

Unsichtbare Gruppen – sichtbare Effekte: Eine Untersuchung potenzieller Auswirkungen von Untererfassungen in deutschen Melderegisterdaten und Meldepflichtsbefreiungen auf gängige visuelle Darstellungen von Bevölkerungsgrößen in SDSS

Julia Mayer

(Julia Mayer, Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Trippstadter Straße 122, 67663 Kaiserslautern, julia.mayer@dfki.de)

1 ABSTRACT

Datenbasierte räumliche Entscheidungsunterstützungssysteme (Spatial Decision Support Systems, kurz: SDSS) besitzen ein hohes Potenzial kommunale Planungs- und Entscheidungsprozesse effizienter zu gestalten. Ihr Einsatz ermöglicht eine umfängliche Erfassung, automatisierte Analysen und objektive Bewertungen von Kriterien – allerdings nur, wenn die zugrunde liegenden Daten von hoher Qualität und frei von Verzerrungen sind, und der Algorithmus an allen kritischen Stellen sorgfältig konzipiert wurde (Sugumaran und DeGroot, 2010).

Aggregierte kommunale Melderegisterdaten sind hierbei eine häufig genutzte und zuverlässige Quelle mit überwiegend standardisierten Attributen, die auf verschiedenen räumlichen Granularitäten Informationen über die lokale Bevölkerung bereithalten (Mayer und Memmel, 2026a). Diese Daten werden von den kommunalen Stellen sorgfältig gepflegt und verarbeitet, sind jedoch aus praktischen und systembedingten Gründen nicht völlig frei von Verzerrungen. Eine solche Verzerrung ist der sogenannte Untererfassungsbias, der bedeutet, dass nicht alle Einwohnerinnen und Einwohner einer Kommune im Melderegister erfasst sind. Dies betrifft häufig marginalisierte Gruppen wie wohnungslose Menschen oder Personen ohne legalen Aufenthaltsstatus (BMAS, 2022 und Fisher, 1999), aber auch NATO-Angehörige (BfJ, 2020).

Diese Arbeit untersucht am Beispiel der Stadt Kaiserslautern, ob und wie der Untererfassungsbias quantifiziert werden kann. Unter Anwendung einer softwarebasierten Exploration werden die potenziellen Auswirkungen weiterer Verarbeitungslogiken oder Konzepte auf die Genauigkeit und Verlässlichkeit von SDSS-Visualisierungen betrachtet. Zudem wird analysiert, wie Datenaggregation auf verschiedenen räumlichen Ebenen diese Effekte verstärken oder abschwächen. Häufig genutzte Darstellungsformen wie Balkendiagramme und Choroplethenkarten werden herangezogen, um diese Dynamiken zu illustrieren.

Keywords: Bias, Spatial Decision Support System, Bevölkerungsdaten, Visualisierungen, Untererfassungen

2 HINTERGRUND

Im Zuge der digitalen Transformation finden auch in Behörden immer mehr datenbasierte SDSS Anwendung. Ziel dieser Systeme ist es, kommunale Akteurinnen und Akteure bei einer effizienten, fundierten sowie möglichst objektiven und transparenten Entscheidungsfindung zu unterstützen. Insbesondere in partizipativen und adaptiven Planungsprozessen erweisen sie sich als hilfreiches Werkzeug: Im Kontext der Raumplanung aufgefasst als heuristischer und wissensgenerierender Prozess können sie den Austausch zwischen den Beteiligten fördern, Handlungsoptionen systematisch zu prüfen und einen zweiseitigen Lernprozess ermöglichen (Nollert, 2013), auch wenn ihre Ausgaben teilweise mit Unsicherheiten behaftet sind.

2.1 Spatial Decision Support Systems

Räumliche Entscheidungsunterstützungssysteme umfassen technische Systeme, die auf der Verarbeitung georeferenzierter Daten basieren und mithilfe unterschiedlicher Technologien entscheidungsrelevante Ergebnisse generieren. Der resultierende Output reicht von der strukturierten Aufbereitung und Visualisierung heterogener Informationen über Analyse- und Prognosewerkzeuge bis hin zu hochkomplexen, interdependenten Berechnungen und Szenarioanalysen (Sugumaran und DeGroot, 2010).

Diese kommen potenziell in nahezu allen kommunalen Handlungsfeldern zum Einsatz, etwa bei der Verkehrssteuerung, der Standort- und Flächenplanung oder beim Monitoring von Klimaanpassungsmaßnahmen. Abhängig von Anwendungsdomäne und Zielsetzung greifen diese SDSS auf sehr unterschiedliche Datenquellen zurück, die von demographischen Informationen (z.B. Altersverteilung, über Fahrplandaten des ÖPNV bis hin zu Echtzeit-Sensordaten reichen.

Ein weiterer zentraler Aspekt bei der Betrachtung von SDSS sind die eingesetzten Technologien. Diese umfassen unter anderem Verfahren zur Datenaufbereitung und- integration, Visualisierungstechniken, mathematische und statistische Modelle, Optimierungsalgorithmen sowie Methoden des maschinellen Lernens und des DeepLearnings.

2.1.1 Visuelle Darstellungen

Eine häufig genutzte und zugleich intuitiv verständliche Form der Visualisierung raumbezogener Daten stellen Kartendarstellungen dar. Insbesondere Choroplethenkarten finden Anwendung, wenn quantitative Daten auf definierte räumliche Flächeneinheiten aggregiert vorliegen. Choroplethenkarte visualisieren Merkmalsausprägungen, indem begrenzte Gebiete entsprechend der Merkmalsausprägung farblich (oder durch Muster) differenziert dargestellt werden (Dent et al. 2009).

Im vorliegenden Beitrag liegt der Fokus auf dieser Darstellungsform in Kombination mit einem ergänzenden Balkendiagramm (vgl. Abb. 1), das eine vergleichende Einordnung der räumlichen Ausprägungen ermöglicht. Neben der Wahl geeigneter Klassifikationsverfahren kommt der Auswahl der Farbkodierung eine besondere Bedeutung zu, da sie maßgeblich die Wahrnehmbarkeit, Interpretierbarkeit und potenzielle Verzerrung der dargestellten Informationen beeinflusst.

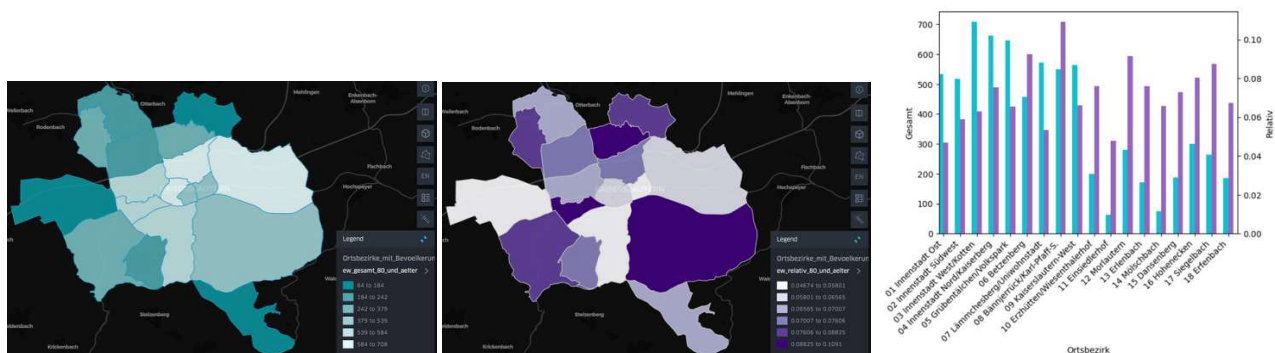


Abb. 1: Gängige Visualisierungen von Bevölkerungszahlen. Hier die über 80-Jährigen in Kaiserslautern auf Ortsbezirksebene im Jahr 2022: Gesamtanzahl (türkis) und relativ zur Gesamtbevölkerung des jeweiligen Ortsbezirks (lila) als Choroplethenkarten und Barchart mit zwei Y-Achsen. (Eigene Darstellung basierend auf von der Stadt Kaiserslautern bereitgestellten Daten.)

