

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Investigating the risk of spatial re-identification of health geodata from ESDA outputs of confidential versus masked geodata.

verfasst von | submitted by

Valeria Wallach BA

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt |
Degree programme code as it appears on the
student record sheet:

UA 066 856

Studienrichtung lt. Studienblatt | Degree
programme as it appears on the student
record sheet:

Masterstudium Kartographie und Geoinformation

Betreut von | Supervisor:

Ass.-Prof. Dr. Ourania Kounadi BSc MSc

Acknowledgements

First, I would like to thank my supervisor Dr. Ourania Kounadi. Thanks to your specialist knowledge and advice, this thesis came about.

I would also like to thank the University of Vienna for the educational opportunities it has offered me.

Especially I would like to thank my family who supported me during my studies. A heartfelt thanks to my sons Jerome and Benjamin, who showed great understanding for my studies.

At this point, thanks to all medical experts who took part in the survey.

Abstract

Spatial analyses are a part of health geography and takes place at an aggregated level. So far, no standardized geomasking method has been approved by the GDPR to be able to perform spatial analysis at the city level and smaller. Health data is highly sensitive data that can only be mapped in scientific journals if it does not pose a risk of spatial reidentification. This makes it all the more important to find a suitable masking method that retains the spatial pattern of the original data while complying with the GDPR. In this thesis, discrete health data were at the city level of Vienna simulated. In the next step, the simulated data (3) was geomasked using Adaptive Voronoi Masking (AVM) and Adaptive Areal Masking (AAM). Finally, the masked points were examined for their suitability in medical research using mainly traditional ESDA-techniques. Based on the ESDA results, the risk of spatial re-identification was visually assessed. Based on the point maps with masked points, the AVM method retains the spatial pattern better than the AAM method while providing better protection against spatial re-identification. The results of the cluster outlier analysis (Anselin) at the point level have enabled detailed data analysis. Based on the AVM-Outliers visualization in the area with lower density, a connection between spatial anonymity and the road network and building density could be assumed. The results of the Kernel Density Estimation and Hotspot Analysis, Choropleth maps, and Cluster-Outlier analysis at the registration district level, have shown that the risk is lower compared to point maps but is still present. For this reason, it was recommended to perform the spatial analyses with masked data. For health geography, the AVM method is recommended. This is because the AAM method spreads the point pattern widely and increases the risk of false re-identification. However, that doesn't mean that the AAM geomask doesn't have any benefits in itself. But a certain degree of self-criticism must remain because health data is highly sensitive information. As part of this work, a qualitative online survey was carried out in Viennese health care facilities. According to the results (less than expected), more public information could be available via spatial analysis and geomasking, which could improve public health.

Keywords: Health-Data-Simulation, Location Privacy, Health Geography, Risk of Reidentification, Spatial K-Anonymity, Geomasking, Adaptive Voronoi Masking, Adaptive Areal Masking, ESDA Outputs Visualization.

Zusammenfassung

Die räumliche Analysen sind ein Teil der Gesundheitsgeographie und findet auf aggregierter Ebene statt. Bisher wurde keine standardisierte Geomasking-Methode von der DSGVO genehmigt, um räumliche Analysen auf Stadtebene und kleiner durchführen zu können. Gesundheitsdaten sind hochsensible Daten, die nur dann in wissenschaftlichen Fachzeitschriften abgebildet werden können, wenn sie kein Risiko einer räumlichen Reidentifikation darstellen. Umso wichtiger ist es, eine geeignete Maskierungsmethode zu finden, die das räumliche Muster der Originaldaten beibehält und gleichzeitig die DSGVO einhält. In dieser Arbeit wurden discrete Gesundheitsdaten auf Wiener Stadtebene simuliert. Im nächsten Schritt wurden die simulierten Daten (3) anhand Adaptive Voronoi Maskierung (AVM) und Adaptive Areal Maskierung (AAM) geomaskiert. Letztendlich wurden die maskierte Punkte mit vorwiegend traditionellen ESDA-Techniken auf ihre Eignung in der medizinischen Forschung getestet. Basierend auf den ESDA-Ergebnissen wurde das Risiko einer räumlichen Re-Identifizierung visuell untersucht. Anhand der Punktkarten mit maskierten Punkten behält die AVM-Methode das räumliche Muster besser bei, als die AAM-Methode und bietet gleichzeitig einen besseren Schutz vor der räumlichen Reidentifikation. Die Ergebnisse der Cluster-Ausreißer-Analyse (Anselin) auf Punktebene haben eine detaillierte Datenanalyse ermöglicht. Basierend auf der AVM-Außreiser Visualisierung im Gebiet mit geringerer Dichte konnte ein Zusammenhang zwischen räumlicher Anonymität mit dem Straßennetz und der Bebauungsdichte angenommen werden. Die Ergebnisse der Kerndichte Schätzung und Hotspotanalyse, der Choropleth-Karten und der Cluster-Ausreißer-Analyse auf der Ebene der Registrationsbezirke haben gezeigt, dass das Risiko im Vergleich zu Punktkarten geringer ist, aber immer noch vorhanden ist. Aus diesem Grund wurde empfohlen die räumlichen Analysen mit maskierten Daten durchzuführen. Für die Gesundheitsgeographie empfiehlt sich die AVM-Methode. Dies liegt daran, dass die AAM-Methode das Punktmuster streut und das Risiko einer falschen Re-Identifizierung erhöht. Dies bedeutet jedoch nicht, dass die AAM-Geomaske keine Vorteile an sich hat. Aber ein gewisses Maß an Selbstkritik muss jedoch bleiben, denn bei den Gesundheitsdaten handelt es sich um hochsensible Informationen. Im Rahmen dieser Arbeit wurde eine qualitative Online-Befragung in Wiener Gesundheitseinrichtungen durchgeführt. Den Ergebnissen zufolge (weniger als erwartet) könnten mehr öffentliche Informationen über räumliche Analysen und Geomaskierung verfügbar sein, womit die öffentliche Gesundheit verbessert werden könnte.

Table of Contents

Abstract	3
Zusammenfassung	4
List of Figures	7
List of Tables	8
1. Introduction	9
2. Literature Review	11
2.1. Geoinformation in Health Geography	11
2.2. Protection of Health Information	12
2.2.1. Electronic Health Records-ELGA	14
2.3. Exploratory Spatial Data Analysis (ESDA)	16
2.3.1. Spatial Analysis Tools	18
2.4. Risk of Spatial Reidentification	19
2.5. Geomasking Methods.....	22
2.5.1. Spatial K-Anonymity	23
2.5.2. Adaptive Geomasking	23
2.5.2.1. Adaptive Areal Elimination (AAE).....	24
2.5.2.2. Adaptive Areal Masking (AAM)	26
2.5.2.3. Adaptive Voronoi Masking (AVM)	27
2.6. Selected Geomasking-Methods and ESDA.....	28
2.6.1. Adaptive Geomasking Methods.....	28
2.6.2. Spatial Visualization-Techniques.....	29
2.6.3. Hotspot Detection on Kernel Density Estimation	30
2.6.4. Spatial Autocorrelation Analysis (Moran's I).....	30
2.7. Software Overview	31
3. Methodology	32
3.1. Study Area of Vienna.....	32
3.2. Online-Survey on GIS-Analysis in Health Facilities.....	34
3.3.1. An Excerpt of the e-mail Correspondence	35
3.3.2. The Questions of the Survey on spatial GIS-Analysis.....	37
3.3. Population-Data Simulation	39
3.3.1. Simulated Dataset Analysis	42
3.3.2. General Considerations on Dataset Choice.....	45
3.4. Adaptive Geomasking Methods.....	47
3.4.1. Adaptive Areal Masking (AAM).....	48
3.4.2. AAE-Polygons (RoRi-Value 100 and 1.000).....	50
3.4.3. Adaptive Voronoi Masking (AVM)	53

3.5. Spatial Visualization-Techniques	55
3.5.1. Point Pattern Maps	55
3.5.2. Choropleth Maps	55
3.6. Hotspot Detection on Kernel Density Estimation	56
3.6.1. Kernel Density Estimation (KDE)	56
3.6.2. Optimized Hotspot Analysis	56
3.7. Spatial Autocorrelation Analysis (Moran's I)	57
3.7.1. Global Moran's I	57
3.7.2. Cluster-Outliers Analysis (Anselin Local Moran's I)	58
3.7.3. Average Nearest Neighbor	58
4. Results	59
4.1. Evaluation of the Survey Results (of 15.10.2023)	59
4.1.1. Conclusion of the Survey Evaluation	62
4.2. Spatial Visualization-Point Pattern	63
4.2.1. Overview of the Point Pattern	66
4.2.2. Point Displacement-Techniques.....	68
4.2.3. Masked Points on Orthophotos	71
4.2.4. Reidentification of Point X,Y Coordinates	74
4.2.5. Choropleth Maps at Registration District	75
4.3. Hotspot Detection on Kernel Density Maps	78
4.3.1. Re-identification Risk of Hotspot in Low-Density Area	81
4.4. Spatial Autocorrelation - Global Moran's I	84
4.4.1. Cluster-Outliers (Anselin) at Registration-District	85
4.4.2. Cluster-Outliers and Hotspot Maps.....	86
4.4.3. Risk of Reidentification and Outlier's	89
5. Discussion	92
6. Literature References	99
6.1. Online References	103
6.2. ArcGIS References	104
6.3. Orthophotos References.....	105
6.4. Data-Simulation References.....	105
7. Appendix	107
7.1. AVM Python Script adapted from Polzin & Kounadi.....	107
7.2. AAM Python Script adapted from Laura Charleux	110
7.3. Lime Survey Questionnaire	113
7.4. Survey Responses	129
7.5. Cluster- and Outliers Point Maps at Vienna City	148

List of Figures

Figure 1 Study Area Vienna ©ArcGIS Pro.....	32
Figure 2 Population Density 2012 ©ArcGIS Pro.....	33
Figure 3 Statistic of Data Simulation ©QGIS.....	40
Figure 4 Overview Simulated Points ©QGIS.....	41
Figure 5 Registration District 2223 ©QGIS.....	42
Figure 6 Registration District 1103/1102 ©QGIS.....	43
Figure 7 Registration District 1204 ©QGIS.....	44
Figure 8 Registration District 0304 ©QGIS.....	44
Figure 9 Simulated Points ©ArcGIS Pro.....	46
Figure 10 AAM Points (RoRi-100) ©ArcGIS Pro.....	48
Figure 11 RoRi Population 2012 and Simulated Points ©ArcGIS Pro.....	50
Figure 12 AAE-Polygons RoRi 100 ©ArcGIS Pro.....	51
Figure 13 AAE-Polygons RoRi 1.000 ©ArcGIS Pro.....	52
Figure 14 AVM, Voronoi Polygons ©ArcGIS Pro.....	53
Figure 15 AVM Points ©ArcGIS Pro.....	54
Figure 16 Diagram of Survey Participants ©LimeSurvey.....	59
Figure 17 Diagram of Spatial Analysis in Healthcare ©LimeSurvey.....	61
Figure 18 Point Pattern 1102 ©ArcGIS Pro.....	63
Figure 19 Point pattern 1103 ©ArcGIS Pro.....	64
Figure 20 Point Pattern 2223 ©ArcGIS Pro.....	64
Figure 21 Point Pattern 0304 ©ArcGIS Pro.....	65
Figure 22 Point Pattern 1204 ©ArcGIS Pro.....	65
Figure 23 Points Pattern AVM – Simulated ©ArcGIS Pro.....	66
Figure 24 Points Pattern AAM – Simulated ©ArcGIS Pro.....	67
Figure 25 Simmering 1102/1103 ©ArcGIS Pro.....	68
Figure 26 AVM Point Displacement K=100 ©ArcGIS Pro, Geoland Basemap.....	69
Figure 27 AAM Point Displacement K=1.000 ©ArcGIS Pro, Geoland Basemap.....	70
Figure 28 Study Area 1204 Orthophoto ©ArcGIS Pro.....	71
Figure 29 Study Area 1103 Orthophoto ©ArcGIS Pro.....	72
Figure 30 Study Area 2223 Orthophoto ©ArcGIS Pro.....	72
Figure 31 Vienna's Outskirts - Orthophotos ©ArcGIS Pro.....	73
Figure 32 X, Y Coordinates-Geomasking FID-1669, Orthophotos ©ArcGIS Pro.....	74
Figure 33 Choropleth Map Simulated (1:220.000) ©ArcGIS Pro.....	75
Figure 34 Choropleth Map AAM, RoRi 1.000 (1:220.000) ©ArcGIS Pro.....	76
Figure 35 Choropleth Map AVM, RoRi 100 (1:220.000) ©ArcGIS Pro.....	77
Figure 36 KDE and Hotspot Simulated-Points 1:100.000 ©ArcGIS Pro.....	78
Figure 37 KDE and Hotspot AVM-Method 1.1:100.000 ©ArcGIS Pro.....	79
Figure 38 KDE and Hotspot AAM-Method (RoRi 100) 1:100.000 ©ArcGIS Pro.....	80
Figure 39 Hotspot Area 1103-1102 ©ArcGIS Pro.....	81
Figure 40 Hotspot AVM 1102, Orthophoto ©ArcGIS Pro.....	82
Figure 41 Hotspot areas at scale 1:3.000 - 1:5.000 ©ArcGIS Pro.....	83
Figure 42 Cluster-Outliers Region Districts of Vienna ©ArcGIS Pro.....	85

Figure 43 Cluster-Outliers and Getis-Ord Gi* Maps Simulated ©ArcGIS Pro.....	86
Figure 44 Cluster-Outliers and Getis-Ord Gi* Maps AVM ©ArcGIS Pro.....	87
Figure 45 Cluster-Outliers and Getis-Ord Gi* Maps AAM ©ArcGIS Pro.....	88
Figure 46 AVM Outliers HL, Orthophotos ©ArcGIS Pro.....	90
Figure 47 Simulated Points Cluster-Outliers (Anselin) ©ArcGIS Pro	148
Figure 48 AAM Points Cluster-Outliers (Anselin) ©ArcGIS Pro	149
Figure 49 AVM Points Cluster-Outliers (Anselin) ©ArcGIS Pro	149

List of Tables

Table 1 Evaluation of the Survey Results ©Lime Survey	60
Table 2 Summary Statistics of the Mean (SKA) ©ArcGIS Pro.....	70
Table 3 Coordinates of Simulated-Point FID 1669	74
Table 4 Spatial Autocorrelation - Global Moran's I ©ArcGIS Pro.....	84
Table 5 Outliers Results – Anselin Local Moran's I ©ArcGIS Pro	89
Table 6 Outliers Z-Value ©ArcGIS Pro.....	90
Table 7 Average Nearest Neighbor ©ArcGIS Pro	91

1. Introduction

ESDAs consist of techniques for describing spatial distributions, identifying spatial outliers, detecting spatial association patterns, and clustering (Anselin and Bao, 1997). In addition, these techniques are suitable for investigating the strengths and weaknesses of geographical representations with masked datasets (Kounadi and Leitner, 2014). Health data includes data at different scales and levels of detail (Kounadi and Leitner, 2014). In order to be able to process and publish simple maps with point data or cluster analyses with disaggregated location-based health data, other protection methods such as geographic masking methods are necessary in addition to data anonymization. According to Kounadi and Leitner (2014), geomasking methods should be used to protect the privacy of health data in scientific articles (Kounadi and Leitner, 2014). Geomasking can be used to prevent the risk of re-identification of confidential (discrete) geospatial data (Kounadi and Leitner, 2014). When the location data is anonymized, the address data is converted into geocoordinates, and all personal data is deleted from the data set (Delmelle and Kanaroglou, 2015).

In practice, it is known that in 2014 confidential data were published in publications without prior protection (Kounadi and Leitner, 2014), but also that a few years later, a detailed spatial view of health is limited by lack of data access and privacy policies (Koller et al., 2020). This is due to the EU General Data Protection Regulation, which was introduced in 2018 (GDPR). This is because, due to their sensitivity, health data constitutes a special category of personal confidential data pursuant to Art. 9 (1) GDPR (BMASGK), the processing of which is subject to strict data protection regulations in Austria (BMASGK 2019). This also includes location-based information from patients (BMASGK 2019). In practice, a distinction is made between anonymized and pseudonymized data and their use (BMASGK 2019).

The health geography maps cover the distribution of a disease by regional units. Likewise, data at the city level are only available to a limited extent (Koller et al., 2020, p. 1112). On the one hand, because health services are more accessible regionally, and because user-friendly software solutions are available at the regional level (Koller et al., 2020). However, the statistical data that occurs at the district level is not as meaningful as the geodata at the point level. It can be assumed that better access to small-scale data could improve health care. This would also make accessibility analyses possible (Koller et al., 2020). It turns out that the data sets that contain location-based patient data and their disease are particularly important for health geography research. By analyzing location data, localized disease clusters can be identified, and their causes investigated, thus improving public health (Zhang et al., 2017). Spatial geographic analyses are part of public health and epidemiology (Koller et al., 2020).

Spatial analysis methods can be used to create disease distribution maps that provide meaningful information about the direction, intensity, and likelihood of disease spread.

Therefore, GIS and ESDA-techniques are well suited for vector-borne disease surveillance (Delmelle and Kanaroglou, 2015). Thus, ESDA is more advantageous compared to EDA, as it can be used to detect the spatial pattern (Zakaria et al., 2021).

The aim of this thesis is to investigate different forms of exploratory spatial data analysis (ESDA) that can be applied to city-level health data. The methodological part, the choice of optimal ESDA-techniques and the geomasking method, the approach to determining the risk of spatial re-identification, were first examined based on the scientific literature. It was important to find appropriate Geomasking and ESDA-techniques that do not violate data protection. It would be important that plausible conclusions for disease analyses can be drawn from the results, while the risk of spatial re-identification remains low. Since the focus is on spatial re-identification, it is necessary to examine the results of original and geomasked data for spatial re-identification risk. The risk of re-identification represents the risk of identifying the personal locations from a sensitive geospatial data set. The aim is to clarify which location-based spatial analyses are safest with regard to the risk of spatial re-identification in the context of data protection. On the basis of the results of the scientific literature and methodological part, it should be possible to answer the research questions (and the supplementary questions):

Research:

1. What techniques are commonly used by healthcare facilities to visualize patients' spatial data?
2. What other ESDA techniques are less popular but very relevant and informative?
3. What geomasking technique can be used for sensitive geospatial data that enables the application of a number of ESDA best techniques and minimizes or eliminates the spatial risk of re-identification?
4. How can map attackers tell which geomasking method was used?

The additional questions would be:

What are the implications of data protection for research? How high is the risk of spatial reidentification with the chosen techniques? Does data protection restrict the disclosure and use of health geodata in research or to inform the public?

In this thesis, confidential¹ location-based health data were simulated, based on the population density in Vienna 2012. An online survey of healthcare facilities in Vienna/the surrounding area was also conducted. The question of the extent to which healthcare institutions are aware of spatial analyses of patient data, or also carry out where potential barriers are located.

¹ The simulated data represents discrete address data of "household"

2. Literature Review

In this chapter, the topic was examined in more detail on the basis on the literature. This concerned the confidentiality and accessibility of health data as well as the risk of spatial re-identification. In addition, the spatial protection measures, including the selection of appropriate geomasking methods and ESDA techniques.

2.1. Geoinformation in Health Geography

According to Delmelle, Kanaroglou, health geography, or spatial epidemiology, deals with two fundamental questions: "Where and when do diseases occur? Why are there such patterns?" (Delmelle and Kanaroglou, 2015). It can be seen from the literature that the concept of an individual's place has increased in importance in recent years, as well as that the field of health geography has grown (Kwan 2012). Advances in geographic modelling techniques have made it possible to perform spatial analyses (Delmelle and Kanaroglou, 2015). Access to health professionals, which every patient should have equally, is challenging. Finally, travel impedance prevents the use of health services (Delmelle and Kanaroglou, 2015). It follows that the accessibility of health services in health geography is of great importance, that accessibility is a critical element in the health system (Delmelle and Kanaroglou, 2015). Accessibility can be examined based on accessibility analyses to be able to take necessary measures. This includes information on care inequalities at the local and national levels, as well as nationwide analyses (Koller et al., 2020). In addition , location analyses have been developed to support strategic decisions for the location of healthcare facilities (Delmelle and Kanaroglou, 2015).

The term health geography was already introduced in the German-speaking world by Leonhard Ludwig Finke, in which he described the first health geography map at the end of the 18th century (Koller et al., 2020). About 70 years later, medical historian August Hirsch published a three-volume handbook on historical-geographical pathology (Koller et al., 2020). The maps include classic information on disease ecology, such as regional spread of a disease. The temporal relationships between space and health have already been investigated, as well as access to care facilities (Koller et al., 2020).

The use of geographic information systems (GIS) has increased since the last few decades, resulting in the emergence of online health atlases such as the "Dartmouth Atlas of Health Care" (Koller et al., 2020). Nowadays, geographical methods can mainly be observed in (real-time) dashboards (Koller et al., 2020). Thanks to access to regional health data and user-friendly software solutions (GIS), geospatial analysis has become a part of health geography research (Koller et al., 2020).

Depending on technical possibilities and data availability, health data can be study using ESDA techniques and the results can therefore be viewed in map displays (Koller et al., 2020). In this way, new insights can be gained about the raw data sets. The literature of Koller et al. shows that the methodological spectrum in health geography ranges from map representations at different aggregation levels to complex methods, such as spatiotemporal ESDA representations (Koller et al., 2020). Geographic analyses at the municipal level include accessibility analyses, network analyses, as well as the walkability index (Koller et al., 2020). In addition, Modifiable Area Unit Problem (MAUP) are observed at the municipal level², which can be mitigated by geographically specific methods (Koller et al., 2020).

According Koller et al., another important step in health geography would be access to small-scale, location-based health data to be able to carry out accessibility analyses at a city level (Koller et al., 2020). This is because spatial analyses are severely limited by data protection and the lack of data access to small-scale information (Koller et al., 2020). The weaknesses do not lie in spatial statistics, or methods and analyses. But because the data are collected on different geographical units that are not compatible with each other, or the primary studies do not have a sample size that allows small-scale analyses (Koller et al., 2020). Ultimately, in practice, the small-scale data are aggregated at the regional level and finally analyzed (Koller et al., 2020). It can be judged that detailed spatial analyses in health geography are limited by data protection guidelines (Koller et al. 2020, p1115).

2.2. Protection of Health Information

Geomasking methods are used to protect sensitive patient data in map representations from spatial re-identification and correct person-household assignment. Because a patient's address can be used to infer identity, address fields are removed as part of the patient's confidential information before the data is released for research purposes (Stinchcomb, 2004). Several levels of anonymization policies are used, which depend on the recipient as well as the intended use (Benitez, Malin 2010). It is possible to geo-reference the cases, geocode the address data or aggregate them to the specific areas. In the process of geocoding, address data is converted into latitude and longitude (Allshouse et al. 2010). Anonymization methods are often used in practice to protect the privacy of data, but there are concerns about the protective effect of these methods

² Planning at the municipal level is based on administrative boundaries that do not extend beyond the city limits, resulting in different boundaries within an area that are not congruent (Koller et al., 2020), the size of the units is referred to as MAUP.

(Benitez and Malin, 2010). This is because the HIPAA³ (Benitez and Malin, 2010) privacy guidelines of the U.S. Department of Health and Human Services anonymize patient identifiers, which are applied without being aware of the risk of re-identification (Benitez and Malin, 2010). Sweeney (1996) was able to merge the publicly available, anonymized hospital discharge records with registration records (of identified voters) based on the common fields (date of birth, gender, zip code) and thereby re-identify the medical record of the Governor of Massachusetts. In this way, uncover the reason for a mysterious hospitalization (Benitez and Malin, 2010). As a result, it is not enough to rely on data anonymization. There is a need to protect health data through further protection measures such as geomasking methods. The importance of data protection for geodata is very important, some health information may be considered more sensitive than the other: A distinction must be made between influenza, where the risk of endangering privacy is low, and the opposite of the very sensitive data of HIV patients (Kounadi and Leitner, 2014).

According to Lang et al., due to the high potential for misuse, sensitive data contains a higher level of protection. Their processing is in the field of tension between technical possibilities and the legal data protection framework (Lang et al., 2019). As a result, any conclusions that could point to individual persons must be excluded (Lang et al., 2019). According to Art. 9 (1) GDPR, they constitute a special category of personal data due to their sensitivity (Lang et al., 2019). Since 25.5.2018, the EU Regulation of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data has been in force in Austria. This resulted in the repeal of Directive 95/46/EC (GDPR) and the Federal Act on the Protection of Natural Persons with regard to the Processing of Personal Data (FADP), Federal Law Gazette I No. 165/1999. As a result, any processing of personal data requires a lawful (Art. 9) basis (Lang et al., 2019). GDPR classifies pseudonymized, but not anonymized data, as personal data. Pseudonymization (Art. 4 No. 5 GDPR) refers to the processing of personal data, according to which the personal data can no longer be assigned to a specific data subject without the use of additional information (Lang et al., 2019). This is subjected to the Federal Act on Documentation in Health Care No. 100/2018 (Lang et al., 2019).

Anonymizing data can lead to a reduction in data usefulness (Kounadi and Polzin, 2021). If the personal reference has been completely deleted and the data has been anonymized, it is no longer subject to data protection law and can be processed without legal difficulties (Zeger et al., 2023). In the health geography research, it can be understood as recording the number of hospitalized patients rather than evaluating individual cases (Zeger et al., 2023). However, epidemiological studies differ from clinical, because they are observational studies that focus on risk factors and their distribution (Degelsegger-Márquez et al. 2022).

³ HIPAA Health Insurance Portability and Accountability Act von 1996

The report by Degelsegger-Marquez et al. on the current situation in Austria shows that the legal bases, as well as the qualitative and technical requirements for the use of health data databases, are not always met. There are different pseudonyms and area-specific personal identifiers, as well as unclear requirements regarding the anonymization of health data (Degelsegger-Márquez et al., 2022).

§ 2 Documentation and Information System for Analyses in the Health Care System (DIAG) applies. The pseudonym used is an industry-specific personal identifier of the health documentation (bPK GH-GD), which is assigned to the responsible Federal Minister of Health (§ 5a, § 6c Doku-G) (Degelsegger-Márquez et al. 2022).

The report also found that patient consent (Art. 9(2) GDPR) was the most common legal basis for the processing of personal data by healthcare institutions (Degelsegger-Márquez et al., 2022). Health data is receiving increased attention in political discourse (Degelsegger-Márquez et al., 2022). Their ease of use is mentioned, but there are also privacy concerns that point to the sensitivity of health data. However, the challenges of the Corona pandemic brought with them new perspectives on the processing of health data (Degelsegger-Márquez et.al., 2022).

2.2.1. Electronic Health Records-ELGA

According to the Federal Ministry of Social Affairs, Health, Care and Consumer Protection (BMSGPK), the health data of healthcare facilities throughout Austria are stored in electronic information systems of the e-health-card ELGA. As a result, this data is available across different institutions, regardless of time and place. The system enables complete traceability of all accesses to ELGA health data (BMSGPK). The ELGA healthcare facilities only have access to the data of the treating patients. Electronic processing of ELGA health data is prohibited by the Telematics Act No. 111/2012 (BMSGPK).

With ELGA, the level of data protection and data security of health data has been significantly increased. In total, 25 federal laws form the legal framework for the processing of health data in Austria (Zeger et al., 2023). This should improve health care (Zeger et al., 2023). However, there are also concerns that the ELGA concept is intended to facilitate the misuse of health data, to "legalize" the uses for marketing purposes (Zeger et al., 2023). The data is stored centrally and de-centrally, which results in insecure data processing. Access is always centrally via the Internet. The patient key can be used to open countless health databases and thus extract confidential data. This results in some vulnerabilities (Zeger et al., 2023). Due to the critical statements of the Austrian Society for Data Protection (ARGE), the project that the ELGA data should also be used for serious market research was ultimately stopped (Zeger et al., 2023).

According to the “Vienna Medical Association”, computer systems in hospitals are up to 30 years old, and training is required to enable the technical knowledge and availability of new-generation equipment. In addition, the vulnerabilities in internet-based services can have various causes. They can increase the risk of data theft or direct manipulation of devices. This is because criminal acts also affect the health sector: The darknet is a particularly well-sealed area of the network that serves as a starting point for such digital attacks. There, encryption, and anonymization are guaranteed, and the secret marketplaces can serve as platforms for the sale of stolen data (Vienna-Medical-Association, 2023). But in general, health data is highly sought after by criminals, and the outbreak of the Corona pandemic in particular has exacerbated cybercrime. The Vienna Medical Association offers a free online questionnaire, the so-called "Cyber Security Health Check"⁴, to check the security situation of the healthcare facility. A special warning comes from ransomware attacks, which have become a serious risk to the healthcare system (Vienna-Medical-Association, 2020)

With regard to data trading in the marketing sector, it should be mentioned that permission to store and share data on the Internet is hidden. Such privacy policies are often unknowingly accepted by the user (Louisa, 2023).

It can be assumed that there is a correlation between risks of spatial reidentification from ESDA analyses and the accessibility of health data. If it is possible to restore anonymized personal data by linking from other data sources, it is even more important to use geomasking methods so that it is not possible to determine the households' address coordinates and restore the patient data. A risk assessment must be carried out in advance to protect against possible re-identification risks (GDPR). For this reason, every controller should ask himself the question: What dangers arise in cases of spatial reidentification for the individual participants in the map? Who is interested in the target group data? It is necessary to be aware of what is possible with health data. What consequences can result from spatial reidentification to be able to assess the risk of spatial reidentification in advance and to be able to take the right countermeasures in good time. Especially by choosing suitable visualization techniques and geomasking methods. It can be assumed that the density of the study area has a high influence on the obfuscation method, especially for the vulnerable target groups. From the dangers of data trafficking and the black market to the protection of individuals' confidential data and the DNA database, there are countless reasons why health data needs to be protected. Accordingly, the term "sensitive" data is justified.

⁴ <https://www.aekwien.at/documents/263869/331939/190624-Cyberkriminalit%C3%A4t-Top-10-Sicherheitstipps-f%C3%BCr-%C3%84rzte+.pdf/ef2f1b27-8342-86a0-4e9f-4774152269ef?t=1575038514586>

2.3. Exploratory Spatial Data Analysis (ESDA)

Based on the literature of Delmelle and Kanaroglou, geographers, statisticians and health experts focus on the identification of spatial disease clusters and on the relationships between natural factors and the built environment, health, and individual well-being (Delmelle and Kanaroglou, 2015). Exploratory spatial data analysis (ESDA) is a set of techniques for describing and visualizing spatial distributions (Anselin, 1996). The concept of spatial autocorrelation applies, in which the similarity of the location goes hand in hand with the similarity of the value (Delmelle and Kanaroglou, 2015). Spatial autocorrelation consists of the geostatic perspective of the natural sciences and the lattice perspective of the social sciences (Anselin, 1996). Can be estimated using the Global Moran's I statistics (Moran 1950) to find out if neighboring data are interdependent. Local statistical calculations can be used to determine where clusters of similar observations (related of the individual location) are located (Delmelle and Kanaroglou, 2015). The Local Moran's I can be used to determine outliers in datasets (Anselin, 1996).

Many ESDA techniques have been developed only recently (Anselin, 1996), and can be used to identify spatial or spatiotemporal patterns, depending on the data (Delmelle and Kanaroglou, 2015). Geographic information systems and spatial analysis can be used to collect timely information about the occurrence of a disease at a location and at a specific time (Delmelle and Kanaroglou, 2015). It turns out that point pattern analysis is one of the most basic concepts of spatial analysis. Here, statistical methods are used to test spatial patterns to see if there is a significant difference to a random spatial pattern (O'Sullivan and Unwin, 2003). The Kernel Density Estimation⁵ (KDE) is used in the visualization and analysis of spatial data to identify the event patterns (Smith et al., 2015). For example, during dengue fever (2010) in Colombia, spatial-temporal visualization methods of KDE were used to detect outbreaks of infectious diseases (Delmelle and Kanaroglou, 2015). However, KDE can also help visualize accessibility in healthcare (Delmelle and Kanaroglou, 2015). There are several statistical methods for identifying the hotspots (Delmelle and Kanaroglou, 2015), all based on density. Predictions can be made by hotspot detection (Kounadi and Leitner, 2016). The exploratory methods, interpolation and clustering techniques are well suited to study the spatial pattern for infectious disease outbreaks (Delmelle and Kanaroglou, 2015).

According Kalinic and Krisp, the KDE- and Hotspot-Analysis (G_i^*) can be performed with point data, allowing for point pattern analysis (Kalinic and Krisp, 2018). Summary statistics (number, mean, median, standard deviation) can also provide valuable information about a pattern (Kalinic and Krisp, 2018). However, there is a difference between KDE and G_i^* Hotspot Analysis. Although they can give the same thing visually, "Kernel Density Estimation" (Silverman, 1986) and "Getis-Ord G_i^* statistics" (Getis and

⁵ KDE – Kernel Density Estimation

Ord, 1992) are different from each other (Kalinic and Krisp, 2018). The KDE can be very useful in detecting hotspots (Kalinic and Krisp, 2018). KDE defines clusters with high values within the data. However, KDE-values can also be grouped into low-density ranges and still be identified as hotspots of Getis-Ord G_i^* statistics. This depends on the bandwidth settings (Kalinic and Krisp, 2018). G_i^* Statistics can be used to identify statistically significant spatial clusters with high (hot spots) and low values (cold spots) (Kalinic and Krisp, 2018).

Cluster analyses are statistical methods that group cases into relatively coherent clusters. The hierarchical clustering dendrogram "inverted tree diagram" can be applied in any point distribution (Polzin, 2020). Spatial clustering is one of the most important epidemiological lessons in a pandemic (Zakaria et al., 2021). Mapping based on the combination of geographical and socio-economic characteristics can support the study of infectious disease spread (Zakaria et al., 2021). As a result, ESDA techniques, bar charts, box plots, scatter plots, as well as spatial-temporal representations, can detect a hidden pattern and anomalies in data (Vienken, 2021). The 2021 Zakaria study applied ESDA techniques to analyze the existence of spatial Covid-19 clusters to map the spread pattern of the COVID-19 pandemic in Malaysia (Zakaria et al., 2021). The spatial autocorrelation was calculated using Global Moran's I and Local Moran's Index (Zakaria et al., 2021). Moran's I was used to measure the local spatial association (LISA) (Zakaria 2021).

The Corona pandemic has shown that especially in the case of new disease outbreaks, it is important to examine the neighborhood areas, especially in high-risk areas (Zakaria et al., 2021). ESDA techniques can be used in health geography but also in criminology to analyze the spatial clustering of cases (Zakaria et al., 2021). In general, it can be seen that ESDA is more advantageous compared to exploratory data analysis EDA because it can detect hidden spatial patterns (Zakaria et al., 2021). Particularly Moran's, I scan statistics are good for exploring event data and detecting spatial clustering (ArcGIS(d), 2024).

2.3.1. Spatial Analysis Tools

Spatial analysis tools SatScan and CrimeStat have been developed to detect clusters of point events and perform simulations, as well as to assess the statistical significance of the clusters (Delmelle and Kanaroglou, 2015).

CrimeStat 3.3⁶ is a spatial statistics program for the investigation of crime events. Allows performing ESDA assessment analysis, such as nearest neighbor hierarchical cluster analysis (Kounadi and Polzin, 2021).

SaTScanTM software can be used to analyze spatial, temporal, and spatiotemporal data using the scan statistics (Kulldorff, 2024). It is suitable for conducting geographic surveillance of diseases to detect spatial or spatiotemporal disease clusters and determine whether they are statistically significant. SaTScan has been in continuous development for years (Harvard University). SaTScan can be used, for example, to test whether a disease is randomly distributed or not. Possible disease outbreaks can be detected at an early stage based on monitoring. SaTScanTM software⁷ can be used in many disciplines, including criminology or genetics (Kulldorff, 2024).

⁶ <https://nij.ojp.gov/topics/articles/crimestat-spatial-statistics-program-analysis-crime-incident-locations#program>

⁷ <https://www.satscan.org/download.html>

2.4. Risk of Spatial Reidentification

Geocoded address data can be displayed in the form of a point map, but depending on the size of the scale, the latitude and longitude of a particular point can be estimated. For this reason, they belong to sensitive information (Stinchcomb, 2004). This is because there are also software tools that can perform reverse geocoding and thus generate an approximate address based on latitude and longitude (Stinchcomb, 2004).

In health geography, point maps are used to represent the distribution of “discrete” patient addresses over an area. But this is exactly where the problem of spatial reidentification becomes apparent: If scientists want to include point distributions, they must consider the risk of re-identification, especially regarding population density and map scale (Kounadi and Leitner, 2014). It is possible to replace maps with point distributions with hot-spot maps to avoid possible scientific misconduct. However, not all spatial analyses are then possible. To detect inequalities in data sets, spatial distributions of diseases are studied with fine resolutions (Kounadi and Leitner, 2014).

Kounadi and Leitner analyzed the published articles with unmasked and masked confidential information for spatial re-identification (Kounadi and Leitner, 2014). They compared the map scale of the published MCI⁸ articles (16) and the UCI⁹ articles (41). They determined the map scale of the published articles with masked data (MCI-16) and the articles with unmasked data (UCI-41). They analyzed MCI articles on a scale of 1:42 to 1:29,000 with a population density of 21/km² -12.507/km². And UCI articles with a density of 21/km² -13.000/km² and with a map scale range of 1:56 to 1:153.846. Results, the UCI articles exposed over 68.000 addresses containing confidential information (Kounadi and Leitner, 2014). From this study, it can be seen that the larger the scale, the easier it is to re-identify a location. As well as that a low population density facilitates the re-identification process. At a map scale of 1:56 it is possible to identify households, when unmasked data, but also when masked data is shared (Kounadi and Leitner, 2014).

Leitner et al. investigated that even people with little knowledge of geospatial technologies can successfully re-identify the addresses of residential buildings (Leitner et., 2007). Brownstein et.al examined the risk of re-identification of two hypothetical maps. They examined a low-resolution and a high-resolution map with geocoded patient addresses (Brownstein et. al, 2006). To get the results, they scanned and rectified the maps. They then extracted the locations of the patients using supervised classification. It follows that reducing the resolution of the map does not provide sufficient protection against spatial re-identification (Kounadi and Leitner, 2014). It is therefore possible to identify a high proportion of addresses at a map scale of 1:100.000 (Kounadi and Leitner, 2014). Even in densely populated areas, location data

⁸ MCI Articles contain maps with masked confidential information

⁹ UCI Articles contain maps with unmasked confidential information

must be masked. Regardless of whether enough people live at one address. This is not sufficient to maintain the anonymity of people. Thus, the risk of spatial reidentification represents the probability of determining a person's whereabouts (Kounadi and Leitner, 2014).

According Kounadi and Leitner, the risk can be assessed by re-identifying locations from published maps that contain unmasked discrete locations and also by estimating the actual locations from masked data (Kounadi and Leitner, 2014). Elderly and single populations are considered easy targets of potential attacks, especially sick and weak people (Kounadi and Leitner, 2014). The risk of re-identification remains high, particularly in areas with low population density, but also in purely residential areas. The buildings and property types can be easily identified, especially if the address information in such an area is clearly displayed on a map (Kounadi and Leitner, 2014). Re-identifying privacy remains a risk. The protection of the rights of the individual should come first (Kounadi and Leitner, 2014). It is justified to judge that the collecting confidential data is a difficult process. Therefore, the data should be properly protected when visualizing and disseminating the results. This problem could be solved by plausible geomasking method (Kounadi and Leitner, 2014). If unmasked data is visualized, there is a high risk of the coordinates being reassigned to households, especially in lower density areas. If masked data is visualized but the geomasking method is not suitable for the study area and cannot ensure spatial anonymity, false reidentification or even re-identification of the person may occur. To be able to determine the entire risk of spatial reidentification, it is necessary to know sensitive factors of spatial reidentification, which should always be treated confidentially. However, this also includes the study area, the subject of the geomasking, the number of people studied, the map scale, the population density of the area, the method of site protection and the characteristics of the methodology (Kounadi and Leitner, 2014). The "Gaussian distortion" (Cassa et al., 2008), as a "re-identification method" can calculate the average (by calculating the spatial average) of the masked points of each release. For each origin point, the average coordinate value of the multiple shares, the masked points, and the origin point was recorded on a map. Consequently, the Euclidean distance of all points to the point of origin was calculated (Kounadi and Leitner, 2014).

As a result, the disclosure of masked metadata, especially information about the type of geographic mask and its characteristics, leads to a reduction in confidentiality: The combination of this information can increase the risk of spatial reidentification (Kounadi and Leitner, 2014).

In contrast to the risk that can arise from displaying confidential discrete point maps, spatial aggregation of point data to a (0.5km) grid ensures low risk of disclosing sensitive information (Curtis et al., 2011). Spatial aggregation is used to maintain confidentiality. This ensures that the subject is sufficiently hidden in a larger population and thus its identity is also protected (Allshouse et al. 2010). By means of spatial aggregation at the regional level, individual locations are combined into one regional

area (Kounadi and Leitner, 2014). An example would be the Choropleth map, in which the areas are shaded according to the measurement of a variable. In point aggregation, the transferred points are moved to a new position where there are actual observations or where they cannot exist, which can ultimately lead to false re-identification (Kounadi and Polzin, 2021).

It is clear from the literature of Kounadi and Leitner (2014) and Kounadi and Polzin (2021) that the assessment of the re-identification risk must be carried out prior to visualization release. The potential risk of reidentification can be in paper or digital form, knowingly or unknowingly. In the event of re-identification, this can lead to enormous consequences for the person concerned. The literature also shows that there are some ESDA methods for re-identifying individuals in practice (Kounadi and Polzin, 2021).

Based on Kim et al. study (2020) on the subjective perception of disclosure risk, it was found that the risk increased as location information and participant information increased, was influenced by the map scale. On the other hand, the risk decreased when the degree of aggregation and the distance of the displacement increased in the geomasking methods (Kim et al., 2020). It turned out that to prevent the risk of re-identification, the point-to-polygon aggregation methods were better than the point-to-point "relocation" methods. The perceived risk of disclosure of a map that depicted information on vulnerable groups was significantly higher (Kim et. al. 2020).

