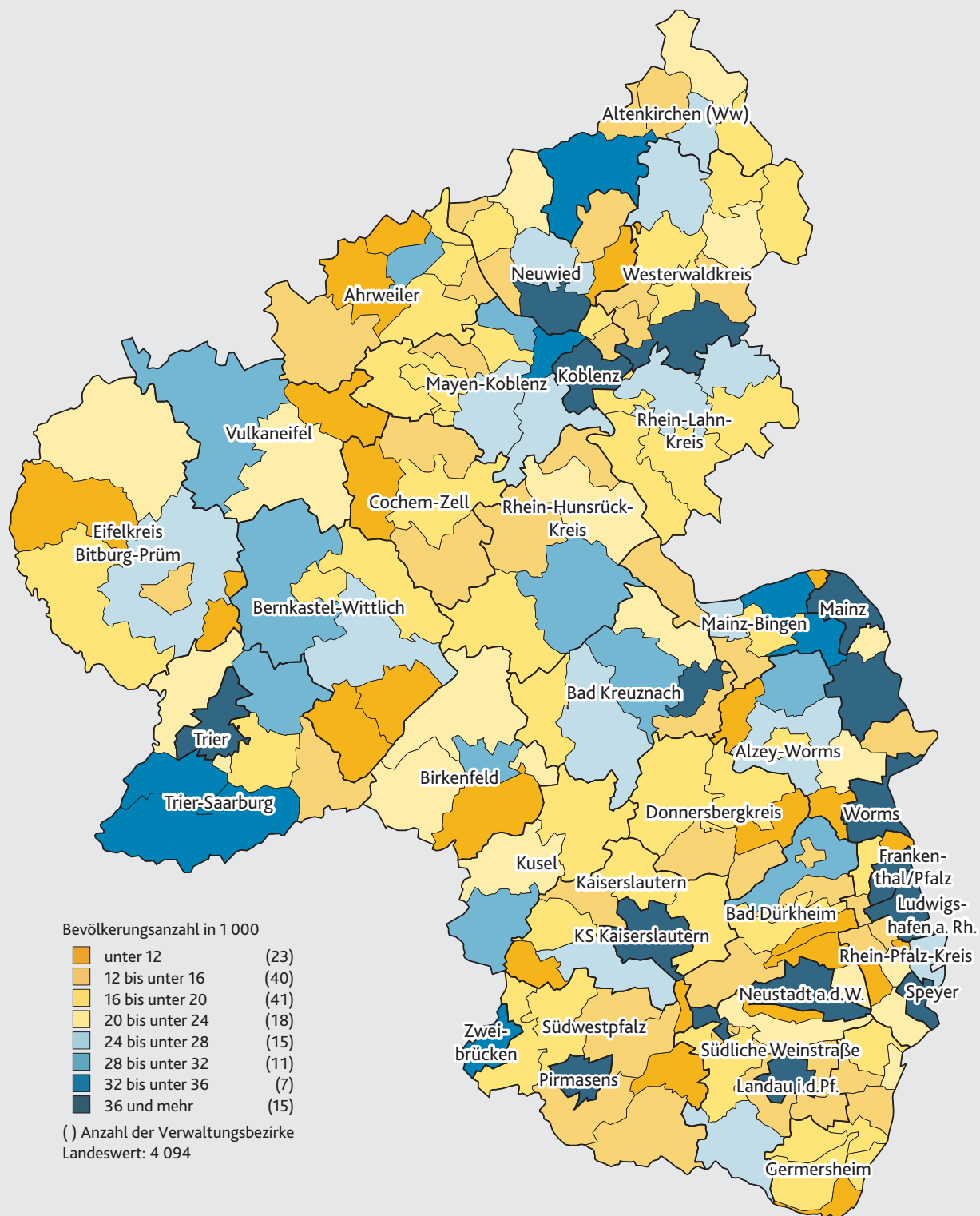
Positive und negative Intervalle

Bei dieser Methode geht es darum, klar zu definieren, welche Werte im positiven und welche im negativen Bereich liegen. Im Gegensatz zur symmetrischen Methode wird nicht der Mittelwert, sondern der Wert 0 als Grenze gewählt. Da dann die Elemente in den meisten Fällen nicht gleich-

Klassifikation
führt zu klarer
Unterschei-
dung von
positiven und
negativen
Werten