Unabhängig von der konkreten Visualisierungsform lässt sich festhalten, dass die Aussagekraft und Qualität der erzeugten Outputs wesentlich von der Verfügbarkeit, Nutzbarkeit und Güte der zugrunde liegenden Daten abhängt (Mayer und Memmel, 2026a).

2.1.2 Datenbeschaffung und -integration

Datenbasierte räumliche Entscheidungsunterstützungssysteme (SDSS) greifen in Abhängigkeit vom Anwendungskontext auf eine Vielzahl unterschiedlicher Datenquellen zurück. In kommunalen Anwendungsszenarien stammen diese Daten sowohl aus internen als auch aus externen Quellen, die sich hinsichtlich Format, Aktualität und räumlicher Granularität erheblich unterscheiden. Ein zentraler Zugangsweg ist die automatisierte Bereitstellung über standardisierte Programmierschnittstellen (Application Programming Interfaces, APIs), die eine regelmäßige oder auch echtzeitnahe Aktualisierung von Daten, etwa aus Sensor- oder Fachsystemen, ermöglichen. Ergänzend erfolgt die Datenbereitstellung häufig projekt- oder anlassbezogen durch kommunale Fachämter oder externe Institutionen. Die Integration dieser heterogenen Datenquellen erfordert insbesondere die Harmonisierung räumlicher und zeitlicher Bezugssysteme, die Vereinheitlichung von Attribut- und Bedeutungsstrukturen sowie die Berücksichtigung von Metadaten und rechtlicher Rahmenbedingungen. (Mayer und Memmel 2026a).

2.1.3 Datenqualität und Transparenz

Die Qualität der Ergebnisse von SDSS hängt maßgeblich von der Qualität der zugrunde liegenden Daten ab. Die Beurteilung der Datenqualität setzt ein hohes Maß an Transparenz hinsichtlich der Datenerhebung, der Vorverarbeitung sowie der angewandten Transformations- und Aggregationsschritte voraus. Informationen über Erhebungsmethoden, Erhebungszeitpunkte, Vorverarbeitungsschritte sowie über potenzielle Fehlerquellen sind notwendig, um die Aussagekraft und Belastbarkeit der Daten einschätzen zu können. Fehlen derartige Angaben oder sind sie nur unzureichend dokumentiert, stellt dies selbst eine Form von

Unsicherheit dar, die das Risiko von Fehlinterpretationen erhöht. Transparenz über Datenherkunft und -verarbeitung ist daher eine zentrale Voraussetzung für die Nachvollziehbarkeit, Vergleichbarkeit und verantwortungsvolle Nutzung datenbasierter Analysen und Entscheidungsunterstützungssysteme (Mayer und Memmel 2026b).

Nicht in allen Fällen ist es Entwicklerinnen und Entwicklern von SDSS möglich, belastbare Aussagen zur Datenqualität zu treffen, insbesondere dann, wenn Informationen zur Datenerhebung und -vorverarbeitung nicht oder nur unzureichend vorliegen. Auf diesen Fall fokussiert sich der vorliegende Beitrag, da auch das Wissen über und mögliche Auswirkungen durch fehlende oder unvollständige Informationen selbst eine relevante Form von Information darstellen. Diese sollten den Nutzenden eines SDSS transparent kommuniziert werden, um einen angemessenen Umgang mit den Outputs zu ermöglichen (Montello et al., 1998). Das Visualisieren von Unsicherheiten setzt jedoch voraus, dass diese selbst messbar oder in irgendeiner anderen Form systematisch bewertet werden können (MacEachren et al., 2005). In vorliegenden Fall wird jedoch auf Grundlage einer versucht sehr geringen Informationsbasis auf Attributebene versucht, potenzielle alternative Realisierungen zu erzeugen. Damit soll herausgestellt werden, dass die dargestellte Realität lediglich eine mögliche Ausprägung darstellt, da die Datengrundlage mit Unsicherheiten behaftet ist.

2.2 Bevölkerungsdaten

Bevölkerungsdaten stellen einen zentralen Indikator in vielen SDSS dar und werden auf unterschiedlichen räumlichen Aggregationsebenen genutzt. Häufig erfolgt ihre Verwendung in Kombination mit weiteren demografischen Merkmalen, wie etwa Altersstruktur, Haushaltszusammensetzung oder Nationalität. Diese Daten liefern wesentliche Informationen über die Zusammensetzung der Bevölkerung und ermöglichen es, innerhalb einer Kommune räumliche Muster zu identifizieren sowie Standorte mit spezifischen demografischen Charakteristika miteinander zu vergleichen.

Die Analyse der Altersverteilung bildet beispielsweise eine wichtige Grundlage für Prognosen des Bedarfs an Kindertagesstätten oder schulischer Infrastruktur. Ebenso ist demografisches Wissen über ältere Bevölkerungsgruppen für Anwendungen im Katastrophenschutz von Bedeutung, etwa zur Identifikation potenziell besonders vulnerabler Personengruppen. Darüber hinaus finden Bevölkerungsdaten auch in der Verkehrs- und Mobilitätsplanung Anwendung, beispielsweise zur Abschätzung von Nachfragepotenzialen oder zur Priorisierung infrastruktureller Maßnahmen.

2.2.1 Kleinräumige Bevölkerungsdaten

Bevölkerungsdaten auf gesamtkommunaler Ebene werden in Deutschland im Rahmen der amtlichen Statistik durch die statistischen Landesämter bereitgestellt und sind über entsprechende Veröffentlichungen und Datenportale zugänglich. Diese Daten basieren auf bundes- und landesrechtlich geregelten Erhebungen und dienen insbesondere der Vergleichbarkeit zwischen Kommunen sowie der Abbildung übergeordneter demografischer Entwicklungen.

Sollen Bevölkerungsdaten in kleinräumigerer Auflösung innerhalb einer Kommune genutzt werden – etwa auf Ebene von Ortsbezirken, statistischen Bezirken oder bis hin zur Baublockebene –, ist in der Regel eine direkte Anfrage bei den jeweiligen Kommunen erforderlich. Die Bereitstellung solcher kleinräumigen Daten fällt in den Aufgabenbereich kommunaler Statistikstellen. Diese haben den Auftrag, zur Unterstützung kommunaler Planungsprozesse und politischer Entscheidungsfindung räumlich differenzierte statistische Informationen aufzubereiten und zu analysieren. Um Mehrfacherhebungen zu vermeiden, greifen kommunale Statistikstellen dabei überwiegend auf Daten zurück, die bereits auf Grundlage bundesstatistischer oder landesgesetzlicher Regelungen erhoben wurden (Leischner und Khalil 2026).