2.5. Geomasking Methods

According to Kounadi and Leitner, the crime or health-related information of individuals should be treated as confidential (Kounadi and Leitner, 2014). In general, a distinction is made between confidential topics with private, criminal or health information. In the case of health information, the points on the maps represent the patients addresses, while in the case of criminal topics, the points represent criminal incidents. Regarding location protection: The sensitive, confidential information in the form of points in the point maps is at risk, as is the information about individual mobility patterns in the trajectory maps of individuals (Kounadi and Leitner, 2014). Given the risks associated with mapping sensitive data, methods have been developed to protect the location data and enable valid spatial analysis (Kounadi and Leitner, 2014). Geomasking is therefore used on the one hand to ensure data protection for individual address information and at the same time to maintain spatial resolution for mapping purposes. Geomasking refers to methods in which a person's geographic location is changed to protect that confidentiality. At the same time, attempts are being made to preserve the spatial relationship between geomasked data and the original point data (Kounadi and Leitner, 2014). The history of geomasking is marked by Armstrong et al. (1999), who produced a comprehensive summary of various traditional masking techniques that has been mentioned in many scientific literatures. These include affine transformation, random perturbations, point aggregation, and spatial aggregation (Allshouse et al., 2010). The affine transform shifts original positions by a fixed increment (translation), by a scaling constant (scaling), or by a fixed angle by rotating a pivot point (rotation). A random perturbation shifts positions around a random increment and in a random direction (Kounadi and Leitner, 2014). Random interference methods (donut method) use population density to determine the distance a person must travel to achieve a certain level of privacy (Allshouse et al., 2010). In contrast, the point aggregation method replaces a series of points on the map that are geographically similar with compound coordinates (Armstrong et al., 1999). Spatial aggregation into large areas impairs the ability to recognize clusters and other spatial relationships (Charleux and Schofield, 2020). In the research of Kounadi and Leitner, random perturbation was the most used geomask that had the least effect on spatial analysis and point pattern (Kounadi and Leitner, 2014).

Geographers are constantly working to develop suitable site protection methods for discrete point maps (Kounadi and Leitner, 2014). It can be assumed that each geomasking method has its advantages and disadvantages, but since 1999 to the present day (2024), no geomasking technique recommended as a "standard" in health geography or has been approved by GDPR (EU) to protect the location-based data of individuals. The HIPAA guidelines (USA) concern the anonymization of geodata (Benitez and Malin, 2010). Not the obfuscation techniques of geomasking methods.

2.5.1. Spatial K-Anonymity

If data is available at the address level, the protection of the privacy of the address data and the spatial resolution of the map should be considered first (Allshouse et al., 2010). Non-spatial aggregation methods, such as age categories, are used to prevent individuals from being identified in the dataset (Allshouse et al., 2010). How well the protection method works is defined by the degree of masking. The privacy provided by the geomask can be determined by the concept of K-Anonymity as a measure of success (Allshouse et al., 2010). The greater the K-Anonymity, the better privacy is protected, and the more errors are introduced into the point pattern (Leitner and Curti, 2006). K-Anonymity is defined as the number of cases from which an anonymized case cannot be distinguished (Allshouse et al., 2010). Represented as the number of households within a circle whose point is the original household location and whose radius is the Euclidean distance between the original and masked location (Allshouse et al., 2010). Spatial K-Anonymity can be chosen as a measure to protect privacy. To be able to calculate this, the original and the masked location must be known (Allshouse et al., 2010).

In terms of geospatial data, K-anonymity ensures that any place such as “household, building, address, location” of individuals cannot be distinguished from K-1 locations (Kounadi and Polzin, 2021). Spatial K-Anonymity SKA describes the probability of identifying a location that can be assigned to a person by reverse geocoding, this is necessary to assess the level of privacy and measure the degree of change (Kounadi and Polzin, 2021).

2.5.2. Adaptive Geomasking

To visualize sensitive data in publications, scientists prefer geographic masking techniques called isomasks (Kounadi and Leitner, 2014). The geographic “Isomasks” ensure the protection of privacy by shifting original points within the areas of uncertainty created by the masks (Kounadi and Leitner, 2016). Adaptive geomasks adjust for the displacement error by considering the underlying population density (Kounadi and Leitner, 2016). This allows them to define a desired level of K-anonymity, where each record within a masked dataset cannot be identified by at least K-1 datasets (Kounadi and Leitner, 2016). The study area Vienna is characterized by its different population density, 0-1416 addresses per km² (Kounadi and Leitner, 2016). Seems like a good example of heterogeneous areas. Although donut masking (DM) has the best retention area pattern, it is not suitable for heterogeneous study areas: The random perturbations algorithms are based on the assumption of homogeneously distributed populations to calculate the displacement distances (Kounadi and Leitner, 2016). The use of DM in heterogeneous areas could lead to insufficient protection of individual

address data (Kounadi and Leitner, 2016). According to the literature, the methods of adaptive areal elimination (AAE), adaptive surface masking (AAM) and adaptive voronoi masking (AVM) belong to adaptive geomasking methods (isomasks) and are preferred in scientific papers. The application of these isomasks has already been investigated in scientific papers, which will be described in more detail in the next step to be able to point out the strengths and weaknesses of these methods and to be able to decide on suitable geomasking methods. These geomasks are among the latest geomasking methods developed. They can¹⁰ be obtained from GitHub. Below you will find sufficient information regarding the isomasks, as well as the AAE and AAM toolbox (ArcGIS) and the AVM Python script.

It is recommended to consider data usefulness and confidentiality protection when selecting the appropriate geomasking method (Kounadi and Polzin, 2021). For spatial analyses in health geography, it is important that the geomasking method maintains the spatial point pattern and at the same time sufficiently protects the discrete point data from renewed spatial reidentification (Kounadi and Polzin, 2021). Subsequently, the AAE, AAM and AVM methods were examined for these properties according to the literature.

2.5.2.1. Adaptive Areal Elimination (AAE)

According Kounadi and Leitner, in adaptive geographic masking, the level of masking is adjusted based on the density of existing risk information for re-identification (RORI)¹¹(Kounadi and Leitner, 2016). The following applies to the fact that a lower density requires a higher degree of masking. It can be assumed that the finer the resolution of the RoRi file, the more k-anonymized areas are generated. This also reduces the overall spatial error of the masking process (Kounadi and Leitner, 2016). Defining the RoRi file (Risk of Reidentification) is necessary, the method does not work without RoRi (Kounadi and Leitner 2016). The principle of the AAE method is to create areal with a minimum of K-anonymity. In the last step, there are options to aggregate the discrete points to center of the resolved polygons (adaptive point aggregation APA) or to randomly shift the discrete points within the resolved polygons - adaptive random perturbation ARP (Kounadi and Leitner, 2016). Because each masked point is supposed to lie within its K-anonymized polygon, the displaced masked points must be "conditioned" at the boundary of each polygon (Kounadi and Leitner, 2016). One advantage of this method is that the APA and ARP k-anonymized areas can be safely exposed without increasing the risk of spatial re-identification (Kounadi and Leitner, 2016). The effectiveness of the

¹⁰ <https://kithub.com/okunadi/kioprivasi>

¹¹ RoRi corresponds to K-Anonymity, it can be the number of people, households, residential addresses, and others (Kounadi and Leitner, 2016).

AAE method has been investigated in several scientific studies. In the study by Kounadi and Leitner (2016), the geomasking methods were examined using four statistical techniques: The spatial error and the shifted distance (in meters) were analyzed. The spatial error was determined using pattern retention. In addition, the Pearson correlation coefficient of the density areas, as well as the divergence of the hotspots and the specificity of the clusters (in percentage) were calculated (Kounadi and Leitner, 2016). Most recently, the ARP method of random perturbations was recommended, which has fewer spatial errors than point aggregation (Kounadi and Leitner, 2016). According to the study, the AAE method is suitable for geomasking confidential health or crime data (Kounadi and Leitner 2016), because the AAE method allows the disclosure of masked discrete points. The discrete locations in K cases cannot be identified in a successful reengineering process (Kounadi and Leitner, 2016). Provided that the disclosure information is available to RoRi and that it is accurate (Kounadi and Leitner, 2016).

However, the study by Charleux and Schofield (2020) showed concerns about the AAE fusion process (Charleux and Schofield, 2020). This was done due to the excessive loss of information, the aggregation process, and the processing time. On the one hand, because some points are shifted more than necessary, and on the other hand, because all polygons in the study area are part of the aggregation process, even if they do not contain points to be anonymized (Charleux and Schofield, 2020). This could be a problem at the state level (Charleux and Schofield, 2020). This is where the adaptive areal masking (AAM) method would be preferable. After all, the AAE method can be seen as a solution to the geomasks, which cannot guarantee the expected level of K-anonymity (Charleux and Schofield, 2020). To ensure privacy, masked points must be moved further into sparsely populated areas. The adaptive methods are based on the concept of K-Anonymity to be able to determine the local radius of the shift (Charleux and Schofield, 2020). The points should be moved within the neighborhoods large enough to achieve K-Anonymity, but the shift should be minimized (Charleux and Schofield, 2020). In the AAE method, the point displacement became larger than necessary, resulting in a loss of information in the discrete point dataset (Charleux and Schofield, 2020).

The study by Kounadi and Polzin (2021) also showed concerns about this method. Again, the problem with moving the masked data over longer distances became apparent. This was done for reasons of merging polygons, where the neighbor with the longest boundary was selected. As a result, the K-anonymized areas were larger than necessary to still be able to guarantee spatial K-anonymity (SKA) (Kounadi and Polzin 2021). It turned out to be advantageous that it is possible to guarantee K-anonymity even if the geomasking method and its parameters are known (Kounadi and Polzin, 2021). In summary, however, there are more disadvantages than advantages for the AAE method.

2.5.2.2. Adaptive Areal Masking (AAM)

According Charleux and Schofield, the AAM method is a modified version of AAE and is based on the principle of SKA (rori). This means that a certain Rori value must be present in a polygon for the masked points to have sufficient SKA protection. The AAM ensures a true level of spatial k-anonymity (Charleux and Schofield, 2020). The AAM method generates individual anonymization polygons for each point, which can overlap each other (Charleux and Schofield, 2020), also known as the interactive aggregation process of polygons. For each point, an anonymization polygon is generated, which is aggregated with the population polygon whose center of gravity is closest to the Euclidean distance (Charleux and Schofield, 2020). This is repeated until all points are masked. The polygons that do not have a simulated point are not aggregated, and the aggregation process is terminated when the K-Anonymity threshold is reached. The smaller polygons are joined first. (Charleux and Schofield, 2020). In the last step, the original data is randomly shifted within the merged polygons: the point coordinates are shifted by a random distance in a random direction (Kounadi and Polzin, 2021). Therefore, anonymization polygons generated by the AAM process must not be shared with end users. Unlike AAE. This also affects the characteristics of the population and area, as this information increases the risk of re-identification. Especially if population density data is also publicly available (Charleux and Schofield, 2020).

The study by Charleux and Schofield (2020) evaluated the effectiveness of the AAE-ARP and AAM method using statistical measurements with data sets of different sizes. Geographic performance was evaluated based on displacement distance, density estimation, and spatial accessibility (Charleux and Schofield, 2020). Three methods were used, the AAE method of aggregating original polygons to anonymization polygons, followed by the ARP method of randomly shifting the points within the anonymization polygons, and finally the AAM method (Charleux and Schofield, 2020).

At the same level of K-anonymity, AAM provides higher correlation than AAE-ARP (Charleux and Schofield, 2020). Overall, the AAM method can be better seen as the AAE method (Charleux and Schofield, 2020). Especially in terms of computing power (ease of use) and geographical performance (Charleux and Schofield, 2020). The AAM preserves the information better than AAE (Charleux and Schofield, 2020). Both AAE and AAM method shift the points, leading to "local uncertainty" (Charleux and Schofield, 2020). During the anonymization algorithm, AAE-ARP method results in larger average point shifts than AAM method (Charleux and Schofield, 2020). It is due to the differences in population density, rather than the number of polygons, in each area (Charleux and Schofield, 2020). However, the AAM method is preferable for accessibility reasons (Charleux and Schofield, 2020).

2.5.2.3. Adaptive Voronoi Masking (AVM)

Based on Literature Kounadi and Polzin, the Adaptive Voronoi masking is based on the concepts of the AAE method and Voronoi masking. By defining K-anonymity areas and creating Voronoi polygons, the underlying population density is taken into account by stitching polygons such as AAE together and replacing the original data based on the Voronoi masking concept (Kounadi and Polzin, 2021). Unlike other geomasks, AVM's method takes into account the underlying topography and moves the data points to the nearest road intersections where there are more buildings in the area (Kounadi and Polzin, 2021). This makes it impossible to move points directly to another place of residence, thus preventing false re-identification (Kounadi and Polzin, 2021). Thus, it also prevents displacement to places such as bodies of water or forests (Kounadi and Polzin, 2021), thus minimizing the risk of false re-identification (Kounadi and Polzin, 2021).

The AVM study by Polzin and Kounadi investigated the effectiveness of AVM, VM, AAE and donut masking methods. These methods were compared in terms of spatial K-anonymity (SKA), false reidentification, and data utility (Kounadi and Polzin, 2021). ESDA methods were applied to both the original and masked data to investigate the effects of the geomasks on the original data to detect the differences in the loss of spatial information and the preservation of the original data pattern. ESDA-techniques applied: including Point patterns, central tendency, Ripley-K function, next neighbor hierarchical cluster analysis (Kounadi and Polzin, 2021).

Visualizing the point pattern made it clear whether the masked data was moved to other residential locations or empty locations (lake), whether this increased the risk of false re-identification (Kounadi and Polzin, 2021). Based on the central tendency, the displacement distances between the masked and the original center and median centers, it became apparent that the lower the density of the points, the more they vary (Kounadi and Polzin, 2021). The euclidean distance (between original and unmasked points) has been used to indicate the distance error (Kounadi, 2024). Using the "Ripley function", it was shown that the VM and AVM methods produced a more clustered pattern than the original data, while the opposite was true for the AAE and DM methods (Kounadi and Polzin, 2021). Using hierarchical cluster analysis with the nearest neighbor, it was possible to examine the clustering pattern (Everett, 1974) of the discrete data in terms of number of clusters, size, orientation, and density and compare it with the pattern of masked data (Kounadi and Polzin, 2021). According to the results, the AVM method had the closest average points per cluster and the most similar cluster density (Kounadi and Polzin, 2021). Based on the results of the statistical analyses, the AVM and AAE methods obtain the spatial K-anonymity SKA, but the AVM is the only one that does not move points to illogical locations, thus significantly minimizing the risk of false re-identification (Kounadi and Polzin, 2021). The AVM is the only method that prevents false re-identification (Kounadi and Polzin, 2021).

The VM method is efficient in maintaining the spatial point pattern of the original data (Kounadi and Polzin, 2021). But the VM process does not guarantee that the parts of the Voronoi polygons will not cross residential devices, so VM method cannot reduce the risk of false reidentification (Kounadi and Polzin, 2021). The DM method is a popular technique, maintains the masked data pattern that is most similar to the original data, followed by AVM method, while AAE method distorts the pattern the most (Kounadi and Polzin, 2021). DM method provides best data utility, but only partially preserves a certain degree of spatial K-anonymity as it is assumed that the population is homogeneously distributed (Kounadi and Polzin, 2021). According to Kounadi and Polzin (2021), AVM method also AAE method retain a certain degree of spatial K-anonymity SKA (Kounadi and Polzin, 2021).

According to the study results, the AVM method is an advantageous technique for masking geospatial data (Kounadi and Polzin, 2021). Based on ensuring spatial K-anonymity, reducing the risk of false re-identification, and maintaining the original point pattern (data utility), AVM is the best way to protect location privacy (Kounadi and Polzin, 2021).

2.6. Selected Geomasking-Methods and ESDA

After an extensive review of the literature, the selection of suitable geomasking methods as well as ESDA techniques was made. It was important to choose methods and techniques that can be used in spatial analysis in the health geography.

2.6.1. Adaptive Geomasking Methods

In the previous subchapter, three geomasking methods were described in detail. It can be seen that the geoprivacy studies apply the ESDA methods on original and masked data to investigate the performance of the geomasks (Polzin, 2020). The point shifts of the unmasked and masked data of AAE, AAM, and AVM methods were investigated for spatial pattern retention in the scientific studies of Kounadi and Leitner (AAE study), Charleux and Schofield (AAM study), Kounadi and Polzin (AVM study). Here it became apparent that AAM is a better choice compared to AAE (ARP) method, as well as that AVM is also better compared to AAE method. As a result, the further application of the AAE method was ruled out. It was assumed that the AVMs with Voronoi's adaptive masking, and the adaptive areal masking AAM would also be suitable, as both methods are based on spatial K-anonymity (SKA) (Kounadi and Polzin, 2021). As a result, the research question regarding the choice of the appropriate geomasking method can already be answered here based on the scientific studies of Kounadi and Leitner, as well as Kounadi and Polzin and the results that evaluate the AVM method. The focus of

this scientific work has now been limited to AAM and AVM methods. But the AAE anonymization polygons were needed to perform the AVM method.

In this study, it is not expected that the AAM will be able to outperform the AVM method. It is incumbent upon the AAM and AVM method to investigate both the right and the false re-identification using ESDA techniques at the risk of spatial re-identification. To date, no comparative study has been conducted between the AAM and the AVM method. In the literature, there are different ways to analyze the effectiveness of geomasks. Because this involves simulated health data, it was necessary to examine the chosen geomasking methods using ESDA techniques for the spatial risk of re-identification and for the spatial retention of the point pattern. From this perspective, it was important to select appropriate ESDA techniques.

2.6.2. Spatial Visualization-Techniques

According to the literature, geocoded, disaggregated data in point maps can only be aggregated or geomasked at a certain level in scientific journals for privacy reasons (Delmelle and Kanaroglou, 2015). For this reason, it was necessary to investigate the risks of spatial reidentification from original (simulated) and masked points. It can be assumed that an ESDA output of masked data does not pose a risk of re-identification compared to the same output of unmasked data (Kounadi and Leitner, 2014). Point pattern visualization can be applied to compare the scope of confidential data points with the geo-masked ones. But also, to see if the geomasked data points have been moved to other places of residence or to the void (Polzin, 2020). The authors, Leitner and Curtis (2006), as well as Kwan et al. (2004), analyzed the visualization of the point pattern (Kounadi and Leitner, 2016). Therefore, this method was chosen here. To be able to visually examine the changes in the point pattern to get a better understanding of the point pattern density. Likewise, to be able to visually investigate the risk of spatial reidentification from orthophotos (Kounadi and Leitner, 2016). As already mentioned in the literature of Kounadi and Leitner, it is particularly important to take into account the way illustrations are presented in scientific papers. The study by Kounadi and Leitner showed that confidential information in UCI articles can be disclosed in a zoom view of 1:56 (Kounadi and Leitner, 2014). Consequently, the visualization of the point patterns was chosen in order to be able to investigate the displacement distance of the original (simulated) and masked points and the displacement technique in more detail. This was done to gain a better understanding of the geomask methods used. To recognize what risks of spatial reidentification (x, y coordinates) could emanate from the masked point locations. The Choropleth maps were selected to visually determine the differences in the number of points per registration district. The Choropleth maps, also called ratio maps, can be used to visualize values over an area, consequently, to examine for pattern changes (Severino, 2024). There is a difference between the

density of the examination area and the density of the point pattern. The density in the registration district is determined based on the points per area. The results of the point pattern density represents the spatial distribution of points in the study area (ArcGIS(b), 2024).

2.6.3. Hotspot Detection on Kernel Density Estimation

Kernel Density Estimation can be used to study the distribution of points (distribution of cases). It was chosen to compare the point pattern density of the original (simulated) and masked dataset. As well as to investigate the suitability of the masked datasets for spatial analyses in health geography. Estimating kernel density can determine areas of high density, but not determine whether they are statistically significant (Kalinic and Krisp, 2018). The combination of Kernel Density Estimation (Silverman, 1986) and G_i^* Hotspot-Analysis (Getis and Ord, 1992) is intended to increase the efficiency of interpreting the results (Kalinic and Krisp, 2018).

Local statistics G_i^* enable the identification of spatial association areas in health geography (Kalinic and Krisp, 2018). The GI Statistics Global Indicators of Spatial Association measure the spatial pattern of the data. The Getis-Ord G_i^* statistics (Getis and Ord, 1996) were selected to statistically identify significant hotspots and cold spots in the three datasets (Kalinic and Krisp, 2018). It determines which areas have above-average high (or low) values and whether the clusters are significant. According to ArcGIS, the Optimized version of the Hotspot-Analysis (Getis-Ord G_i^* statistics) is suitable for studying the point datasets (ArcGIS(a), 2024). Also to describe the risk of spatial reidentification of the (ESDA) results.

2.6.4. Spatial Autocorrelation Analysis (Moran's I)

Spatial autocorrelation can be measured based on point positions and attribute values using Moran's I statistics (ArcGIS(d), 2024). Spatial Autocorrelation (Global Moran's I) provides an overall view of the spatial pattern and returns a value for the entire dataset (ArcGIS(d), 2024). Using Global Moran's I, the spatial autocorrelation of the points can be determined. This method is suitable for pattern recognition and is used in health geography (ArcGIS(f), 2020). The Local Moran's Index (Anselin), on the other hand, is well suited to examine the clusters and patterns at the local level in datasets (Delmelle and Kanaroglou 2015). Therefore, each position is given an I-value, a variance z-value, the expected I, and the variance of I (ArcGIS(f), 2020). This method was chosen to analyze clusters and outliers in the study area on a City Vienna level.

To be able to assess possible risks of spatial re-identification based on anomalies. Compared to the G_i^* statistics, the Local Moran's I differs. The G_i^* looks at the distribution of points over the entire data set. The Local Moran's Index measures local spatial autocorrelation based on selected neighborhood parameters (ArcGIS(e), 2024).

To measure the average distance to the nearest neighbor of the points in the original (simulated) dataset and masked datasets, the Average Nearest Neighbor method was chosen (ArcGIS(e), 2024). This allows the spatial distribution of points in an area to be characterized in point patterns (ArcGIS(f), 2020). The metric measures the Euclidean distance between a point and its nearest neighbor and calculates an average value, this can also determine whether the datasets are clustered or randomly distributed (ArcGIS(c), 2024).

2.7. Software Overview

Lime Survey Free Version¹² was used to create the online survey. The Point data was simulated in QGIS 3.28.4.¹³

The geomasking methods are done in ArcGIS Pro¹⁴ using ArcGIS Toolbox and Python Script. The Python script was processed in the word processor Notepad ++ .

Also used PyCharm¹⁵ software.

The ESDA-techniques have been performed in ArcGIS Pro (3.0).

The diagrams (LimeSurvey) and the tables (spatial analysis) were created in Microsoft Excel.

¹² <https://www.limesurvey.org/de>

¹³ <https://www.qgis.org/de/site/>

¹⁴ <https://www.arcgis.com/index.html>

¹⁵ <https://www.jetbrains.com/de-de/pycharm/>

3. Methodology

This chapter describes the methodological part of the thesis and the justification of the methods used. First, the discrete health data was simulated¹⁶. During this time, an online survey was conducted in healthcare facilities in Vienna. After that, the AAM and AVM method were performed. In addition, it was necessary to create AAE polygons. Finally, ESDA techniques and spatial statistics were performed.

3.1. Study Area of Vienna

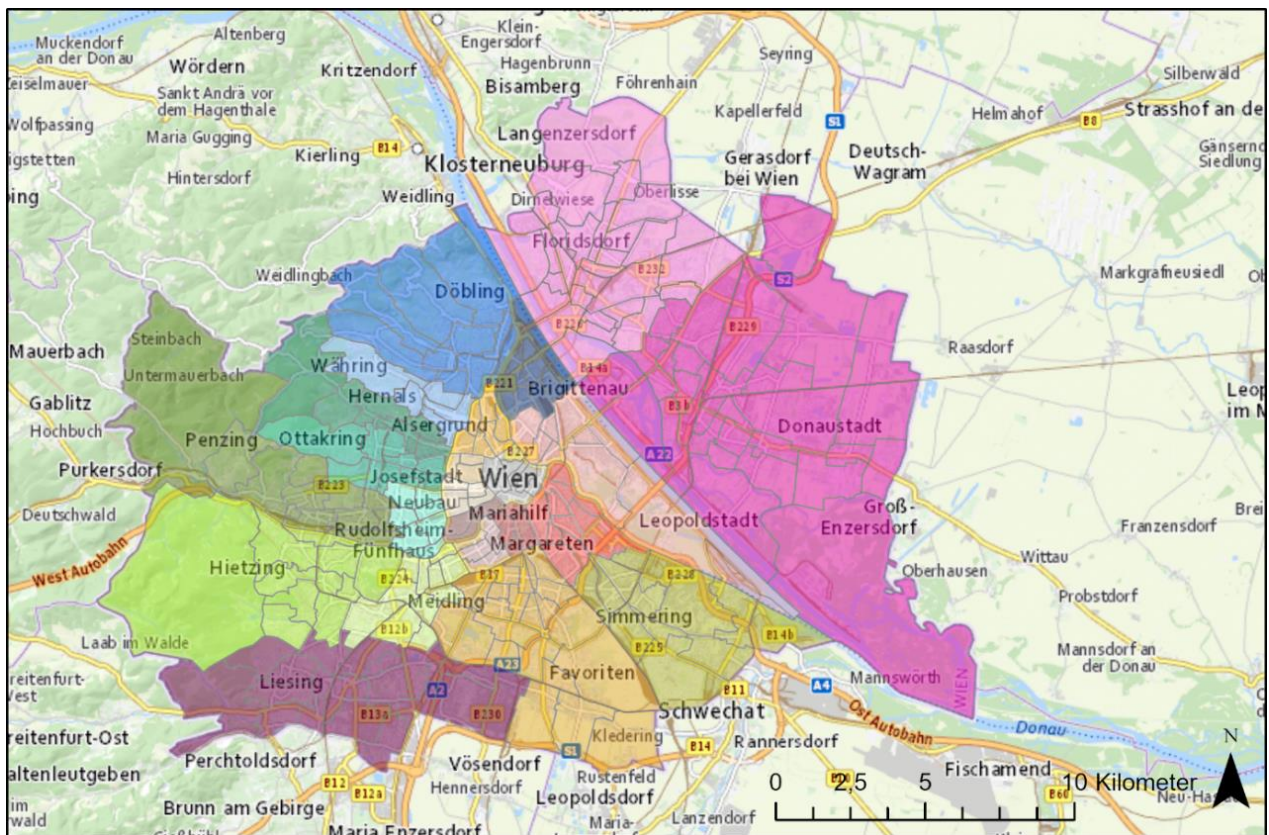


Figure 1 Study Area Vienna ©ArcGIS Pro

The study area is Vienna, the capital of Austria. Vienna comprises 23 districts and has a population of 2.006.134 (as of 2024). In 2012, this figure was 1.717 million inhabitants (Statistic-Austria, 2024).

The building structure in Vienna's districts is characterized by the residential buildings of the wilhelminian period. Until the middle of the 19th century, the buildings were erected along the main transport lines. In the wilhelminian period, due to the increase in the Viennese population, the parcelling of green spaces began. Thus, the agricultural

¹⁶ In Chapters 3. and 4., the simulated data are referred to as "simulated points"

areas were also developed according to the typical grid construction (City of Vienna, 2020). Vienna City Geodetic-Reference-Point WGS84: Length: 16° 22' 22.49" East, Width: 48° 13' 6.75" North (Federal Office of Metrology and Surveying, 2024). Study area Vienna in ArcGIS Pro, in decimal degrees:

S 48,1184° N 16,4095° O O 48,1509° N 16,5787° O

N 48,3232° N 16,4198° O W 48,1716° N 16,1830° O

For reasons of immigration from other EU countries and third countries, Vienna is the fastest growing city compared to the largest EU cities (City of Vienna, 2020). The highest values (as of 2023) are 25.000 inhabitants per km². This is a considerable difference compared to the data from 2012 (Statistic-Austria, 2024). To get a better overview of the spatial situation, the population data were visualized according to the density of the respective registration districts. A Vienna district consists of several registration districts.

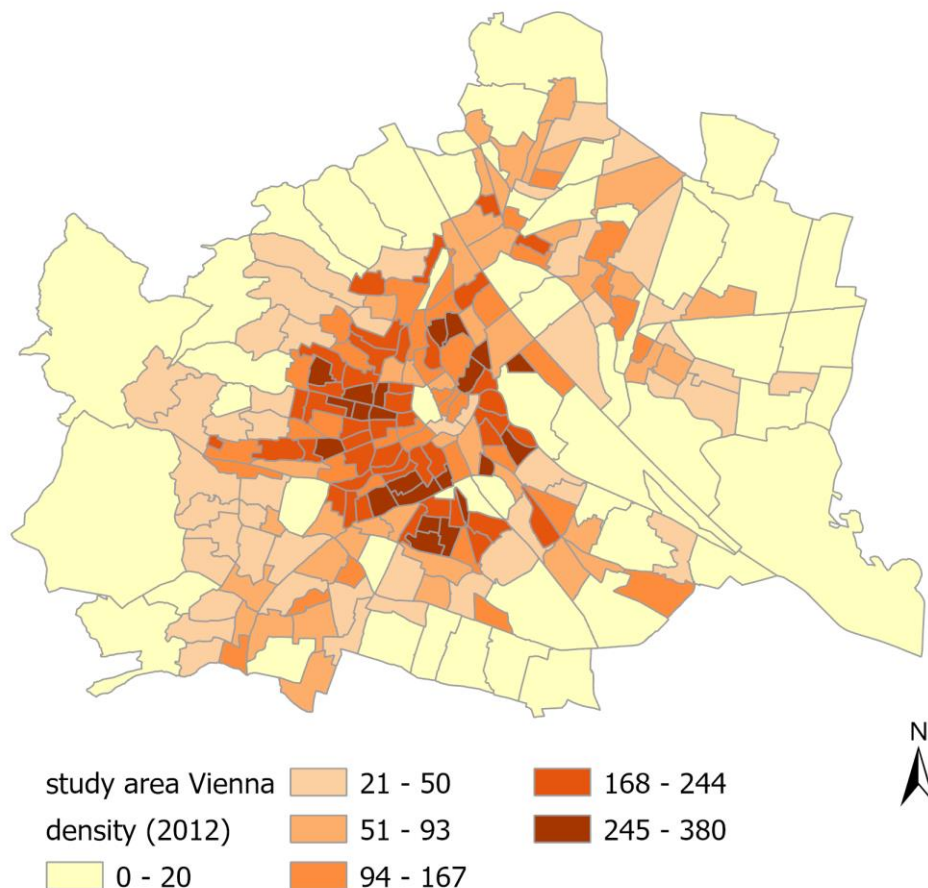


Figure 2 Population Density 2012 ©ArcGIS Pro

Based on the density of population data from 2012, it can be deduced that the areas with the highest population density are in “Wilhelminian” areas outside the belt and in

2 and 20 districts. Significantly less densely populated urban areas, on the other hand, lie outside the Vienna Wilhelminian period (City of Vienna, 2020).

3.2. Online-Survey on GIS-Analysis in Health Facilities

The aim of the survey¹⁷ was to find out to what extent spatial GIS analyses are carried out in Viennese health care facilities, how strongly “medical and geoinformation research” are linked, where possible obstacles arise, and which health facilities carry out spatial analyses. What knowledge do the doctors, private clinics or hospitals have with regard to spatial analyses in Vienna/Vienna surroundings? Originally, it was planned to actively conduct the survey from 1.9.2023 to 15.9.2023, but due to a lack of feedback, this period was extended. The results were evaluated at the end of October 2023.

On the basis of the online survey, it should be possible to understand the extent to which spatial geoinformation analyses are carried out in the Viennese healthcare system, whether public information is missing and what role is attributed to data protection on the part of healthcare institutions. In general, it should be clear to what extent “medical and geoinformation research” complement each other. The list of participants has been anonymized for reasons of data protection, but it can be summarized that practice doctors, private clinics, and a Viennese hospital, as well as rescue houses were surveyed. The target group mainly concerned Viennese health care facilities, as geodata was simulated in Vienna. The survey participants were searched via the Google search engine and the contact was made in several steps.

On 17.7.2023, all participants in the survey were contacted by telephone and the responsible data protection officer office was queried with e-mail contact details. A small problem that arose when contacting by phone was the GIS abbreviation for Geographic Information Systems, which is similar to the *ORF-GIS* fees, which unfortunately did not give a good overall picture. This is because some of the participants (without prior geographical knowledge) did not understand the relationship between ORF-GIS fees and spatial analysis. After all, healthcare professionals have subject-specific medical-expertise, not geographical expertise. After that, the request was sent to the email address of the survey participants (healthcare facilities).

It was already evident that some private clinics were not interested in participating, but also that the management of the Vienna Hospital confirmed their participation in the survey. Also, the ambulance houses, as well as the doctors were very nice. The survey was conducted through the free version, so the survey was limited to a maximum of 25 possible responses. However, since only 12 responses were returned, this software

¹⁷ <https://raeumliche-gis-analysen-in-gesundheitseinrichtungen.limesurvey.net/715718?lang=de-easy>

limitation had no effect on the survey. Regarding the invitation to the survey, hardly any responses have come back. Finally, the survey link was sent to all participants from the participant list. Subsequently, on 1.9.2023, emails were sent to all participants in the survey with a survey link, a copy was sent to the master's thesis supervisor Ass.-Prof. Dr. Ourania Kounadi, BSc MSc "in CC". A total of eight private clinics were contacted, one Viennese hospital, three ambulances (from Vienna) and also two general practitioners (from Vienna/ Vienna surroundings). The Vienna Regional Health Insurance Fund was invited to take part in the survey, but there was no response.

3.3.1. An Excerpt of the e-mail Correspondence

Including renewal e-mail dated 15.10.2023 and invitation link dated 1.9.2023:

Dear Ladies and Gentlemen.

I would like to ask you again to carry out the short online survey on spatial GIS analysis in the health sector.

Unfortunately, I have not received a single response to date. For this reason, the survey is still activated until 30.9.2023.

The link to the survey would be: <https://raeumliche-gis-analysen-in-gesundheitseinrichtungen.limesurvey.net/715718?lang=de-easy>

Thank you in advance for your time and help.

Sincerely

----- Original message -----

Subject: Survey on Spatial GIS Analysis in Healthcare Facilities

Date: 01.09.2023 12:39

Dear Ladies and Gentlemen.

I am writing a master's thesis at the University of Vienna, on spatial GIS analyses in the health sector: Investigating the risk of spatial re-identification of health geodata from ESDA outputs of confidential versus masked geodata.- Supervisor: Ass.-Prof. Dr. Ourania Kounadi, BSc MSc.

<https://geographie.univie.ac.at/arbeitsgruppen/kartographie-und-geoinformation/publikationen/masterarbeiten/>

As part of my master's thesis, I would like to conduct a short, qualitative survey. I would like to ask you to take part in this survey. In total, about ten health care facilities in Vienna are surveyed.

Link to the survey would be: <https://raeumliche-gis-analysen-in-gesundheitseinrichtungen.limesurvey.net/715718?lang=de-easy>

The survey takes about 10 minutes to complete and these are open-ended, not really complicated questions. Your answers are anonymous and will only be considered for the evaluation of the survey and will then be deleted. The poll is active until 15 September 12:00.

I do not want to receive any data or information about patients from you. I just want to collect the information to what extent you, as a healthcare institution, are aware of the spatial (GIS) analysis and geomasking methods, whether you have already done them. The aim is to find out how far this research has progressed in Austria, whether there are certain obstacles here, and what role is attributed to the data protection of individuals (patients).

This information would help me a lot to get an up-to-date overview of this field. In the following, I would also like to summarize my master's thesis results in an online StoryMap¹⁸ in order to give the public a better overview of the latest GIS visualization methods and their re-identification risks in the health sector. I would also like to send you the link to this online StoryMap.

Thank you in advance for your help.

Sincerely

¹⁸ Too few answers came back. As a result, the Story Map was not created. Also, it is assumed that the thesis will be available online.

3.3.2. The Questions of the Survey on spatial GIS-Analysis

01: What is your health sector? Please provide a brief description in the box below.

02: Have you heard of spatial GIS (Geographic Information System) analysis? Please choose an answer.

03A: If yes, please briefly explain where your spatial (GIS) information comes from in the box below.

03B: If not, would you like to know more about spatial GIS analysis? Explain briefly in the box below.

04: Have you heard of the geomasking methods? Please choose an answer.

05A: If yes, please briefly explain where your information about the geomasking methods comes from in the box below.

05B: If not, would you like to know more information about the geomasking methods? Please briefly explain in the box below.

06: It's assumed that as a healthcare facility, you have location-based patient data, what happens to that data? Please briefly explain in the box below.

07: Are there any problems with the storage and maintenance of patient data? Please briefly explain in the box below.

08: Have you already performed spatial GIS analyses with patient data? Please choose an answer.

Answer YES: 09A: What spatial GIS analyses did you perform? Please briefly explain in the box below.

09B: Why did you perform spatial GIS analyses? Please briefly explain in the box below.

09C: What problems have been identified, what do you think is necessary to be able to perform spatial GIS analyses? Please briefly explain in the box below.

09D: Would you recommend spatial GIS analysis to others? Please give a brief explanation of your answer.

Answer NO: 09E: Why didn't you do any spatial (GIS) analysis, and what do you think are the obstacles? Please briefly explain in the box below.

09F: Do you think further measures are needed to be able to perform spatial GIS analyses with patient data? (For example, data anonymization, etc.). Please explain in the box below.

09G: Would you want to do spatial GIS analysis? Please give a brief explanation of your answer.

10: Do you think there is enough public information available on the topic of spatial GIS analyses and visualization methods with patient data in Austria? Please give a brief explanation of your answer.

11: In your opinion, is geoinformation and medical research in Austria possible without restrictions? Please give a brief explanation of your answer.

12: In your opinion, what influence does the "data protection of patient data" have on this research regarding the possible risk of reidentification of the location data from the cartographic visualizations? Please explain briefly.

13: This survey will be processed anonymously, but please indicate which institution you represent and what your professional position is. You are under no obligation to answer this question.

14: Is there anything else you would like to tell me; do you have any suggestions for improvement to be able to support geoinformation and medical research in Austria even better? Please briefly explain in the box below.

3.3. Population-Data Simulation

In this scientific work, neither confidential nor regional health data was used, but discrete health data was simulated. The data simulation was generated in QGIS and is based on Open Government Data of the City of Vienna¹⁹.

The simulated point data represent a hypothetical health dataset that have similar distribution properties.

First, the preparation of the building polygons was carried out. The description of the WMS Geoserver GetCapabilities link to all meta data of the MZK “1:1000” has been downloaded. The Vector data were downloaded in several parts from the multi-purpose area map MZK of the City of Vienna. To do this, by moving the vertical axes 5 times without overlapping, 5 bounding boxes with the coordinate systems were downloaded: West: -11500, North: 354000; South: 331000; East: 18500. Eventually, these 5 layers were merged into a vector polygon. As a result, the coordinate system was transformed from EPSG 4326 to 31256. Subsequently, the duplicates, the double geometries removed, and the polygonal components (roof, terrace, chimney) of a building were combined into a polygon.

In step II, the building points²⁰ of residential use were prepared. The point data set of the building information was obtained from the City of Vienna via WFS, and the coordinate system was transformed from EPSG 4326 to 31256 during import. Here, each point represents a building (point) or a parcel of land. This is continuously updated building data from the City of Vienna. The data was reduced from 58.360 to 47.595 points after usage. Here, each point stands for residential building or residential plot.

In step III, the dataset on population images and density for Vienna (2011- 2012) “by registration district” was downloaded from the GitHub website²¹. The coordinates data were inserted into a pre-generated text file and saved as a Geojson. The file was then added as a vector layer and exported as an SHP file, as well as the coordinate system was transformed from EPSG 4326 to 31256.

In step IV, the building polygons with “residential use” were further processed. The building polygons have been linked to the Points Housing file by crossmatching, reducing the building polygons from 200.842 to 36.723. All buildings that did not have a point were removed, limiting the data to residential buildings.

¹⁹ <https://www.wien.gv.at/stadtentwicklung/stadtvermessung/geodaten/fmzk/index.html>

²⁰ <https://www.data.gv.at/katalog/en/dataset/gebaeudeinformation-wien>
Identifier: 7a8aae59-71a4-4500-b38b-bdf15c7f627f

²¹ <https://www.data.gv.at/application/bevolkerungszahlen-und-dichte-in-wien-nach-zahlbezirken/#datasets>

In step V., the information of the registration district polygons was mapped to the individual polygons of residential buildings. In step VI, the polygons of the residential buildings were combined with the population data from the registration district to create a building object for each registration district.

The information of the building polygons (with residential use) was assigned to the registration district polygons. In the process, a temporarily merged layer was created, which then had to be exported as .shp in order to be able to store the data. As a result, the buildings of a registration district form a polygon (ID). The goal was for the building polygons to contain the population information: "Bez, Zbez, pop2011, pop2012, popden12.

In step VII, the "basic statistics of the 2012 population data" layer ("residential_pop_regDistrict.shp") were first calculated using the "Basic Statistics for Fields" to get a better overview. Using the field calculator, the points representing "households' with sick residents" were then randomly distributed as points on the combined residential polygon.

```
<html><head>
<meta http-equiv="Content-Type" content="text/html;
<p>Analysiertes Feld: pop2012</p>
<p>Anzahl: 179</p>
<p>Eindeutige Werte: 177</p>
<p>Fehlende Leerwerte (NULL): 0</p>
<p>Minimalwert: 453.0</p>
<p>Maximalwert: 25280.0</p>
<p>Bereich: 24827.0</p>
<p>Summe: 1431913.0</p>
<p>Mittelwert: 7999.513966480447</p>
<p>Median Wert: 7618.0</p>
<p>Standardabweichung: 4539.682450634431</p>
<p>Variationskoeffizient: 0.5674947840152043</p>
<p>Minderheit (am seltensten vorkommender Wert): 453.0</p>
<p>Mehrheit (am häufigsten auftauchender Wert): 8435.0</p>
<p>Erstes Viertel: 4391.5</p>
<p>Drittes Viertel: 10796.0</p>
<p>Interquartilabstand (IQR): 6404.5</p>
</body></html>
```

Figure 3 Statistic of Data Simulation ©QGIS

The points were placed within the polygons, it was assumed that a registration district with a low population density would have fewer points than a registration district with a high population density.

Six data sets were created according to density parameters with random point distribution in all Viennese registration's districts:

- if (pop2012 < 7999, then 30, else 50) $\ddot{y}$ ²² 7999 is the average, result: 7090.
- if (pop2012 < 7999, then 20, else 100) $\ddot{y}$ 7999 is the average, result: 10460.
- if (pop2012 < 7999, then 10, else 150) $\ddot{y}$ 7999 is the average, result: 13830.
- if (pop2012 < 4539, then 30, else 50) $\ddot{y}$ 4539 is the standard deviation, result: 8010.
- if (pop2012 < 4539, then 20, else 100) $\ddot{y}$ 4539 is the standard deviation, result: 14140.
- If (pop2012 < 4539, then 10, else 150) $\ddot{y}$ 4539 is the standard deviation, result: 20270.

The following are the point patterns of the simulated data sets 1-6. It can be judged that these resemble the same pattern and follow the point pattern of the Vienna residential buildings data. The distribution of points shows more points in the districts close to the city center of Vienna than on the outskirts, because the point pattern is based on Vienna population density 2012.