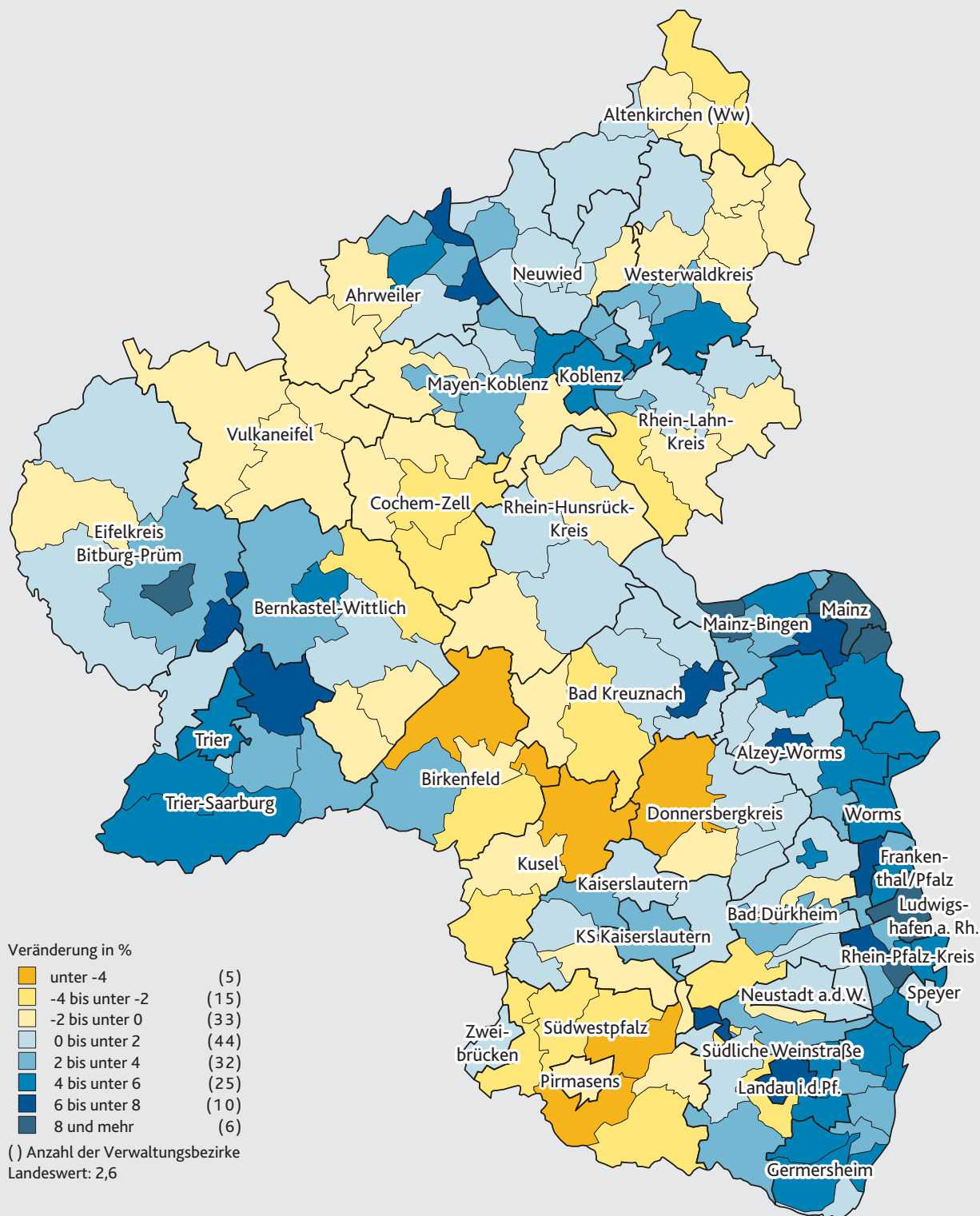


K3 Beispiel einer symmetrischen Klassifikation: Bevölkerung der Verbandsgemeinden in Rheinland-Pfalz am 31. Dezember 2019





K4 Beispiel einer +/- Klassifikation: Veränderung der Bevölkerung in den Verbandsgemeinden von Rheinland-Pfalz vom 31. Dezember 2011 bis 31. Dezember 2019





verteilt sind, wird die Anzahl der Klassen entsprechend dem Verhältnis der Elemente zugeordnet. Die Anzahl der Klassen im negativen Bereich wird durch Multiplikation der gesamten Klassenanzahl mit dem Quotienten aus den im negativen Bereich liegenden Elementen und der Gesamtzahl der Elemente berechnet:

$$k_{\text{negativ}} = k * \frac{n_{\text{negativ}}}{n}$$

Sollte diese Zahl kleiner als die Hälfte der gesamten Klassen sein, wird sie aufgerundet, ansonsten abgerundet. Die übrigen Klassen fallen in den positiven Bereich. Die Berechnung der Klassenbreite findet nun ähnlich zur symmetrischen Methode statt. Bei den Abständen des Wertes 0 zum Minimum und zum Maximum muss jedoch auf das Verhältnis geachtet werden. Ausreißer werden vor der Berechnung der Klassenbreite herausgefiltert.

In Karte 4 wurde diese Methode angewendet und die Klassengrenzen als ganzzahlig definiert. Die Karte differenziert deutlich

zwischen negativen und positiven Entwicklungen der Bevölkerung der kreisfreien Städte, verbandsfreien Gemeinden und Verbandsgemeinden.

Fazit

Die vorgestellten Klassifizierungsmethoden stellen einen Ansatz dar, um statistische Daten möglichst harmonisch zu klassifizieren. Die Methoden sollen die Kartenproduktion vereinfachen und beim anstehenden Relaunch des Online-Kartenangebots umgesetzt werden. Einen perfekten Automatismus für die Vielfalt an statistischen Indikatoren zu finden, ist unmöglich. Es bleibt aber das Ziel, in Zukunft weitere Automatismen zu finden bzw. die bestehenden Automatismen weiterzuentwickeln.

Kevin Bullenkamp, Angewandter Geoinformatiker (Master of Science), ist Mitarbeiter im Referat „Presse, Auskunftsdienst, Vertrieb“