2.2.2 Melderegisterdaten

Bei den kommunalen Meldebehörden liegen bereits detaillierte personenbezogene Daten vor, die sensible Informationen wie Geschlecht, Alter, Religionszugehörigkeit und Meldeadresse enthalten. Eine unmittelbare Nutzung dieser Daten in SDSS – selbst in anonymisierter Form – ist aus rechtlichen und ethischen Gründen nicht zulässig. Stattdessen müssen geeignete Verfahren zur datenschutzkonformen Aufbereitung eingesetzt werden, um den Schutz personenbezogener Informationen sicherzustellen und gleichzeitig eine sinnvolle Nutzung innerhalb eines SDSS zu ermöglichen.

Unsichtbare Gruppen – sichtbare Effekte: Eine Untersuchung potenzieller Auswirkungen von Untererfassungen in deutschen Melderegisterdaten und Meldepflichtbefreiungen auf gängige visuelle Darstellungen von Bevölkerungsgrößen in SDSS

Das Melderegister selbst ist jedoch nicht fehlerfrei. Es können Übererfassungen auftreten, etwa durch nicht aktualisierte Melderegistereinträge („Karteileichen“) nach Wegzug aus der Kommune, ebenso wie Untererfassungen, etwa durch nicht gemeldete Personen, die aber dauerhaft dort leben. Darüber hinaus bestehen gesetzliche Besonderheiten, die bestimmte Bevölkerungsgruppen von der Meldepflicht befreien, wie beispielsweise Angehörige der NATO-Streitkräfte oder Bewohnerinnen und Bewohnern von Seniorinnen- und Seniorenheimen. Andere Gruppen, etwa Personen ohne legalen Aufenthaltsstatus oder ohne festen Wohnsitz, melden sich meist nicht. Gleichwohl können sich auch diese Personen freiwillig melden; Wohnungslose mit dem Vermerk „OFW“ (ohne festen Wohnsitz).

Für die Erstellung vergleichbarer statistischer Auswertungen treffen die Statistikstellen bei der Datenaufbereitung bestimmte Annahmen. Personen mit Nebenwohnsitz innerhalb der Kommune, deren Hauptwohnsitz jedoch außerhalb liegt, sowie Personen, die nicht meldepflichtig sind, werden üblicherweise nicht berücksichtigt. Auf dieser Grundlage wird die sogenannte statistische Bevölkerung gebildet (vgl. Abb. 2). Sie dient der besseren Vergleichbarkeit zwischen Kommunen und verhindert Doppelzählungen, indem insbesondere Personen mit Nebenwohnsitzen ausgeschlossen werden (Deutscher Städtetag, 2019).

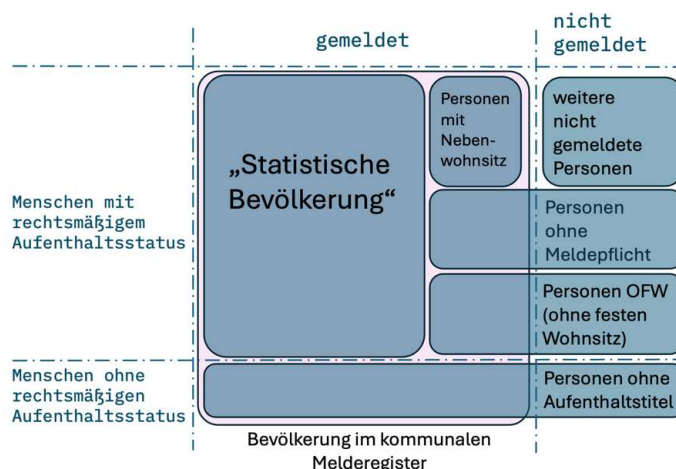


Abb. 2: Die "Statistische Bevölkerung" einer Kommune umfasst standardgemäß nur gemeldete Personen unter bestimmten Voraussetzungen. (Eigene Darstellung nach Deutscher Städtetag, 2019, und Molter, 2024)

2.2.3 Quantifizierung der Untererfassungen

Zur Untersuchung von Über- und Untererfassungen werden in ausgewählten Gemeinden und Gebäuden Stichprobenerhebungen sowohl bei den Meldebehörden als auch direkt bei den Einwohnerinnen und Einwohnern durchgeführt. Die Ergebnisse dieser Stichproben dienen der Hochrechnung registerbedingter Abweichungen und bilden eine wesentliche Grundlage für die statistische Korrektur der Melderegisterbestände. Ergänzend zur Haushaltsstichprobe erfolgt an Adressen mit Wohnheimen und Gemeinschaftsunterkünften eine Vollerhebung. Zusätzlich werden sogenannte Nullanschriften berücksichtigt, d. h. Adressen mit Wohnraum, an denen laut Melderegister keine Personen gemeldet sind. Diese werden gezielt in die Stichprobenziehung einbezogen, um vollständige Fehlbestände zu identifizieren und einer systematischen Untererfassung entgegenzuwirken (Statistisches Bundesamt, 2024).

Auch die Methoden des Zensus unterliegen bestimmten Einschränkungen, stellen jedoch die einzige verlässliche Vergleichsquelle dar. Zensusdaten sollten daher nicht als absolute „Wahrheit“ interpretiert werden, sondern als qualitätsgesicherte Referenz, die für die Einordnung, Validierung und Unsicherheitsabschätzung kommunaler Bevölkerungsdaten genutzt wird.

In ganz Deutschland gibt es nach Auskunft des Statistischen Bundesamts 549.103 Untererfassungen. Das entspricht 0,007% der erfassten Bevölkerung (siehe Tab. 2). In dieser Zahl sind wiederum nur Personen enthalten, die einer Meldepflicht unterliegen. Veröffentlichungen zu fehlerhaften Erfassungen erfolgen üblicherweise nur auf kommunaler Ebene. Zur Wahrung des Rückspielverbots werden keine Informationen zu Fehlerfassungen auf kleinräumiger Ebene herausgegeben.

Zusammenfassend lässt sich an diesem Punkt festhalten, dass aufgrund unzureichender Kenntnisse der zugrunde liegenden Erhebungsverfahren sowie der vorgelagerten Vorverarbeitungsschritte Bevölkerungsdaten in SDSS integriert werden können, die für den jeweiligen Anwendungszweck nur

eingeschränkt geeignet sind. Insbesondere bei der Analyse von Krankheitsinzidenzen ist die Berücksichtigung der gesamten relevanten Wohnbevölkerung erforderlich, da andernfalls erhebliche Verzerrungen der Ergebnisse auftreten können. Gleichzeitig sind die hierfür benötigten Informationen nicht in allen Fällen vollständig verfügbar, sodass methodische Ansätze erforderlich sind, die eine transparente und für die Anwenderinnen und Anwender verständlichen Nutzung der vorhandenen Daten ermöglichen.

3 METHODIK

Diese Arbeit untersucht die potenziellen Auswirkungen von Untererfassungen in Melderegistern auf gängige Visualisierungen in SDSS. Da in diesem Fall – wie in Abschnitt 2 dargelegt – kaum Vergleichsgrößen vorliegen, die eine quantitative Überprüfung der Daten erlauben, und lediglich Vermutungen über die räumliche Verteilung der Untererfassungen angestellt werden können, ist eine mathematische Analyse, beispielsweise anhand von Best- und Worst-Case-Szenarien, zwar grundsätzlich möglich, jedoch aufgrund der Vielzahl beteiligter Variablen nur sehr vage. Aus diesem Grund wurde ein explorativer Ansatz gewählt, der mithilfe eines in Python implementierten Tools die potenziellen Auswirkungen von Untererfassungen auf Choroplethenkarten und Balkendiagramme visualisierbar macht.