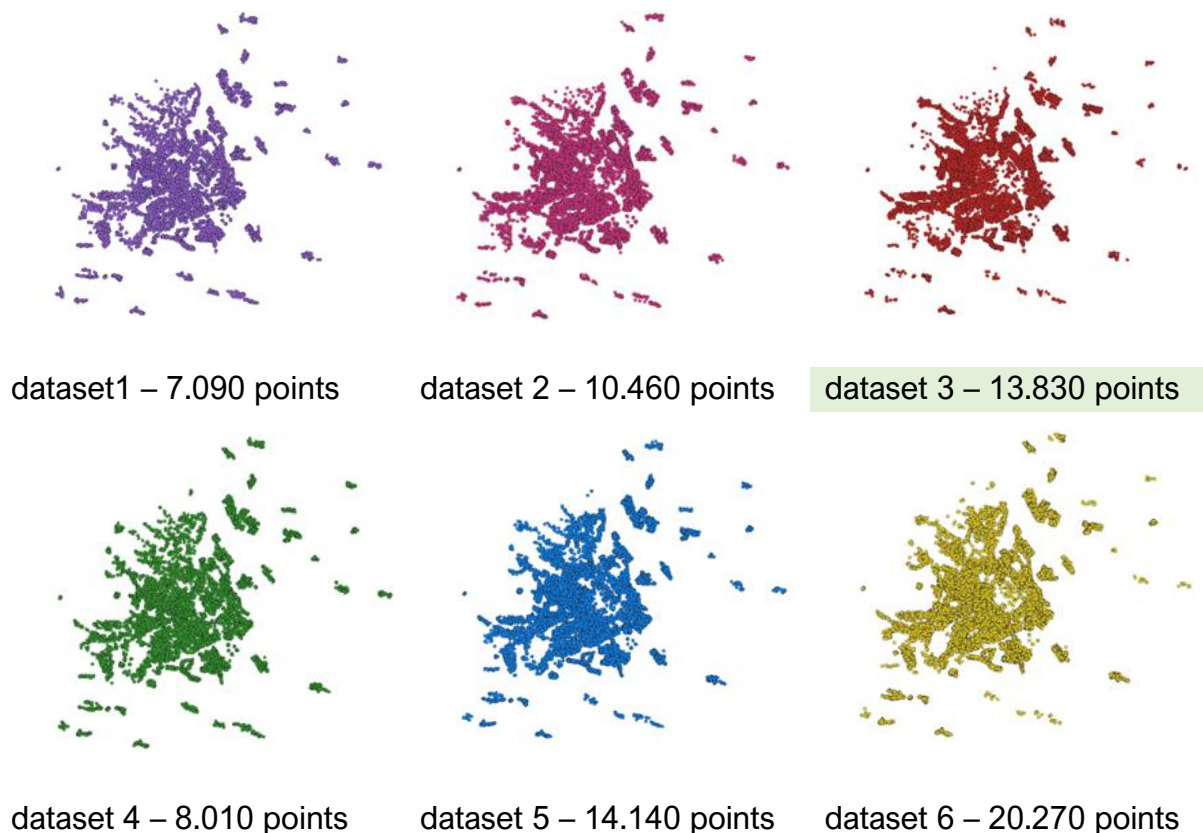


Figure 4 Overview Simulated Points ©QGIS

²² ($\ddot{y}$)* Standard deviation 4539 or mean MTV 7999

3.3.1. Simulated Dataset Analysis

Vienna is characterized by a different heterogeneous population density of 0-1416 addresses per km² (Kounadi, Leitner 2016). This also results in a heterogeneous overall picture of the simulated data sets, as they are based on the number of inhabitants of Vienna (2012). In this step, a dataset was selected for spatial analysis. Several study areas in the center and on the outskirts of Vienna were chosen in order to be able to examine the simulated data sets in the registration districts in more detail. The chosen size of the sections depended on the actual size of the registration district, the density, and the visibility of the point data. The smaller section was chosen, the larger the points could be displayed. Preference was given to study areas that include larger streets but also residential areas, as it can be assumed that the structure of the buildings depends on the road network. Subsequently, building data with the population data 2012 and simulated data sets 1-6 (right) were displayed.

3.3.1.1. Field of study 2223 Aspern

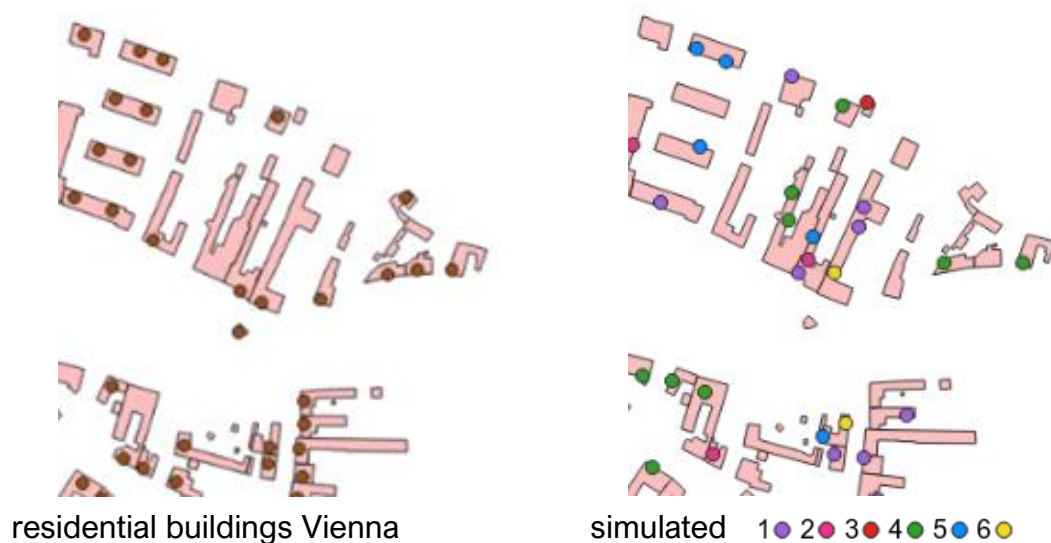


Figure 5 Registration District 2223 ©QGIS

In this suburb of Vienna, you will find Aspernstraße, the Siegesplatz, the Groß-Enzersdorferstraße. The urban development in Aspern took place mainly thanks to the U2 subway line (City of Vienna). The registration district 2223 covers 1.096.084 m² with a population density of 40 persons per km² (as of 2012) and is in 1220. With the low population of 4.428 inhabitants, it is below the calculated mean (7999) and standard deviation (4539) of the baseline data (2012). Therefore, thanks to six simulation parameters, a total of 120 points could be generated here. It can be seen that the

generated points have been evenly distributed in the registration district due to the large number of residential polygons. However, due to the number of inhabitants and the size of the registration district, only a few points can be observed, as the density is very low compared to other registration districts. Only 10 Points were simulated in this area.

3.3.1.2. Field of study 1103 / 1102 Simmering

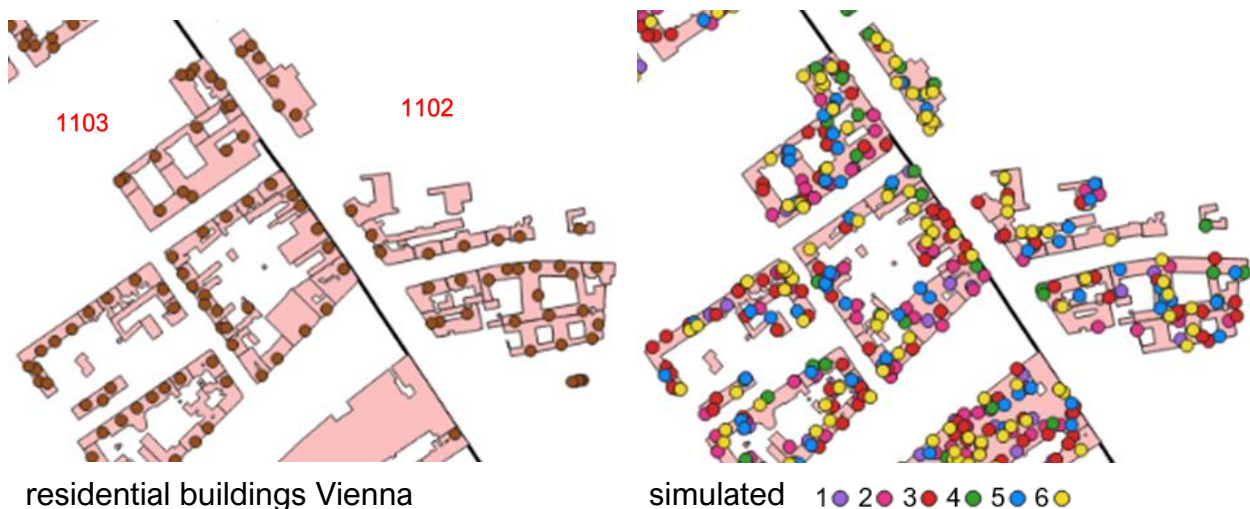


Figure 6 Registration District 1103/1102 ©QGIS

This section of Simmering includes two registration districts. In the registration district 1102 with an area of 1.094.043 m² and 13.584 inhabitants (2012) and population density of 124 persons per km² there are few residential polygons, so the simulated points are scattered around the point positions of the former residential buildings (brown points). The registration district 1103 covers 861.120 m² and 16.231 inhabitants and thus has a population density of 189 persons per km². Above the black line (polygon boundary) is 1102. The population in these registration districts is above the calculated average and above the standard deviation.

Compared to Aspern, the number of inhabitants is higher here, which could generate more points. If one were to compare the two registration districts 1102 and 1103, which have two different population densities, the thresholds of the mean would differ greatly from each other. The “Simmeringer Hauptstraße” is well-known shopping street and is heavily frequented.

3.3.1.3. Field of study 1204 Wilhelmsdorf

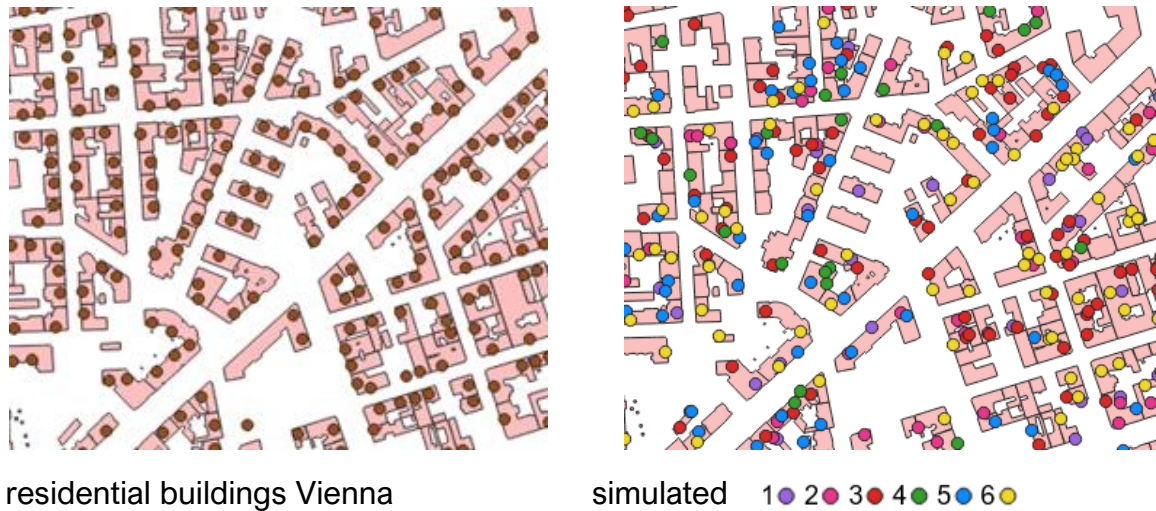


Figure 7 Registration District 1204 ©QGIS

The registration district 1204 is located near the Vienna Belt. Here, as in the 3rd and 11th, the building structure is shaped by the Wilhelminian period (City of Vienna, 2020). Wilhelmstraße, Eichenstraße, Niederhofstraße are also located here. The population density here is 225 persons per km², with an area of 376.162 m². The population of the 1204 population (8.474) is thus above the calculated average of the base population data of 2012 (7.999) and above the standard deviation (4.539). Up to 600 points could be simulated based on 6 data sets in this registration district.

3.3.1.4. Field of study: 0304 Fasangasse

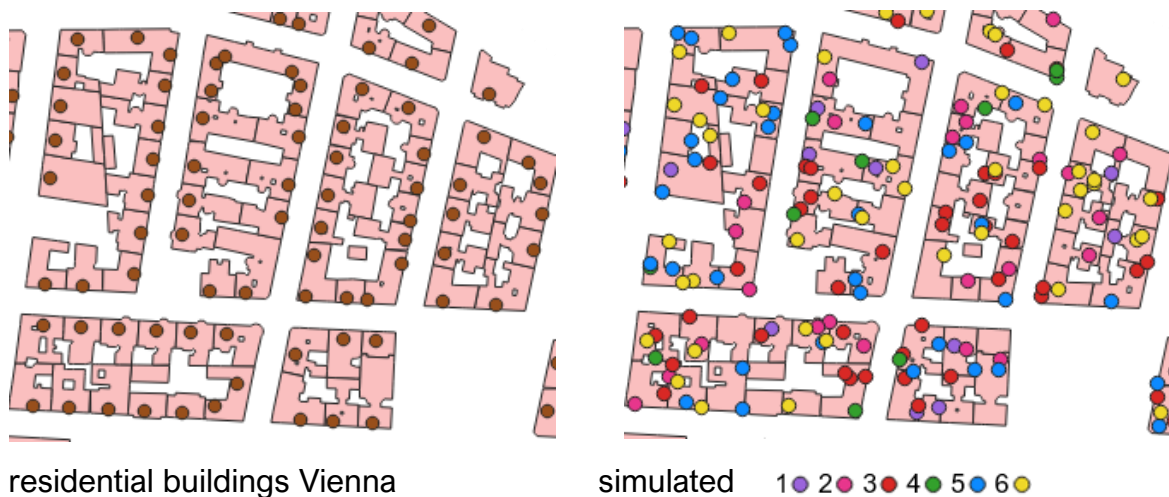


Figure 8 Registration District 0304 ©QGIS

Again, the more points, the higher the zoom insight was necessary. Because the points overlap. Due to the high density, the points are very close to each other, distributed over the entire study section. This section is located in Vienna's 3rd district. The study area is located between Landstraße Gürtel, Fasangasse and Rennweg, Aspangstraße. The third district of Vienna is adjacent to the First District of Vienna. The population of the registration district population (10.869) is above the calculated average and standard deviation. The registration district 0304 has an area of 285.813 m². Accordingly, many points were awarded here according to the parameter of the simulation. There is a nice arrangement of the simulated point data, sometimes with several points per residential polygon. Again, the higher the population density 2012 (380 people per km²), the higher the number of simulated points per residential polygon. A residential polygon can have several households with sick people.

3.3.2. General Considerations on Dataset Choice

The number of inhabitants of Vienna (as of 2012) in the respective registration district forms the basis for the calculation of the simulated data sets. The analyses show that some of the brown points (residential buildings) occur in places without a residential polygon. This is due to the data reduction process. This has been fixed during the simulation. The simulated datasets have retained the point pattern of the "Population Data 2012" because the figures show that the simulated points mostly occur at the former locations of the residential buildings. In the area with a higher density and a reasonable proportion of residential polygons, the point simulation showed a different picture than in registrations districts with fewer residential polygons and many simulated points. The simulated points were randomly distributed among the possible residential polygons.

In general, the peripheral districts have a lower density compared to the city center because the population is lower compared to the city center and because the peripheral registration districts have a larger area. It turns out that the number of inhabitants and the area of the registration district, as well as the number of residential polygons, have an influence on the distribution of points.

The different population sizes and density of the registration districts lead to heterogeneous patterns of the simulated data. Most registration districts have a large population and are above the averages and standard deviations of²³ the 2012 population data. In the outskirts, there are some differences in the data simulation compared to the city center: The more points there are in a registration district, the more confusing it becomes, as the points overlap. A higher resolution is required to

²³ The standard deviation indicates how far the data deviates from the average.

visually view the data in higher density areas in display mode. As a result, only small sections of the registration districts could be depicted. Compared to the print media, the digital version allows the data to be analyzed in higher resolution.

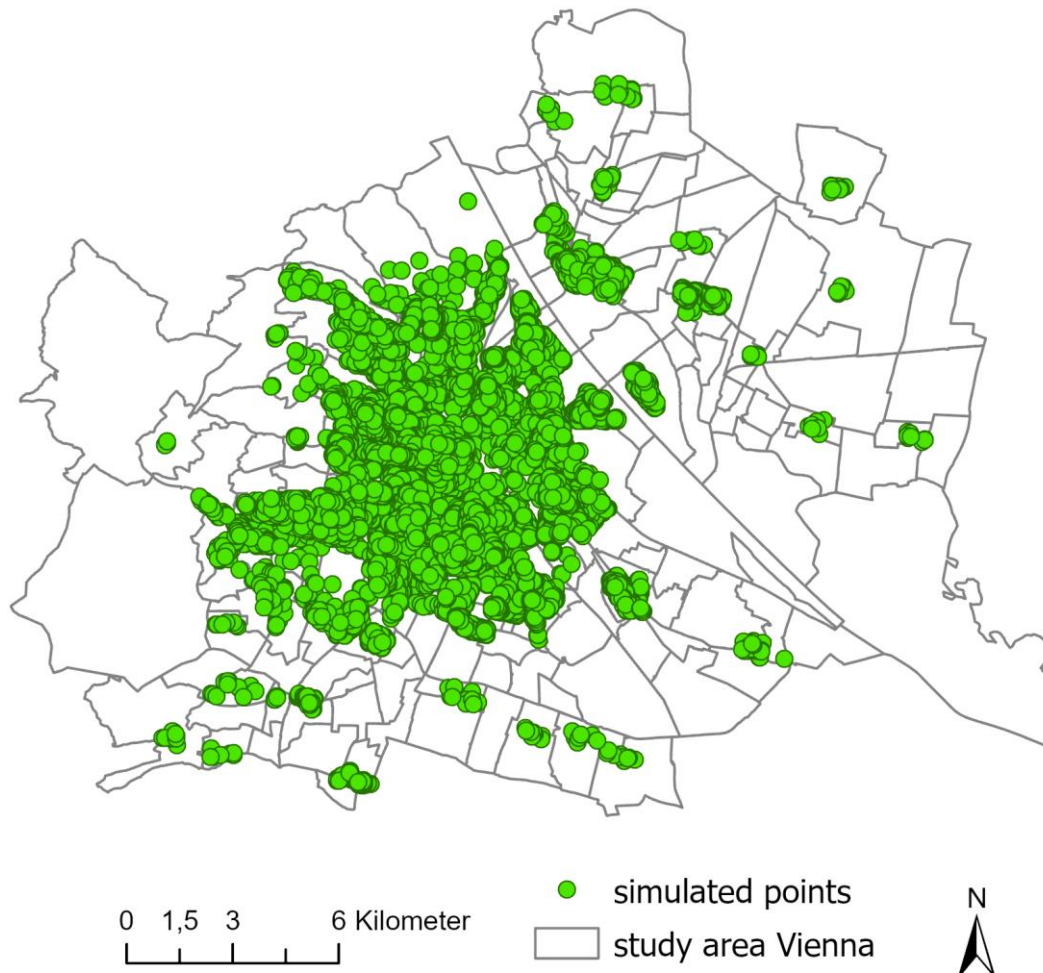


Figure 9 Simulated Points ©ArcGIS Pro

A total of 6 data sets are available, but data set no. 3, with the parameter "If (pop2012 < 7999, then 10, else 150) ÿ average, result: 13.830 points, was chosen for further analysis. Because one of the datasets was needed. After all, the datasets are similar, it is possible to use each dataset or even several together. According to the results, the data sets have the lowest number of simulated points with 7090 (a) and the highest number with 20270 (f). The calculated average of all simulated points is 12.300 points. The number of simulated points is based on the selected simulation parameters of the data set and the individual registration district data. In the next step, geomasking methods were applied to the selected dataset.

The selected simulated points (3) was still considered as the "original" dataset of "discrete health data" from Vienna. The ESDA results were visualized on a small or

large scale, depending on the analysis. For reasons of better visibility, it was therefore not always possible to visualize the entire study area. The point pattern was study with the study areas 1204, 1102/1103, 0304, 2223. The geomask displacement techniques and the masked locations of the points were also examined using these registration districts.

3.4. Adaptive Geomasking Methods

In order to be able to perform the geomasking methods, a new ArcGIS Project would have to be created and the GEOLAND basemap of Austria with the MGI coordinates system MGI Austria GK East would have to be added to the project.

Basic data created in QGIS were added to the project: the simulated points (3), population_regDistrict (population registrations district polygons = rori) and the road lines. The coordinate system of all features was projected from WGS to Austria GK East. It was not possible to carry out the geomasking process with geodetic WGS system.

Data reductions and changes in the attributes of the points features were necessary: The X, Y coordinates were added to the point record. Also, a column CID has been created and the field of CID has been recalculated $CID = ID + 100$ to be able to perform the AVM method.

In the street²⁴ shapefile attributes, unnecessary data columns have been removed. After a close examination of the street lines, it was found that the attributes generally included important roads and this number was not reduced.

In the population polygon (population_regDistrict), the attribute data has also been changed. Added column "rori" = Pop2012 data.

The Area column has been recalculated and the column "id_text (values = id)" has been added. Unnecessary items have been deleted. Finally, the symbolization of the population polygon area was made transparent in order to continue to be able to view only the boundaries of the study area.

²⁴ https://www.data.gv.at/katalog/dataset/stadt-wien_straengraphwien#resources

3.4.1. Adaptive Areal Masking (AAM)

In order to be able to perform the AAM method, it was necessary to download the toolbox from AAM²⁵ and import it into ArcGIS Pro via "Add Toolbox".

The chosen parameters were Points do Displace Simulated Points (3), Population Polygon (population_regDistrict) and finally Polygon ID Field (rori = Population 2012), with the size of the Population Threshold 100 and 1.000.

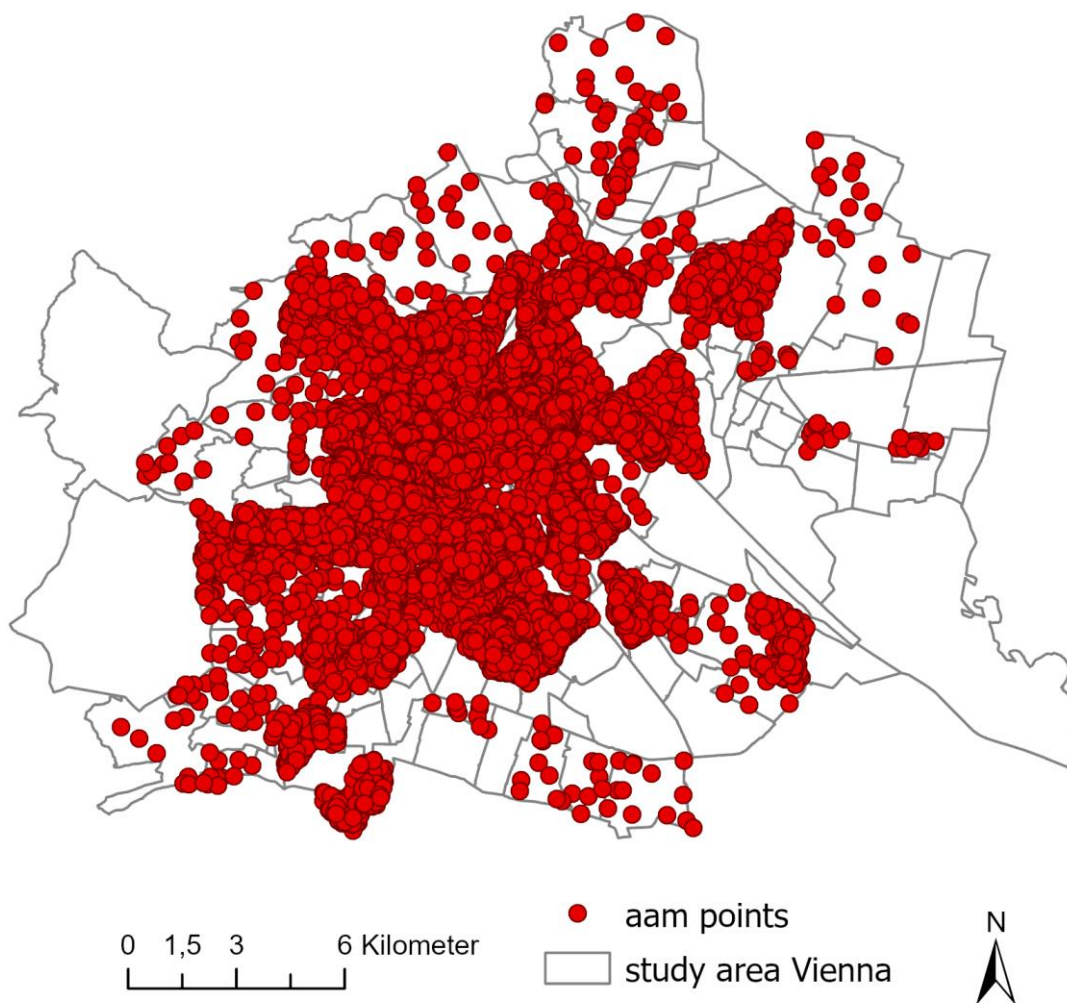


Figure 10 AAM Points (RoRi-100) ©ArcGIS Pro

The K-anonymity threshold has a large impact on the average point shift (Charleux and Schofield, 2020).

²⁵ AAM (and AAE) Toolbox

<https://www.arcgis.com/home/item.html?id=3ef11e690c1841c78df70433a2720724>

Next, the distance "Generate Start-Destination Links"²⁶ was evaluated. This calculated the mean distances between simulated points and the geomasked points. Here it was important to select the (CID) parameter in the data attributes (CID = p_3):

- Origin Features (simulated_3), Field Origin Group (CID), Target Features (AAM Points), Field Target (simulated_p_3), in the Meter Distance Unit.

The mean statistic (average) were used from the results of the Dist-Links performed. The distance between the simulated points and the new masked positions was measured: value: 540 m.

The lowest RoRi value in the study area is 264. Thus, more polygons were merged with RoRi 1.000 than with RoRi K=100. In the AAM method, there was a difference between RoRi 100 and 1.000 in the AAM geomasking process. For this reason, AAM masking was performed with RoRi 1.000 to examine the results for the displacement of the masked points.

In the next step, the AAM results were examined using AAE polygons (RoRi) to maintain spatial anonymity (SKA).

²⁶ <https://pro.arcgis.com/de/pro-app/latest/tool-reference/analysis/generate-origin-destination-links.htm>

3.4.2. AAE-Polygons (RoRi-Value 100 and 1.000)

The anonymized AAE polygons were then created. This was done using the AAE toolbox in ArcGIS Pro: A point file and a spatial dataset including RoRi were used. The RoRi attribute represents the Viennese population data (2012).

The selected parameters: Population Polygons (population_regDistrict), Polygon ID Field (id_text), Population Field (rori), Population Threshold (100 and 1.000).

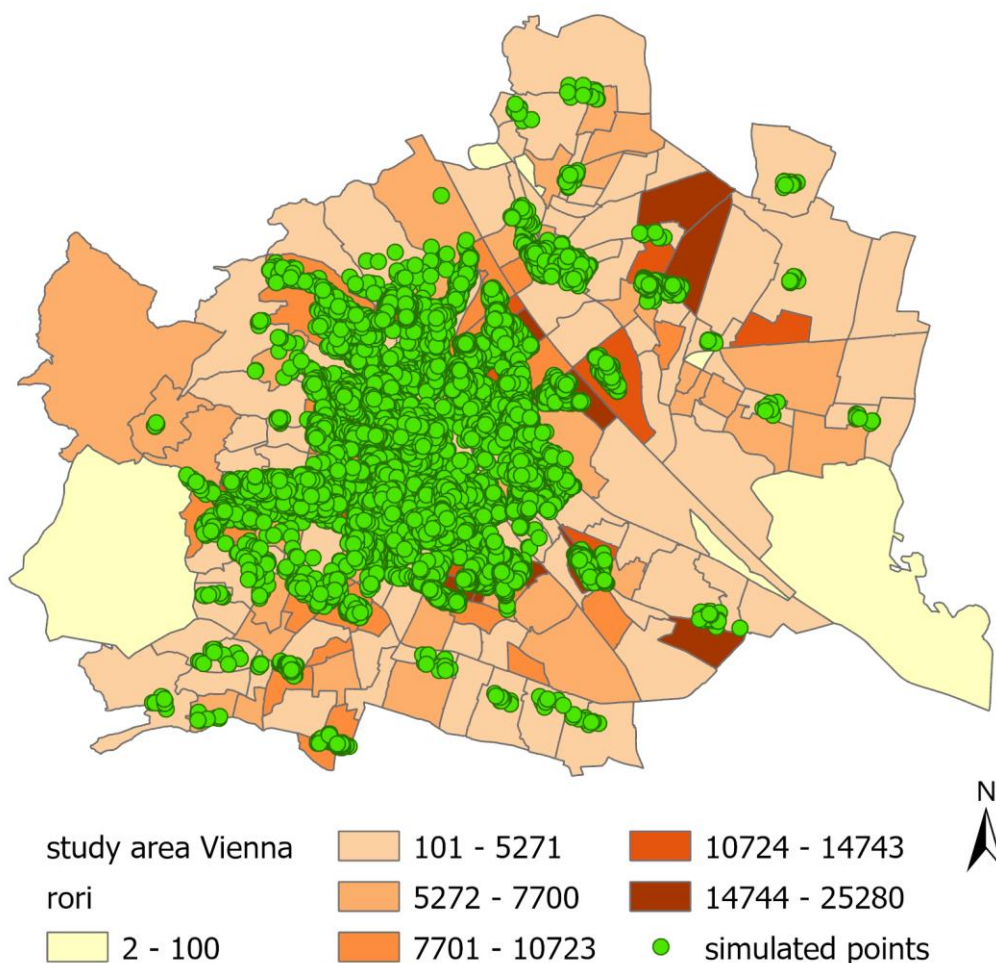
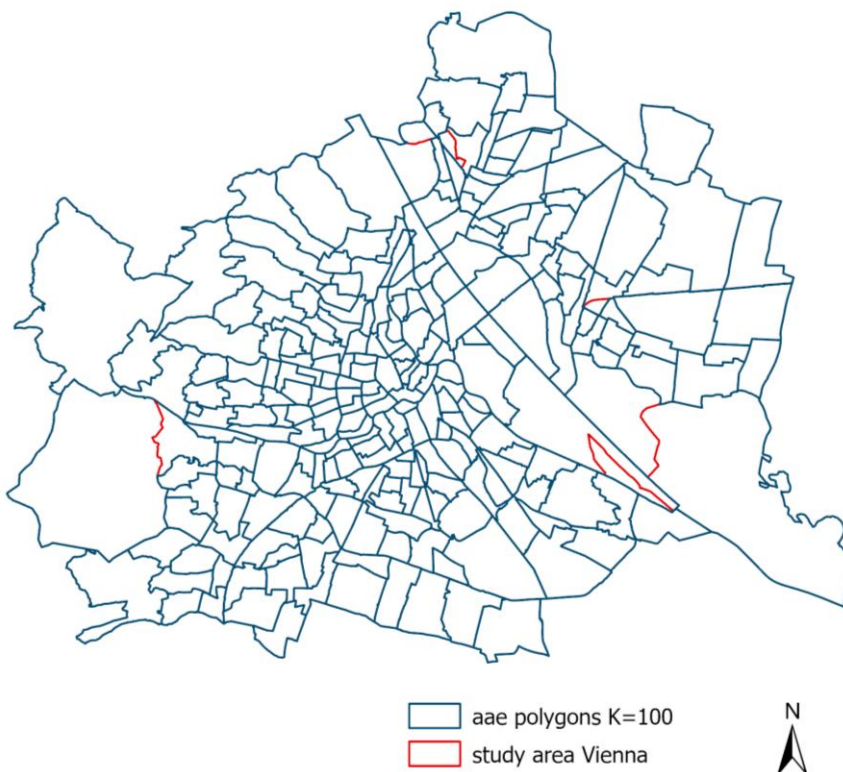


Figure 11 RoRi Population 2012 and Simulated Points ©ArcGIS Pro

The registrations district with values below 100 can be observed in the figure. In principle, anonymization polygons were merged by disclosure value (RoRi) (Kounadi and Polzin, 2021). The disclosure value was defined as 100, a value that each polygon registration district had to have. During editing, the AAE polygons were stored in the *.gdb of the ArcGIS Project. Based on this, any polygon whose RoRi value was less than



100 should actually be resolved with its neighbor until all polygons have the RoRi values equal to or greater than 100 (Kounadi and Leitner, 2016).

To choose the RORI value: First a value of 100 was used. Upon visual inspection of the AAM method results, it was found that no masked points were moved outside the registration area ($\text{RoRi} < 100$). Because in the registration district areas with low RoRi values (< 100), there were no simulated points. For this reason, an analysis was also carried out on next step with RoRi 1.000 to investigate this. In the comparison of the AAE anonymization polygons to the population polygons (2012), these were reduced from 250 to 244.

Thus, six polygons had not reached the 100 values before anonymization. To ensure an identical anonymization threshold, anonymization polygons should be larger in rural areas than in urban areas (Charleux and Schofield, 2020).

The analysis with the RoRi value 100: Spatial anonymity was retained. Comparing the results with the positions of the simulated points it was shown that all simulated points are in AAE polygons with > 100 RoRi.

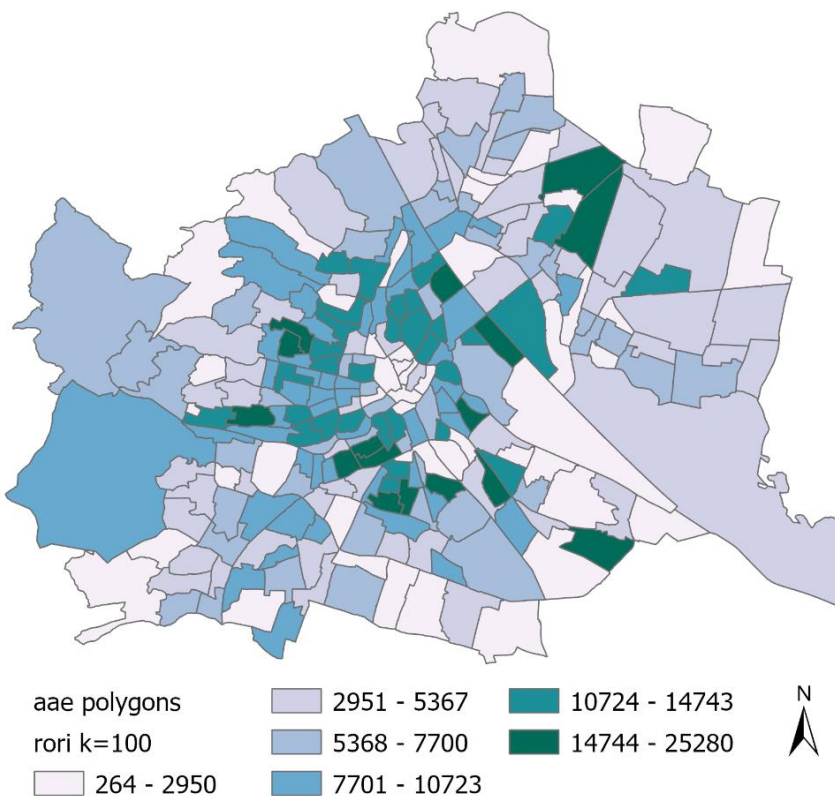


Figure 12 AAE-Polygons RoRi 100 ©ArcGIS Pro

Based on the illustrations, the dissolved border of the Vienna study area can be recognized by the red lines. Several analyzes were carried out to investigate this.



The Rori 200 (6), 300 (13), 500 (13), 750 (13) showed no really visible changes in the points shifts to other registrations districts. Only with RoRi 1.000 (18). It should also be noted that RoRi represents a population value that must be present in each registration district to maintain spatial anonymity.

The results of the map visualization show the number of residents (RoRi) per registration district. In order to check whether the spatial anonymity of the generated AAE polygons was maintained, the AAE polygons were mapped on the basis of RoRi.

If you look at the lowest values in the figure, you can see that there are no polygons left < 1.000. This is how anonymization worked. The result is that maintaining SKA is a very important point in geomasking.

In the next step, the AAE polygons RoRi 100 were used to perform the AVM.

In the case of AAM, the importance of RoRi was highlighted.

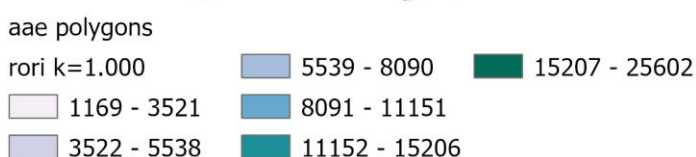
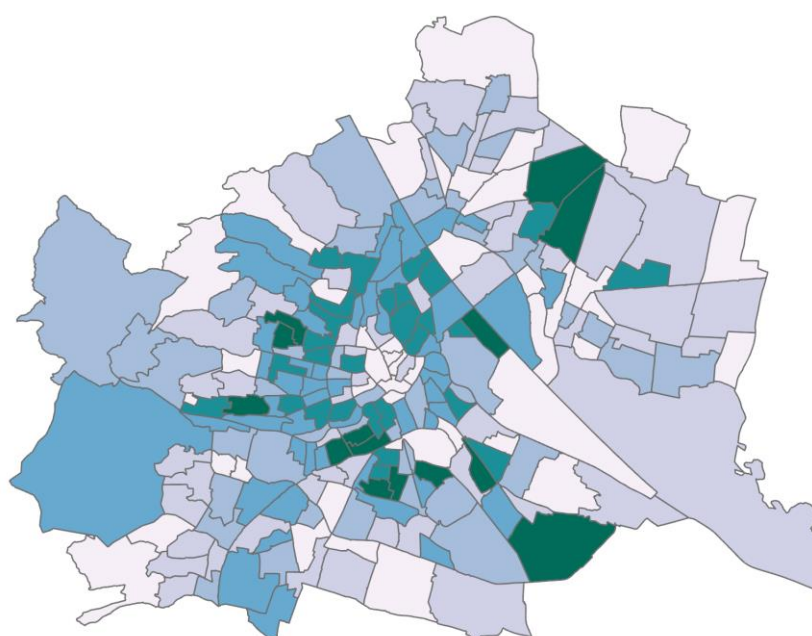


Figure 13 AAE-Polygons RoRi 1.000 ©ArcGIS Pro

3.4.3. Adaptive Voronoi Masking (AVM)

In order to be able to perform this method, it was necessary to adapt external and internal parameters of the ArcGIS Project to the Python script ²⁷ of Prof. Kounadi GitHub. It was necessary to create the AAE polygons in advance, as well as to adjust the attribute field of CID. This is not confidential data that depicts a 1:1 distribution of the disease, but simulated data. Thus, the CID (CID+100) could be changed manually without significantly changing the score record.

According to the literature, AVM method is based on the concept of adaptive areal elimination (AAE) and Voronoi masking (VM). Takes population density into account by defining the areas of minimal K-anonymity where Voronoi polygons are created (Kounadi and Polzin, 2021). By shifting the points to intersections, the topography is taken into account (Kounadi and Polzin, 2021).

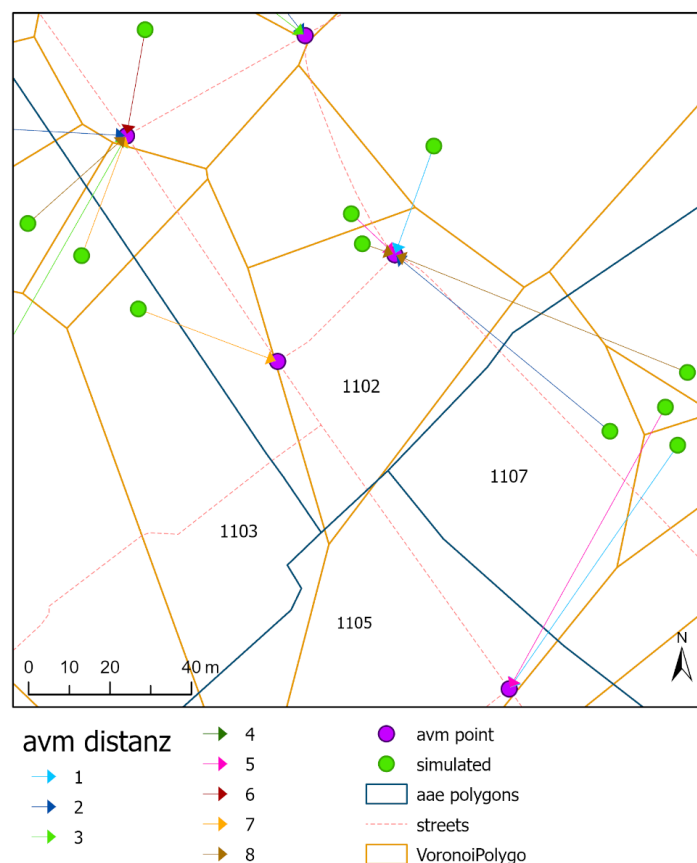


Figure 14 AVM, Voronoi Polygons ©ArcGIS Pro

According to Polzin the AVM geomasking process consists of several stages: The polygons under consideration are resolved using a selected disclosure value (DV) (Polzin, 2021). Then, all data points that are within a polygon fused with at least one other data point form a Voronoi polygon, to whose next boundaries the points are

²⁷ <https://github.com/okounadi/Geoprivacy>

shifted. If the edge of the Voronoi polygon is outside the spatial K-anonymity (SKA) polygon, the data point is transferred to the outer boundary of the connected polygon. The masked data points are spatially dependent on the simulated data points, they are not randomly distributed (Polzin, 2020). Points that were not moved by the Voronoi mechanism were randomly distributed within their SKA polygons. Finally, all data points were aligned at the nearest intersection (Polzin, 2020).

