universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„A survey on re-identification of geographical masked data using human cognition“

verfasst von / submitted by

Isabella Noll, BSc

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

Master of Science (MSc)

Wien, 2023 / Vienna 2023

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 855
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Geographie
Betreut von / Supervisor:	Ass.-Prof. Dr. Ourania Kounadi, BSc MSc

Table of contents

Abstract

Some geomasking techniques introduce spatial error to reduce the risk of re-identification and protect the confidentiality of spatial data on a point level. Studies have been conducted to test geographical masked points by reverse geocoding for their re-identification risk, which excluded human ability to re-identify masked points. Seidl et al. (2019) were the first who tested human perception with a topological framework for reducing the risk of correct and false re-identification. Further, it has never been investigated if other map elements, such as different basemaps and disclaimers with different information of the masking technique have an influence on human perceived confidence and the actual risk of correct or false re-identifications. A survey was conducted with 49 participants to investigate the varying confidence during a re-identification task with street masked points and donut masked points using four different basemaps and two different disclaimers on geomasking techniques. Results show that confidence levels vary, especially between the maps with different basemaps. The orthophoto basemap stands out regarding high confidence levels compared to the other basemaps, while confidence decreased with the intermodal reference system as a basemap. On the other hand, maps with disclaimers including information of the masking parameters had higher correct re-identification rates, especially with donut masked points. Finally, the results from the survey suggest that confidential data should not be published on orthophotos to reduce the risk of correctly and falsely re-identified spatial data, and that information on geomasking parameters should be avoided, especially when using donut masked points.

Zusammenfassung

Einige Techniken zur geographischen Maskierung setzen räumliche Fehler ein, um das Risiko der Re-identifikation zu reduzieren und die Vertraulichkeit von räumlichen Daten auf Punktebene zu schützen. Studien wurden bereits durchgeführt, um geografisch maskierte Punkte durch Rückwerts-Geocodieren auf ihr Risiko der Re-Identifikation zu testen, wobei die Fähigkeit von Menschen geographisch maskierte Punkte zu re-identifizieren nicht betrachtet wurde. Seidl et al. (2019) waren die ersten, die die Wahrnehmung von Menschen mit einem topologischen Ansatz zur Reduzierung des Risikos der richtigen und falschen Re-Identifikation in einer Studie getestet haben. Darüber hinaus wurde nie untersucht, ob andere Kartenelemente, wie verschiedene Grundkarten und Zusatzinformationen zur Maskierungsmethode, Einfluss auf die von Menschen wahrgenommene Zuversicht und das tatsächliche Risiko der richtigen oder falschen Re-identifikation haben. Es wurde eine Umfrage mit 49 TeilnehmerInnen durchgeführt, um die Zuversicht der wahrgenommenen Re-identifikation von maskierten Wohnungspunkten, die mittels zwei verschiedenen geographischen Maskierungsmethoden (Donutmaskierung und Straßenmaskierung) maskiert wurden, untersucht. Die kartographischen Elemente haben sich in den vier verschiedenen Grundkarten, sowie zwei verschiedenen Zusatzinformationensebenen zur jeweiligen Maskierungsmethode unterschieden. Die Ergebnisse zeigen, dass die Zuversicht insbesondere zwischen den Karten mit verschiedenen Grundkarten variiert. Die Orthophoto-Grundkarte stach durch die hohe Zuversicht im Vergleich zu den anderen Grundkarten hervor, während die Zuversicht mit dem intermodalen Referenzsystem als Grundkarte abnahm. Andererseits hatten Karten mit Zusatzinformationen zu den Maskierungsparametern höhere richtige Re-identifikationsraten, insbesondere bei den Donut-maskierten Punkten. Schließlich legen die Ergebnisse der Umfrage nahe, dass vertrauliche Daten nicht auf Orthophotos veröffentlicht werden sollten, um das Risiko der richtigen und falschen Re-identifikation von räumlichen Daten zu reduzieren, und dass Informationen zu den Parametern von geographischen Maskierungen vermieden werden sollten, insbesondere bei Verwendung von Donut-maskierten Punkten.

List of Tables

List of Figures

1. Introduction

As Kounadi and Leitner (2015b) state, scholars employ two components in academics regarding geoprivacy: One component is about publishing spatial information or datasets which can lead to the exposure of private information. Secondly, they deploy the protection of the confidential or private information regarding spatial data. This chapter is going to extend the first component to demonstrate the need for protection.