Die zugrunde liegende Idee besteht darin, dass eine definierte Anzahl von Untererfassungen nach bestimmten Kriterien auf den städtischen Raum verteilt wird. Jede Person, die möglicherweise nicht gemeldet oder in der Statistik erfasst ist, erhält hierbei einen zufällig zugewiesenen Wohnort. Auf diese Weise kann der von der Kommune bereitgestellte Datensatz direkt mit dem alternativ generierten Datensatz visuell verglichen werden. Für die Erstellung der alternativen Verteilungen müssen jedoch verschiedene Annahmen getroffen werden.

3.1 Annahmen und Vorgehensweise

Es bestehen Hinweise darauf, dass Personen, die nicht im Melderegister erfasst sind, nicht gleichmäßig, peripher oder zentral im Stadtraum verteilt leben, sondern, dass sich räumliche Ballungs- oder Clusterzentren bilden. Diese Annahme wird durch Vollerhebungen des Zensus in Wohnheimen sowie durch rechtliche Besonderheiten, wie die fehlende Meldepflicht in Seniorinnen- und Seniorenheimen, gestützt.

Die methodische Vorgehensweise basiert auf der Erzeugung zufälliger Punkte innerhalb der Stadtgrenzen von Kaiserslautern, die die Wohnorte potenziell nicht erfasster Personen repräsentieren. Die Punkte können beliebig im Stadtraum liegen, einschließlich Grünflächen, Straßen, Gewässern oder öffentlichen Gebäuden. Für die Analyse ist jedoch ausschließlich der übergeordnete räumliche Bezirk – entweder der Ortsbezirk oder der statistische Bezirk – relevant, sodass diese Unschärfe vernachlässigt werden kann.

Zur Untersuchung werden zwei räumliche Verteilungen berücksichtigt. Bei der (räumlichen) Gleichverteilung werden die Punkte gleichmäßig über das Stadtgebiet verteilt, wobei größere Flächen aufgrund der Verteilungsmethodik automatisch stärker berücksichtigt werden (siehe Abb. 2). Bei der Cluster-Verteilung wird zunächst eine definierte Anzahl zufälliger Clusterzentren generiert, um die herum die zufälligen Punkte lokal konzentriert und zufällig verteilt werden. Hierbei werden um die x- und y-Koordinaten der Clusterzentren gleichverteilte Zufallswerte im Bereich von -0.01 bis $+0.01$ um die x- und y-Koordinaten der Clusterzentren erzeugt, anschließend wird überprüft, ob die generierten Punkte innerhalb der Stadtgrenzen liegen. Für jeden erzeugten Punkt wird das jeweilige räumliche Gebiet ermittelt, um auf dieser Grundlage eine angepasste Bevölkerungsstatistik zu erstellen. (Hierbei wird jedes Mal zufällig ein Clusterzentrum gewählt. Clusterzentren, die am Rand liegen, fallen somit weniger ins Gewicht, da zufällig erzeugte Punkte im Bereich um diese häufig außerhalb der Stadtgrenzen liegen.)

Nach der Generierung aller zufälliger Punkte wird bestimmt, in welchen räumlichen Gebieten sie liegen und Visualisierungen (Choroplethenkarte und Barchart) zum direkten Vergleich erzeugt. Für die Darstellungen wird eine feste farbliche Skalierung verwendet, um die Vergleichbarkeit der resultierenden Karten sicherzustellen.

3.2 Analyseumgebung und Implementierung

Zur explorativen Analyse wurde ein Prototyp in Python umgesetzt, der mehrere spezialisierte Bibliotheken nutzt, um unterschiedliche funktionale Anforderungen abzudecken:

- Dash¹ dient als Framework zur Erstellung einer interaktiven Webanwendung und ermöglicht die benutzerfreundliche Darstellung sowie Steuerung der Analyse.
- Plotly² wird für die interaktive Visualisierung der Daten eingesetzt und erlaubt dynamische, explorative Grafiken.
- GeoPandas³ erweitert die Datenanalyse um räumliche Aspekte und unterstützt die Verarbeitung und Analyse von Geodaten.
- Shapely⁴ wird für geometrische Operationen wie das Berechnen, Prüfen und Manipulieren von Geometrien verwendet wird.
- Random⁵ kommt zur Erzeugung von (Pseudo-)Zufallszahlen zum Einsatz. In diesem Fall die Erzeugung der Lage der Clusterzentren und möglichen Wohnorte der untererfassten Personen.

Gemeinsam bilden diese Bibliotheken die technische Grundlage für eine flexible, interaktive und räumlich orientierte explorative Datenanalyse.

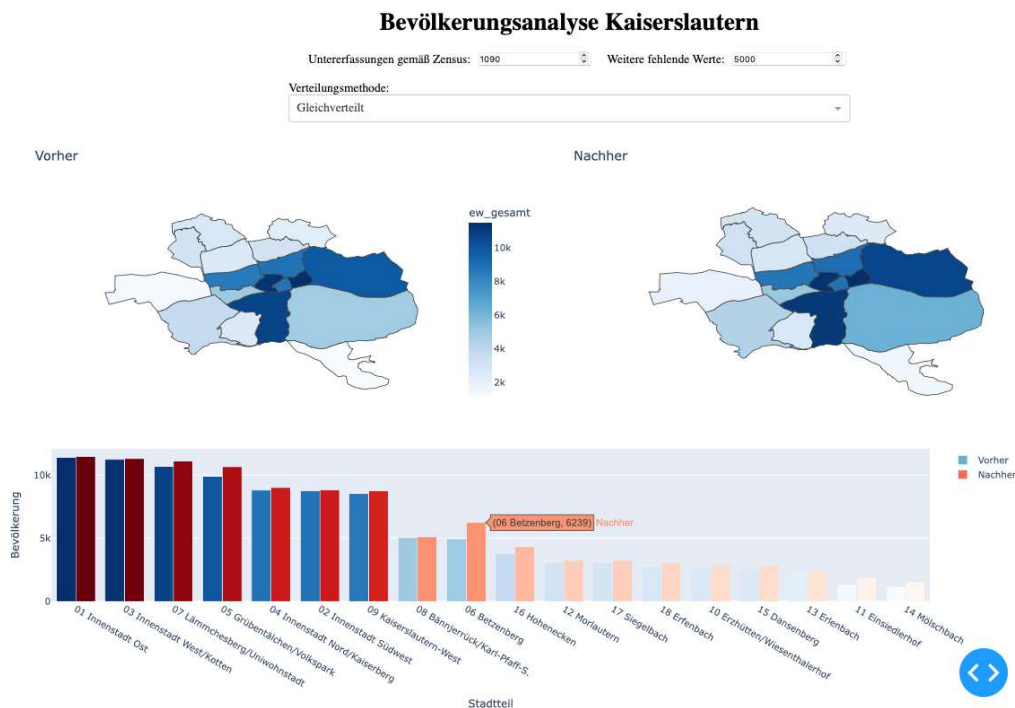


Abb. 3: Oberfläche des Tools. In diesem Fall wurden die zufälligen Punkte räumlich gleichverteilt. Man sieht, dass große Gebiete anteilig mehr Zuwachs bekommen. Die Farbwahrnehmung der Karte ändert sich kaum.