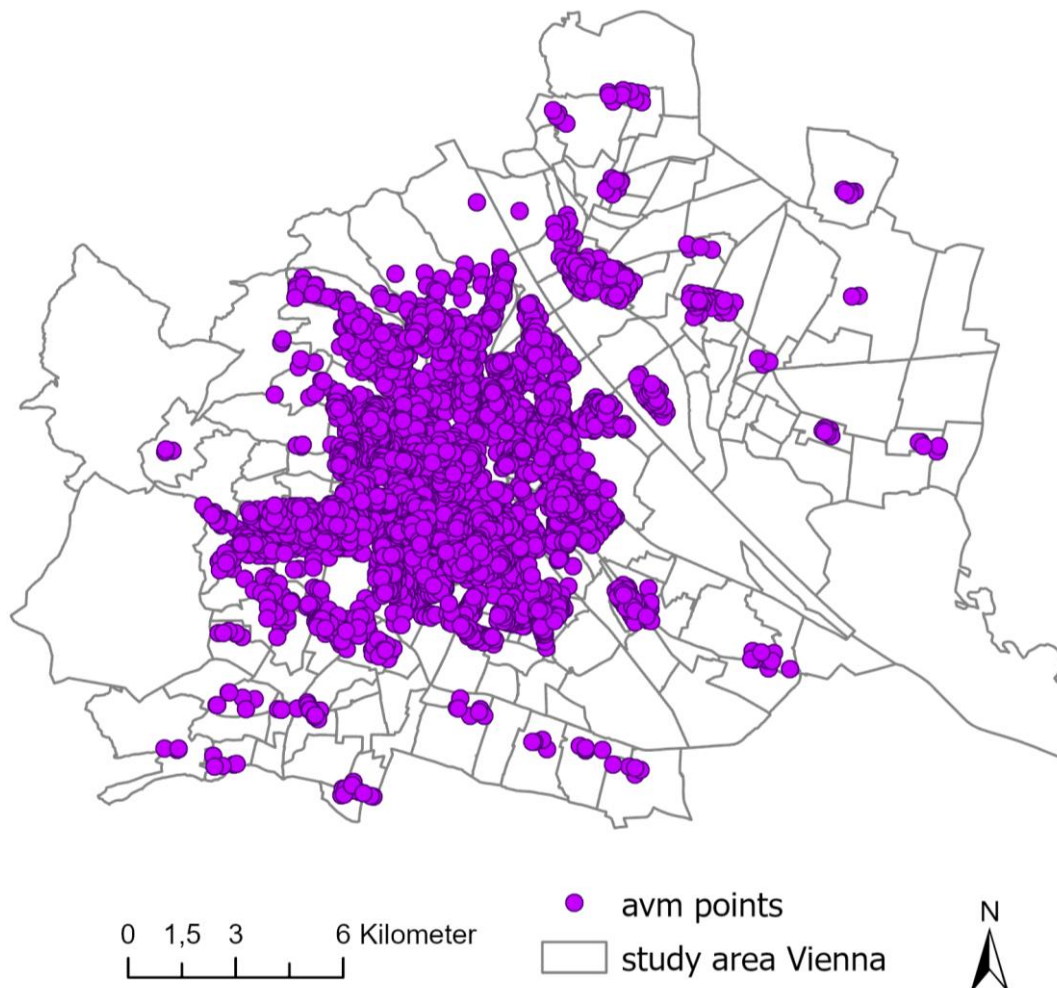


Figure 15 AVM Points ©ArcGIS Pro

The AVM method was also done with 1.000 RoRi to see if there would be any differences. The results were identical, which is why further ESDA analyses were performed with RoRi 100. At the end of the AVM process, the distance of the AVM points to simulated points was calculated in ArcGIS Pro using the Python script. The thesis would be that the AVM method is suitable for performing visualizations with health data, while maintaining the spatial pattern of the points and not violating the privacy of the participants. At present, it is not yet possible to assess the benefits of the AVM method in health geography. It is necessary to investigate this using ESDA techniques, especially for the spatial re-identification risk and spatial point pattern density change assessment.

3.5. Spatial Visualization-Techniques

In this step, the results of the simulated points and geomasking points were visualized. The displacement techniques of the geomasks was also investigated. The new positions of the masked points were visually evaluated with the help of orthophotos. Different zoom views were used and the risk of spatial reidentification was assessed.

3.5.1. Point Pattern Maps

The visualization of the point pattern, geomasking results, and the simulated dataset was explored using orthophotos in ArcGIS Pro. The point pattern of the original (simulated) data was compared with the point pattern of the geo-masked data (AAM and AVM method) for changes in point scattering. Based on the statistical value, the geomasked data points were examined for the displacement distance.

The AAM statistics data were calculated using the ArcGIS Pro "Generate Start-Destination Links". In the AVM method, on the other hand, it was executed in the last step of the Python script. The average values of the displacement distance can be obtained in ArcGIS under Statistics (Dist_link) of the masked dataset. The results of the postponements were summarized in a table.

In order to examine the masked points for displacement distance and new locations, it was necessary to examine the points in higher resolution. Especially in areas with higher density. This was done on the basis of the four areas of study. Subsequently, a section of a region with lower density (2223) was selected to be examined for spatial reidentification using the X,Y coordinates. The coordinates of a point were displayed with the orthophotos (ArcGIS Pro) and the Geoland basemap (ArcGIS) in different zoom views. This was followed by a visual assessment of the risk (Polzin, 2020).

3.5.2. Choropleth Maps

The Choropleth maps were created based on the points per registration district (ArcGIS Pro). It was necessary to combine the points with the polygon of population data, to get the number of points per registration district. For this reason, a spatial join (contains) was performed. As a result, the spatial information of the points was attributed to the polygon. This was done for original (simulated) and masked data sets. This was followed by a representation of the points per registration district of original (simulated) and geomasked points, in the same color scale, to see which geomasking method changed the proportion of points per registration district. The data was displayed proportionally using Choropleth map symbology. The Choropleth maps were categorized into the same color scale to better compare the data. A Choropleth map

with AAM points RoRi 1.000 was also created to compare the data. A separate density evaluation was not carried out. Because it can be assumed that the density depends on the occurrence of the points in the respective registration district.

3.6. Hotspot Detection on Kernel Density Estimation

Both analyses were performed in ArcGIS Pro spatial tool and were performed based on the points. For this purpose, the tools were obtained from ArcGIS Pro: Kernel Density Estimation²⁸, and Optimized Hotspot-Analysis²⁹.

3.6.1. Kernel Density Estimation (KDE)

Based on the density estimate (in ArcGIS Pro), the areas with the highest and lowest point distribution could be identified. The density of the point pattern was examined on the basis of the point characteristics. The analyses results of the surface density which search radius of 500 km and 750 km were different. This is because the information displayed depends on the choice of bandwidth and the size of the output grid (Kalinic and Krisp, 2018). For further, KDE analysis with parameters of a bandwidth of 500 km was used. The calculated grid standard value of the Pop2012 study area was 91.08 m. The spatial mask of the output grid was calculated to the size of the 2012 population data polygon. The calculated standard value of the cell size of the individual analyses was 79.27 m (simulated), 78.85 m (AVM), 89.51 m (AAM). The cell size (5) defined the output grid that was created so that the KDE results were not so pixelated. The classification of the KDE of the output grid (Krisp and Spatenkova 2009) was visualized according to natural fractions Jenks (1967) in 6 classes, which made it possible to display similar data values in the same class (Kalinic and Krisp, 2018).

3.6.2. Optimized Hotspot Analysis

The spatial pattern of the simulated and geomasked datasets was examined for significant clusters with high and low values. The Optimized Hotspot-Analysis was also performed without an analysis field in order to be able to examine the spatial pattern of

²⁸ <https://pro.arcgis.com/de/pro-app/3.1/tool-reference/spatial-analyst/kernel-density.htm>

²⁹ <https://pro.arcgis.com/de/pro-app/latest/tool-reference/spatial-statistics/how-optimized-hot-spot-analysis-works.htm>

the points and compare the results. The tool aggregated the points to determine the number of points and be able to use them as an analysis field (Kalinic and Krisp, 2018).

In the G_i^* statistics, the local sum and the statistical z-score were calculated. This statistic identified statistically significant hotspots. These have a high value and are surrounded by other points of high value (Kalinic and Krisp, 2018). The higher values have formed G_i^* positive (hotspot) and the lower values G_i^* negative (cold spot). Based on p- and z-values, the statistically significant clusters were determined with a certainty of 99%, 95% or 90% (ArcGIS). The null hypothesis that there is no relationship between a certain characteristic of a region and its neighbors (Getis and Ord. 1996) thus had to be contradicted.

3.7. Spatial Autocorrelation Analysis (Moran's I)

In this step, several spatial autocorrelation analyses were performed. First, the Spatial Autocorrelation tool of ArcGIS (Global Moran's I) was used.

After that, the spatial tools Local Moran's I (Anselin) and Average Nearest Neighbor, were carried out.

3.7.1. Global Moran's I

Global Moran's I^{30} was run to see how the points correlate with each other. The following parameters have been selected (in ArcGIS Pro):

- Simulated points with masked data sets were examined. The input parameters were (analysis field Points, CID 13.830) and "conceptualization of spatial contexts- K-Nearest Neighbors" (8).

This was done for all three datasets. The results were summarized in a table.

The statistic values were used to evaluate whether the pattern is grouped, distributed, or random (ArcGIS, 2024).

³⁰ <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/spatial-autocorrelation.htm>

3.7.2. Cluster-Outliers Analysis (Anselin Local Moran's I)

To get a better overview of the data, Local Moran's I ³¹ (according to Anselin) was performed (ArcGIS Pro). Local Moran's I was created with registration districts and also with point datasets (attribute field: CID).

- 1. Analysis parameter on the registration districts: The polygons - spatial join - were selected as input features, analysis field Count Area (250). Parameter: "conceptualization of spatial contexts K-Nearest Neighbors"(8), 999 Permutation.

The individual 13.830 points were connected to 250 registrations tracts using spatial join. The points did not change the location as a result. Here, the results of the ArcGIS analysis were automatically generated in maps at the registration district level.

- 2. Analysis parameter on the point level: Points (CID), "conceptualization of spatial contexts K-Nearest Neighbors" (8), 999 Permutation.

Here, 4 runs were performed and the results of the cluster outlier analysis of the masked and simulated points were summarized in a table. Also here, the results of the analysis were automatically generated – but here in maps at the points level.

Local Moran's I maps of the points were visually mapped with GI* maps to see the differences between the two statistical methods. The results of the outliers were summarized in a table. This includes the z-value. The outliers were also examined for possible anomalies and possible risks of spatial reidentification using orthophotos.

3.7.3. Average Nearest Neighbor

Also, the spatial statistic tool to calculate the average nearest neighbor (ANN) ³² was performed in ArcGIS Pro. The input was the point data sets. The average nearest neighbor index was calculated based on the average (centroid) distance of each feature to the nearest neighbor feature (ArcGIS(c), 2023).

This was done with original (simulated) and masked data sets. In order to be able to compare the results. Also here, the results were summarized in a table. If the index < 1 (ratio for mean nearest neighbor) it is a cluster patterns (ArcGIS). When determining the randomness of clusters, the p-value was important.

³¹ <https://pro.arcgis.com/de/pro-app/latest/tool-reference/spatial-statistics/cluster-and-outlier-analysis-anselin-local-moran-s.htm>

³² <https://pro.arcgis.com/de/pro-app/latest/tool-reference/spatial-statistics/h-how-average-nearest-neighbor-distance-spatial-st.htm>

4. Results

In this chapter, the results of the analyses of Chapter 3 have been explained. ESDA techniques were chosen, which were used to investigate the geomasks as well as the simulated points in more detail. First of all, the results of the online survey were described, as they offer a brief insight into the current situation in Austria. Most of this chapter consists of ESDA techniques results, and research into the effectiveness of geomasks. This includes the investigations of the point displacement, the KDE and Hotspot-Analysis, the statistical evaluation of the neighbors and the risk of re-identification of hotspots.

4.1. Evaluation of the Survey Results (of 15.10.2023)

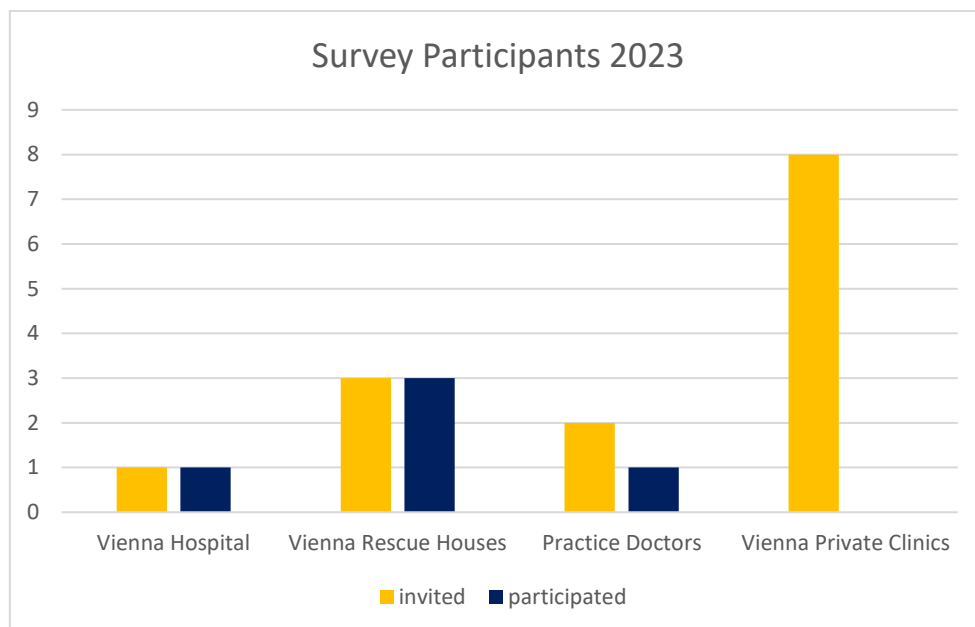


Figure 16 Diagram of Survey Participants ©LimeSurvey

Most of the invited participants came from private clinics in Vienna, and thus made up the highest proportion of the survey participants. Followed by emergency services (3), general practitioners (3), and a Viennese hospital, a total of 15 health facilities were invited to participate in the survey. The Wienerberg Health Insurance Fund was also invited. There were 12 responses, but after a detailed evaluation of the responses, it turned out that of these 12, only 5 complete responses were available. Overall, the complete responses came back in full of a Viennese hospital, emergency services (3), and a practice doctor (1). On the part of the private clinics, no relevant replies were received (0/8), the answer sheets were sent blank. The survey was primarily intended to be qualitative and is nevertheless considered sufficient of the few responses. Finally, relevant information was shared.

SURVEY 2023	HOSPITAL	PRAXIS	AMBULANCE (1)	AMBULANCE (2)	AMBULANCE (3)
Knowledge of spatial GIS-Analysis? (Q02)	Yes	No	No	Yes	Yes
Origin of the information? (Q03A)	Active since 1998	-	-	University of Vienna dept. Natural Sciences	General Knowledge
Would you like more information? (Q03B)	-	NO, thank you	Yes	-	-
Knowledge of geomasking methods? (Q04)	Yes	No	No	No	Yes
Origin of the Information? (Q05A)	Literature	-	-	-	Google
Would you like more information? (Q05B)	-	No, thank You	Yes	Yes, of course,	-
What happens to the data? (Q06)	are analyzed	no, thank you	not relevant	are collected	internal use
Problems with data retention? (Q07)	No	No	Each assignment requires manual elevation	Fixed processes predefined	ISO-certified organization, strict guidelines given with handling of personal data
Have you performed spatial GIS analyses? (Q08)	Yes	No	No	No	No
Reasons for no spatial GIS analyses? (Q09B)	-	No time	I do not know	No confrontation necessary on aa daily basis	No need
Do you need to take further measures in advance? (Q09F)	-	No	Strongly assume that	I don't know	No need, but data available
Interested in performing spatial GIS analyses? (Q09G)	Have been carried out since 1998	-	Yes	-	-
Leave it to the experts?	-	Yes	-	Yes	Yes, but no need
Sufficient publicly available information? (Q10)	Yes, in the network	Yeas, I don't know	Yeas, I've never heard of GIS-Analysis	I don't know	I don't know
Is research possible without restrictions? (Q11)	No	I don't know	No	I don't know	I don't know
Influence of data protection? (Q12)	Great, they are not allowed to evaluate individual cases	I don't know	Habits can be derived	Data protection is super important	I don't know

Table 1 Evaluation of the Survey Results ©Lime Survey

As a result, 5 participant answers could be evaluated. The anonymized responses (any personal information has been removed) are attached. The table shows that since 1998 the Vienna Hospital has been carrying out spatial analyses based on scientific information.

A rescue service house (2) obtains its knowledge from the University of Vienna, Dept. Natural Sciences, while another rescue service house (3) sees the knowledge of spatial GIS analyses as "general knowledge". With regard to the knowledge of geomasking methods, it can generally be seen that the necessary knowledge is only partially available and that there are information interests.

The results show that only the Vienna Hospital analyzes the patient data. In the case of the emergency services, the answers are very different, and the application of the data seems to differ according to the main activity of the hospitals. On the one hand, the data is collected by rescue hospitals according to strict ISO guidelines and used for internal purposes. But also, the data would be available in principle, but not considered relevant. There were no difficulties on the part of the general practitioner in storing the data (according to Chapter II, health data is stored by ELGA), and the Vienna Hospital agrees.

On the part of the Vienna Hospital, it was communicated that they use location, penetration, and route analyses. The spatial analyses are carried out for reasons of performance planning. It is important to properly organize the process of spatial analysis from data entry to processing of the analyses. The spatial analyses can be recommended because they allow a better view of the data and thus offer new solutions to overcome complexity.

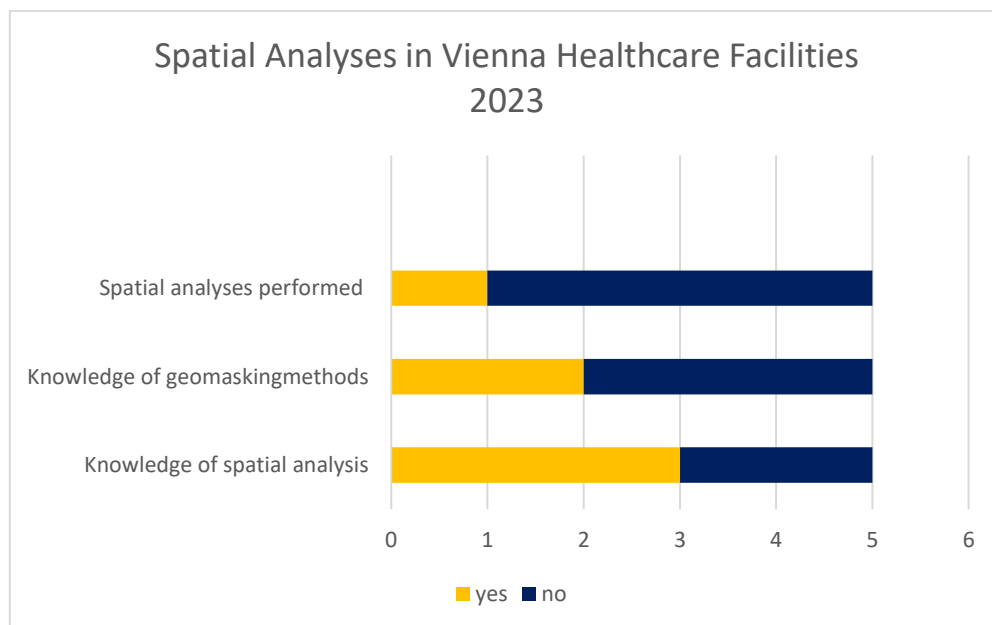


Figure 17 Diagram of Spatial Analysis in Healthcare ©LimeSurvey

Using the graph, 4 out of 5 participants did not perform spatial GIS-Analysis. The lack of time and knowledge about spatial analysis seemed to be the reason for this.

4.1.1. Conclusion of the Survey Evaluation

Overall, data is available, but not spatially analyzed. Participants were better informed about spatial analysis than about geomasking methods. As a result, it can be recommended that more information be disseminated regarding spatial analyses in health facilities in Vienna. In general, there is interest in carrying out spatial GIS analyses, and the majority of participants (except Vienna hospitals) would opt for experts.

With regard to data protection, the result is that no individual cases may be evaluated that data protection is "super important". In general, the participants knew that data protection has a high priority on research, and that this is not possible without restrictions. Healthcare facilities are therefore well informed about the privacy of patient data, but they do not have any further information on how to analyze their data, for what reason, and it is assumed that they should contact if they are interested.

In particular, the emergency services showed interest in further information on spatial analyses and geomasking methods. Which was very surprising. The Vienna hospital seems to have very good literature and expertise in spatial GIS analyses. Their answers complement the topic of this master's thesis very well, and not only with regard to data protection. Overall, it can be seen that the most relevant answers came from the Vienna Hospital (unfortunately the name may not be mentioned for data protection reasons).

Several private clinics were invited to participate in the survey, which were specially selected in advance because they were expected to invest more time in analyzing patient data than public hospitals as private hospitals. Overall, no responses came back from the private clinics. It is likely that the survey was not properly understood or considered sufficiently qualified by the participants. Similarly, a large hospital will have more information than a smaller healthcare facility. Provided that it has not been explicitly searched for.

The practice physicians were also personally invited to participate in the survey, as there was no feedback from the private clinics. From a personal conversation with a doctor participating in the survey, it could be inferred that there is interest in spatial analyses, but due to the need for informed consent from all participating patients (including older people), he is not able to do so due to lack of time and the high effort. However, more information would be very supportive ©LimeSurvey, 2023.

The aim of the survey was to find out what knowledge the health facilities have. Not to ask for information about patient data. Unfortunately, it has not been possible to survey all health facilities. As a result, it could be recommended to carry out more surveys with regard to spatial analyses and to motivate health care facilities, perhaps on the part of the Vienna-Medical-Association, to carry out spatial analyses. After all, spatial analyses are a part of health geography research and can improve the in-house management.

There is a lack of the necessary know-how. The fact that the private clinics did not provide any answers was very disappointing. On the other hand, the responses from practice doctors, emergency hospitals and the Vienna Hospital showed that it at least made sense to conduct the survey ©LimeSurvey, 2023.

Based on these results, it can be recommended to inform healthcare facilities about current geomasking options and to provide comprehensible information to healthcare facilities. The five participant responses suggest that more public information on spatial analysis in health geography would be good. In order to get a better overview, an expert interview can be recommended.

4.2. Spatial Visualization-Point Pattern

In this step, the point pattern of the study areas (0304, 1204, 1102, 1103, 2223) was visualized and subsequently examined for differences.

The scattering, displacement, and displacement distance of the AAM and AVM points in comparison to simulated points was investigated, as well as a reidentification of point coordinates using Orthophoto.



Figure 18 Point Pattern 1102 ©ArcGIS Pro

Registration district 2223 has a density of 40 person per km², 1204 has a density of 226 person per km² and 0304 has a density of 380 person per km². Looking at the registration districts individually, the AVM point pattern is similar to simulated points. The AVM method shows fewer points than simulated, while the AAM method scatters the points up to polygon boundaries.

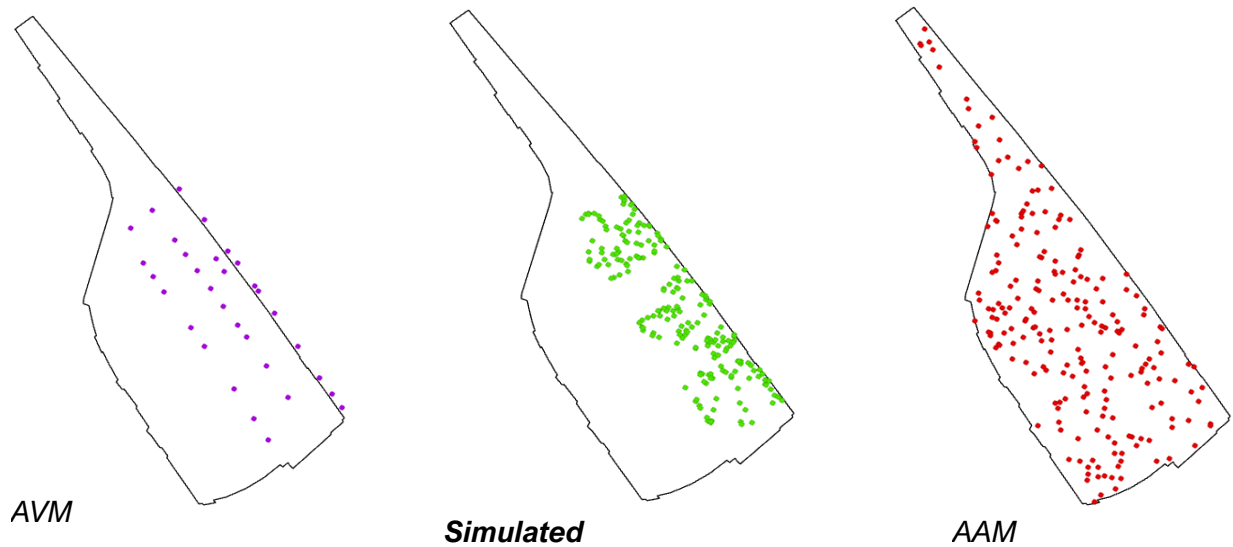


Figure 19 Point pattern 1103 ©ArcGIS Pro

In the 1103 from Simmering, there is a higher concentration of points than in the 1102. In terms of the number of points, the AVM method also shows fewer points than simulated ones. In the 1103 section you can see AVM points outside of registration district. The AAM method shows the same result as in the previous district, the AAM points are scattered throughout the polygon during obfuscation.

The registration district 2223 has a very low density. This is also where the AAM point dispersion can be observed. Again, the point pattern was best obtained from AVM method. On the other hand, the AAM method seems to like the large area.

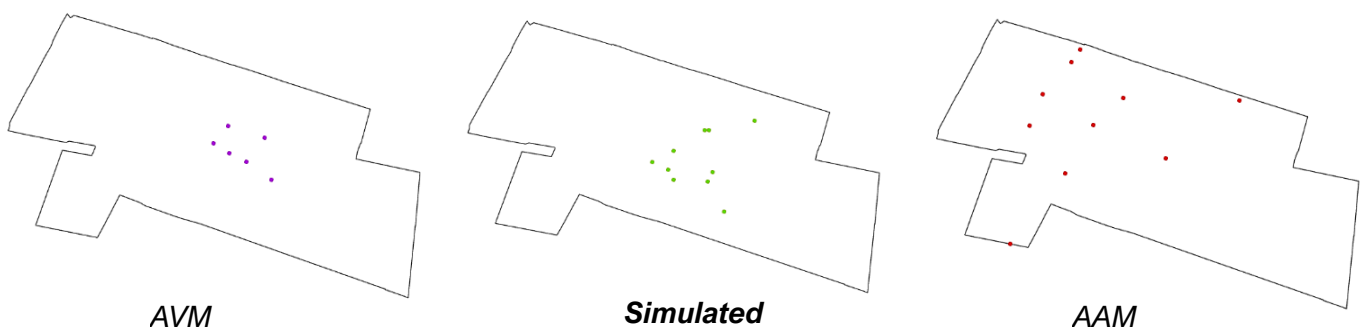


Figure 20 Point Pattern 2223 ©ArcGIS Pro

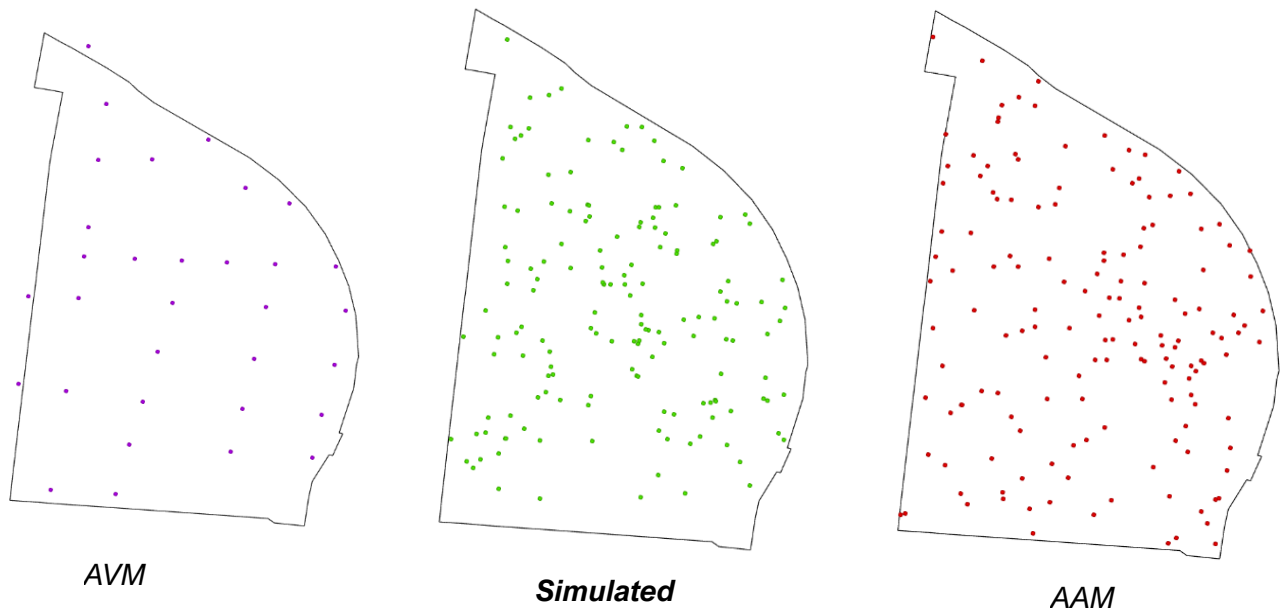


Figure 21 Point Pattern 0304 ©ArcGIS Pro

In the very high-density study areas 0304, the point pattern is similar to the results from previous high-density study areas. With the AVM method are fewer points visible compared to simulated points and AAM points because several simulated points have been masked in an AVM point. This can be observed in all study areas.

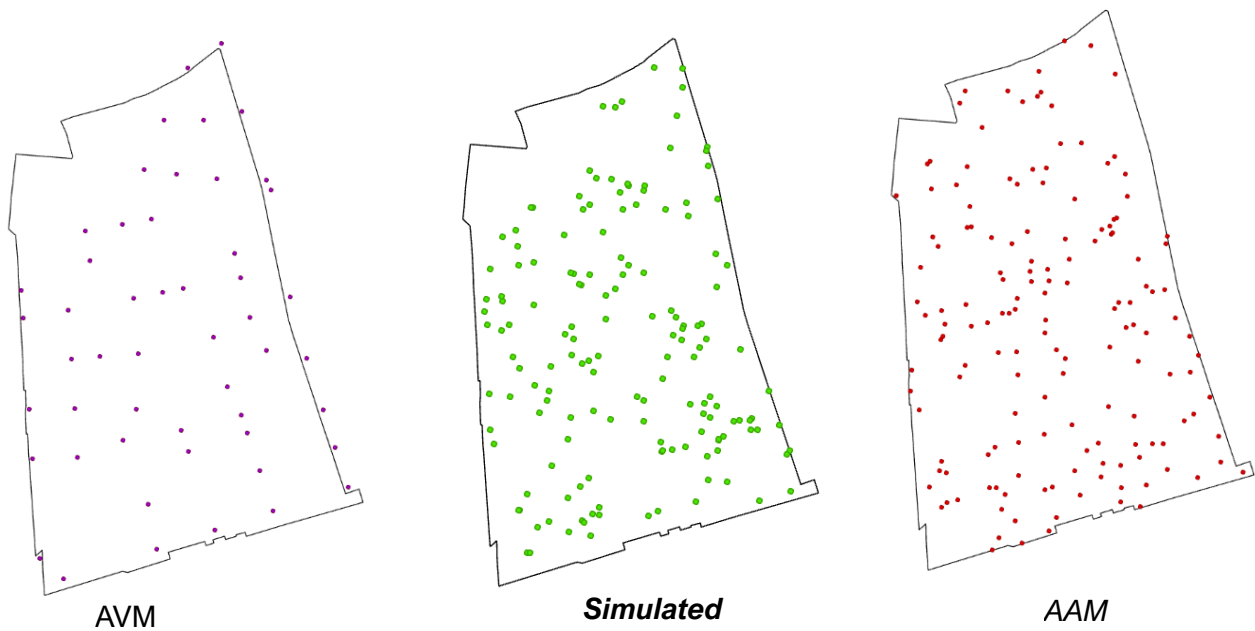


Figure 22 Point Pattern 1204 ©ArcGIS Pro

4.2.1. Overview of the Point Pattern

To get an overall impression of the point pattern, the geomasking results were compared to the simulated points using point pattern visualization.

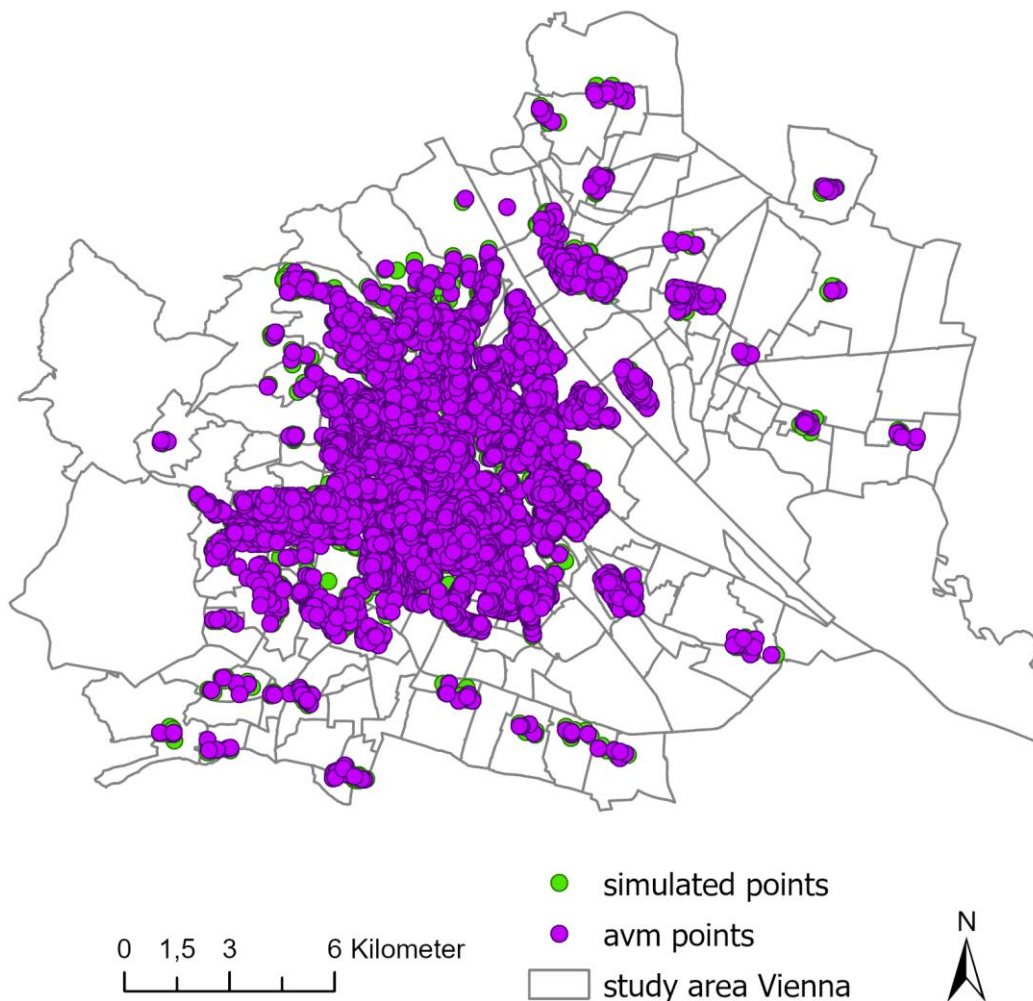


Figure 23 Points Pattern AVM – Simulated ©ArcGIS Pro

Here it can be clearly seen that the AVM points overlap with the simulated points, indicating a similar location. This should show that the AVM points are very close to the original (simulated) points. The visualization was done with the same symbol parameters. Based on the results of the AVM point patterns, this method moves the points to nearby intersections. In this way, the data was not scattered far from simulated points and the point pattern was preserved.

For this reason, further investigations such as Kernel Density Estimation were necessary.

The point pattern map shows that the AAM points look very scattered compared to the simulated points. This comparison is the same number of points: 13.830.

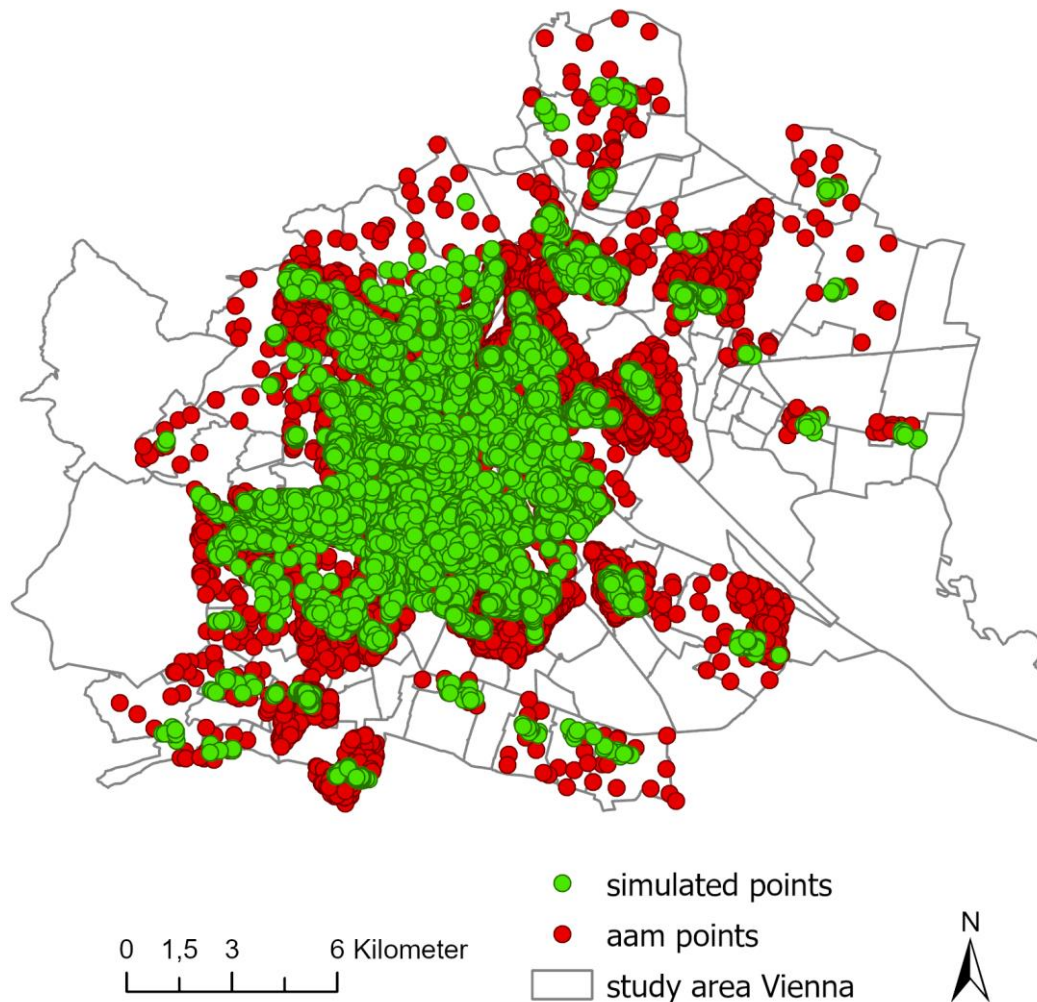


Figure 24 Points Pattern AAM – Simulated ©ArcGIS Pro

In health geography research, it is important that the health data remains meaningful even after geomasking, as well as that the spatial pattern is not distorted. Based on the results, it can be estimated that the AAM method would not be suitable in the healthcare sector, for example, to carry out accessibility analyses.

However, if one does not want to represent an original (simulated) data pattern truthfully and the false identification would be irrelevant, the AAM method could be used (preferably not in the health geography research).

4.2.2. Point Displacement-Techniques

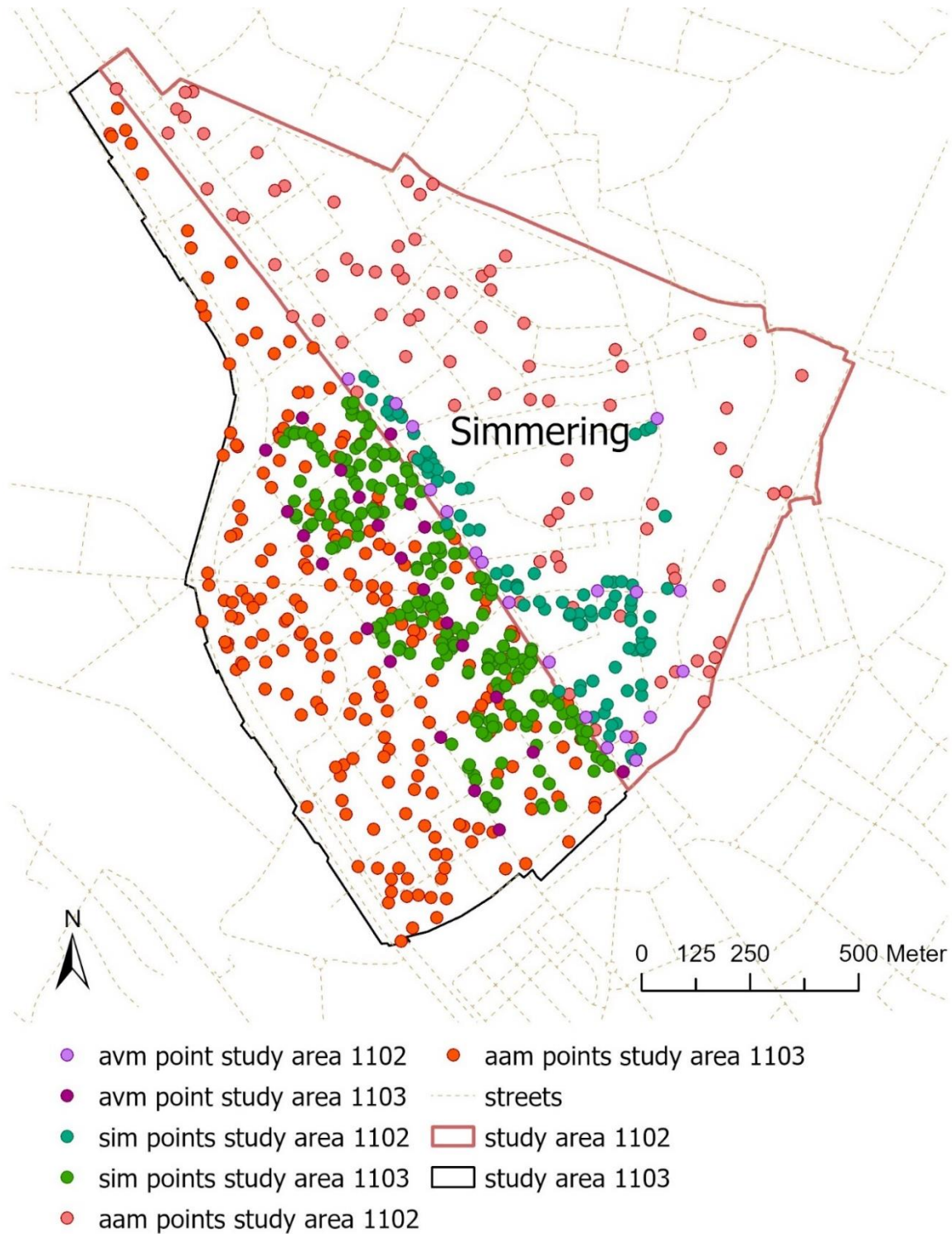


Figure 25 Simmering 1102/1103 ©ArcGIS Pro

Based on the study area of Simmering, the characteristics of the two geomasking methods can be visualized. The registration district 1102 (1103) has a density of 124 person per km² (198 p./km²) with a population of 13.584 (16.231). The differences in density of the registration districts 1102 and 1103 show two different results of the

points dispersion. There are two geomasking methods with the same SKA (RoRi) parameters. However, they produced different displacement distances. With the AVM method, the points were moved to the nearest intersection point, even outside the registration district. Already with the lower (RoRi 100) SKA parameters. For reasons of Voronoi polygons and road crossing.