3.3 Parameter

In der Eingabemaske können mehrere Parameter direkt durch die Nutzerinnen und Nutzer festgelegt werden. Dazu zählt die Anzahl der Untererfassungen gemäß Zensus 2022, die standardmäßig mit dem für Kaiserslautern bekannten Wert von 1090 vorbelegt ist. Zusätzlich steht ein weiteres Eingabefeld zur Erfassung zusätzlicher Untererfassungen zur Verfügung, das initial mit einem Beispielwert von 1 000 befüllt ist. Darüber hinaus kann gewählt werden, ob die zufällig generierten Punkte gleichmäßig über das Stadtgebiet verteilt werden sollen, oder konzentriert um zufällig erzeugte Clusterzentren auftreten; voreingestellt ist die Gleichverteilung.

Weitere Parameter werden innerhalb des Codes definiert. Hierzu gehören die räumliche Granularität der Analyse, die entweder auf Ebene der Ortsbezirke oder der statistischen Bezirke erfolgen kann, sowie die Anzahl der Clusterzentren, die in Abhängigkeit von der jeweiligen Art der Untererfassung festgelegt wird.

¹ <https://dash.plotly.com>

² <https://plotly.com/python/>

³ <https://geopandas.org/>

⁴ <https://shapely.readthedocs.io/>

⁵ <https://docs.python.org/3/library/random.html>

3.4 Auswertungen

In der Untersuchung wurden die in Tabelle 1 dargestellten Merkmalskombinationen analysiert und bewertet. Dabei wurden zum einen die Anzahl der räumlichen Einheiten erfasst, die gegenüber dem Ausgangswert eine Veränderung von mehr als 20% aufweisen, und zum anderen das jeweilige Maximum der Abweichungen bestimmt.

Raumbezug	Verteilung	Untererfassungen gemäß Zensus 2022	Weitere fehlende Werte	Anzahl der Clusterzentren je Art der UE
Ortsbezirke	Clusterzentren	1090	1000	5
Statistische Bezirke			5000	20
			10000	

Tabelle 1: Gewählte Parameterausprägungen für die initiale Untersuchung. Diese werden in allen möglichen Kombinationen betrachtet.

Da die Ergebnisse in hohem Maße zufallsabhängig sind und die Untersuchung einen explorativen Charakter aufweist, wird ergänzend eine vergleichende Beschreibung der visuellen Wahrnehmung der beiden Choroplethenkarten vorgenommen.

4 UMSETZUNG AM BEISPIEL DER STADT KAISERSLAUTERN

Zum Stichtag 15. Mai 2022 lebten in der Stadt Kaiserslautern gemäß Zensus knapp 100.000 Personen. Die Bevölkerungsstruktur der Stadt ist durch einen Faktor geprägt, der für die vorliegende Untersuchung von besonderer Relevanz ist: Die Nähe zur Air Base Ramstein, der größten militärischen Einrichtung der USA im Ausland, führt seit mehreren Jahrzehnten zu einer ausgeprägten Präsenz von Angehörigen der US-Streitkräfte sowie deren Familien im Stadtgebiet und im umliegenden Landkreis. Für Stadt und Landkreis Kaiserslautern wird die Zahl der in der Region ansässigen US-Amerikanerinnen und US-Amerikaner, einschließlich aktiver Soldatinnen und Soldaten, ziviler US-Beschäftigter, ehemaliger Mitglieder der Streitkräfte sowie deren Angehöriger, auf etwa 50.000 Personen geschätzt. Die Zahl der offiziell in Deutschland gemeldeten US-Staatsbürger ist hingegen deutlich geringer, da für NATO-Angehörige keine Meldepflicht besteht. Für den Landkreis Kaiserslautern werden im Jahr 2023 rund 1.185 US-amerikanische Staatsangehörige ausgewiesen, während in der Stadt Kaiserslautern im Jahr 2023 etwa 910 US-Bürgerinnen und US-Bürger registriert waren (Statistisches Landesamt Rheinland-Pfalz 2024). Die US-Community stellt somit einen signifikanten demografischen und strukturellen Faktor dar, der auch die Stadt Kaiserslautern prägt.

Darüber hinaus existieren in der Stadt neun von der Stadtverwaltung gelistete Pflegeheime⁶ sowie sieben Gemeinschaftsunterkünfte für Geflüchtete⁷, die für die vorliegende Betrachtung ebenfalls von Bedeutung sind. Der Zensus geht für die Stadt von 1.090 Untererfassungen aus, was einer Quote von 0,01 % entspricht und somit über dem Durchschnitt von Rheinland-Pfalz sowie dem Bundesdurchschnitt liegt (siehe Tab. 2).

Raumbezug	Einwohner gesamt	Stationierte US-Streitkräfte	Übererfassungen gesamt	Übererfassungen anteilig (gerundet)	Untererfassungen gesamt	Untererfassungen anteilig (gerundet)
Deutschland	82.719.540	37.547	1.790.259	0,022	549.103	0,007
Rheinland-Pfalz	4.094.169	18.959	68.153	0,017	24.977	0,006
Stadt Kaiserslautern	99.689	n.n.	2.198	0,022	1.090	0,01

Tabelle 2: Über- und Untererfassungen ausgewählter Raumeinheiten gemäß Zensus 2022, Anfrage an Destatis und Deutscher Bundestag 2022

⁶ https://www.kaiserslautern.de/sozial_leben_wohnen/soziales_und_gesellschaft/aeltere_mitbuerger/angebote_und_hilfen/Altenpflegeheime/index.html.de

⁷ <https://www.asz-kl.de/fluechtlingsunterkuenfte/> und <https://www.drk-kl.de/angebote/migration-integration-fluechtlingshilfe/>

Es wurden sämtliche möglichen Kombinationen der gewählten Merkmalsausprägungen für die Stadt Kaiserslautern untersucht (vgl. Tab. 2). Die hierfür erforderlichen kleinräumigen Bevölkerungsdaten wurden freundlicherweise von der Stadtverwaltung zur Verfügung gestellt.

4.1 Ergebnisse in Kaiserslautern

Tabelle 3 zeigt die Ergebnisse einer explorativen Analyse verschiedener Kombinationen von Merkmalen im Hinblick auf Untererfassungen in Kaiserslautern. Als Ausreißer werden dabei räumliche Einheiten definiert, deren aktualisierter Wert mehr als 20 % über dem ursprünglichen Wert liegt. Dies entspricht Bezirken, in denen die Anzahl der Einwohnerinnen und Einwohner um mehr als 20 % höher ist als in den städtischen Ausgangsdaten.

Die Ergebnisse weisen mehrere erkennbare Muster auf:

- Abhängigkeit von der Anzahl fehlender Werte: Mit zunehmender Anzahl zusätzlicher Untererfassungen steigt sowohl die Zahl der Ausreißer als auch die maximale Veränderung.
- Einfluss der Anzahl der Clusterzentren: Eine größere Anzahl an Clusterzentren verteilt die Untererfassungen stärker über die Bezirke, wodurch die maximalen Veränderungen bei niedrigen zusätzlichen Untererfassungen reduziert werden. Bei hohen zusätzlichen Untererfassungen bleiben Ausreißer jedoch weiterhin signifikant.
- Räumliche Granularität: Auf der Ebene der kleineren statistischen Bezirke zeigen sich deutlich höhere Maximalwerte und mehr Ausreißer. Dies verdeutlicht, dass kleinere räumliche Einheiten besonders sensibel auf Untererfassungen reagieren.

Insgesamt lässt sich festhalten, dass die Anzahl und Ausprägung von Ausreißern sowohl mit der Zahl der zusätzlichen Untererfassungen als auch mit der Anzahl der Clusterzentren zunimmt. Extreme Ausreißer, wie ein Zuwachs von 245 %, könnten bereits bei einer Untererfassung von etwa 1,1 % der Gesamtbevölkerung auftreten. Moderate Ausreißer treten hingegen vor allem dann auf, wenn viele Clusterzentren vorliegen und die Anzahl fehlender Werte gering ist.

Fall	Raumbezug	Verteilung	Untererfassungen gemäß Zensus 2011	Weitere fehlende Werte	Anzahl der Clusterzentren je Feld	Anzahl der Ausreißer >20%	Maximale Veränderung (zum Ausgangswert)
1	Ortsbezirke	Clusterzentren	1090	1000	5	0	-
2	Ortsbezirke	Clusterzentren	1090	5000	5	3	+98%
3	Ortsbezirke	Clusterzentren	1090	10000	5	5	+77%
4	Ortsbezirke	Clusterzentren	1090	1000	20	0	-
5	Ortsbezirke	Clusterzentren	1090	5000	20	1	+27%
6	Ortsbezirke	Clusterzentren	1090	10000	20	4	+63%
7	Statistische Bezirke	Clusterzentren	1090	1000	5	5	+69%
8	Statistische Bezirke	Clusterzentren	1090	5000	5	10	+151%
9	Statistische Bezirke	Clusterzentren	1090	10000	5	14	+245%
10	Statistische Bezirke	Clusterzentren	1090	1000	20	5	+41%
11	Statistische Bezirke	Clusterzentren	1090	5000	20	12	+122%
12	Statistische Bezirke	Clusterzentren	1090	10000	20	18	+156%

Tabelle 3: Untersuchte Merkmalsausprägungen und zugehörige Ergebnisse. Für die grau hinterlegten Zeilen werden in diesem Beitrag exemplarisch graphische Darstellungen bereitgestellt.

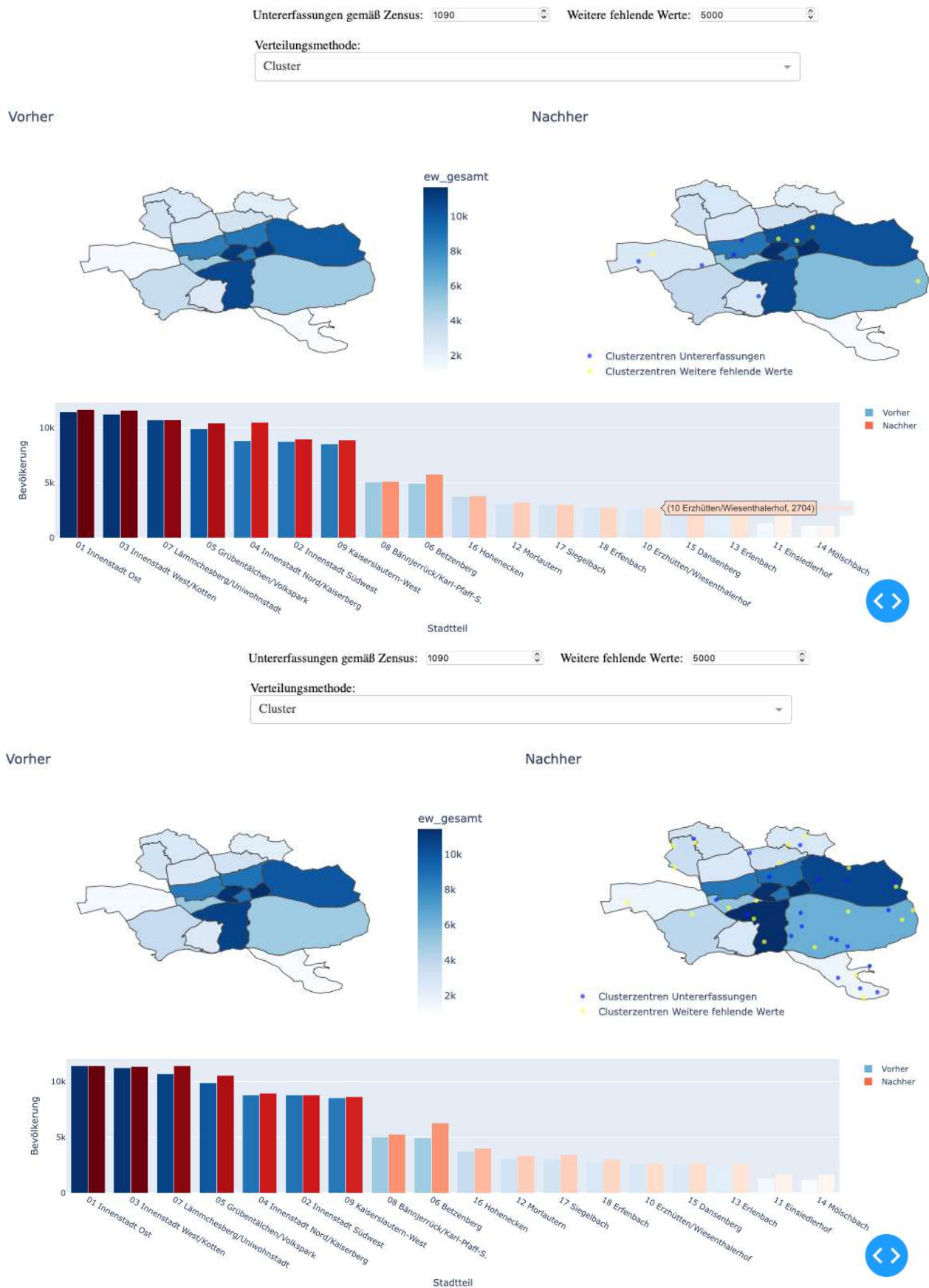


Abb. 3: Vergleich zweier Ergebnisse auf Ortsbezirksebene. Oben sind insgesamt 10 Clusterzentren (Fall 2) und unten 40 (Fall 5).

Unsichtbare Gruppen – sichtbare Effekte: Eine Untersuchung potenzieller Auswirkungen von Untererfassungen in deutschen Melderegisterdaten und Meldepflichtsbefreiungen auf gängige visuelle Darstellungen von Bevölkerungsgrößen in SDSS