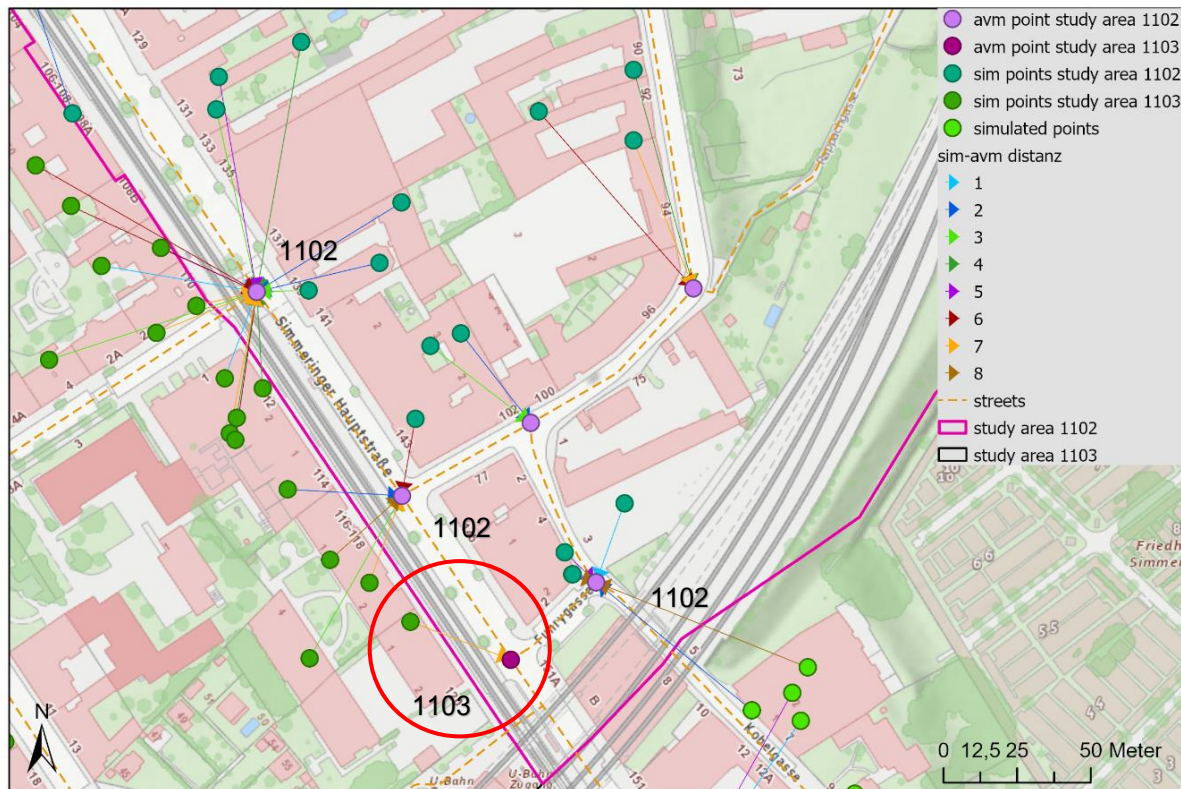


Figure 26 AVM Point Displacement K=100 ©ArcGIS Pro, Geoland Basemap

Depending on the simulated points (registration district) and the location of the intersection where the points were moved (registration district), the AVM points were assigned to the registration district numbers. Depending on whether they shared this intersection with other points. According to the masking results: If the point 1103 at the intersection was moved to 1102, the AVM point was linked to 1103. But if other points had been moved to the same intersection, this AVM point would have belonged to 1102. If points from both registration districts meet there, they belong to the registration district depending on their spatial location.

Visual inspection of the displacement distances revealed that the masking process also moved the simulated points into registration districts where no points previously existed. In low-density areas multiple simulated points (10) per (1) AVM point were masked. Especially in areas with lower density it is believed that it can minimize the risk of spatial re-identification. In the AVM method, the masked points have been moved beyond the boundaries of the registration district (RoRi 100), which means that the proportion of AVM points per registration district has changed.

The AAM method points (RoRi 100) were distributed within the registration district. For this purpose, further AAM analyses (also AVM for comparison) were carried out. Since the AAM method moves the points randomly, each run resulted in a different AAM point pattern that was very similar to the RoRi 100. The Result of the AAM masking SKA 1.000 RoRi showed how the mask shifts the points in other registrations districts:

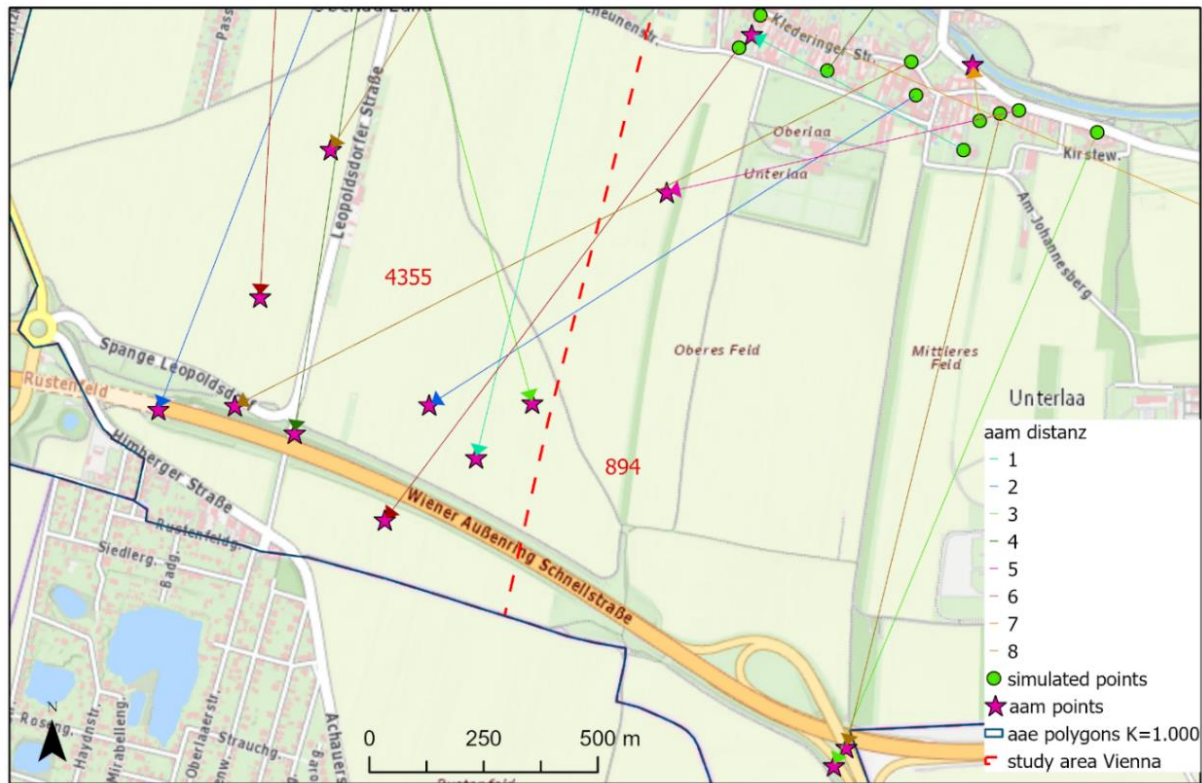


Figure 27 AAM Point Displacement K=1.000 ©ArcGIS Pro, Geoland Basemap

The Result for AAM was, that a higher RoRi-K value was required to be able to observe the shifts of the points outside the registration districts. Also, it was observed that the lower the density, the more the displacement distance of the points varied. As a result, the total point pattern appears scattered compared to the original (simulated) data set.

Displacement Based on RoRi	AAM-Method (100)	AAM-Method (1.000)	AVM-Method (100)	AVM-Method (1.000)
Average	540,60 m	543,10 m	49 m	49 m
Median	449,20 m	451,40 m	44,50 m	44,50 m
Std. Deviation	396,80 m	399,60 m	27,00 m	27,60 m

Table 2 Summary Statistics of the Mean (SKA) ©ArcGIS Pro

In the latter case, the AAM method shows differences in displacement distance compared to the AVM method. The AAM displacement distance increased minimally when the masking process was performed with RoRi 1.000. Based on the results of the statistical values, the AAM points were shifted 540 m and AVM points 49 m on average

from the simulated points. It is assumed that in areas with lower density, there is better protection with the AVM method compared to the AAM method. Similarly, the spatial location and the point pattern are retained because the AVM points are close to the previous locations of the simulated points. One might assume that the closer the masked points are moved to the locations of the original (simulated) points, the higher the risk of re-identification. But here, several simulated points have been moved to an intersection, which means that the AVM points are protected in multiple ways (Polzin, 2020).

4.2.3. Masked Points on Orthophotos



Figure 28 Study Area 1204 Orthophoto ©ArcGIS Pro

Subsequently, the results from the study areas were presented in descending order of density 1204, 1103, 2223. The orthophotos (Federal Office of Metrology and Surveying) show the positions of the points. Thus, there is a very high risk of spatial reidentification with the original (simulated) points and risk of false reidentification with the AAM method: The study area in 1204 has a higher density than the 1103 area.

While the AAM points are randomly distributed on the street, in the park, on residential roofs. The AVM points are only located at intersections. The road crossings are uninhabitable areas. The same applies to Simmering on the basis of the Orthophoto. As a result, the method minimizes the risk of re-identification of the persons.



Figure 29 Study Area 1103 Orthophoto ©ArcGIS Pro

Here it can be seen that the AAM points were also distributed to sports areas. The purple points on the other hand, follow a beautiful net-like image. Based on the orthophoto from Aspern, it can be seen that the AAM method has moved the points to the green area and to the street.

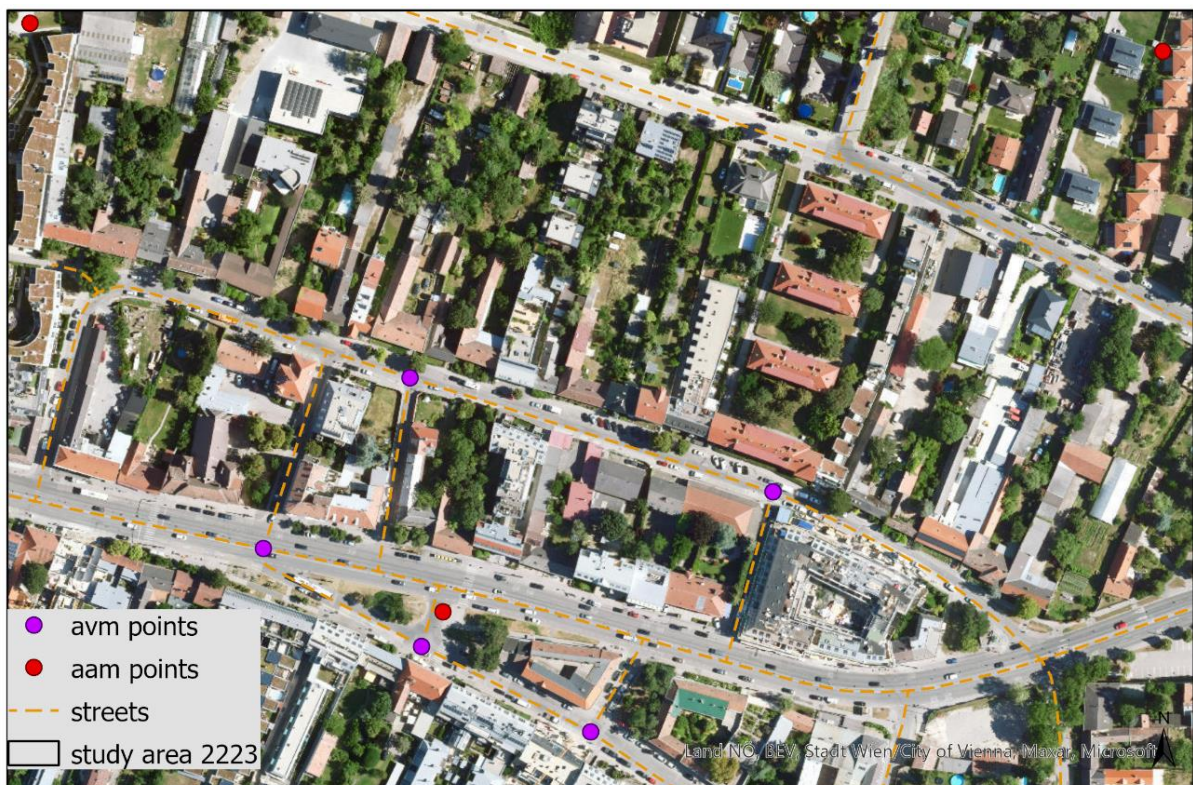


Figure 30 Study Area 2223 Orthophoto ©ArcGIS Pro

The AAM points seem to have been randomly distributed everywhere. The more free space, the fewer AAM points can be found on residential plots. This can be a problem in a high-density urban area. AAM points are also in unrealistic locations: This allows Map Viewer to see that the data has been geomasked. The simulated points were moved from residential building to near the road, but also to other residential buildings. However, if points are moved to other parcels, the risk of false re-identification increases. As a result, these people are included in a spatial analysis and thus influence the analysis but are also affected by the consequences of false re-identification to a "disease". The results make it clear that spatial analysis in health geography with AAM point is not recommended.

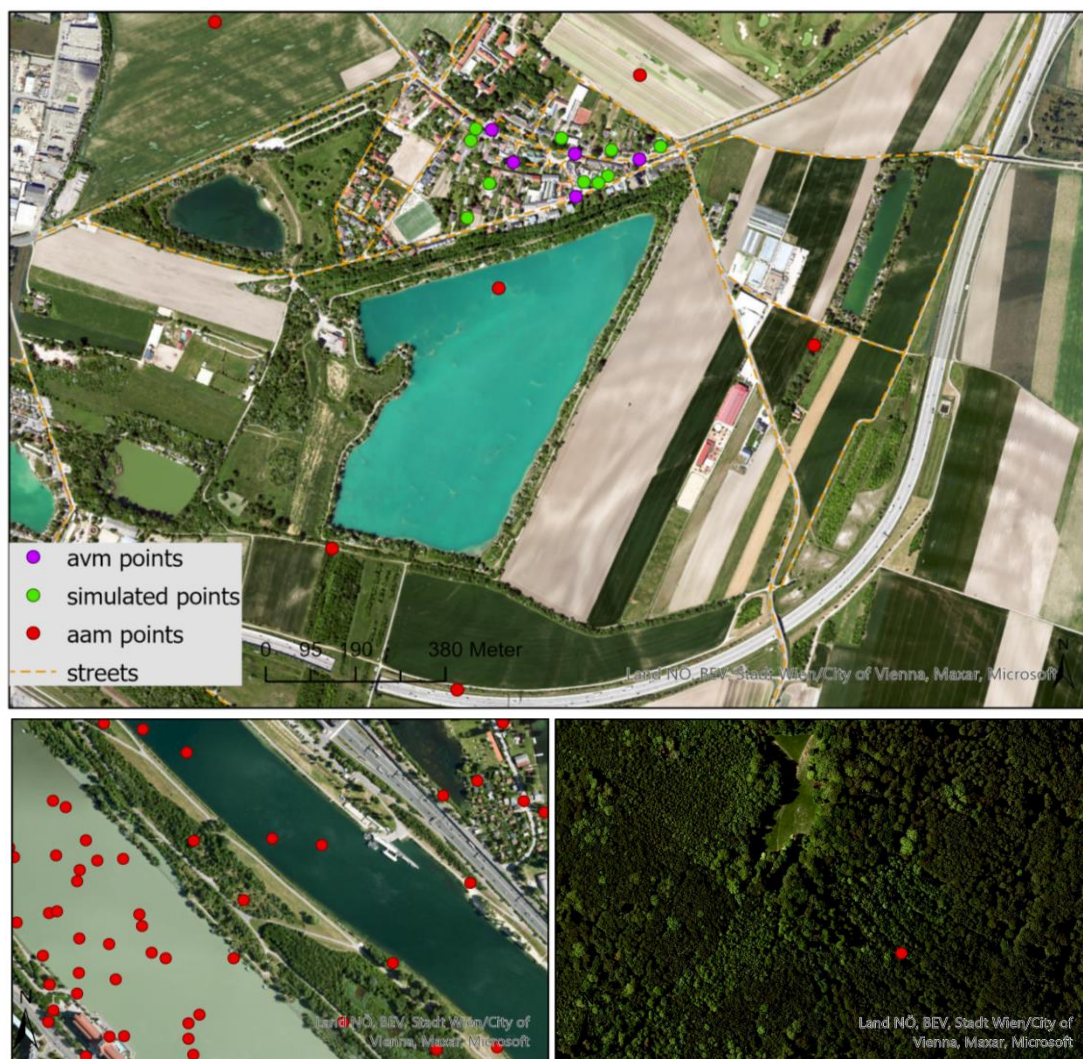


Figure 31 Vienna's Outskirts - Orthophotos ©ArcGIS Pro

On the outskirts of Vienna there are low-density areas with few residential buildings and a lot of agricultural land. Based on the Orthophotos, it can be seen that the AAM points are mainly located on fields, bodies of water, the Danube, also in the forest and on dirt roads. Only simulated points and the AVM points are here located in the urban area. It follows that the AAM method have the simulated points far removed from their original (simulated) position.

4.2.4. Reidentification of Point X,Y Coordinates

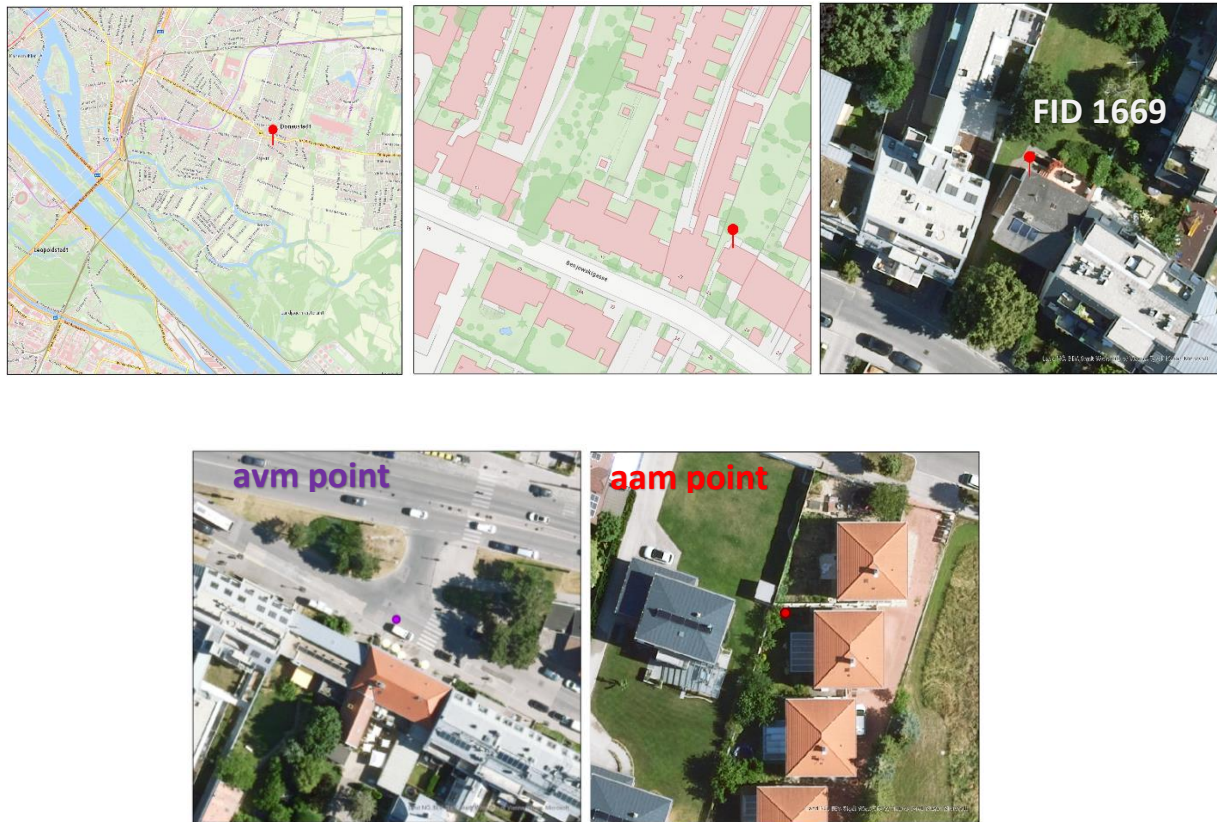


Figure 32 X, Y Coordinates-Geomasking FID-1669, Orthophotos ©ArcGIS Pro

Points	X Coordinates	Y Coordinates	Displacement distance
FID:1669	11471,969752 m	341979,017986 m	Original position
AVM	11536,778007 m	342087,962691 m	127 m
AAM	11845,4487 m	342334,9256 m	515,90 m

Table 3 Coordinates of Simulated-Point FID 1669

In this series of images, a simulated point can be seen with its coordinates and on masked locations, on a road intersection (AVM method) and on a foreign residential parcel (AAM method). The simulated point FID 1669 is located in study area 2223 with a very low density: According the Orthophotos results, it is possible to easily recognize the POI (point of interest), which results in an increased risk of spatial reidentification of the location of households of sick people. In the city center, on the other hand, the point data is superimposed, and an exact location of the point coordinates is only visible in very high digital resolution. It can be estimated that a spatial re-identification of point maps in the area with higher density could be easier in digital (scientific journals) than in print media. Because the points overlap. For this reason, point maps containing sensitive health data may not be published at the city level without first masking the locations of the original (simulated) data points. There would be no possibility of reidentification at the AVM point (Polzin, 2020) although false identification could occur at the AAM point.

4.2.5. Choropleth Maps at Registration District

The results of the Choropleth Map show the number of points per registration district. The illustration contains the symbolization of the Join_count field - points of the respective registration district - in color scale (6). The result of simulated points per registration district can be viewed here.

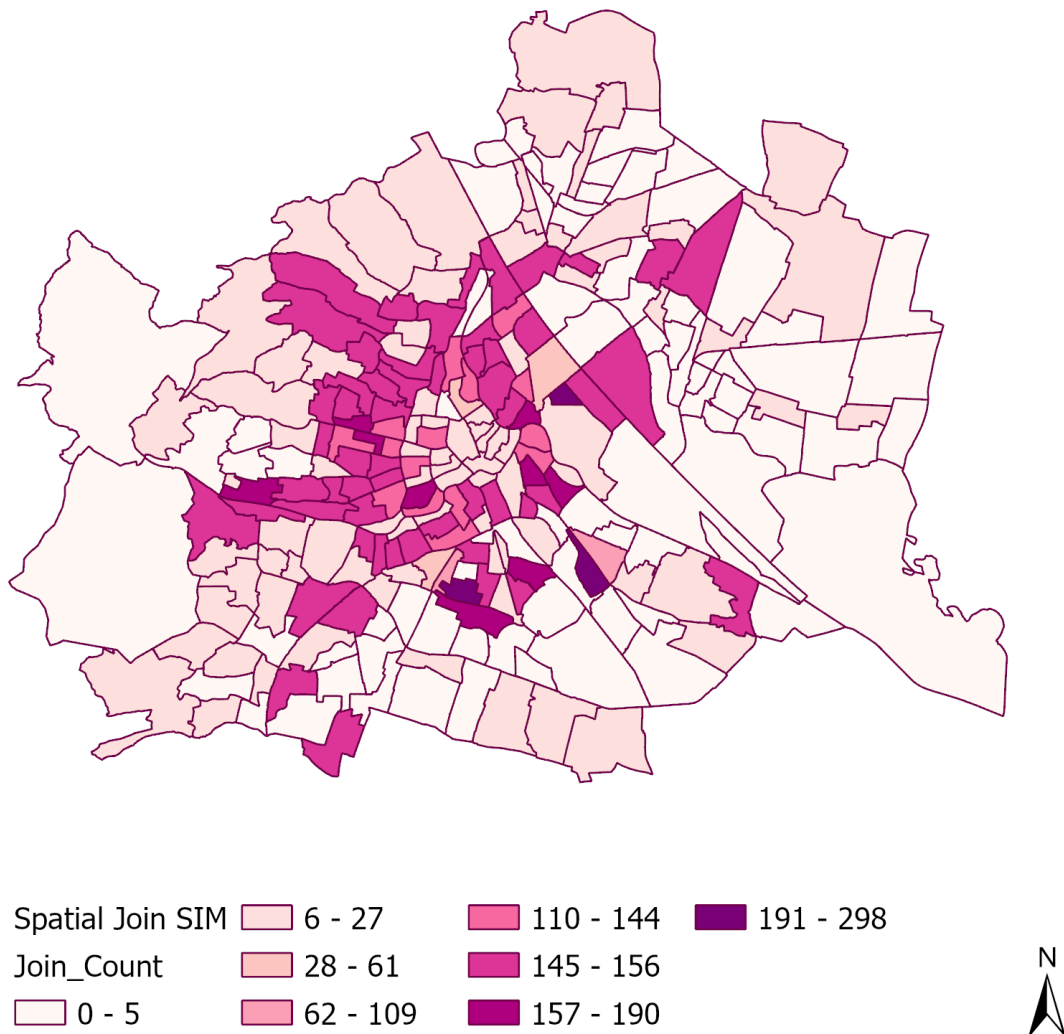


Figure 33 Choropleth Map Simulated (1:220.000) ©ArcGIS Pro

The Choropleth map of the simulated points and AAM points (RoRi 100) showed no difference. Because with AAM RoRi 100, the point ratio per registration district has not changed.

In contrast to the point maps, the data aggregated for each registration district displayed here. On the Choropleth map, it can be seen that the highest score per registration district with simulated points is 298.

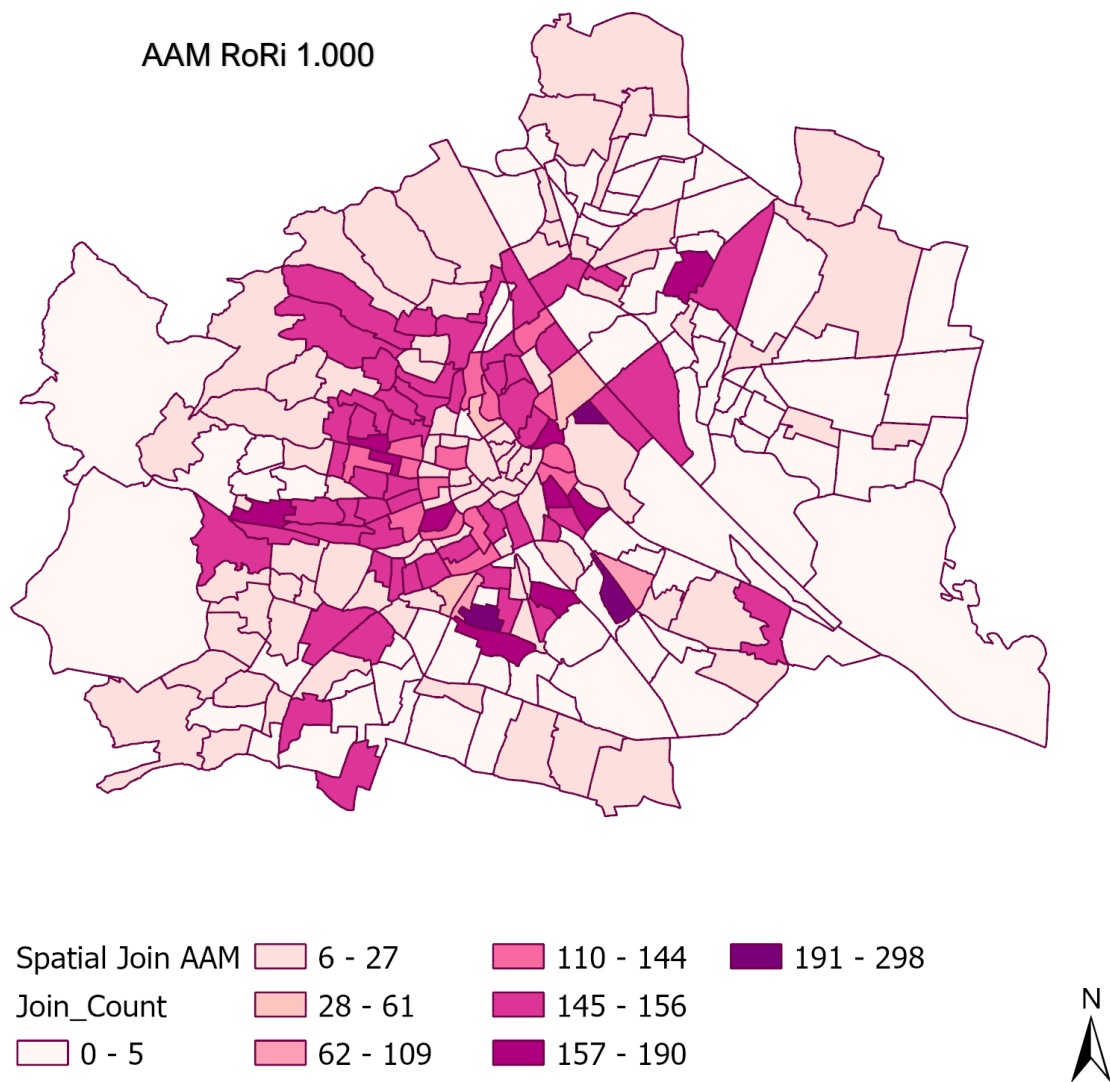


Figure 34 Choropleth Map AAM, RoRi 1.000 (1:220.000) ©ArcGIS Pro

The results show that it is important to choose the right RoRi value for the AAM method. There is a visible difference in the results of the Choropleth map AAM (RoRi 100) and AAM (RoRi 1.000).

After comparing the results with simulated points, minimal differences were found. Also here, it can be seen that the highest score per registration district is 298.

The overall picture of the study area has changed only slightly on the basis of this mapping (2012).

In contrast, AVM's Choropleth map already shows a clear difference in spatial anonymity RoRi 100.

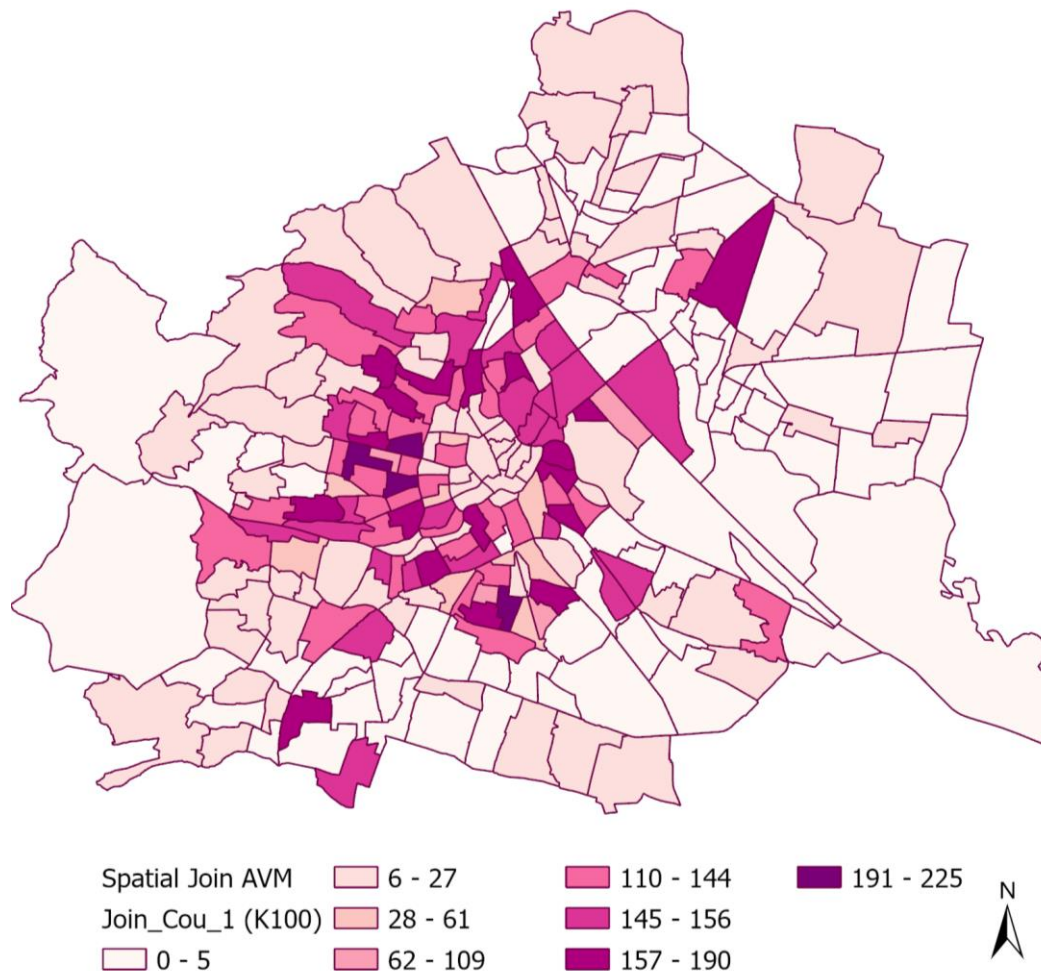


Figure 35 Choropleth Map AVM, RoRi 100 (1:220.000) ©ArcGIS Pro

The Choropleth map shows that the maximum AVM points share per registration district is 225 points. The AVM method moves points (several at once) to other registration districts, which changes the number of points in the registration districts.

Compared to Kernel Density Estimation, the Choropleth maps can be used to see how the number of points of each dataset per registration district has changed. This also changed the density in the registration districts. It allows a closer look at the ratios in the data set.

No direct points from the Choropleth map are displayed. However, the number of points per registration district indicates which registration district has the most, or fewest points, at which a risk of spatial re-identification could arise.

It is not clear from these results whether these points form a high or low density or whether they have similar values.

4.3. Hotspot Detection on Kernel Density Maps

In this step, the results of the Kernel Density Estimation (KDE) and Hotspot Analysis of the simulated dataset with the AAM and AVM datasets were explained. In this case, the KDE shows the density surface of the point pattern in study area. Gi technique belongs to the group of LISA statistics like Local Moran's I. The KDE and Hotspot Analysis are used together to study the data points for high concentration and significance.

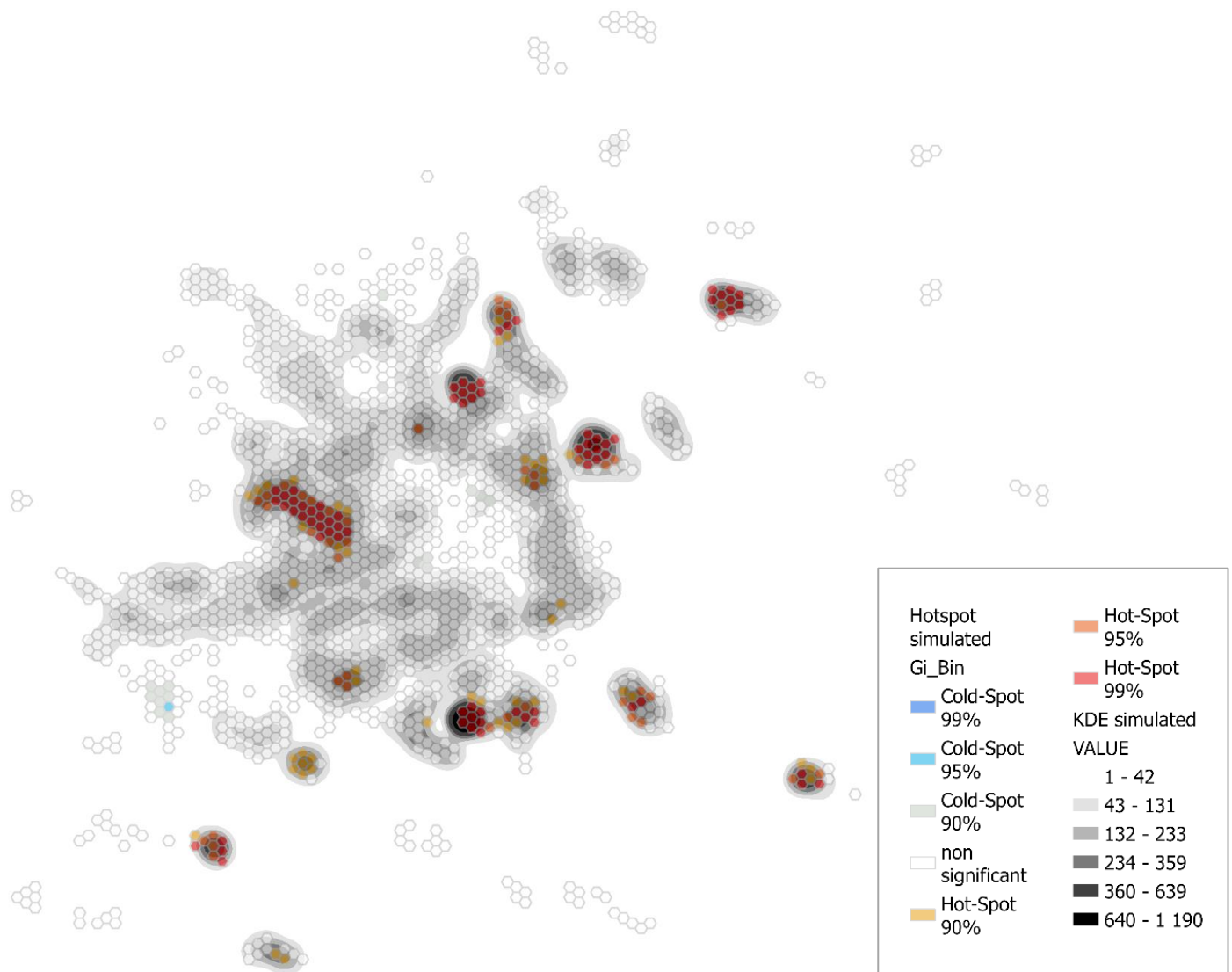


Figure 36 KDE and Hotspot Simulated-Points 1:100.000 ©ArcGIS Pro

The KDE and Hotspot Analysis results as a Map were presented at a scale of 1:100.000 and in a color scheme that distinguishes maximum values (black) and low values (white) of the point pattern densities (Kalinic and Krisp, 2018). Subsequently, the results of the kernel density analysis with 500 km bandwidth and cell size (5) were presented in 6 classes visualization according to Jenks. In order to be able to visually check the significance of the density surface of the point pattern, the results of Optimized Hotspot

Analysis (Hexagon) were transparently (50%) overlaid on top of the KDE results and color adjusted. The hotspot results in orange-red areas indicate a high incidence of positive values and a blue hotspot indicates a high incidence of low values. The gray color indicates the non-significant. The results of the Hotspot Analysis without specifying the field of analysis showed where point clustering is statistically significant, pronounced or low (Kalinic and Krisp, 2018).

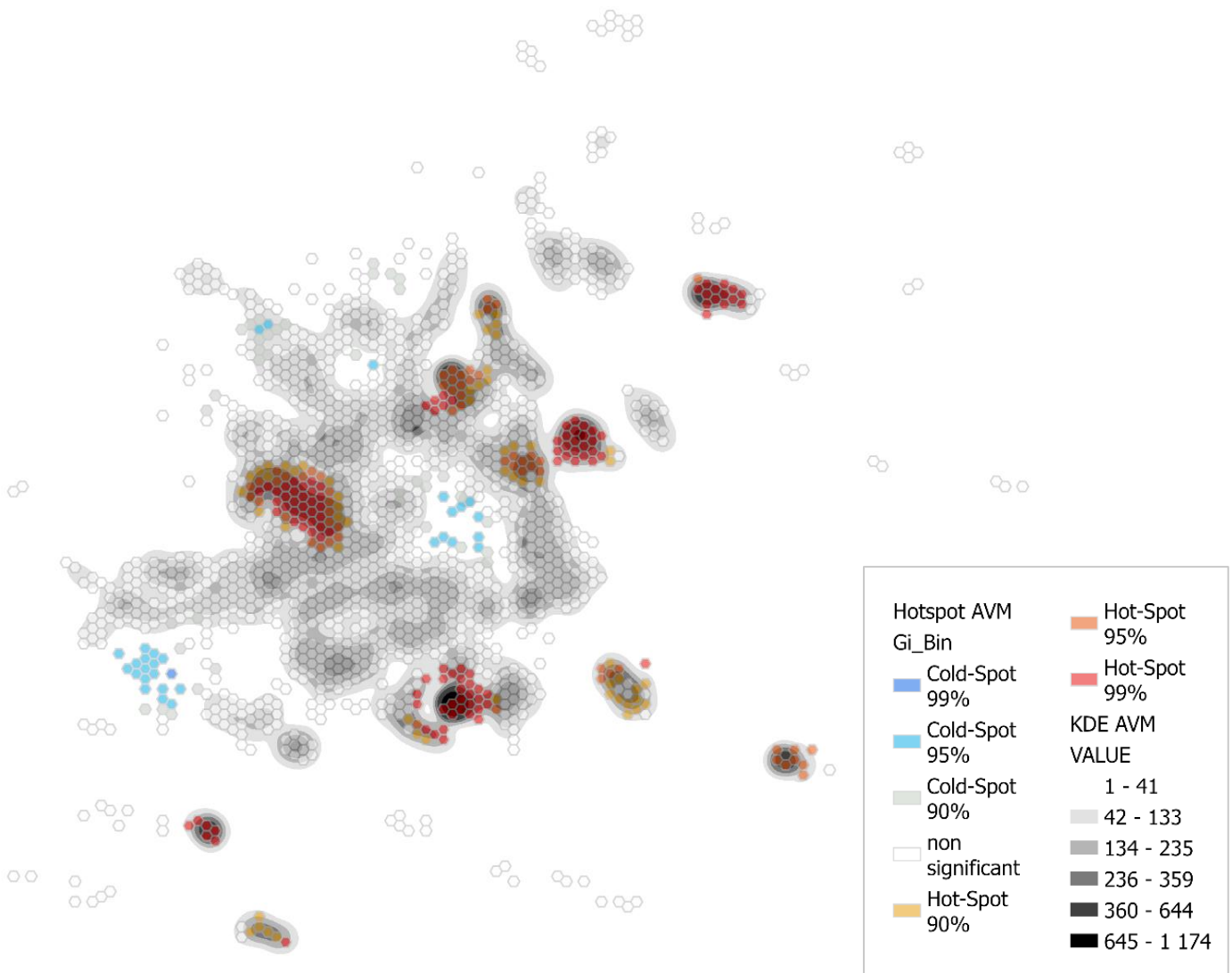


Figure 37 KDE and Hotspot AVM-Method 1.1:100.000 ©ArcGIS Pro

Based on the results, the AVM method preserves the point pattern better. Most of the high-density areas of the point pattern KDE and Hotspot map overlap. The results of the AVM points show a strong similarity to the results of the simulated points. Clusters with high values are also visible. Compared to simulated dataset, there is more cold spot in AVM data. In the last step, the analyse results for AAM are described. The AVM method has maintained the density of the point pattern because the AVM points have been moved close to the simulated points. On the one hand, according to the

Choropleth maps, the number of points and thus also the density in the individual registration districts has been (slightly) changed. But not the density of the point pattern. For location analyses, the AVM method would therefore be preferable.