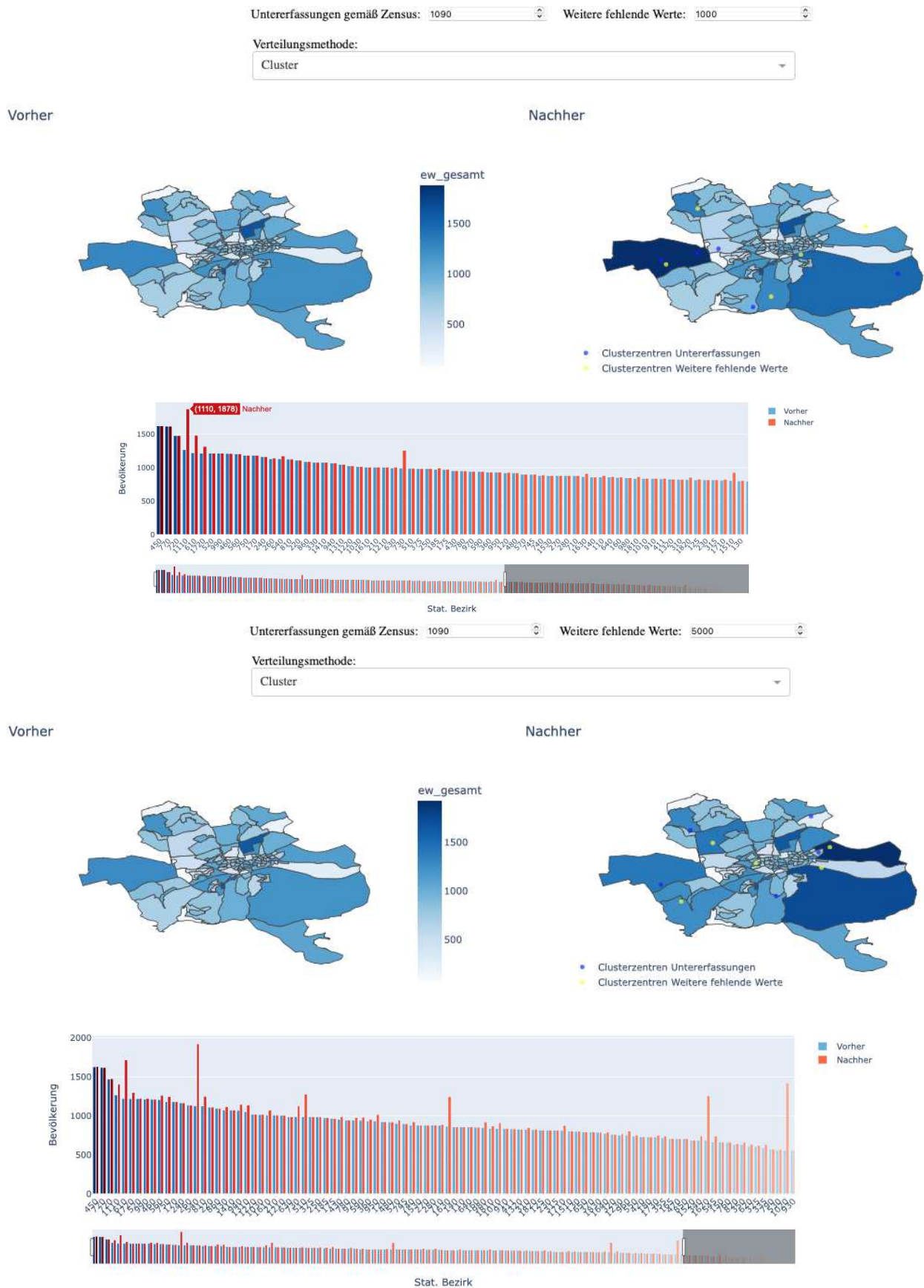


Abb. 4: Vergleich zweier Ergebnisse auf Ebene der statistischen Bezirke. Oben sind 1000 weitere fehlende Werte (Fall 7) und unten sieht man die Ergebnisse für 5000 fehlende Werte (Fall 8).

5 ERGEBNISSE

Die Analyse zeigt erwartungsgemäß, dass die Auswirkungen auf die Choroplethenkarten von der gewählten Verteilungsart, der Anzahl und Lage der Clusterzentren sowie der Größe der betrachteten räumlichen Einheiten abhängen. Bei einer räumlich gleichverteilten Zuweisung von Untererfassungen werden größere Ortsbezirke stärker berücksichtigt, da sie proportional mehr Fläche einnehmen (siehe Abb. 2).

Bei der Cluster-basierten Verteilung zeigt sich eine deutliche Abhängigkeit von der Anzahl der Clusterzentren: Mit zunehmender Anzahl der Clusterzentren nehmen die maximalen Abweichungen in den Bezirken ab. Ortsbezirke, die kein Clusterzentrum enthalten, weisen nur geringe Veränderungen auf, während Bezirke mit Clusterzentren signifikante Unterschiede zeigen (siehe Abb. 3). Zudem ist zu beobachten, dass die Auswirkungen stärker ins Gewicht fallen, je kleiner die betrachtete räumliche Einheit ist (siehe Abb. 4).

Es ist zu beachten, dass die Ergebnisse zufallsbasiert sind; eine erneute Durchführung der Simulation kann daher zu abweichenden Verteilungsmustern führen.

6 KRITIK

Trotz der gewonnenen Erkenntnisse weist der gewählte Ansatz mehrere methodische Einschränkungen auf. Die Ergebnisse sind stark abhängig von den getroffenen Annahmen sowie der zufallsbasierten Generierung der Punkte, wodurch eine hohe Variabilität und eingeschränkte Reproduzierbarkeit besteht. Validierungsdaten für die tatsächliche räumliche Verteilung von Untererfassungen sind nahezu nicht verfügbar, insbesondere für marginalisierte Gruppen oder spezifische Bevölkerungssegmente wie Angehörige der US-Streitkräfte. Darüber hinaus bleibt unklar, in welchem Umfang die betrachteten Gruppen für Planungsprozesse relevant sind und ob die Visualisierungen tatsächlich einen praktischen Mehrwert liefern. Auch die Betrachtung von Übererfassungen oder Korrelationen zu anderen Merkmalen ist aufgrund fehlender Vergleichsdaten nur sehr begrenzt möglich. Insgesamt handelt es sich daher um einen explorativen Ansatz, der eher ein qualitatives Verständnis der potenziellen Auswirkungen von Untererfassungen vermittelt, während eine quantitative Validierung derzeit nicht realisierbar ist.

Aktuelle ergibt sich kein Unterschied durch die beiden Eingabefelder außer die Verschiedenfarbigkeit der erzeugten Punkte, zukünftig ist allerdings eine Kombination aus zwei verschiedenen Verteilungen und eine Integration von Übererfassungen denkbar.

7 FAZIT UND AUSBLICK

Die Untersuchung zeigt, dass Untererfassungen in Meldedaten einen deutlichen Einfluss auf kleinräumige Darstellungen wie Choroplethenkarten und Balkendiagramme haben können. Die Auswirkungen sind insbesondere dann ausgeprägt, wenn die betrachteten räumlichen Einheiten klein sind oder die Untererfassungen in wenigen Clusterzentren konzentriert auftreten. Aufgrund der hohen Unsicherheit in der räumlichen Verteilung der nicht erfassten Personen und der eingeschränkten Verfügbarkeit valider Vergleichsdaten auf kleinräumiger Ebene kann die Analyse nur explorativ erfolgen. Die Ergebnisse liefern dennoch wertvolle Hinweise darauf, wie sich Untererfassungen in unterschiedlichen Szenarien auswirken können und ermöglichen ein besseres Verständnis der potenziellen Verzerrungen in kommunalen Daten.

Auch wenn die Auswirkungen nicht immer sichtbar werden, können selbst ausbleibende Effekte wertvolle Einblicke in die Robustheit von SDSS gegenüber Datenlücken liefern. Diese Untersuchung unterstreicht die Notwendigkeit einer transparenten Kommunikation sowohl von Annahmen, Verarbeitungslogiken und bekannter Besonderheiten von Datengrundlagen, um fundierte Entscheidungen zu ermöglichen und ein kritisches Verständnis datenbasierter Anwendungen zu fördern. Diese Arbeit bildet eine Grundlage für zukünftige Forschung zu Methoden der Unsicherheitskommunikation und zur Gestaltung entsprechender Visualisierungsstrategien in SDSS.