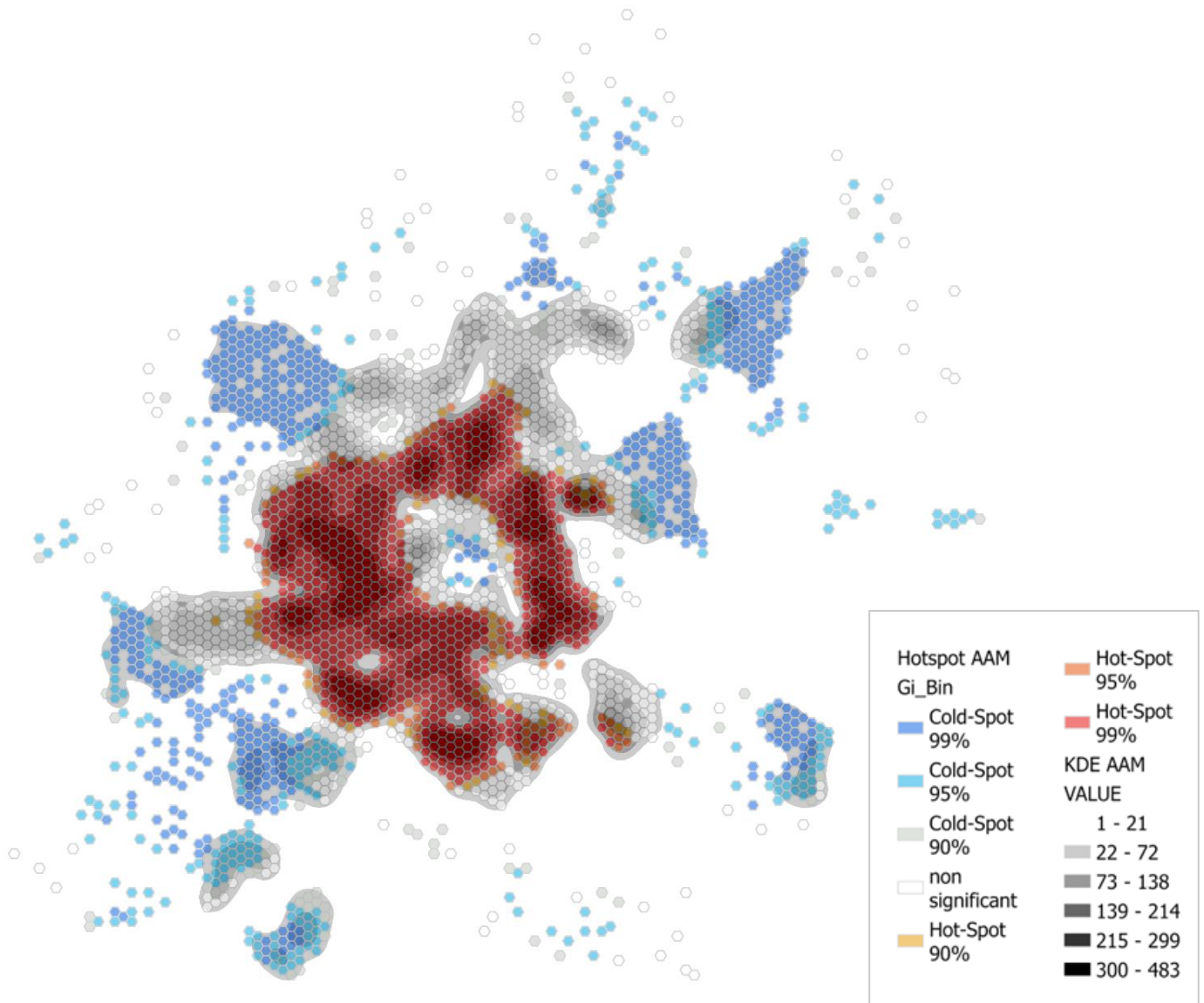


Figure 38 KDE and Hotspot AAM-Method (RoRi 100) 1:100.000 ©ArcGIS Pro

By comparing the results of the density of the point pattern of all three datasets, it became clear that the AAM points patterns are very different from the others. On the outskirts of Vienna, there is a low density of the point pattern (KDE), but there are many cold spots (Hotspot Analysis) to be observed. According to the result, there is a strong AAM dispersion on the outskirts of the city. The hotspots are located around the center. Based on these results, the density of the AAM point pattern was changed compared to other data within the registration districts and eventually to the total point pattern. The results of Optimized Hotspot Analysis: Neighbors of each point - Simulated points are 19, AVM points are 55, and AAM points are 61 neighbors. From the results of the

KDE and hotspot map, the risk of spatial re-identification can be estimated based on the visible hotspots. Especially in areas with low population density (Kounadi and Leitner, 2014). Based on the results of the KDE and Hotspot Analysis, the point pattern of AAM points in the study area of 2223 has a very strong cold spot compared to simulated and AVM points. The results of this analysis confirm the assumption from the literature that the better a geomasking method maintains the density of the point pattern, the better it can preserve hotspots (Charleux, Schofield 2020). However, this way, the masked hotspot area is close to the original (simulated) hotspot area. This makes it possible to map the geographical location of the hotspot or cold spot area more accurately. The AAM scatters the density of the point pattern, so it cannot be assigned a more accurate hotspot area compared to AVM, as it defines an entire area as a hotspot or cold spot.

4.3.1. Re-identification Risk of Hotspot in Low-Density Area

These results show the differences in KDE and Hotspot in 1102 and 1103 registrations districts. In this step, the risk of spatial reidentification from the results of the Hotspot Analyses with simulated and masked data points was investigated.

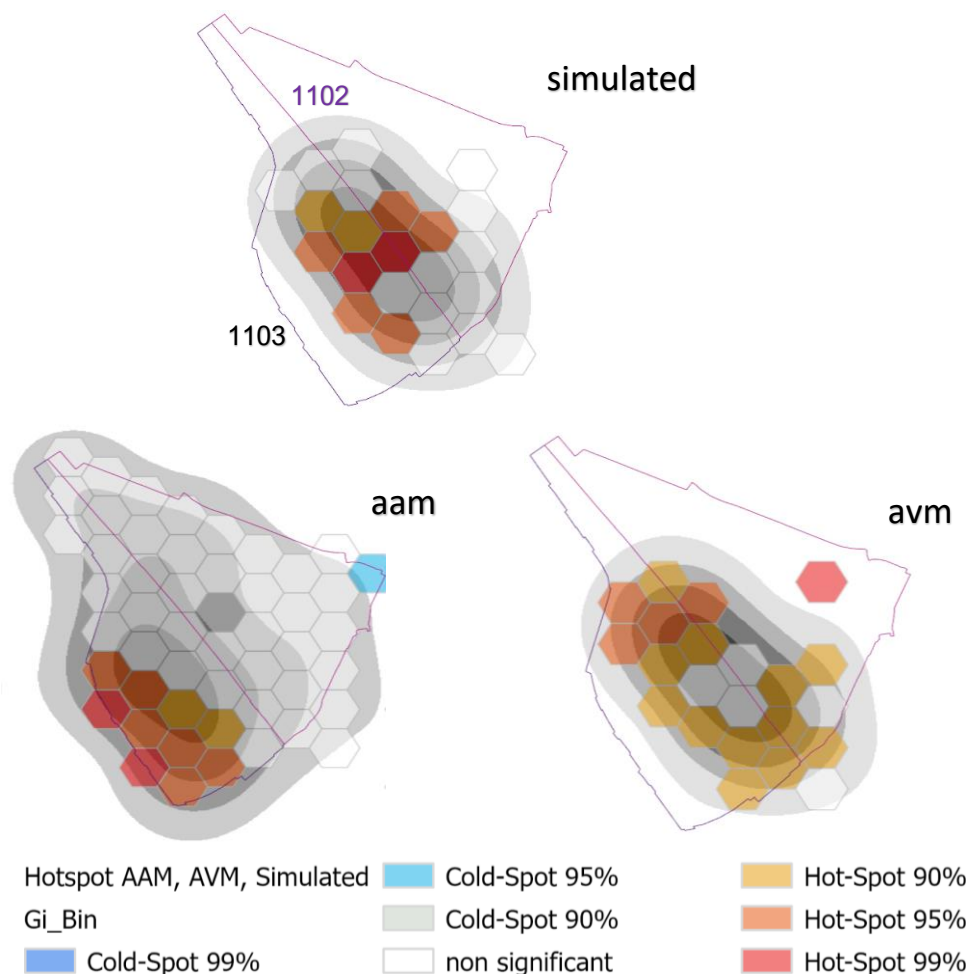


Figure 39 Hotspot Area 1103-1102 ©ArcGIS Pro

The result of the simulated data set showed how high the density of the points in the study area 1103/1102 originally was. As expected, 1103 shows a higher density of points and thus the occurrence of hotspots. The result of the avm masking method shows a higher incidence of 90% hotspot in the polygons border area of the two registration districts. AAM shows similar hotspot values to simulated points: 95% and 99% hotspot. In 1102, the hotspots are located from AAM and AVM points. The results of the Geomasks are different. But AVM shows almost similar results with simulated points.



Figure 40 Hotspot AVM 1102, Orthophoto ©ArcGIS Pro

The AVM method has identified a hotspot in a low-density area on 1102. Probably because there are 3 masked points at an intersection that have no other neighbor nearby. Although this area have a low KDE-based density, there is a 99% hotspot here. Hotspot identifies the concentration of points that have high values in this case.

The higher the z-score, the more intense the accumulation of high values (hot spots). For negative z-scores, it was the other way around: the lower the z-score, the more intense the accumulation of low values (cold spot) (Kalinic and Krisp, 2018).

There is a difference in the risk of spatial reidentification between Results Visualization with Hotspot-Area and individual points. The KDE analysis results (grid), and hotspot (hexagon) have “anonymized” the points location. It is not possible to reidentify the points, but it is possible to determine their geographical location, especially in areas of lower density.

It is important to consider whether the location of the Hotspot-Area should be mapped correctly, which masking method should be used so as not to increase the risk of spatial re-identification and false re-identification.

In next step, the points were placed below the hexagon grid and visualized with a higher zoom view (scale 1:3.000): The excerpts consist of three different hotspot areas within the study area of Vienna.



Figure 41 Hotspot areas at scale 1:3.000 - 1:5.000 ©ArcGIS Pro

According to the results, there is a difference between a 1:3.000 and a 1:100.000 scale map. Here, it can be seen that the higher the scale, the greater the risk of spatial reidentification (Kounadi and Leitner, 2014). The results of the hotspot analysis are based on the selection of the dataset. At the original (simulated) points, the positions would again be clearly visible and identifiable in display mode 1:1.300 (Simulated-0304). The Hotspot-AAM-0304 shows that the AAM points have occurred on residential buildings, which could lead to false re-identification. The next Hotspot-AVM-1103 shows that the purple AVM points occurred at intersections in several hotspots, so there was no risk of spatial re-identification. It can be assumed that a direct re-identification based on the hexagon results is not possible. Based on the advancing technology, it would be advisable to perform hotspot analyses with masked data.

Especially in Hotspot areas with lower density, it may be possible to identify the persons by linking further information from other databases (Benitz and Malin, 2010). There is also always a certain risk of re-identification (Kounadi and Leitner, 2014).

4.4. Spatial Autocorrelation - Global Moran's I

Based on these results it was possible to examine the datasets for spatial autocorrelation. The results of the Spatial Autocorrelation - Global Moran's I (K-Nearest-Neighbor, 8) were created in an HTML document and the information was summarized in Table for a better view. The Global Moran's I provided one value (index) per data set. Analysis results with point dataset, analysis field CID:

Global Moran's I	avm points (orig_cid)	aam points (random_p3)	Simulated point (cid)
Morans-Index	0,957318	0,939511	0,976730
Expected Index	-0,000072	-0,000072	-0,000072
Deviation	0,000015	0,000016	0,000016
Z-Value	249,4326600	232,258914	242,444851
p-Value	0,000000	0,000000	0,000000
Results	Clustered	Clustered	Clustered

Table 4 Spatial Autocorrelation - Global Moran's I ©ArcGIS Pro

Result of AVM points: "Assuming the z-score is 249.4326600, there is a less than 1% chance that this grouped pattern is a random outcome" (ArcGIS Pro). The same result came out for AAM- and simulated points.

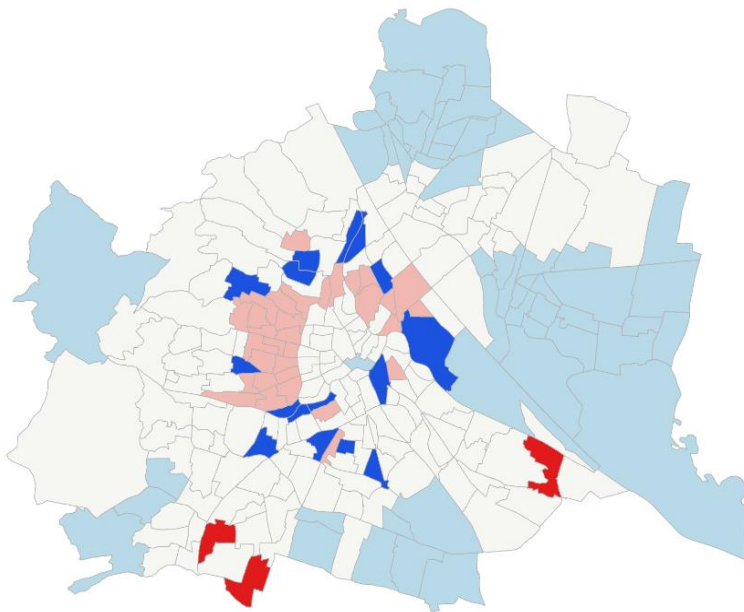
The result means that it is extremely unlikely that the clusters were formed by chance. There is a high Moran value and a low p-value (0), so clustering is significant and not random. The standard value for all three data sets was 3053.7879 meters.

It can be assumed that the three dataset points were not randomly distributed (ArcGIS, 2024). There is a spatial Autocorrelation. The points were identified in clusters.

A comparison of the results across all datasets shows that the AVM method is an outlier in terms of the highest z-value. It can also be assumed that there is a strong spatial autocorrelation in the datasets, as a high positive z-value can be seen. And that similar values are next to each other. Especially with the AVM.

4.4.1. Cluster-Outliers (Anselin) at Registration-District

Simulated



The results show the correlation and outliers values in study area on basis the 250 registrations districts. Unlike the points, here the data was grouped into registration districts using spatial join and the values for registration district were calculated.

The analysis was carried out with 8 neighbors' parameters. Thus, the districts that have different values compared to their neighbors were recognized as outliers.

Cluster-Outlieranalysis
HH Cluster
HL Outliers
LH Outliers
LL Cluster
non significant

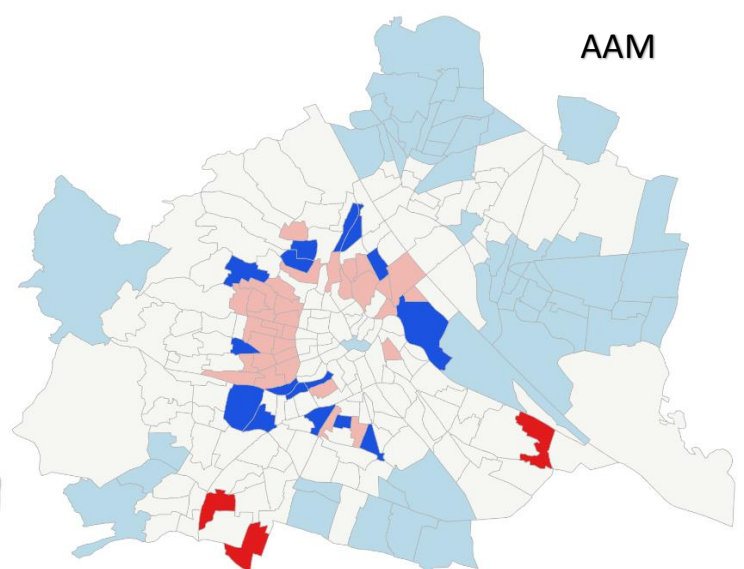
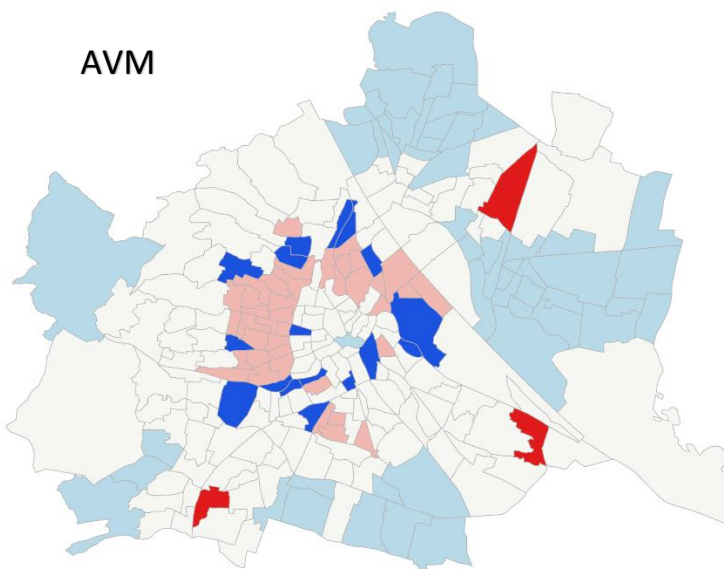


Figure 42 Cluster-Outliers Region Districts of Vienna ©ArcGIS Pro

This ESDA result can be used to identify particularly sensitive areas in comparison to the neighborhood. The AAM and AVM results differ slightly from original (simulated) data. You can easily observe differences in the region district outliers. From the point of view of the spatial reidentification risk, no individual point data are shown here.

4.4.2. Cluster-Outliers and Hotspot Maps

The results of the autocorrelation of the statistical spatial tool local Moran's I (according to Anselin) are point maps. These results also made it possible to identify outliers on the maps. The results were compared with Optimized Hotspots (Getis-Ord-Gi statistic) to show differences between these methods.

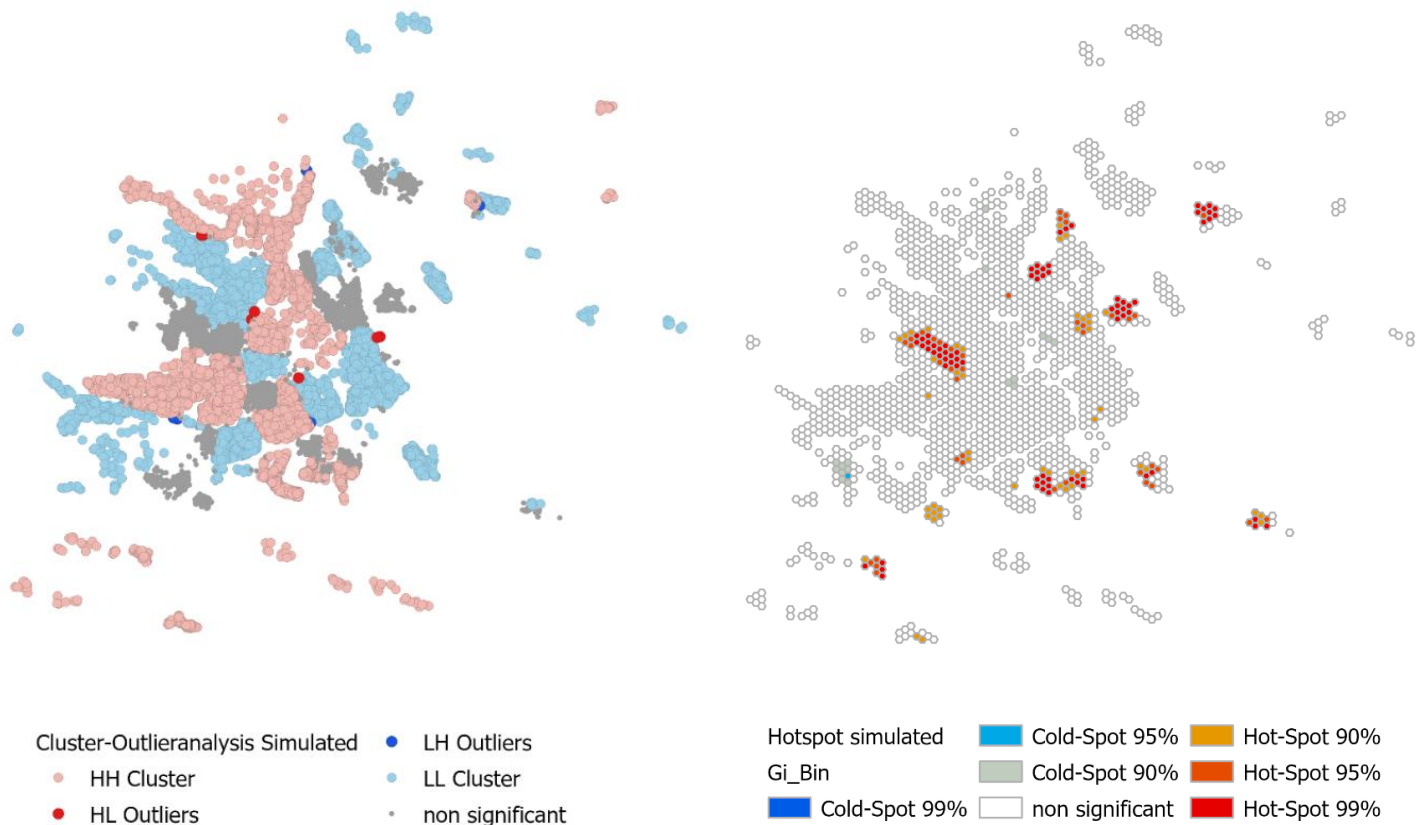


Figure 43 Cluster-Outliers and Getis-Ord Gi* Maps Simulated ©ArcGIS Pro

Outliers are points in the area whose positions do not correspond to the typical pattern of the other points in the data set, a distinction is made between global and local. The local outliers differed from their neighbors in the cluster, among other things by density position (ArcGIS(g), 2024).

It can be observed that the simulated data also contains outliers. The results of the analysis were presented in point maps - therefore the risk of spatial re-identification with original (simulated) datasets would be (depending on zoom view) high.

The Hotspot Analysis indicates where hotspots are located (significant) in the point dataset. The results of the analyses can be used to see where the data correlates with neighbors in the cluster. When high values in the cluster are together, but a point in the

spatial pattern is far away from them, that point has been identified as an AAM-Outlier. Depending on whether the clusters are pink or light blue and whether the values (CID) are low or high, red HL or blue LH outliers have formed.

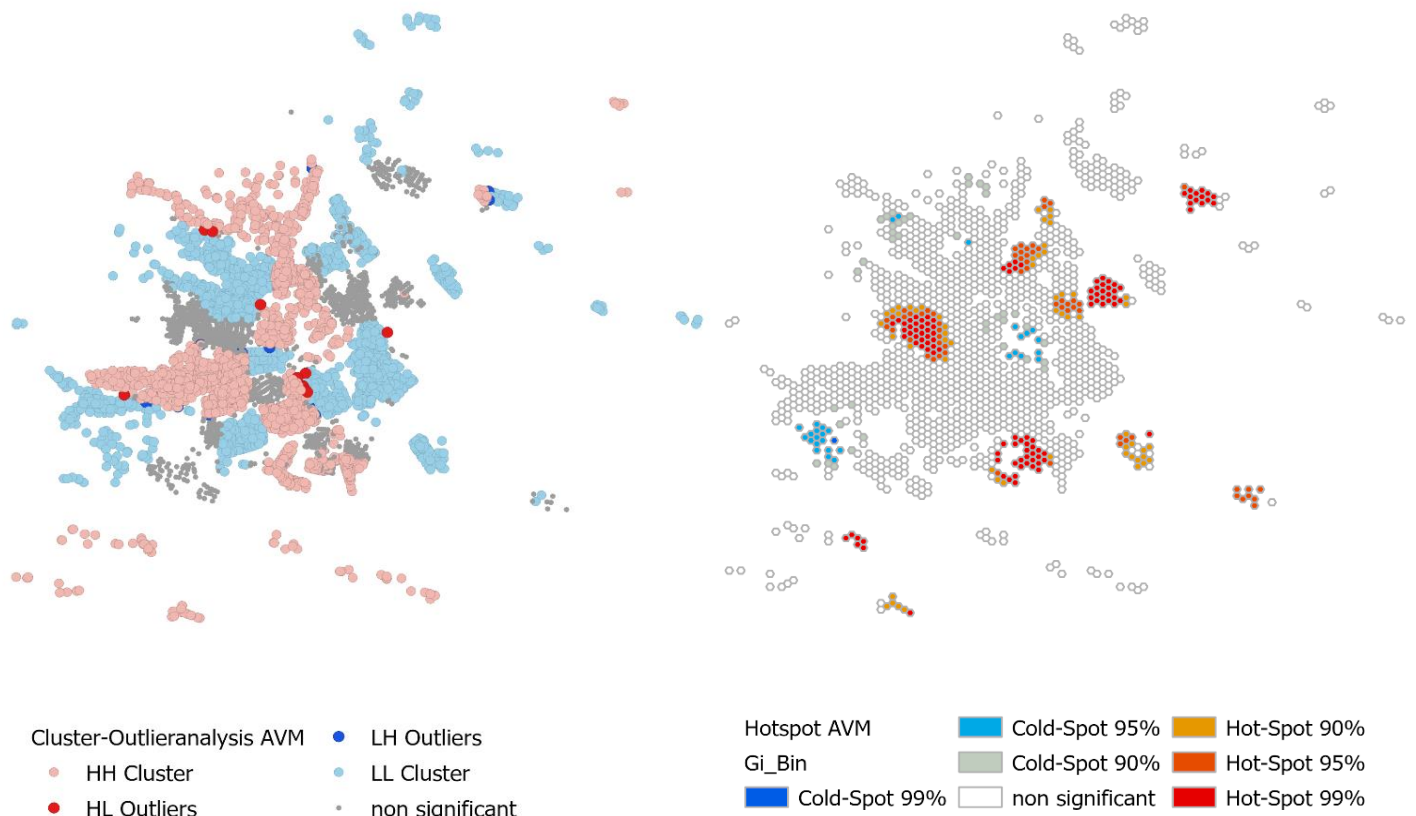


Figure 44 Cluster-Outliers and Getis-Ord Gi* Maps AVM ©ArcGIS Pro

The AVM and the simulated points show also similar patterns here. This cluster outlier analysis was used to examine masked health data at the point level – in the city of Vienna. The CID (ID) attributes of the points were used here. The results therefore concern this distribution. If the attribute were instead a disease of patients, these outliers would have indicated abnormalities related the disease:

The cluster and outlier analysis Locale Morans (Anselin) identifies the points in clusters based on their CID (ID) Number and neighborhood parameters (8). It is a detailed statistical method to be able to examine the data. In health geography, this method can be particularly advantageous in the case of a pandemic investigation.

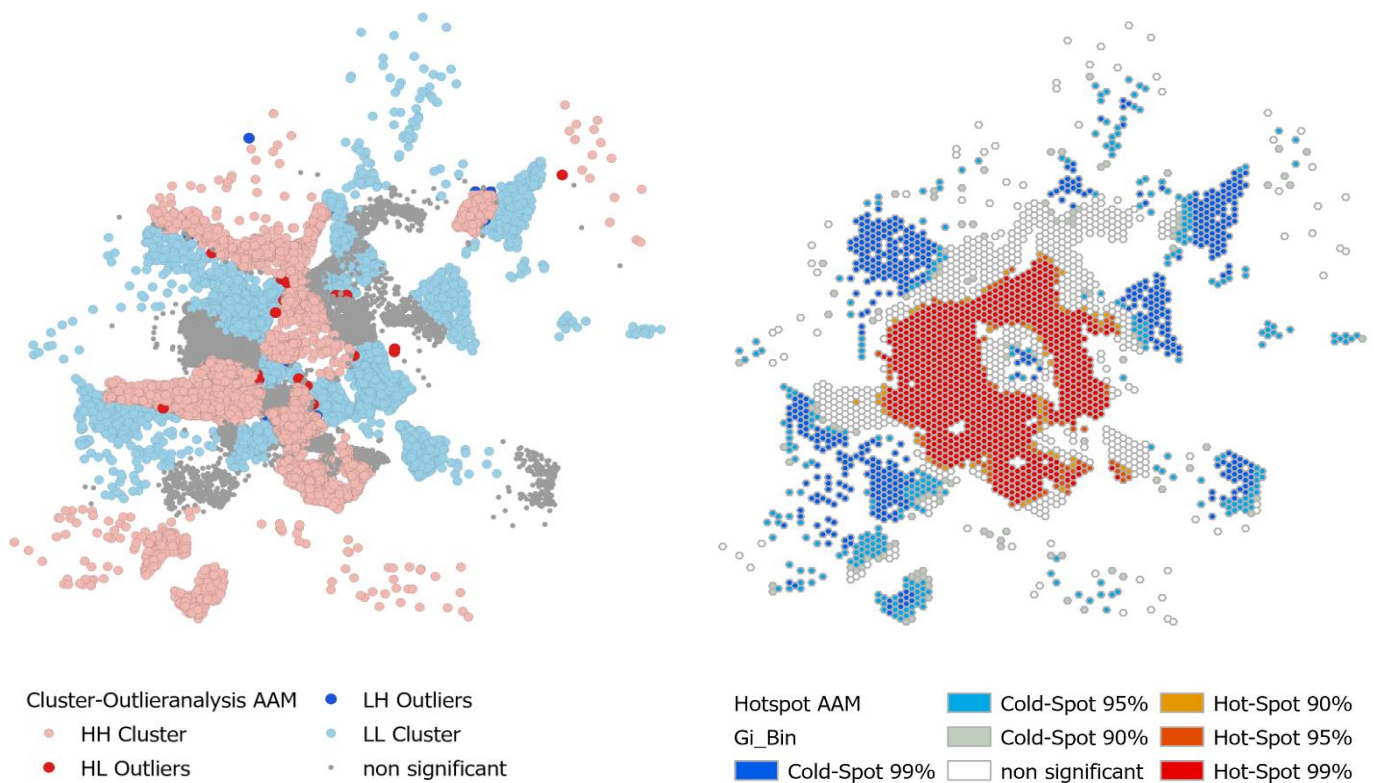


Figure 45 Cluster-Outliers and Getis-Ord Gi* Maps AAM ©ArcGIS Pro

In the hotspot result, no points are displayed, but hexagons. However, the Cluster-Outlier analysis examines the points based on the selected parameters – in this case, 8 neighbors.

The risk of spatial re-identification would be high here, as with other point maps, depending on the geomask and display mode used. Since the data is even more precisely defined and sensitive information is presented together, there is a high risk of spatial re-identification. This analysis yielded very revealing results. This point-level analysis should only be performed with masked data – in this case, AVM.

In the next step, the results of the AAM and AVM outliers were compared with original (simulated) points outliers.

4.4.3. Risk of Reidentification and Outlier's

The cluster and outlier analysis was carried out several times (4) to control the results. It turned out that the results are slightly different from each other but follow the same pattern. The number of LHs was stable in AVM but varied in AAM. In contrast, HL results for AAM were stable again.

Outliers' Analysis (Anselin)				
COType KNN 8 RS Permutation 999	999-1	999-2	999-3	999-4
aam points				
HH Cluster	4.594	4.600	4.604	4.596
HL Outliers	18	21	18	21
LH Outliers	19	19	19	19
LL Cluster	4.433	4.453	4.434	4.446
avm points				
HH Cluster	4.671	4.680	4.672	4.673
HL Outliers	17	17	17	17
LH Outliers	34	36	35	35
LL Cluster	4.855	4.836	4.844	4.838
simulated points				
HH Cluster	4.566	4.552	4.561	4.546
HL Outliers	7	7	7	7
LH Outliers	11	11	11	11
LL Cluster	4.573	4.564	4.569	4.573

Table 5 Outliers Results – Anselin Local Moran's I ©ArcGIS Pro

AAM's visual analysis showed that AAM outliers were shifted to greater distances than usual. The runaways were located on the outskirts of Vienna in the forest, above a tunnel. In urban areas, on subway or tram tracks and residential buildings.

With the AVM method, a better understanding could be derived. The outliers are close together in the area with the higher density - in narrow alleys. It should be mentioned that these outliers were mostly defined as non-significant and sometimes as outliers in the cluster and outliers' analysis with simulated data. In the lower-density area, these AVM points had few buildings around the intersection. Or the masked points (Outliers) were moved to a faraway intersection.

As a result, it was found that the outliers are a good method to investigate anomalies in the masking data set. Since the intersections are close to each other in densely populated areas. On the outskirts of the city, these are usually further away. The AVM method is based on intersections - so the street-network is important. It could be seen from the outliers, where intersections with few buildings around them are located (Building site), where spatial anonymity could potentially be improved. Especially in the area with lower density.



Figure 46 AVM Outliers HL, Orthophotos ©ArcGIS Pro

Subsequently, two AVM outliers from Table 999-1 were shown with an orthophoto. AVM masked point (Orig_CID 13.300): This is located on the outskirts of the city at an intersection - in front of a building. The second outlier was a masked point in the city center: This can be assumed because there is a "market" here and also not enough buildings.

Outliers	LH (blue)		HL (red)	
	LMiZScore KNN8RS	Object ID	LMiZScore KNN8RS	Object ID
999-1 Z-Value				
Points simulated	-1,99/-2,63	762-4552	-2,34/-3,92	8011-12508
AAM	-1,73/-4,93	56-4552	-1,69/-3,82	7111-13338
AVM	-1,76/-7,27	43-6482	-1,82/-4,06	8012-13300

Table 6 Outliers Z-Value ©ArcGIS Pro

The values of the Outliers z-score are summarized in the table. Compared to other datasets, the outliers of the AVM method have the highest z-scores. A calculation of the non-significant proportion in the three datasets showed that it was lowest for AVM and highest for AAM. The AAM results were similar to those of the simulated dataset.

For this reason, when it comes to health data with "highly sensitive information": The road data set should be adapted in advance, especially on the outskirts of the city. The results show that spatial reidentification risk in point maps is high compared to KDE, Hotspot, Choropleth or Morans maps at the registrations district level. AVM eliminates this risk and retains the pattern best with original (simulated) data. Therefore, it can be recommended to perform this analysis with AVM.

The results of Local Moran's I are part of the ESDA analysis. In order to determine the risk of spatial re-identification on the basis of these, the outliers were examined.

4.4.4. Average Nearest Neighbor (ANN)

The result of the analysis is an HTML report, which was written by ArcGIS in the Project folder and inserted into table for information. Results of the ANN analysis of the Points-datasets.

EUCLIDEAN Distance	avm points	aam points	Simulated points
Relationship with nearest neighbors	0,055079	0,501428	0,292795
Observed mean distance	4,7462 m	45,7071 m	25,3719 m
Expected mean distance	86,1702 m	91,1538 m	86,6539 m
Z-Value	-212,587415	-112,168258	-159,106176
P-Value	0,000000	0,000000	0,000000
Study area	410767560,703257	459654885,906003	415392690,277144
Results	Clustered	Clustered	Clustered

Table 7 Average Nearest Neighbor ©ArcGIS Pro

The index is the ratio between the observed and expected mean distance. The index is below 1 for all datasets, resulting in a clustered result (ArcGIS, 2024). Assuming that the z-score is -212.587415, there is less than a 1% chance that this AVM cluster pattern is a random result (ArcGIS, 2024). Again, the p-value is zero.

Therefore, it is also extremely unlikely that this cluster result was created by chance. According to the results, AAM and simulated point clusters are also not a random result. The AVM method displays the smallest distance to its neighbor but the highest z-value. This may indicate a high spatial structure in the dataset.

According to the results of the analysis, the distance of the average nearest neighbor (Euclidean distance) is the shortest for AVM points, followed by simulated points and AAM points.

This makes it clear that the average distance of the AAM points is wide compared to simulated points. On the one hand, the points in the AAM method were shifted far during masking. In addition, they are far apart, this results in the pattern of scattered AAM points patterns.

5. Discussion

In this step, the results were discussed on the basis of the research questions and the selected additional questions.

1. What techniques are commonly used by healthcare institutions to visualize patients' spatial data ?

ESDA consist of various techniques that are more advantageous compared to descriptive statistics. The health data can be visualized by means of diagrams, box plots, point maps, but also by means of complicated spatiotemporal representations. Spatial statistics such as means, and standard deviations are used to examine the datasets in more detail. Based on the survey results (5 participants) and the literature of Chapter II., it can be judged that point map analysis is the most commonly used ESDA technique. For privacy reasons, point maps with discrete patient data may only be displayed geo-masked or aggregated on one level. Without masking techniques, they violate the GDPR – because re-identification can be done based on the location coordinates (Polzin, 2020). Because the health data are aggregated at the regional level, they are also statistically examined at the regional level. However, the spatial analyses at the regional level are not as informative as the individual data at the city level are (Koller et al., 2020). For a better overview, the aggregated data can be displayed in Choropleth Maps, Kernel Density and Hotspot Maps. In the results of the hotspot analysis (Getis-Ord-GI statistics), significant hot or cold spots can be identified (Kalinic and Krisp, 2018). In particular, the Local Indicators of Spatial Association (LISA) according to Anselin (1995) offer the possibility to recognize spatial structures of data and to understand complex spatial patterns and relationships (Anselin 1995). Patterns can be identified in datasets that cannot be detected using traditional spatial techniques (Delmelle and Kanaroglou, 2015). The Local Moran's Index measures the spatial autocorrelation of a given point compared to its neighbors, thus identifying the outliers at the local level. The result values can be in the form of point maps. Global Morans I, on the other hand, examines the dataset on a global scale. Cluster- and Outliers Analysis (Anselin) is very meaningful in a pandemic and can be used to protect against the spread of the disease (Delmelle and Kanaroglou, 2015).

2. What other ESDA techniques are “less popular”, but very relevant and informative?

Cluster-Outlier Analysis (Anselin) can determine (masked) points that significance differ from their neighbors based on the selected attributes, spatial location (distance, density). This allows the masked address data to be better examined and it is assumed that the masking process can be improved - especially in the area with lower density.

In addition, Cluster-Outlier's combined with the AVM method enables very informative cluster analysis with sensitive data at the point level.

Also, the average nearest-neighbor metric of ArcGIS can also be mentioned. This is well suited for measuring the dataset for clustering and average distance between individual points (ArcGIS(g), 2024). In addition, the Global Moran's I (K_nearest_neighbors) can be mentioned, which is used to evaluate the spatial autocorrelation and clustering datasets, as well as outlier datasets.

As part of the reidentification, ESDA techniques can be used to investigate the effectiveness of the masking methods. A very commonly used method is point pattern analysis. However, other ESDA techniques have also been mentioned, such as the technique of "hotspot divergence" of masked and unmasked data (Kounadi and Leitner, 2014). ESDA techniques such as the cross-functional "Ripley-K function", distance to nearest K neighbor, and hierarchical cluster analysis of the nearest neighbor were also applied in this context (Kounadi and Leitner, 2016).

On the other hand, the "Gaussian distortion" (Cassa et al., 2008) enabled the spatial reidentification of masked data. Another possibility of spatial reidentification of geocoded patient data was carried out by means of supervised classification. The location data were extracted from this literature (Brownstein et al., 2006).

Supervised image classification is based on training data and includes statistical and spectral matching (ArcGIS(i), 2024). Can also be based on the concept of the Convolutional Neural Network (CNNs) from the field of deep learning - machine learning. With these techniques, it is possible to detect clusters and hotspots in images (ArcGIS(j), 2024). Unsupervised classification uses statistical differences in the spectral properties of pixels (ArcGIS(h), 2024).

3. What geomasking technique can be used for sensitive geospatial data that enables the application of a range of ESDA best techniques and minimizes or eliminates the spatial risk of re-identification?

Art. 9 regulates the processing of confidential personal data in the event of the disclosure of health data (GDPR). The analyses based on discrete health data that have not been obfuscated or aggregated are not GDPR compliant (Polzin, 2020). Geomasks that obfuscate health data in a way that ensures complete confidentiality of the data and there is no risk of household-address data being re-identified are GDPR compliant (Polzin, 2020). Thus, spatial analyses using geomasking techniques, which pose an increased risk of re-identification or false identification of the address data, do not comply with the provisions of the GDPR (Polzin, 2020).

This is because the "discrete" address data contains X and Y coordinates. These can determine a person's location, which is why the location is linked here as a "personal

identifier" to the "right to personal data protection" (Polzin, 2020). In practice, there is no Geomask method approved by the Data Protection Authority.

Based on the results of the ESDA techniques, the AVM method can be recommended as a suitable method for spatial analysis in health geography. According to the results, the method retains the spatial point pattern and obfuscates the data using a complex technique that is spatially impossible to identify. The AVM method shifted the simulated points at intersections that do not represent residential address data. At the same time, several simulated points were moved per intersection, "ensuring multiple protection" (Polzin, 2020). The AVM method slightly changes the point pattern density in the study area, but not the point pattern. The AVM points are close to the previous positions of the original (simulated) points. Since the masked points at intersections have also been moved to other registration districts, the number of points per registration district has changed, which means that the density of each registration district has changed. Thus, the AVM hotspots are similar to the unmasked (simulated hotspots) data. This method is therefore suitable especially for location analyses, or detailed spatial cluster observations in a city. The result of the AVM method forms a beautiful net-like overall image of the masked AVM points over the entire study area. Also, the AVM method was studied in detail in the study by Kounadi and Polzin (2021).

The AAM method changed the density of the point pattern when analyzed with RoRi $k=100$. In the AAM method, the points are moved over long distances, including to other parcels. The point pattern seems to be highly scattered. Especially on the outskirts of the city. The AAM points are also located in bodies of water or forests, which means that a site analysis in health geography with very sensitive information would not be recommended. In addition, this method carries the risk of false re-identification. With regard to location-based analyses, the AVM would be preferable. In health geography, a geomask is intended to ensure the privacy of the individual and maintain spatial similarity to original, discrete health data (Polzin, 2020). It means choosing appropriate masking methods to publish the data or conduct further research analysis. The results show that the AAM points were observed on other people's residential buildings in the urban area. Therefore, it can be assumed that this method would not be suitable for visualizations with patient data in densely populated areas if one does not want to risk spatial false reidentification: Depending on the display mode and map properties, whether it is interactive online point maps (web maps, dashboards) or printed maps on a scale of 1:100.000. According to Kounadi and Leitner, confidential information in scientific articles can already be disclosed in the display mode of 1:56 (Kounadi and Leitner, 2014).

4. How can map attackers tell which geomasking method was used?

In general, knowledge of geomasking is a prerequisite for the spatial identification of geomasked points. Each geomask uses a different obfuscation method. The traditional geomasks move points randomly, while the adaptive methods also take into account the density of the study area and the spatial K-Anonymity SKA (Kounadi and Leitner, 2014). With sufficient knowledge, a map viewer can recognize the geo-masked technology based on the locations. If the masked points (household address data) are displayed at road intersections, Map Viewer could use the AVM method or the street masking method (Swanlund et al., 2020) as the masking method. There are reidentification techniques and software that can identify a geomasking method. Point maps in particular should be mentioned in this context. With the KDE hotspot maps, it is not possible to know which masking method was used without additional information. On the other hand, in point-level cluster analysis, as in other point maps, the masked method can be seen on a larger scale. According to the results, the AAM method shifts the simulated points to green spaces, dirt roads, sports areas, bodies of water, sidewalks, roads, and worst of all, in terms of spatial misidentification, to other people's residential buildings. It is believed that a normal user cannot accept the AAM points in water bodies as a place of residence. This allows Map Viewer to assume that the locations of the households' have been manipulated and that a geomasking method has been applied. If a map viewer overlooks the fact of geomasking because the masked data can be seen on (foreign) residential building's (as with AAM in the densely populated area of Vienna), it can assume the masked data as actual address data. An incorrect spatial assignment of a disease to the address data of households can have serious consequences for the incorrectly identifiable persons and falsify the analysis evaluation (Kounadi and Leitner, 2014).

Additional questions:

What are the implications of data protection for research?

According to the results (5 participants) of the lime survey, data protection has a major influence on spatial analyses in "medical and geoinformation research". It is not permissible to look at individual persons. Spatial analysis and statistics are common in health geography at the regional level because the location data is aggregated at the regional level. In the event of a new disease outbreak, as was the case with Covid-19, accurate tracking of the spread of the disease in a city or municipality based on location-based health data of individuals is theoretically possible, but not permissible in practice. During the Covid-19 pandemic (2021), contact tracking was carried out to monitor the spread of the infection in the area and to isolate the infected persons, on

the one hand on the basis of the risk of infection itself and the contact data. The data was not analyzed based on the locations of the infected people. In the case of the Corona pandemic, the health data was evaluated regionally and treated as pandemic research according to the GDPR. The patient data is sensitive data and very well protected by the GDPR, for this reason it must not be made available to the public. In Austria, health data is stored electronically in ELGA. Only the persons concerned have access to the ELGA (doctor/patient). Spatial analyses depend on the accessibility of the data, which means that public spatial analyses with health data are not possible in health geography, or they are carried out "internally". Based on the current data protection regulations, spatial analyses are only permitted in aggregated form.

With the AVM geomasking method, it is quite possible to perform a cluster analysis with points without increasing the risk of spatial reidentification. One suggestion would be to propose the AVM method as the standard masking method of the Medical-Association and to have further spatial investigations carried out by the Vienna-Medical-Association for effectiveness. It is necessary to implement a geomask that complies with the GDPR and enables the exchange of data and the publication of results of the spatial analysis in health geography. It would be advisable to implement a standard way (EU GDPR) to be able to protect health data uniformly in spatial analyses. According Polzin, the AVM method is "GDPR-compliant" (Polzin, 2020).

However, the risk of spatial reidentification should continue to be viewed critically - highly sensitive (health) data is involved.

How high is the risk of spatial reidentification with the chosen techniques?

According to the results the original points (simulated) in the points maps can be re-identified using x, y coordinates. As a result, it is important to disguise the coordinates in such a way that the address data of the households cannot be traced. The adaptive geomasking methods take population density into account and refer to spatial K-anonymity (SKA). The RoRi file was used to create the anonymized polygons AAE. In this study, the RoRi value with 100 person per registration district (Population data 2012) was chosen: In general, there are registration districts in the city of Vienna that guarantee this number, because on average the population is more than 100 per registrations district. But in registration districts that do not meet this value, the anonymization polygons have been merged - the polygons have been reduced from 250 to 244. However, there were no simulated points in these registration polygons, so no point shifts were observed. For this reason, the analysis at AAM method was again conducted with 1.000 to show the differences in masking technology. At AVM method, it was already evident from RoRi 100 that the points had been displacement outside the registration district.

Based on the point map analyses, AAM points were shifted to third-party residential buildings, especially in densely populated areas. Accordingly, there is a high risk of false re-identification here. In areas with lower density, they were shifted to green spaces, bodies of water and fields or sidewalks. The more open space there was, the further away they were from the residential area. The results of the Cluster-Analysis according to Anselin have defined outliers for both masking Methods. Based on the AAM result of the Cluster-Outlier's analysis, the AAM outliers were shifted very far during masking process. Since the masked AAM points are randomly shifted, an understanding could only be determined by the z-value. The points in the city center were mainly located at rails and stations.

With the AVM method, on the other hand, there is no risk of spatial reidentification from the point data, according to Polzin (2020). Neither direct spatial re-identification nor false identification of households' address data (x, y- coordinates) should be possible. According to Polzin (2020), due to the very complex procedure, it is difficult to find out the original location and thus identify the people. The spatial K-anonymity of several data points per intersection further minimizes the risk (Polzin, 2020). The probability of determining the right house from "all" buildings around the intersection – without knowing whether the right building is located at this intersection – is not possible (Polzin, 2020). Here it was assumed that an AVM-Outliers point in the area with low density indicates the importance of the road network density (area density) and the spatial anonymity (SKA). Here it can be recommended to carefully examine the road network for gaps in advance – especially in areas with low population density. For the connection between buildings and the road network in order to minimize the outliers in the data. Also, perhaps the area could be examined for current (building demolition) construction sites on the basis of satellite images.

The results of the KDE and Hotspot Analysis can be viewed in Hexagon Netz. The KDE grid result has been smoothed out by adjusting the cell size (5): This makes the KDE result less pixelated. Through these analyses, the points were "anonymized", so to speak. Therefore, there is better protection against spatial reidentification for Hotspots on Kernel Density maps, Cluster maps based on registrations districts, and Choropleth maps. Therefore, the parameters of the geomask, such as the displacement parameters, density, scale, should not be made available to the public. This is because linking this data with data from other databases could lead to a re-identification of the households' address data. For this reason, it is important to carry out the anonymization procedure in advance. The results of Global Moran's I (spatial autocorrelation) and the average analysis of the nearest neighbor are summarized as index values in a table. This allows for a general understanding of the data displayed. Depending on the scale, the Cluster maps of the points can increase the risk of re-identification. It is recommended to perform this analysis with masked data: the AVM method is recommended here.

Does data protection restrict the disclosure and use of health geodata in research or to inform the public?

The health data is stored in ELGA in Austria, is available at the city level (and smaller) but is not accessible to the GDPR for data protection reasons. Access to ELGA data is restricted. There are understandable reasons why the data is not publicly available. Spatial analysis requires all information indicating a person's location to be anonymized in accordance with the GDPR. This removes all relevant information about a person except for the X,Y coordinates. Therefore, it is necessary to carry the coordinates using a geomasking method. Therefore, in health geography, it is not possible to map discrete health data on a detailed spatial level. It can be assumed that health workers are experts in health geography and not geographical analysis. Certainly, in healthcare facilities could be improved by more information about geomasking methods and their application in spatial analyses. This would also improve public health.

Especially from the survey results (5 participants) it can be estimated that there are gaps in knowledge about geomasking methods and spatial analyses in the healthcare facilities in Vienna. The survey showed that the participants were well informed about data protection. Although the survey was also sent to eight private clinics, there was no response from them. For this reason, there were only five completed survey questionnaires. The majority of the participants consisted of Viennese private clinics, as it was assumed that they carry out spatial analyses. There is no recognized GDPR geomasking method for health data that allows the publication of results of spatial analyses in public health. Especially in the event of a crisis, in the event of an outbreak of a pandemic, it would be advantageous to have an established geometry method to be able to carry out detailed analyses such as Local Moran's I (Anselin) at the city level and smaller and to visualize the results in the form of point maps in scientific articles. Without increasing the risk of spatial re-identification of the persons.

6. Literature References

(APA)

Allshouse W.B., Fitch M.K., Hampton K.H., Gesink D.C., Doherty I.A., Leone P.A., Serre M.L., & Miller W.C. (2010). Geomasking sensitive health data and privacy protection: an evaluation using an E911 database. NIH Public Access. 1.10.2010, 25(6), (pp 443-452). <https://www.tandfonline.com/doi/abs/10.1080/10106049.2010.496496>

Armstrong MP, Rushton G, Zimmerman DL.(1999). Geographically masking health data to preserve confidentiality. Statist Med 1999.18.497–525.
[https://doi.org/10.1002/\(SICI\)1097-0258\(19990315\)18:5<497::AID-SIM45>3.0.CO;2-%23](https://doi.org/10.1002/(SICI)1097-0258(19990315)18:5<497::AID-SIM45>3.0.CO;2-%23)

Anselin, L. (1996). Interactive Techniques and Exploratory Spatial Data Analysis. Regional Research Institute working papers. West Virginia Universität,
https://researchrepository.wvu.edu/rri_pubs/200

Anselin, L., & Bao, S. (1997). Explorative räumliche Datenanalyse zwischen SpaceStat und ArcView. In Recent Developments in Spatial Analysis: Spatial Statistics, Behavioral Modeling, and Computational Intelligence. (pp.35-59) Berlin, Heidelberg.
<https://www.researchgate.net/publication/246146496>

Benitez, K., & Malin, B. (2010). Evaluating re-identification risks with respect to the HIPAA privacy rule. Journal of the American Medical Informatics Association, 17(2), (pp. 169–177). <https://doi.org/10.1136/jamia.2009.000026>

Bundesministerium für Soziales, Gesundheit, Pflege und Konsumentenschutz (2019). E-Health in Österreich.
https://www.oesterreich.gv.at/themen/gesundheit_und_notfaelle/elga_elektronische_gesundheitsakte/Seite.3110001.html

Bundesministerium für Soziales, Gesundheit, Pflege und Konsumentenschutz (2019). Schutz sensibler Daten. Position der Gesundheitssektionen VIII und IX des BMASGK
https://www.sozialministerium.at/dam/jcr:99d679b8-3676-455e-abfc-a6ec27c4d64d/Schutz_sensibler_Daten_-_Position_der_Gesundheitssektionen_VIII_und_IX_.pdf

Cassa, CA, Wieland, SC, & Mandl, KD (2008). Re-identification of home addresses from spatial locations anonymized by Gaussian skew. International Journal of Health Geographics.7.(45).
<https://ij-healthgeographics.biomedcentral.com/articles/10.1186/1476-072X-7-45>

Charleux, L., & Schofield, K. (2020). True spatial k-anonymity: adaptive areal elimination vs. adaptive areal masking. *Cartography and Geographic Information Science* 47(6), (pp. 537-549) (pp.1-13).
<https://doi.org/10.1080/15230406.2020.1794975>

Curtis, A., Mills, JW, Agustin, L., & Cockburn, M. (2011), Confidentiality risks in fine scale aggregations of health data. *Computers Environment and Urban Systems* 35(1):57-64. <https://doi.org/10.1016/j.compenvurbsys.2010.08.002>

Degelsegger-Márquez A., Gruböck A., & Fidon I.K (2022). Gesundheitsdaten in Österreich – ein Überblick. Bericht über die Ergebnisse. *Gesundheit Österreich*. Im Auftrag des Bundesministeriums für Soziales, Gesundheit, Pflege und Konsumentenschutz. Wien. https://www.sozialministerium.at/dam/jcr:46a49dd4-dac0-4f67-b933-7c7285b70e5b/Gesundheitsdaten%20in%20%C3%96sterreich_bf_final.pdf

Delmelle, E., & Kanaroglou, P., (2015). Einführung: Räumliche Analyse in der Gesundheitsgeographie. Kapitel I. DOI: <https://doi.org/10.4324/9781315610252>

Hascher T., (2020). Zitieren nach den Richtlinien der American Psychological Association APA (7th ed.). Institut für Erziehungswissenschaft. Abteilung Schul- und Unterrichtsforschung (ASU). Universität Bern.
https://www.edu.unibe.ch/unibe/portal/fak_humanwis/philhum_institute/inst_paed/content/e39/e268485/e268367/pane268376/e441386/ASU_Zitierrichtlinien_August20_ger.pdf

Kalinic M., & Krisp J.M. (2018). Kernel Density Estimation KDE vs. Hot Spot Analysis - Detecting Criminal Hot Spots in the City of San Francisco. *Agile 2018 -Lund, 12-15. Augsburg*.[https://agile-gi.eu/images/conferences/2018/documents/shortpapers/66%20Kernel%20Density%20Estimation%20\(KDE\)%20vs.%20Hot-Spot%20Analysis%20-%20Detecting%20Criminal%20Hot%20Spots%20in%20the%20City%20of%20San%20Francisco_UPDATE.pdf](https://agile-gi.eu/images/conferences/2018/documents/shortpapers/66%20Kernel%20Density%20Estimation%20(KDE)%20vs.%20Hot-Spot%20Analysis%20-%20Detecting%20Criminal%20Hot%20Spots%20in%20the%20City%20of%20San%20Francisco_UPDATE.pdf)

Kim, J., Kwan, M., Levenstein, M.C., & Richardson, D.B., (2020). How do people perceive the disclosure risk of maps? Examining the perceived disclosure risk of maps and its implications for geoprivacy protection
<https://doi.org/10.1080/15230406.2020.1794976>

Koller, D., Wohlrab, D., Sedlmeir, G., & Augustin, J. (2020). Geographische Ansätze in der Gesundheitsberichterstattung. *Bundesgesundheitsblatt. Gesundheitsforschung. Gesundheitsschutz*. Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie - IBE, (9), Ludwig-Maximilians-Universität München.
<https://doi.org/10.1007/s00103-020-03208-6>

Kounadi, O., & Leitner, M. (2014). Why Does Geoprivacy Matter? The Scientific Publication of Confidential Data Presented on Maps. *Journal of Empirical Research on Human Research Ethics*. 9(4). <https://doi.org/10.1177/1556264614544103>

Kounadi, O., & Leitner, M. (2016). Adaptive areal elimination (AAE): A transparent way of disclosing protected spatial datasets. *Computers Environment and Urban Systems* 56:59-67 https://www.researchgate.net/figure/Adaptive-area-elimination-AAE-the-steps-of-the-method_fig2_292993122

Polzin, F., & Kounadi, O. (2021). Adaptive voronoi masking: A method to protect confidential discrete spatial data. In 11th International Conference on Geographic Information Science (GIScience 2021)-Part II. Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

Kulldorff, M. (1997). A Spatial Scans Statistics. *Communications in Statistics – Theory and Methods*. 26 (6). 1481-1496.

Lang, E.C., Pfandlsteiner, E., LL.M., Satzinger, G., Scharinger, R., Schmeisl, B., & Wörel, T. (2019). Schutz sensibler Daten 35 von 37. Position der Gesundheitssektionen VIII und IX des Bundesministeriums für Arbeit, Soziales, Gesundheit und Konsumentenschutz. BMASGK. [https://www.sozialministerium.at/dam/jcr:99d679b8-3676-455e-abfc-a6ec27c4d64d/Schutz sensibler Daten - Position der Gesundheitssektionen VIII und IX .pdf](https://www.sozialministerium.at/dam/jcr:99d679b8-3676-455e-abfc-a6ec27c4d64d/Schutz_sensibler_Daten_-_Position_der_Gesundheitssektionen_VIII_und_IX_.pdf)

Leitner, M. & A. Curtis (2006). "Ein erster Schritt in Richtung A Framework for Presenting the Location of Confidential Point Data on Maps – Results of an Empirical Perceptual Study." *Internationale Zeitschrift für Geoinformationswissenschaft* 20(7). 813-822. Doi:10.1080/13658810600711261

Leitner, M., Mills, JW, & Curtis, A. (2007). Can Novices to Geospatial Technology Compromise Spatial Confidentially? *Journal of Cartography and Geographic Information* 57(2):78-84. DOI: [10.1007/BF03544029](https://doi.org/10.1007/BF03544029)

O'Sullivan, D. & Unwin, J.D. (2003). Point Pattern analysis. In *Geographic Information Analysis*. New Jersey, USA. 121-154.

O'Sullivan, D. & Unwin, J.D. (2010). *Geographic Information Analysis*. Second Edition. ISBN 978-0-470-28857-3

Polzin, F.S., (2020). Adaptive Voronoi Masking. A method to protect confidential discrete spatial data (Master-Thesis). Geographical, Information, Management, Application. GIMA. Universität Nederland. https://studenttheses.uu.nl/bitstream/handle/20.500.12932/39480/20200810_MasterThesis_FionaPolzin%20%281%29_compressed.pdf?sequence=1&isAllowed=y

Schmidt, R. (2007). DNA Daten. Öffentliche Sicherheit. Die Zeitschrift des Innenministeriums. Meilenstein der Kriminologie, (9/10). <https://www.bmi.gv.at/>

Seidl, D.E, Paulus, G., Jankowski, P. & Regenfelder, M. (2015). Spatial obfuscation methods for privacy protection of household-level data. Applied Geography. (63). (253–263). <https://doi.org/10.1016/j.apgeog.2015.07.001>

Seidl, D.E., Jankowski, P., & Clarke, K., (2017). Privacy and False Identification Risk in Geomasking Techniques. Geographical Analysis. 50(6). Oktober 2017 (9). <https://doi.org/10.1111/gean.12144>

Smith, M.-J., Goodchild, M.-F., & Longley, P.-A. (2015). Geospatial analysis: a comprehensive Guide to Principles, Techniques and Software Tools. Winchelsea, UK, ISBN: 1905886608, 9781905886609.

Stinchcomb, D. (2004): Procedures for Geomasking to Protect Patient Confidentiality. ESRI-International Health GIS Conference. Washington DC. 19.10.2004. <https://proceedings.esri.com/library/userconf/health04/papers/pap3012.pdf>

Swanlund D., Schuurman N., Zandbergen P., & Brussoni M. (2020). Street masking: a network-based geographic mask for easily protecting geoprivacy. Canada. <https://doi.org/10.1186/s12942-020-00219-z>

Wong, D.W. (1999). Several fundamentals in implementing spatial statistics in gis: Using Centographic measures as examples. Geographic Information Sciences, 5(2), 163-174.

Zakaria, S., Zaini, NE, Abdul Malik, SM., & Wan Alwi, WS (2021). Exploratory spatial data analysis (ESDA) on Covid-19 cases in Malaysia. UTM Journal-Teknologi. Fakultät für Meerestechnik und Informatik, Universität von Malaysia Terengganu, Jurnal Teknologi, 83(6), pp 83-94. Online 20. Oktober 2021. <https://doi.org/10.11113/jurnalteknologi.v83.17106>

Zeger, H.G., Krenn, M., & Sulzgruber, E. (2023). ARGE DATEN. Österreichische Gesellschaft für Datenschutz. https://www.argedaten.at/php/cms_monitor.php?q=PUB-TEXT-ARGEDATEN&s=34706eta

Zhang, S., Freundsuh, S.M., Lenzer, K., & Zandbergen, P.A. (2017). The location swapping method for geomasking. Cartography and Geographic Information Science. University New Mexico, USA. 44 (1), (pp. 22-34). <https://doi.org/10.1080/15230406.2015.1095655>

6.1. Online References

BEV. Bundesamt für Eich- und Vermessungswesen. "Federal Office of Metrology and Surveying". (2024). Kontrollpunkt für Smartphones in 100 Jahre BEV. Geodätischer Referenzpunkt. Wien. <https://kontrollpunkt.bev.gv.at/100JahreBEV>

Kulldorff, M., (2022). SaTScan TMSoftware for the spatial, temporal, and space-time scan statistics Department of Health geography. Harvard Medical School. USA <https://www.satscan.org/>

Louisa N. (2023). Datenhandel: Kann das legal sein? | Datenschutzerklärung 2023. Beratung zu Handel und V... 08.10.2023 <https://www.datenschutz.org/datenhandel/>

Österreichische Ärztekammer für Wien. „Vienna-Medical-Association“. (2019). Cyberkriminalität in der Medizin: Ärztekammer für Wien (oa. Univ.-Prof. Dr. Thomas Szekeres, Dr. Cornelius Granig, 24.06.19 https://www.aekwien.at/pressekonferenzen/-/asset_publisher/jcmXijhc423X/content/cyberkriminalit%25C3%25A4t-in-der-medizin

Stadt Wien. „Vienna City“. (2020). Karten zum Thema Bevölkerung – Stadtforschung. City of Vienna population 2020. <https://www.wien.gv.at/stadtentwicklung/grundlagen/stadtforschung/karten/bevoelkerung.html>

Statistik Wien. „Statistic Vienna“. (2024). Bevölkerung zu Jahres-/Quartalsanfang: <https://www.statistik.at/statistiken/bevoelkerung-und-soziales/bevoelkerung/bevoelkerungsstand/bevoelkerung-zu-jahres-/-quartalsanfang>

Severino, R. (2024). Choropleth Map. The Data Visualisation Catalogue. <https://datavizcatalogue.com/methods/choropleth.html>

Szekeres T. (2023). Gesundheitsdaten sind bei Kriminellen begehrt. Presse. 28.08.2023 <https://www.diepresse.com/5858845/gesundheitsdaten-sind-bei-kriminellen-begehrt>

Vienken S. (2021). ArcGIS Blog online. Interview 09/07/2021. Explorative spatial Daten Analyse <https://arcgis.esri.de/interview-explorative-raeumliche-datenanalyse/>

6.2. ArcGIS References

Last access 21.05.2024:

AAM ArcGIS Toolbox (2019): Adaptive Areal Anonymization Toolbox.

<https://www.arcgis.com/home/item.html?id=3ef11e690c1841c78df70433a2720724>

Adaptive Geographical Masking AVM - Geographical Masking methods. Applicable to discrete confidential-private-sensitive geodata.

<https://github.com/okounadi/Geoprivacy>

ArcGIS(a). (2024). Kerndichte (Spatial Analyst) <https://pro.arcgis.com/de/pro-app/3.1/tool-reference/spatial-analyst/kernel-density.htm>

ArcGIS(b). (2024). Optimierte Hotspot Analyse <https://pro.arcgis.com/de/pro-app/latest/tool-reference/spatial-statistics/how-optimized-hot-spot-analysis-works.htm>

ArcGIS(c). (2024). Funktionsweise von Mittlerer nächster Nachbar (ANN). <https://pro.arcgis.com/de/pro-app/latest/tool-reference/spatial-statistics/h-how-average-nearest-neighbor-distance-spatial-st.htm>

ArcGIS(d). (2024). Räumliche Autokorrelation (Global Moran's I) und KNN Analyse (räumliche Statistik) <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/spatial-autocorrelation.htm>

ArcGIS(e). (2024). Cluster- und Ausreißer Analyse (Anselin Local Morans I) (Spatial Statistics) <https://pro.arcgis.com/de/pro-app/latest/tool-reference/spatial-statistics/cluster-and-outlier-analysis-anselin-local-moran-s.htm>

ArcGIS(f). (2020). ArcGIS StoryMap. Global VS Local spatial Autocorrelation, Geog 418/518, 3. November.2022. <https://storymaps.arcgis.com/stories/5b26f25bb81a437b89003423505e2f71>

ArcGIS(g). (2024). Funktionsweise des Werkzeugs Räumliche Ausreißer erkennen <https://pro.arcgis.com/de/pro-app/latest/tool-reference/spatial-statistics/how-spatial-outlier-detection-works.htm>

ArcGIS(h). (2024). Überblick über Bildklassifizierung <https://pro.arcgis.com/de/pro-app/latest/help/analysis/image-analyst/overview-of-image-classification.htm#:~:text=Bei%20der%20un%C3%BCberwachten%20Klassifizierung%20wird,Ihrem%20Bild%20vorhanden%20sein%20sollen.>

ArcGIS(i). (2024). Klassifizierung und Mustererkennung in ArcGIS Pro <https://pro.arcgis.com/de/pro-app/latest/help/analysis/image-analyst/classification-and-pattern-recognition-in-arcgis-pro.htm>

ArcGIS(j). (2024). Einführung in Deep Learning. <https://pro.arcgis.com/de/pro-app/latest/help/analysis/deep-learning/what-is-deep-learning-.htm>

ArcGIS(k). (2024). Start-Ziel-Verbindungen erstellen (Analysis) <https://pro.arcgis.com/de/pro-app/latest/tool-reference/analysis/generate-origin-destination-links.htm>

6.3. Orthophotos References

Geoland Basemap.(2024). <https://www.basemap.at/wmts>

Geoland Orthophoto of Austria.(2024). <https://www.basemap.at/wmts>

Orthophotos. ArcGIS Pro.(2024). World imagery (Austria). Satellite-Aerial-Imagery. Land NÖ, Bundesamt für Eich- und Vermessungswesen Wien, BEV.
https://services.arcgisonline.com/ArcGIS/rest/services/World_Imagery/MapServer

6.4. Data-Simulation References

Building polygon Stadt Wien – data.wien.gv.at. The name of the data is Vector data of the multi-purpose area map (MZK) and is downloadable from the WFS:
https://www.data.gv.at/katalog/de/dataset/stadtwnien_flchenmehrzweckkartevektordate
nwien, identifier: 7cf0da04-1f77-4321-929e-78172c74aa0b

WMS Geoserver GetCapabilities to all meta data of the MZK Blattschnitt 1:1000.
<https://data.wien.gv.at/daten/geo?version=1.3.0&service=WMS&request=GetCapabilities>

Buildings of the MZK as a shape file in a bounding box from the API link to all meta data of the (MZK):
https://www.data.gv.at/katalog/api/3/action/package_show?id=7cf0da04-1f77-4321-929e-78172c74aa0b

MZK 5 Link to creates 5 bounding boxes, for the download:

1. <https://data.wien.gv.at/daten/geo?service=WFS&version=1.0.0&request=GetFeature&typeName=ogdwien:FMZKGEBOGD&outputFormat=shapezip&SRS=EPSG:31256&BBOX=-11500,331000,-5000,354000>
2. <https://data.wien.gv.at/daten/geo?service=WFS&version=1.0.0&request=GetFeature&typeName=ogdwien:FMZKGEBOGD&outputFormat=shapezip&SRS=EPSG:31256&BBOX=-5000,331000,0,354000>
3. <https://data.wien.gv.at/daten/geo?service=WFS&version=1.0.0&request=GetFeature&typeName=ogdwien:FMZKGEBOGD&outputFormat=shapezip&SRS=EPSG:31256&BBOX=0,331000,5000,354000>
4. <https://data.wien.gv.at/daten/geo?service=WFS&version=1.0.0&request=GetFeature&typeName=ogdwien:FMZKGEBOGD&outputFormat=shapezip&SRS=EPSG:31256&BBOX=5000,331000,10000,354000>

5. <https://data.wien.gv.at/daten/geo?service=WFS&version=1.0.0&request=GetFeature&typeName=ogdwien:FMZKGEBOGD&outputFormat=shapezip&SRS=EPSG:31256&BBOX=10000,331000,18500,354000>

Building points Stadt Wien with residential use only.

<https://www.data.gv.at/katalog/en/dataset/gebaeudeinformation-wien>

Identifier: 7a8aae59-71a4-4500-b38b-bdf15c7f627f

Building information locations Vienna (from WFS):

<http://data.wien.gv.at/daten/geo/ows?service=WFS&request=GetFeature&version=1.1.0&typeName=ogdwien:GEBAEUDEINFOOGD&srsName=EPSG:4326&outputFormat=shape-zip>

Population data Wien (2012) create Geojson:

<https://www.data.gv.at/application/bevolkerungszahlen-und-dichte-in-wien-nach-zahlbezirken/#datasets>

7. Appendix

7.1. AVM Python Script adapted from Polzin & Kounadi

```
1  # Name: Adaptive Voronoi Masking
2  # Description: Displacing points to Voronoi segments within the SKApoly
  (with exceptions), and then displacing to street intersections.
3  # Author: Fiona Polzin, f.s.
4  # Reference: Adaptive Voronoi Masking: A Method to Protect Confidential
  Discrete Spatial Data.
  https://drops.dagstuhl.de/opus/volltexte/2021/14760/pdf/LIPICs-GIScience-
  2021-II-1.pdf
5  # This script is edited by O.Kounadi, ourania.kounadi@univie.ac.at,
  2022
6
7
8  #The external and internal variables must be set by the user.
9  #Code can be executed in ArcMAP10.x and ArcGISPro.
10 #-----
11 # Import system modules
12 import arcpy
13
14 # EXTERNAL variables to be inputted by the user
15 workspace = r".." # Set the workspace
16 PointFile = r"..."
17 BlockFile = r"..."
18 StreetFile = r".."
19 SearchRadius = "100000 Meters"
20
21 # Step 1: Find points for VM
22 pointsJoin = "points_join"
23 pointsForVM = "Points_VM"
24 pointsNoVM = "Points_No_VM"
25
26 # Step 2: Create Thiessen and split (Union) them based on blocks
27 VoronoiPolygon = "VoronoiPolygo"
28 VoronoiUnion = "VoronoiPoly_Unio"
29
30 # Step 3: Now, snap the pointsForVM to the closest edge of their
  corresponding Voronoi_clip polygon or the SKApoly segment
31 # Step 4: Find the polygons with only data points
32 # Step 5: Create one random points (points not for VM)
33 RandomPoints = "RandomPoint"
34
35 # Step 6: Merge StreetFile
36 MergedStreets = "UnsplitStreets"
37
38 # Step 7: Find street intersections
39 Intersection = "Intersectio"
40
41 # Step 8: Move pointsForVM to closest intersection
42 # Step 9: Move pointsNoVM to closest intersection
43 # Step 10: merge pointsForVM and pointsNoVM to a new layer
44 AVM = "AVM_point"
45
46 # Set workspace
47 arcpy.env.workspace = workspace
48 arcpy.env.overwriteOutput = True # Controls whether the algorithm will
  replace any present output automatically
49 arcpy.env.addOutputsToMap = True # Add output to map automatically
```

```

50
51 # Step 1 - Find points for VM and not for VM
52 arcpy.analysis.SpatialJoin(PointFile, BlockFile, pointsJoin,
53 'JOIN_ONE_TO_ONE', 'KEEP_ALL')
54 arcpy.analysis.Statistics(pointsJoin, "stat", "id COUNT", "id")
55 arcpy.management.CopyFeatures(BlockFile, "Blocks")
56 arcpy.JoinField_management("Blocks", "id", "stat", "ID", "COUNT_ID")
57
58 arcpy.SelectLayerByAttribute_management("Blocks",
59 selection_type="NEW_SELECTION", where_clause='"COUNT_ID">1')
60 arcpy.MakeFeatureLayer_management(pointsJoin, "points_lyr")
61 arcpy.management.SelectLayerByLocation("points_lyr", "INTERSECT",
62 "Blocks", None, "NEW_SELECTION", "NOT_INVERT")
63 arcpy.CopyFeatures_management("points_lyr", pointsNoVM)
64 arcpy.SelectLayerByLocation_management("points_lyr",
65 overlap_type="INTERSECT", select_features="Blocks", search_distance="",
66 selection_type="SWITCH_SELECTION",
67 invert_spatial_relationship="NOT_INVERT")
68 arcpy.CopyFeatures_management("points_lyr", pointsForVM)
69 arcpy.SelectLayerByAttribute_management("Blocks",
70 selection_type="CLEAR_SELECTION")
71 arcpy.SelectLayerByAttribute_management("points_lyr",
72 selection_type="CLEAR_SELECTION")
73
74 # Step 2 - Create Thiessen and split (Union) them based on blocks
75 arcpy.env.extent = BlockFile
76 arcpy.CreateThiessenPolygons_analysis(pointsForVM, VoronoiPolygon,
77 "ALL")
78 arcpy.Union_analysis([VoronoiPolygon, BlockFile], VoronoiUnion)
79
80 # Step 3 - Snap data points to the closest segment (Voronoi or SKApoly)
81 inFeature = pointsForVM
82 snap_feature = VoronoiUnion
83 snap_Type = "EDGE"
84 snap_Distance = SearchRadius
85 arcpy.Snap_edit(inFeature, [[snap_feature, snap_Type, snap_Distance]])
86
87 # Step 4 - Find all polygons which contain data points not transferred
88 by VM
89 arcpy.SelectLayerByAttribute_management("Blocks",
90 selection_type="NEW_SELECTION", where_clause='"COUNT_ID">1')
91 arcpy.management.JoinField('Blocks', 'id', 'Points_No_VM', 'id', 'CID')
92 # add CID of no VM points
93 arcpy.management.AddField('Points_VM', 'cid_VM')
94 arcpy.management.CalculateField("Points_VM", "cid_VM", "!CID!",
95 "PYTHON3", '!', "TEXT")
96 arcpy.management.DeleteField('Points_VM', 'CID')
97
98 # Step 5 - Create random points (for points no VM)
99 arcpy.CreateRandomPoints_management(out_path=workspace,
100 out_name=RandomPoints,
101 constraining_feature_class="Blocks", number_of_points_or_field="COUNT_ID")
102 arcpy.management.JoinField(RandomPoints, 'CID', 'Blocks', 'FID', 'CID')
103 # add CID info to random points
104 arcpy.management.AddField(RandomPoints, 'cid_no_VM')
105 arcpy.management.CalculateField(RandomPoints, "cid_no_VM", "!CID_1!",
106 "PYTHON3", '!', "TEXT")

```

```

95 arcpy.management.DeleteField("RandomPoint", "CID;CID_1")
96 arcpy.SelectLayerByAttribute_management("Blocks",
selection_type="CLEAR_SELECTION")
97
98 # Move all data points to the closest intersection
99 # Step 6 - To assure that the input streetfile is not divided into
multiple substreets causing various streets to intersect but do not produce
"cross" street intersections, streets will be "unsplitted"
100 in_features = StreetFile
101 out_feature_class = MergedStreets
102
103 # Run UnsplittedLine
104 arcpy.UnsplitLine_management(in_features, out_feature_class)
105
106 # Step 7 - Use "Intersect" tool to find street intersection and receive
output as points
107 in_features = MergedStreets
108 out_feature_class = Intersection
109
110 arcpy.Intersect_analysis(in_features, out_feature_class, "", "",
"POINT")
111
112 # Step 8 - Move pointsForVM to the closest intersection
113 in_fc = pointsForVM
114 snapClass = Intersection
115 snapType = "VERTEX"
116 snapDistance = SearchRadius
117 arcpy.Snap_edit(in_fc, [[snapClass, snapType, snapDistance]])
118
119 # Step 9 - Move "Points_Not_For_VM" to the closest intersection
120 in_fc = RandomPoints
121 snapFeature = Intersection
122 snapChar = "VERTEX"
123 snapWidth = SearchRadius
124 arcpy.Snap_edit(in_fc, [[snapFeature, snapChar, snapWidth]])
125
126 # Step 10 - Merge all masked data points into a new layer
127 arcpy.Merge_management('RandomPoint;Points_VM', AVM, "")
128 arcpy.management.AddField(AVM, 'orig_CID')
129 arcpy.management.CalculateField("AVM_point", "orig_CID", "!cid_VM! +
!cid_no_VM!", "PYTHON3", '', "TEXT")
130 print("AVM points are created")
131
132 # Step 11 - Calculate displaced distance - AVM and original points
133 arcpy.analysis.GenerateOriginDestinationLinks(PointFile, AVM,
"AVM_dist", "CID", "orig_CID", "PLANAR", 1, None, "KILOMETERS",
"NO_AGGREGATE", None)
134 print("Displaced distance calculated. Program successfully finished")
135

```

7.2. AAM Python Script adapted from Laura Charleux

(ArcGIS Toolbox Script)

```
1  # Adaptive Areal Masking
2  # by Laure Charleux
3  # lcharleu@d.umn.edu
4  # December 2019
5
6  import arcpy
7  import random
8  import numpy
9  import datetime
10
11  arcpy.env.overwriteOutput = True
12
13  points = arcpy.GetParameterAsText(0)
14  polys = arcpy.GetParameterAsText(1)
15  polysID = arcpy.GetParameterAsText(2) # must be a text field
16  popfield = arcpy.GetParameterAsText(3)
17  threshold = arcpy.GetParameter(4)
18
19  output = arcpy.GetParameterAsText(5)
20
21  # Create numpy array with coordinates of polygon centroids and place
22  # holders for distance and cumulative population
23  xylist = []
24  with arcpy.da.SearchCursor(polys, [polysID, "SHAPE@X", "SHAPE@Y",
25  popfield]) as polyscursor:
26      for polysrow in polyscursor:
27          xylist.append((polysrow[0], polysrow[1], polysrow[2],
28  polysrow[3], None, None))
29
30  names = ['polyID', 'polyX', 'polyY', 'pop', 'dist', 'cumpop']
31  formats = ['U30', 'f8', 'f8', 'i8', 'f8', 'f8']
32  dtype = numpy.dtype({'names': names, 'formats': formats})
33  myarray = numpy.array(xylist, dtype=dtype)
34
35  # Function to list the polygons to be merged to obtain the population
36  # threshold
37
38  def getTargetPolys(myarray, polyid, threshold):
39      index = numpy.where(myarray['polyID'] == polyid)
40      myx = myarray['polyX'][index]
41      myy = myarray['polyY'][index]
42      mypop = myarray['pop'][index]
43
44      myarray['dist'] = ((myarray['polyX'] - myx) ** 2 +
45  (myarray['polyY'] - myy) ** 2) ** 0.5
46      myarray.sort(order="dist")
47      myarray['cumpop'] = numpy.cumsum(myarray['pop'])
48
49      index = numpy.where(myarray['cumpop'] >= threshold)
50      index = index[0][0]
51      targetpolys = myarray['polyID'][0:index + 1]
52      cumpop = myarray['cumpop'][index]
```

```

53  # Create a copy of the points to anonymize, with the ID of the poly
they fall within
54  # Add a field for cumpop, for operator to check level of k-anonymity,
that should be deleted prior to publication of the dataset because it
creates a risk of re-identification
55
56  IDfieldMap = arcpy.FieldMap()
57  fms = arcpy.FieldMappings()
58
59  IDfieldMap.addInputField(polys, polysID)
60  fms.addFieldMap(IDfieldMap)
61
62  join = arcpy.SpatialJoin_analysis(points, polys,
"%scratchGDB%/myjoin", "JOIN_ONE_TO_ONE", "KEEP_COMMON", fms)
63  arcpy.AddField_management(join, "cumpop", "DOUBLE")
64
65  # Iteration over points: gets the list of polys to merge, merge them
and displace the point
66
67
68  start = datetime.datetime.now()
69  with arcpy.da.UpdateCursor(join, ["SHAPE@", polysID, "cumpop"]) as
pointscursor:
70      for pointsrow in pointscursor:
71          pointgeom = pointsrow[0]
72          polyid = pointsrow[1]
73
74          where = arcpy.AddFieldDelimiters(polys, polysID) + "='" +
polyid + "'"
75          with arcpy.da.SearchCursor(polys, ["SHAPE@", polysID], where)
as polycursor:
76              for polysrow in polycursor:
77                  polygeom = polysrow[0]
78                  polypop = polysrow[1]
79
80                  if polypop < threshold:
81                      targetpolys, cumpop = getTargetPolys(myarray, polyid,
threshold)
82                      polypop = cumpop
83
84                      where = arcpy.AddFieldDelimiters(polys, polysID) + " in "
+ str(tuple(targetpolys))
85                      with arcpy.da.SearchCursor(polys, "SHAPE@", where) as
polycursor:
86                          for polysrow in polycursor:
87                              polygeom = polygeom.union(polysrow[0])
88
89                      extent = polygeom.extent
90                      inside = False
91
92                      while inside == False:
93                          myx = random.uniform(extent.XMin, extent.XMax)
94                          myy = random.uniform(extent.YMin, extent.YMax)
95                          tempPoint = arcpy.Point(myx, myy)
96                          inside = polygeom.contains(tempPoint)
97
98                      pointsrow[0] = tempPoint
99                      pointsrow[2] = polypop
100                      pointscursor.updateRow(pointsrow)
101
102  end = datetime.datetime.now()

```

```

103 print("Duration of displacement procedure: {}".format(str(end -
start)))
104 arcpy.AddMessage("Duration of displacement procedure:
{}".format(str(end - start)))
105
106 # Delete poly ID field, join point attributes to displaced points and
save as the final output
107 arcpy.DeleteField_management(join, polysID)
108
109 origpointlayer = arcpy.MakeFeatureLayer_management(points,
"origpointlayer")
110 oid_fieldname = arcpy.Describe(points).OIDFieldName
111
112 maskedpointlayer = arcpy.MakeFeatureLayer_management(join,
"maskedpointlayer")
113 arcpy.AddJoin_management(maskedpointlayer, "TARGET_FID",
origpointlayer, oid_fieldname)
114
115 arcpy.CopyFeatures_management(maskedpointlayer, output)
116
117 arcpy.Delete_management(join)
118
119

```

Umfrage zu räumlichen GIS-Analysen in Gesundheitseinrichtungen

Die geographischen räumlichen Analysen sind ein Teil der öffentlichen Gesundheit und der Epidemiologie geworden. Einerseits können Medizinkarten die Verteilungen einer Erkrankung nach regionalen Einheiten wie Bundesländer abdecken, andererseits sind diese Daten nicht so aussagekräftig, wie die standortbasierten Daten auf der Stadtebene. Die Daten auf Stadtebene sind wiederum kaum verfügbar, werden auch aus Gründen des Datenschutzes nicht erhoben.

In der heutigen Zeit stellt sich die Frage, wie die standortbasierten, sensiblen/ medizinischen, persönlichen Daten unter Berücksichtigung des Datenschutzes kartographisch visualisiert und veröffentlicht werden können. Aus diesem Grund wurden im Rahmen der Masterarbeit standortbasierte Daten für Wien simuliert, um das Risiko räumlicher Reidentifizierung von "Patientendaten" auf einer "Stadtebene" erforschen zu können. Um die Standortdaten besser schützen zu können, bieten sich heutzutage besonders die Geomaskierungsmethoden an.

Ziel ist es, einen sicheren Weg zu finden um die persönlichen standortbasierten Daten analysieren, visualisieren und anschließend veröffentlichen zu können, dabei das mögliche räumliche Reidentifizierungsrisiko auszuschließen, so gering wie möglich halten.

Link:

<https://raeumliche-gis-analysen-in-gesundheitseinrichtungen.limesurvey.net/715718?lang=de-easy>

Sehr geehrte Damen und Herren!