Für die zukünftige Anwendung bietet sich an, die Visualisierungen als ergänzendes Instrument innerhalb von SDSS, in Schulungen zur Datenqualität oder als Grundlage für dynamische Darstellungen von Unsicherheiten zu nutzen. Ein signifikanter Informationsgewinn wäre außerdem zu erwarten, wenn reale Kenntnisse über Clusterzentren oder zusätzliche Daten zu spezifischen Bevölkerungsgruppen in Zusammenarbeit mit Statistikstellen integriert werden könnten. Darüber hinaus könnten Anfragen bei relevanten Stellen, etwa zu Pflegeheimen oder Angehörigen der NATO, dazu beitragen, die unbekannte Zahl

Unsichtbare Gruppen – sichtbare Effekte: Eine Untersuchung potenzieller Auswirkungen von Untererfassungen in deutschen Melderegisterdaten und Meldepflichtbefreiungen auf gängige visuelle Darstellungen von Bevölkerungsgrößen in SDSS

weiterer Untererfassungen grob einzugrenzen. Insgesamt unterstreicht die Analyse den Wert explorativer Ansätze, um potenzielle Verzerrungen zu erkennen, und liefert zugleich Impulse für weiterführende Untersuchungen und methodische Verbesserungen.

8 FÖRDERUNG

Diese Arbeit wird im Rahmen des interdisziplinären Forschungsprojekts „Ageing Smart – Räume intelligent gestalten“, das von der Carl-Zeiss-Stiftung gefördert wird, durchgeführt. Kaiserslautern ist eine von sieben Modellkommunen in diesem Projekt.

9 QUELLEN

- BAHRENBERG, Gerhard; GIESE, Ernst; NIPPER, Jürgen; SCHÖNERT, Martin (Hrsg.): Statistische Methoden in der Geographie. Band 1: Univariate und bivariate Statistik. Schweizerbart Textbooks, 2020.
- BUNDESMINISTERIUM FÜR ARBEIT UND SOZIALES (BMAS): Ausmaß und Struktur von Wohnungslosigkeit – Der Wohnungslosenbericht 2022 des Bundesministeriums für Arbeit und Soziales. Berlin, 2022.
Verfügbar unter: <https://www.bmwsb.bund.de/SharedDocs/downloads/Webs/BMWSB/DE/veroeffentlichungen/pm-kurzmeldung/wohnungslosenbericht-2022.pdf>
- BUNDESAMT FÜR JUSTIZ(BfJ): Streitkräfteaufenthaltsgesetz, 2020. Verfügbar unter: <https://www.gesetze-im-internet.de/skaufg/>
- DENT, Borden D.; TORGUSON, Jeffrey; HODLER, T. W.: Cartography: Thematic Map Design. McGraw-Hill Higher Education, 2009.
- DEUTSCHER BUNDESTAG: Antwort der Bundesregierung auf die Kleine Anfrage der Abgeordneten Sevim Dağdelen, Andrej Hunko, Ina Latendorf, Zaklin Nastic und der Fraktion DIE LINKE – Kosten für und durch in Deutschland stationierte NATO Streitkräfte in den Jahren 2021 und 2022, BT Drucksache 20/7303), 2023. Verfügbar unter: <https://dserver.bundestag.de/btd/20/073/2007303.pdf>
- DEUTSCHER STÄDTETAG: Statistikdatensatz Bevölkerungsbestand, KOSIS-Gemeinschaft HHSTAT, 2019.
- FISHER, Peter F.: Models of uncertainty in spatial data. In: John Wiley & Sons (Hrsg.): Geographical information systems, S.191-205. New York, 1999.
- LEISCHNER, Sonja; KHALIL, Aiman: Informationelle Infrastruktur in der amtlichen Statistik – Der Zugang kommunaler Statistikstellen zu bundesstatistischen Einzelangaben. In: Statistisches Bundesamt (Hrsg.): WISTA 4/2025. https://www.destatis.de/DE/Methoden/WISTA-Wirtschaft-und-Statistik/2025/04/infrastruktur-amtliche-statistik-042025.pdf?__blob=publicationFile&v=3, abgerufen am 10.01.2026
- MACEACHREN, Alan; ROBINSON, Anthony; HOPPER, Susan; GARDNER, Steven; MURRAY, Robert; GAHEGAN, Mark; HETZLER, Elisabeth: Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know. Cartography and Geographic Information Science. 32. S.139-160. 2005. <https://doi.org/10.1559/1523040054738936>
- MAYER, Julia; MEMMEL, Martin: Daten als Schlüssel zur erfolgreichen Entscheidungsunterstützung. In: Spellerberg, A., Ruzika, S. (Hrsg.): Ageing Smart – Digitale Instrumente im kommunalen Kontext. Wiesbaden: Springer Verlag, 2026. https://doi.org/10.1007/978-3-658-48575-7_12
- MAYER, Julia; MEMMEL, Martin: Wenn die KI mitentscheidet – Ein praxisnahes Bewertungskonzept für datenbasierte Smart City Technologien als Grundlage für besseres Verständnis und transparente Kommunikation. In: WANDERWITZ, Maximilian (Hrsg.): Risikoanalyse Künstliche Intelligenz. Berlin: Springer Verlag, im Erscheinen (2026).
- MOLTER, Christopher: Wie beantrage ich eine Meldeadresse, ohne dort zu wohnen? Jurawelt, 2024. <https://jurawelt.com/wie-beantrage-ich-eine-meldeadresse-ohne-dort-zu-wohnen/>, abgerufen am 11.12.2025.
- MONTELLO, Daniel; GOODCHILD, Michael; BEARD, Kate; CLARKE, Keith; KEUPER, Alex; SHORTRIDGE, Ashton; WINDHOLZ, Thomas; TEAGUE, Paul; FOHL, Peter; KYRIAKIDIS, Phaedon; HUNTER, Gary; AGUMYA, Aggrey: Uncertainty in Geospatial Information Representation, Analysis, and Decision Support. A Research Project Funded by the National Imagery and Mapping Agency under the NURI University Research Initiative, 1998. Verfügbar unter: https://escholarship.org/content/qt14v6587w/qt14v6587w_noSplash_6dbb1f4065b42d807637d968e5270adb.pdf
- NOLLERT, Markus: Raumplanerisches Entwerfen – Entwerfen als Schlüsselement von Klärungsprozessen der aktionsorientierten Planung am Beispiel des regionalen Massstabs. Dissertation, ETH Zürich, 2013. <https://doi.org/https://doi.org/10.3929/ethz-a-009937192>
- PAULY, Jonas; ABU QASEM, Rasha; MAYER, Julia; SCHROTH, Christof; MEMMEL, Martin; PALLAGST, Karina: Entscheidungsunterstützungssysteme in der Raumplanung – Ansätze für eine Typisierung. In: Spellerberg, A., Ruzika, S. (Hrsg.): Ageing Smart – Digitale Instrumente im kommunalen Kontext. Wiesbaden: Springer Verlag, 2026. https://doi.org/10.1007/978-3-658-48575-7_10
- STATISTISCHES BUNDESAMT (Destatis): ZENSUS 2022, 2024. Verfügbar unter: <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Zensus2022>
- STATISTISCHES LANDESAMT RHEINLAND-PFALZ: US Wahlen 2024: Zahlreiche US Staatsangehörige in Rheinland Pfalz wahlberechtigt – Wirtschaft und Tourismus eng mit den USA verknüpft, 2024. Verfügbar unter: <https://www.statistik.rlp.de/nachrichten/nachrichtendetailseite/us-wahlen-2024-zahlreiche-us-staatsangehoerige-in-rheinland-pfalz-wahlberechtigt-wirtschaft-und-tourismus-eng-mit-den-usa-verknuepft>
- SUGUMARAN, Ramanathan; DEGROOTE, John: Spatial Decision Support Systems: Principles and Practices. Baton Rouge: Taylor & Francis, 2010. Print. S. 442 ff.