Im Rahmen meiner Masterarbeit an der Universität Wien, aus Geoinformation und Kartographie, aus dem Fachbereich der "Location Privacy" möchte ich Sie bitten, diese kurze Onlineumfrage betreffend der räumlichen GIS Analysen in Wiener Gesundheitseinrichtungen zu beantworten. Die Umfrage dauert etwa 10 Minuten lang. Ihre Antworten werden anonym verarbeitet und nach der Auswertung gelöscht. Die Umfrage ist von 1.9.2023 (12:00) bis 15.9.2023 (12:00) aktiv.

Ziel ist es eine kurze, qualitative Umfrage durchzuführen, die meine Masterarbeit positiv ergänzen sollte. Interessant wäre für mich zu erfahren, wie weit die aktuelle Forschung aus dem Fachbereich der "Geoinformation und Medizinforschung" in Wien fortgeschritten ist, ob sich hier bestimmte Hindernisse zeigen, oder mögliche Verbesserungsvorschläge finden.

Für Ihre Mithilfe bedanke ich mich im Voraus.

In dieser Umfrage sind 22 Fragen enthalten.

Umfrage zu räumlichen GIS Analysen in Gesundheitseinrichtungen

In welchem Gesundheitsbereich sind Sie tätig? Bitte beschreiben Sie kurz im untenstehenden Feld.

*

Bitte geben Sie Ihre Antwort hier ein:

*Haben Sie schon von räumlichen GIS (**G**eographisches **I**nformations **S**ystem) Analysen gehört? Bitte wählen Sie eine Antwort.*

*

Bitte wählen Sie nur eine der folgenden Antworten aus:

☐ Ja

☐ Nein

*Bitte erläutern Sie kurz im untenstehenden Feld, woher Ihre Informationen über räumliche (GIS) Analysen stammen. **

Beantworten Sie diese Frage nur, wenn folgende Bedingungen erfüllt sind:

Antwort war 'Ja' bei Frage ' [G01Q02]' (Haben Sie schon von räumlichen GIS (Geographisches Informations System) Analysen gehört? Bitte wählen Sie eine Antwort.)

Bitte geben Sie Ihre Antwort hier ein:

Möchten Sie mehr Informationen über räumliche GIS Analysen erhalten? Erläutern Sie kurz im untenstehenden Feld.

*

Beantworten Sie diese Frage nur, wenn folgende Bedingungen erfüllt sind:

Antwort war 'Nein' bei Frage ' [G01Q02]' (Haben Sie schon von räumlichen GIS (Geographisches Informations System) Analysen gehört? Bitte wählen Sie eine Antwort.)

Bitte geben Sie Ihre Antwort hier ein:

*Haben Sie schon von den Geomaskierungsmethoden gehört? Bitte wählen Sie eine Antwort. **

Bitte wählen Sie nur eine der folgenden Antworten aus:

☐ Ja

☐ Nein

*Bitte erläutern Sie kurz im untenstehenden Feld, woher Ihre Informationen über die Geomaskierungsmethoden stammen. **

Beantworten Sie diese Frage nur, wenn folgende Bedingungen erfüllt sind:

Antwort war 'Ja' bei Frage ' [G01Q02a]' (Haben Sie schon von den Geomaskierungsmethoden gehört? Bitte wählen Sie eine Antwort.)

Bitte geben Sie Ihre Antwort hier ein:

Möchten Sie mehr Informationen über Geomaskierungsmethoden erhalten? Bitte erläutern Sie kurz im untenstehenden Feld.

*

Beantworten Sie diese Frage nur, wenn folgende Bedingungen erfüllt sind:

Antwort war 'Nein' bei Frage ' [G01Q02a]' (Haben Sie schon von den Geomaskierungsmethoden gehört? Bitte wählen Sie eine Antwort.)

Bitte geben Sie Ihre Antwort hier ein:

Es wird angenommen, als eine Gesundheitseinrichtung verfügen sie über standortbasierte Patientendaten, was passiert mit diesen Daten? Bitte erläutern Sie kurz im untenstehenden Feld.

*

Bitte geben Sie Ihre Antwort hier ein:

Zeigen sich bei der Aufbewahrung und Pflege von Patientendaten Probleme? Bitte erläutern Sie kurz im untenstehenden Feld.

*

Bitte geben Sie Ihre Antwort hier ein:

Haben Sie bereits räumliche GIS Analysen mit Patientendaten durchgeführt? Bitte wählen Sie eine Antwort.

*

Bitte wählen Sie nur eine der folgenden Antworten aus:

☐ Ja

☐ Nein

Welche räumliche GIS Analysen haben Sie durchgeführt? Bitte erläutern Sie kurz im untenstehenden Feld.

*

Beantworten Sie diese Frage nur, wenn folgende Bedingungen erfüllt sind:

Antwort war 'Ja' bei Frage ' [G01Q07]' (Haben Sie bereits räumliche GIS Analysen mit Patientendaten durchgeführt? Bitte wählen Sie eine Antwort.)

Bitte geben Sie Ihre Antwort hier ein:

*Aus welchem Grund haben Sie räumliche GIS Analysen durchgeführt? Bitte erläutern Sie kurz im untenstehenden Feld. **

Beantworten Sie diese Frage nur, wenn folgende Bedingungen erfüllt sind:

Antwort war 'Ja' bei Frage ' [G01Q07]' (Haben Sie bereits räumliche GIS Analysen mit Patientendaten durchgeführt? Bitte wählen Sie eine Antwort.)

Bitte geben Sie Ihre Antwort hier ein:

Welche Probleme haben sich dabei gezeigt, was ist Ihrer Meinung nach notwendig um räumliche GIS Analysen durchführen zu können? Bitte erläutern Sie kurz im untenstehenden Feld.

Beantworten Sie diese Frage nur, wenn folgende Bedingungen erfüllt sind:

Antwort war 'Ja' bei Frage ' [G01Q07]' (Haben Sie bereits räumliche GIS Analysen mit Patientendaten durchgeführt? Bitte wählen Sie eine Antwort.)

Bitte geben Sie Ihre Antwort hier ein:

Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort.

*

Beantworten Sie diese Frage nur, wenn folgende Bedingungen erfüllt sind:
Antwort war 'Ja' bei Frage ' [G01Q07]' (Haben Sie bereits räumliche GIS Analysen mit Patientendaten durchgeführt? Bitte wählen Sie eine Antwort.)

Bitte sagen Sie uns ein bisschen mehr zu Ihrer Antwort. Dafür können Sie etwas in das leere Feld schreiben.

Bitte wählen Sie die zutreffenden Punkte aus und schreiben Sie einen Kommentar dazu:

☐ Ja

☐ Nein

☐ Sonstiges

Aus welchem Grund haben Sie keine räumlichen GIS Analysen durchgeführt, sowie wo liegen Ihrer Meinung nach die Hindernisse? Bitte erläutern Sie kurz im untenstehenden Feld.

*

Beantworten Sie diese Frage nur, wenn folgende Bedingungen erfüllt sind:

Antwort war 'Nein' bei Frage ' [G01Q07]' (Haben Sie bereits räumliche GIS Analysen mit Patientendaten durchgeführt? Bitte wählen Sie eine Antwort.)

Bitte geben Sie Ihre Antwort hier ein:

Denken Sie, es sind weitere Maßnahmen notwendig, um räumliche GIS Analysen mit Patientendaten durchführen zu können? (Zum Beispiel Datenanonymisierung, etc.) Bitte erläutern Sie im untenstehenden Feld.

*

Beantworten Sie diese Frage nur, wenn folgende Bedingungen erfüllt sind:

Antwort war 'Nein' bei Frage ' [G01Q07]' (Haben Sie bereits räumliche GIS Analysen mit Patientendaten durchgeführt? Bitte wählen Sie eine Antwort.)

Bitte geben Sie Ihre Antwort hier ein:

*Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. **

Beantworten Sie diese Frage nur, wenn folgende Bedingungen erfüllt sind:

Antwort war 'Nein' bei Frage ' [G01Q07]' (Haben Sie bereits räumliche GIS Analysen mit Patientendaten durchgeführt? Bitte wählen Sie eine Antwort.)

Bitte sagen Sie uns ein bisschen mehr zu Ihrer Antwort. Dafür können Sie etwas in das leere Feld schreiben.

Bitte wählen Sie die zutreffenden Punkte aus und schreiben Sie einen Kommentar dazu:

☐ Ja, wir benötigen mehr Information

☐ Nein, das überlassen wir lieber den Experten

☐ Kann ich jetzt nicht sagen

☐ Sonstiges

*Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar?
Bitte begründen Sie kurz Ihre Antwort. **

Bitte sagen Sie uns ein bisschen mehr zu Ihrer Antwort. Dafür können Sie etwas in das leere Feld schreiben.

Bitte wählen Sie die zutreffenden Punkte aus und schreiben Sie einen Kommentar dazu:

☐ Ja

☐ Nein

☐ Sonstiges

Ist Ihrer Meinung nach die "Geoinformation und Medizinforschung" in Österreich uneingeschränkt möglich? Bitte begründen Sie kurz Ihre Antwort.

*

Bitte geben Sie Ihre Antwort hier ein:

Welchen Einfluss hat Ihrer Meinung nach der "Datenschutz von Patientendaten" betreffend des möglichen Reidentifikationsrisikos der Standortdaten aus den kartographischen Visualisierungen auf diese Forschung? Bitte erläutern Sie kurz.

*

Bitte geben Sie Ihre Antwort hier ein:

Diese Umfrage wird anonymisiert verarbeitet, doch bitte geben Sie an welche Institution Sie vertreten und welche berufliche Position Sie ausüben.

Sie sind nicht verpflichtet diese Frage zu beantworten.

*

Bitte geben Sie Ihre Antwort hier ein:

Möchten Sie mir noch etwas mitteilen, haben Sie eventuell Verbesserungsvorschläge, um die "Geoinformation und Medizinforschung" in Österreich besser unterstützen zu können? Bitte erläutern Sie kurz im untenstehenden Feld.

*

Bitte geben Sie Ihre Antwort hier ein:

Vielen Dank!

31.07.2024 – 19:59

Senden Sie Ihre Umfrage ein.

Vielen Dank für die Beantwortung des Fragebogens.

7.4. Survey Responses

Umfrageantwort	
Antwort ID	1
Datum Abgeschickt	2023-09-15 09:22:23
Letzte Seite	22
Start-Sprache	de-easy
Zufallsstartwert	-----
Datum gestartet	2023-09-15 09:17:59
Datum letzte Aktivität	2023-09-15 09:22:23
In welchem Gesundheitsbereich sind Sie tätig? Bitte beschreiben Sie kurz im untenstehenden Feld.	Controlling - Krankenanstalt
Haben Sie schon von räumlichen GIS (Geographisches Informations System) Analysen gehört? Bitte wählen Sie eine Antwort.	Ja
Bitte erläutern Sie kurz im untenstehenden Feld, woher Ihre Informationen über räumliche (GIS) Analysen stammen.	Seit 1998 wird damit gearbeitet
Möchten Sie mehr Informationen über räumliche GIS Analysen erhalten? Erläutern Sie kurz im untenstehenden Feld.	
Haben Sie schon von den Geomaskierungsmethoden gehört? Bitte wählen Sie eine Antwort.	Ja
Bitte erläutern Sie kurz im untenstehenden Feld, woher Ihre Informationen über die Geomaskierungsmethoden stammen.	Literatur
Möchten Sie mehr Informationen über Geomaskierungsmethoden erhalten? Bitte erläutern Sie kurz im untenstehenden Feld.	
Es wird angenommen, als eine Gesundheitseinrichtung verfügen sie über standortbasierte Patientendaten, was passiert mit diesen Daten? Bitte erläutern Sie kurz im untenstehenden Feld.	werden analysiert
Zeigen sich bei der Aufbewahrung und Pflege von Patientendaten Probleme? Bitte erläutern Sie kurz im untenstehenden Feld.	Nein
Haben Sie bereits räumliche GIS Analysen mit Patientendaten durchgeführt? Bitte wählen Sie eine Antwort.	Ja

Welche räumliche GIS Analysen haben Sie durchgeführt? Bitte erläutern Sie kurz im untenstehenden Feld.	Standort-, Penetrationsanalysen, Wege
Aus welchem Grund haben Sie räumliche GIS Analysen durchgeführt? Bitte erläutern Sie kurz im untenstehenden Feld.	Leistungsplanung
Welche Probleme haben sich dabei gezeigt, was ist Ihrer Meinung nach notwendig um räumliche GIS Analysen durchführen zu können? Bitte erläutern Sie kurz im untenstehenden Feld.	Prozess von Dateneingabe bis -verarbeitung organisieren
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Ja]	Ja
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Ja][Kommentar]	neue Sicht auf Daten, Lösungsweg zur Bewältigung Komplexität
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Nein]	Nein
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Nein][Kommentar]	
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges]	Nein
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges][Kommentar]	
Aus welchem Grund haben Sie keine räumlichen GIS Analysen durchgeführt, sowie wo liegen Ihrer Meinung nach die Hindernisse? Bitte erläutern Sie kurz im untenstehenden Feld.	
Denken Sie, es sind weitere Maßnahmen notwendig, um räumliche GIS Analysen mit Patientendaten durchführen zu können? (Zum Beispiel Datenanonymisierung, etc.) Bitte erläutern Sie im untenstehenden Feld.	
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Ja, wir benötigen mehr Information]	N/A
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Ja, wir benötigen mehr Information][Kommentar]	
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Nein, das überlassen wir lieber den Experten]	N/A

Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Nein, das überlassen wir lieber den Experten][Kommentar]	
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Kann ich jetzt nicht sagen]	N/A
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Kann ich jetzt nicht sagen][Kommentar]	
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges]	N/A
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges][Kommentar]	
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Ja]	Ja
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Ja][Kommentar]	im Netzwerk
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Nein]	Nein
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Nein][Kommentar]	
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges]	Nein
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges][Kommentar]	
Ist Ihrer Meinung nach die "Geoinformation und Medizinforschung" in Österreich uneingeschränkt möglich? Bitte begründen Sie kurz Ihre Antwort.	Nein

Welchen Einfluss hat Ihrer Meinung nach der "Datenschutz von Patientendaten" betreffend des möglichen Reidentifikationsrisikos der Standortdaten aus den kartographischen Visualisierungen auf diese Forschung? Bitte erläutern Sie kurz.	groß, sie dürfen keine Einzelfälle auswerten
Diese Umfrage wird anonymisiert verarbeitet, doch bitte geben Sie an welche Institution Sie vertreten und welche berufliche Position Sie ausüben. Sie sind nicht verpflichtet diese Frage zu beantworten.	
Möchten Sie mir noch etwas mitteilen, haben Sie eventuell Verbesserungsvorschläge, um die "Geoinformation und Medizinforschung" in Österreich besser unterstützen zu können? Bitte erläutern Sie kurz im untenstehenden Feld.	Danke

Umfrageantwort	
Antwort ID	4
Datum Abgeschickt	2023-09-15 10:28:14
Letzte Seite	22
Start-Sprache	de-easy
Zufallsstartwert	-----
Datum gestartet	2023-09-15 10:24:46
Datum letzte Aktivität	2023-09-15 10:28:14
In welchem Gesundheitsbereich sind Sie tätig? Bitte beschreiben Sie kurz im untenstehenden Feld.	Allgemeinmedizin
Haben Sie schon von räumlichen GIS (Geographisches Informations System) Analysen gehört? Bitte wählen Sie eine Antwort.	Nein
Bitte erläutern Sie kurz im untenstehenden Feld, woher Ihre Informationen über räumliche (GIS) Analysen stammen.	
Möchten Sie mehr Informationen über räumliche GIS Analysen erhalten? Erläutern Sie kurz im untenstehenden Feld.	nein danke
Haben Sie schon von den Geomaskierungsmethoden gehört? Bitte wählen Sie eine Antwort.	Nein
Bitte erläutern Sie kurz im untenstehenden Feld, woher Ihre Informationen über die Geomaskierungsmethoden stammen.	
Möchten Sie mehr Informationen über Geomaskierungsmethoden erhalten? Bitte erläutern Sie kurz im untenstehenden Feld.	nein danke
Es wird angenommen, als eine Gesundheitseinrichtung verfügen sie über standortbasierte Patientendaten, was passiert mit diesen Daten? Bitte erläutern Sie kurz im untenstehenden Feld.	nein danke
Zeigen sich bei der Aufbewahrung und Pflege von Patientendaten Probleme? Bitte erläutern Sie kurz im untenstehenden Feld.	nein
Haben Sie bereits räumliche GIS Analysen mit Patientendaten durchgeführt? Bitte wählen Sie eine Antwort.	Nein
Welche räumliche GIS Analysen haben Sie durchgeführt? Bitte erläutern Sie kurz im untenstehenden Feld.	

Aus welchem Grund haben Sie räumliche GIS Analysen durchgeführt? Bitte erläutern Sie kurz im untenstehenden Feld.	
Welche Probleme haben sich dabei gezeigt, was ist Ihrer Meinung nach notwendig um räumliche GIS Analysen durchführen zu können? Bitte erläutern Sie kurz im untenstehenden Feld.	
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Ja]	N/A
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Ja][Kommentar]	
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Nein]	N/A
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Nein][Kommentar]	
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges]	N/A
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges][Kommentar]	
Aus welchem Grund haben Sie keine räumlichen GIS Analysen durchgeführt, sowie wo liegen Ihrer Meinung nach die Hindernisse? Bitte erläutern Sie kurz im untenstehenden Feld.	keine zeit einfach
Denken Sie, es sind weitere Maßnahmen notwendig, um räumliche GIS Analysen mit Patientendaten durchführen zu können? (Zum Beispiel Datenanonymisierung, etc.) Bitte erläutern Sie im untenstehenden Feld.	nein
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Ja, wir benötigen mehr Information]	Nein
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Ja, wir benötigen mehr Information][Kommentar]	
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Nein, das überlassen wir lieber den Experten]	Ja
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Nein, das überlassen wir lieber den Experten][Kommentar]	keine zeit dafür
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Kann ich jetzt nicht sagen]	Nein
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Kann ich jetzt nicht sagen][Kommentar]	

Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges]	Nein
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges][Kommentar]	
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Ja]	Nein
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Ja][Kommentar]	
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Nein]	Nein
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Nein][Kommentar]	
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges]	Ja
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges][Kommentar]	ich weiß es nicht
Ist Ihrer Meinung nach die "Geoinformation und Medizinforschung" in Österreich uneingeschränkt möglich? Bitte begründen Sie kurz Ihre Antwort.	dont know
Welchen Einfluss hat Ihrer Meinung nach der "Datenschutz von Patientendaten" betreffend des möglichen Reidentifikationsrisikos der Standortdaten aus den kartographischen Visualisierungen auf diese Forschung? Bitte erläutern Sie kurz.	dont know
Diese Umfrage wird anonymisiert verarbeitet, doch bitte geben Sie an welche Institution Sie vertreten und welche berufliche Position Sie ausüben. Sie sind nicht verpflichtet diese Frage zu beantworten.	Allgemeinmedizin
Möchten Sie mir noch etwas mitteilen, haben Sie eventuell Verbesserungsvorschläge, um die "Geoinformation und Medizinforschung" in Österreich besser unterstützen zu können? Bitte erläutern Sie kurz im untenstehenden Feld.	nein leider

Umfrageantwort	
Antwort ID	6
Datum Abgeschickt	2023-09-18 13:25:46
Letzte Seite	22
Start-Sprache	de-easy
Zufallsstartwert	-----
Datum gestartet	2023-09-18 13:20:36
Datum letzte Aktivität	2023-09-18 13:25:46
In welchem Gesundheitsbereich sind Sie tätig? Bitte beschreiben Sie kurz im untenstehenden Feld.	Rettungsdienst
Haben Sie schon von räumlichen GIS (Geographisches Informations System) Analysen gehört? Bitte wählen Sie eine Antwort.	Nein
Bitte erläutern Sie kurz im untenstehenden Feld, woher Ihre Informationen über räumliche (GIS) Analysen stammen.	
Möchten Sie mehr Informationen über räumliche GIS Analysen erhalten? Erläutern Sie kurz im untenstehenden Feld.	Ja
Haben Sie schon von den Geomaskierungsmethoden gehört? Bitte wählen Sie eine Antwort.	Nein
Bitte erläutern Sie kurz im untenstehenden Feld, woher Ihre Informationen über die Geomaskierungsmethoden stammen.	
Möchten Sie mehr Informationen über Geomaskierungsmethoden erhalten? Bitte erläutern Sie kurz im untenstehenden Feld.	Ja
Es wird angenommen, als eine Gesundheitseinrichtung verfügen sie über standortbasierte Patientendaten, was passiert mit diesen Daten? Bitte erläutern Sie kurz im untenstehenden Feld.	Ich bin in einer Notfallrettungseinrichtung tätig. Unsere Patientendaten machen standortbezogen wenig Sinn.
Zeigen sich bei der Aufbewahrung und Pflege von Patientendaten Probleme? Bitte erläutern Sie kurz im untenstehenden Feld.	Jeder Einsatz erfordert eine manuelle Erhebung von Patientendaten.

Haben Sie bereits räumliche GIS Analysen mit Patientendaten durchgeführt? Bitte wählen Sie eine Antwort.	Nein
Welche räumliche GIS Analysen haben Sie durchgeführt? Bitte erläutern Sie kurz im untenstehenden Feld.	
Aus welchem Grund haben Sie räumliche GIS Analysen durchgeführt? Bitte erläutern Sie kurz im untenstehenden Feld.	
Welche Probleme haben sich dabei gezeigt, was ist Ihrer Meinung nach notwendig um räumliche GIS Analysen durchführen zu können? Bitte erläutern Sie kurz im untenstehenden Feld.	
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Ja]	N/A
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Ja][Kommentar]	
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Nein]	N/A
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Nein][Kommentar]	
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges]	N/A
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges][Kommentar]	
Aus welchem Grund haben Sie keine räumlichen GIS Analysen durchgeführt, sowie wo liegen Ihrer Meinung nach die Hindernisse? Bitte erläutern Sie kurz im untenstehenden Feld.	Weil ich nicht weiß, was GIS Analysen sind.
Denken Sie, es sind weitere Maßnahmen notwendig, um räumliche GIS Analysen mit Patientendaten durchführen zu können? (Zum Beispiel Datenanonymisierung, etc.) Bitte erläutern Sie im untenstehenden Feld.	Ich gehe stark davon aus.
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Ja, wir benötigen mehr Information]	Ja
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Ja, wir benötigen mehr Information][Kommentar]	

Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Nein, das überlassen wir lieber den Experten]	Nein
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Nein, das überlassen wir lieber den Experten][Kommentar]	
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Kann ich jetzt nicht sagen]	Nein
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Kann ich jetzt nicht sagen][Kommentar]	
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges]	Nein
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges][Kommentar]	
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Ja]	Nein
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Ja][Kommentar]	
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Nein]	Nein
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Nein][Kommentar]	
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges]	Ja
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar?	Ich habe noch nie von GIS Analysen gehört.

Bitte begründen Sie kurz Ihre Antwort. [Sonstiges][Kommentar]	
Ist Ihrer Meinung nach die "Geoinformation und Medizinforschung" in Österreich uneingeschränkt möglich? Bitte begründen Sie kurz Ihre Antwort.	Nein
Welchen Einfluss hat Ihrer Meinung nach der "Datenschutz von Patientendaten" betreffend des möglichen Reidentifikationsrisikos der Standortdaten aus den kartographischen Visualisierungen auf diese Forschung? Bitte erläutern Sie kurz.	Ich denke es ließen sich Gewohnheiten ableiten.
Diese Umfrage wird anonymisiert verarbeitet, doch bitte geben Sie an welche Institution Sie vertreten und welche berufliche Position Sie ausüben. Sie sind nicht verpflichtet diese Frage zu beantworten.	Berufsrettung Wien (Die berufliche Position wurde hier anonymisiert)
Möchten Sie mir noch etwas mitteilen, haben Sie eventuell Verbesserungsvorschläge, um die "Geoinformation und Medizinforschung" in Österreich besser unterstützen zu können? Bitte erläutern Sie kurz im untenstehenden Feld.	Nein

Umfrageantwort	
Antwort ID	8
Datum Abgeschickt	2023-09-19 12:33:03
Letzte Seite	22
Start-Sprache	de-easy
Zufallsstartwert	-----
Datum gestartet	2023-09-19 12:28:13
Datum letzte Aktivität	2023-09-19 12:33:03
In welchem Gesundheitsbereich sind Sie tätig? Bitte beschreiben Sie kurz im untenstehenden Feld.	Rettungsdienst
Haben Sie schon von räumlichen GIS (Geographisches Informations System) Analysen gehört? Bitte wählen Sie eine Antwort.	Ja
Bitte erläutern Sie kurz im untenstehenden Feld, woher Ihre Informationen über räumliche (GIS) Analysen stammen.	Universität Wien, Dept. f Naturwissenschaften, Kartierung
Möchten Sie mehr Informationen über räumliche GIS Analysen erhalten? Erläutern Sie kurz im untenstehenden Feld.	
Haben Sie schon von den Geomaskierungsmethoden gehört? Bitte wählen Sie eine Antwort.	Nein
Bitte erläutern Sie kurz im untenstehenden Feld, woher Ihre Informationen über die Geomaskierungsmethoden stammen.	
Möchten Sie mehr Informationen über Geomaskierungsmethoden erhalten? Bitte erläutern Sie kurz im untenstehenden Feld.	Ja, gerne
Es wird angenommen, als eine Gesundheitseinrichtung verfügen sie über standortbasierte Patientendaten, was passiert mit diesen Daten? Bitte erläutern Sie kurz im untenstehenden Feld.	Werden in der Leitstelle gesammelt
Zeigen sich bei der Aufbewahrung und Pflege von Patientendaten Probleme? Bitte erläutern Sie kurz im untenstehenden Feld.	Da gibt es fixe Abläufe

Haben Sie bereits räumliche GIS Analysen mit Patientendaten durchgeführt? Bitte wählen Sie eine Antwort.	Nein
Welche räumliche GIS Analysen haben Sie durchgeführt? Bitte erläutern Sie kurz im untenstehenden Feld.	
Aus welchem Grund haben Sie räumliche GIS Analysen durchgeführt? Bitte erläutern Sie kurz im untenstehenden Feld.	
Welche Probleme haben sich dabei gezeigt, was ist Ihrer Meinung nach notwendig um räumliche GIS Analysen durchführen zu können? Bitte erläutern Sie kurz im untenstehenden Feld.	
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Ja]	N/A
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Ja][Kommentar]	
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Nein]	N/A
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Nein][Kommentar]	
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges]	N/A
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges][Kommentar]	
Aus welchem Grund haben Sie keine räumlichen GIS Analysen durchgeführt, sowie wo liegen Ihrer Meinung nach die Hindernisse? Bitte erläutern Sie kurz im untenstehenden Feld.	Ich habe im täglichen Rettungsdienst keine Analysen zu machen
Denken Sie, es sind weitere Maßnahmen notwendig, um räumliche GIS Analysen mit Patientendaten durchführen zu können? (Zum Beispiel Datenanonymisierung, etc.) Bitte erläutern Sie im untenstehenden Feld.	Weiß nicht
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Ja, wir benötigen mehr Information]	Nein

Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Ja, wir benötigen mehr Information][Kommentar]	
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Nein, das überlassen wir lieber den Experten]	Ja
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Nein, das überlassen wir lieber den Experten][Kommentar]	
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Kann ich jetzt nicht sagen]	Nein
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Kann ich jetzt nicht sagen][Kommentar]	
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges]	Nein
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges][Kommentar]	
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Ja]	Nein
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Ja][Kommentar]	
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Nein]	Nein
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in	

Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Nein][Kommentar]	
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges]	Ja
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges][Kommentar]	Weiß ich nicht
Ist Ihrer Meinung nach die "Geoinformation und Medizinforschung" in Österreich uneingeschränkt möglich? Bitte begründen Sie kurz Ihre Antwort.	Weiß ich nicht
Welchen Einfluss hat Ihrer Meinung nach der "Datenschutz von Patientendaten" betreffend des möglichen Reidentifikationsrisikos der Standortdaten aus den kartographischen Visualisierungen auf diese Forschung? Bitte erläutern Sie kurz.	Datenschutz ist superwichtig!!!
Diese Umfrage wird anonymisiert verarbeitet, doch bitte geben Sie an welche Institution Sie vertreten und welche berufliche Position Sie ausüben. Sie sind nicht verpflichtet diese Frage zu beantworten.	
Möchten Sie mir noch etwas mitteilen, haben Sie eventuell Verbesserungsvorschläge, um die "Geoinformation und Medizinforschung" in Österreich besser unterstützen zu können? Bitte erläutern Sie kurz im untenstehenden Feld.	Wieso werde ich gefragt, ob ich eventuell etwas verändern möchte und dann ist es ein Pflichtfeld? Ich denke ich bin die falsche Person für diesen Fragebogen. Bitte entschuldigen sie.

Umfrageantwort	
Antwort ID	10
Datum Abgeschickt	2023-09-22 08:51:57
Letzte Seite	22
Start-Sprache	de-easy
Zufallsstartwert	
Datum gestartet	2023-09-22 08:46:50
Datum letzte Aktivität	2023-09-22 08:51:57
In welchem Gesundheitsbereich sind Sie tätig? Bitte beschreiben Sie kurz im untenstehenden Feld.	Rettungsdienst
Haben Sie schon von räumlichen GIS (Geographisches Informations System) Analysen gehört? Bitte wählen Sie eine Antwort.	Ja
Bitte erläutern Sie kurz im untenstehenden Feld, woher Ihre Informationen über räumliche (GIS) Analysen stammen.	Allgemeinwissen
Möchten Sie mehr Informationen über räumliche GIS Analysen erhalten? Erläutern Sie kurz im untenstehenden Feld.	
Haben Sie schon von den Geomaskierungsmethoden gehört? Bitte wählen Sie eine Antwort.	Ja
Bitte erläutern Sie kurz im untenstehenden Feld, woher Ihre Informationen über die Geomaskierungsmethoden stammen.	Ich habe gegoogelt, nachdem ich das Wort Geo-Masking im Freundeskreis gehört habe.
Möchten Sie mehr Informationen über Geomaskierungsmethoden erhalten? Bitte erläutern Sie kurz im untenstehenden Feld.	
Es wird angenommen, als eine Gesundheitseinrichtung verfügen sie über standortbasierte Patientendaten, was passiert mit diesen Daten? Bitte erläutern Sie kurz im untenstehenden Feld.	Das sind interne Daten, die auch nur intern verwendet werden.
Zeigen sich bei der Aufbewahrung und Pflege von Patientendaten Probleme? Bitte erläutern Sie kurz im untenstehenden Feld.	Nein, als ISO-zertifizierte Organisation gibt es strenge Richtlinien, wie mit personenbezogenen Daten umzugehen ist.

Haben Sie bereits räumliche GIS Analysen mit Patientendaten durchgeführt? Bitte wählen Sie eine Antwort.	Nein
Welche räumliche GIS Analysen haben Sie durchgeführt? Bitte erläutern Sie kurz im untenstehenden Feld.	
Aus welchem Grund haben Sie räumliche GIS Analysen durchgeführt? Bitte erläutern Sie kurz im untenstehenden Feld.	
Welche Probleme haben sich dabei gezeigt, was ist Ihrer Meinung nach notwendig um räumliche GIS Analysen durchführen zu können? Bitte erläutern Sie kurz im untenstehenden Feld.	
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Ja]	N/A
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Ja][Kommentar]	
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Nein]	N/A
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Nein][Kommentar]	
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges]	N/A
Würden Sie räumliche GIS Analysen weiter empfehlen? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges][Kommentar]	
Aus welchem Grund haben Sie keine räumlichen GIS Analysen durchgeführt, sowie wo liegen Ihrer Meinung nach die Hindernisse? Bitte erläutern Sie kurz im untenstehenden Feld.	Es gibt derzeit keine Notwendigkeit für räumliche GIS-Analysen.
Denken Sie, es sind weitere Maßnahmen notwendig, um räumliche GIS Analysen mit Patientendaten durchführen zu können? (Zum Beispiel Datenanonymisierung, etc.) Bitte erläutern Sie im untenstehenden Feld.	nein, wir haben die Daten, es gibt nur keine Notwendigkeit für GIS-Analysen
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Ja, wir benötigen mehr Information]	Nein
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Ja, wir benötigen mehr Information][Kommentar]	

Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Nein, das überlassen wir lieber den Experten]	Nein
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Nein, das überlassen wir lieber den Experten][Kommentar]	
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Kann ich jetzt nicht sagen]	Nein
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Kann ich jetzt nicht sagen][Kommentar]	
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges]	Ja
Würden Sie räumliche GIS Analysen durchführen wollen? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges][Kommentar]	Sehe derzeit keine Notwendigkeit
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Ja]	Nein
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Ja][Kommentar]	
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Nein]	Nein
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Nein][Kommentar]	
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges]	Ja
Denken Sie, es sind öffentliche Informationen zum Thema der räumlichen GIS Analysen und Visualisierungsmethoden mit Patientendaten in Österreich ausreichend verfügbar? Bitte begründen Sie kurz Ihre Antwort. [Sonstiges][Kommentar]	weiß ich nicht

Ist Ihrer Meinung nach die "Geoinformation und Medizinforschung" in Österreich uneingeschränkt möglich? Bitte begründen Sie kurz Ihre Antwort.	weiß ich nicht
Welchen Einfluss hat Ihrer Meinung nach der "Datenschutz von Patientendaten" betreffend des möglichen Reidentifikationsrisikos der Standortdaten aus den kartographischen Visualisierungen auf diese Forschung? Bitte erläutern Sie kurz.	weiß ich nicht
Diese Umfrage wird anonymisiert verarbeitet, doch bitte geben Sie an welche Institution Sie vertreten und welche berufliche Position Sie ausüben. Sie sind nicht verpflichtet diese Frage zu beantworten.	keine Angabe
Möchten Sie mir noch etwas mitteilen, haben Sie eventuell Verbesserungsvorschläge, um die "Geoinformation und Medizinforschung" in Österreich besser unterstützen zu können? Bitte erläutern Sie kurz im untenstehenden Feld.	keine Angabe

7.5. Cluster- and Outliers Point Maps at Vienna City

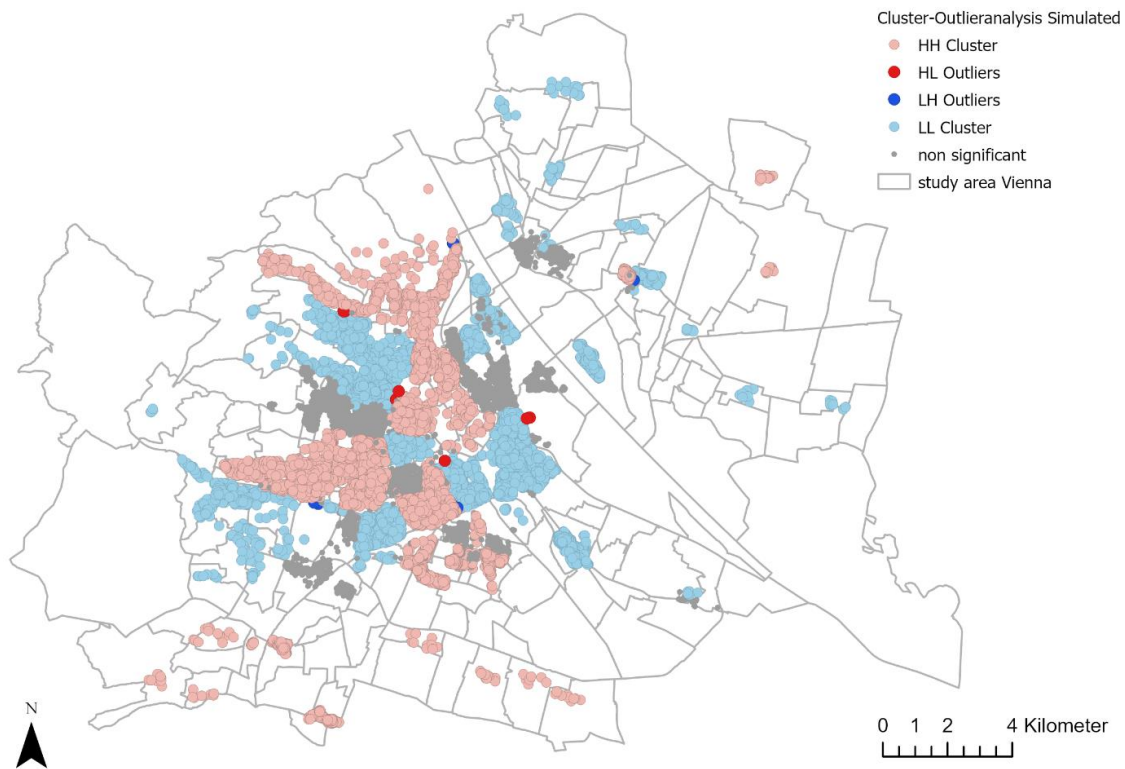


Figure 47 Simulated Points Cluster-Outliers (Anselin) ©ArcGIS Pro

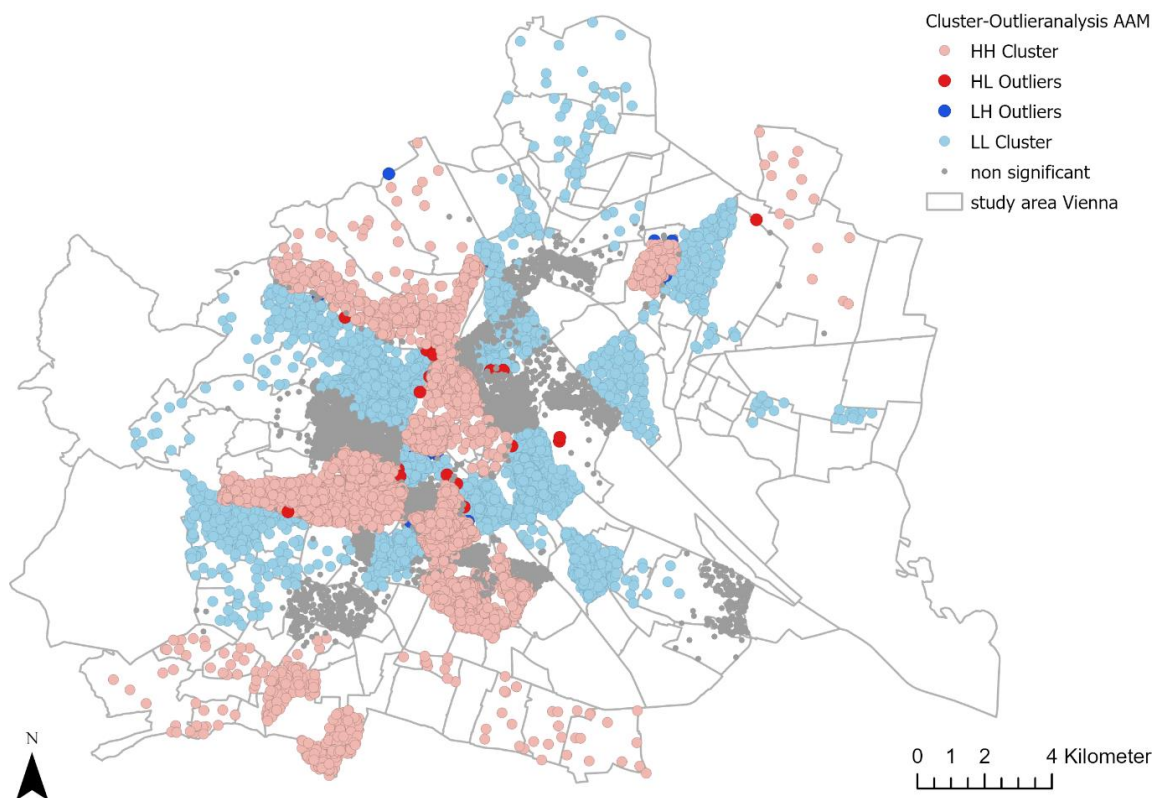


Figure 48 AAM Points Cluster-Outliers (Anselin) ©ArcGIS Pro

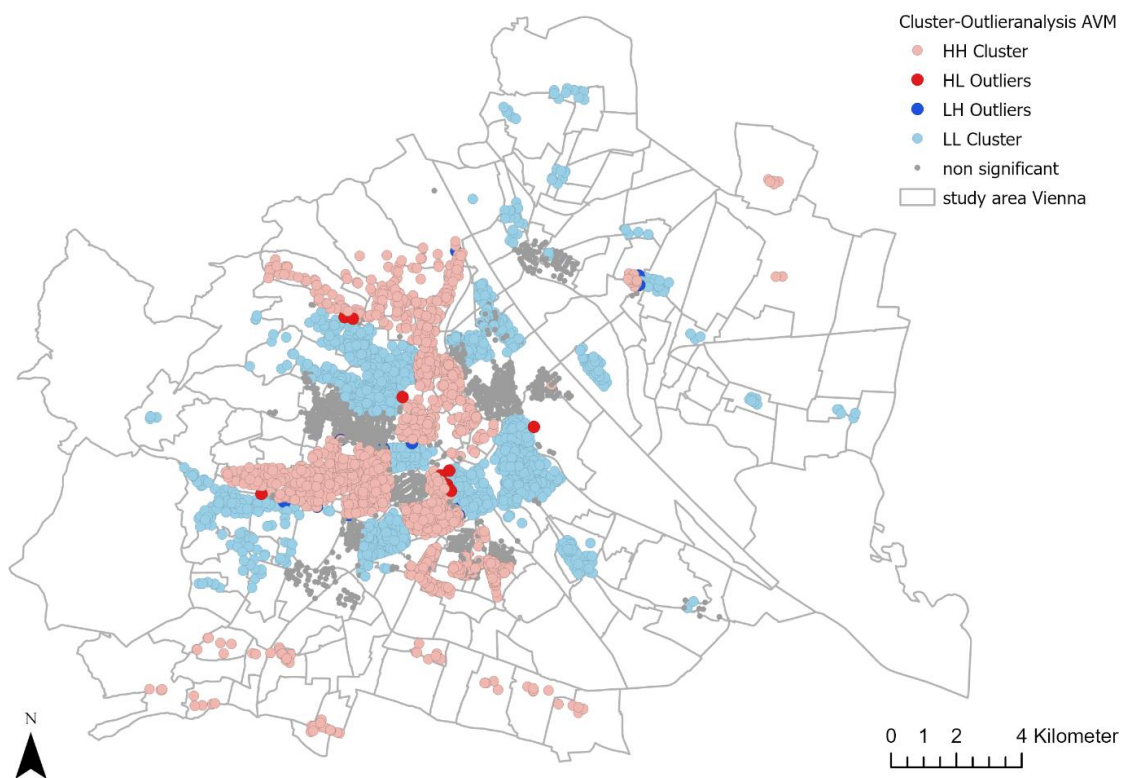


Figure 49 AVM Points Cluster-Outliers (Anselin) ©ArcGIS Pro