Kounadi and Leitner (2014) describe three reasons why geoprivacy is in risk of exposure such as recent technologies used for geographical issues which include geoinformatics systems and location-based services, permissive legislation, and laxness of the contributors who take part in releasing data.

There are several examples of information disclosure due to technical possibilities. Zandbergen (2014) proves in his example with published coordinates of a home address (see Figure 1) that spatial information can be marked on a map and raise the risk of disclosing confidential information using existing spatial technology. This can easily be linked to an address. Even more so if the coordinates are forthwith released on an analog or online map (see Figure 2).

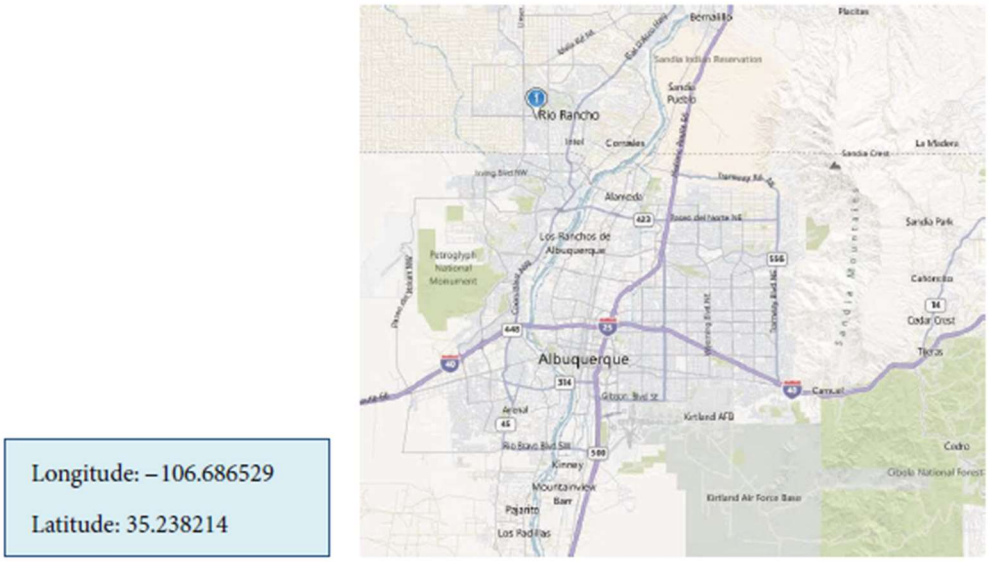


Figure 1 Example of translating coordinates on a map by Zandbergen (2014)

Krumm (2007) analyzed the GPS trajectories of 172 people and simulated a privacy attack by using algorithms to extract home addresses of the trajectories and identifying the individual's name and address through a web service. Despite their low numbers of correctly identified individual's names using location information, the author illustrated the existing risk of exposing a person's identity from their recorded GPS trajectories.

Further technological advances, such as social media also carry another risk for privacy when communicating spatial information (Kounadi and Leitner 2015b). Burdon (2010) lists a case of privacy violation in Japan 2008 where teachers mapped their students home addresses on Google MyMaps but were not able to delete the publicly available information, because it

Figure 2 Example of address visualization by Zandbergen (2014)

was saved on different servers. This example also proved, as mentioned before by Kounadi and Leitner (2014), the laxness of the contributors for releasing the data.

As a broader example, Kounadi and Resch (2018) compared three studies regarding publication of sensitive information on maps led by Brownstein et al. (2006a), Kounadi and Leitner (2014) and Haley et al. (2016). The first study in literature regarding examining publications for private data on maps was led by Brownstein et al. (2006a) who found that over the course of nine years, from 1994 to 2005, 19,000 private addresses have been plotted on maps and published in five medical journals. Another study was conducted by Kounadi and Leitner (2014) who investigated scientific journals between 2005 and 2012 and found 57 articles with confidential information on maps from which 41 articles had maps with unmasked private information. Haley et al. (2016) distinguished in their study the granularity of the private, social spatial data published on 78 maps which were shared between the start of 2013 and September 2015. 22 maps showed aggregated data to a unit which had less than or equal 30,000 people. 23 maps included point data, where 10 maps were masked, 8 maps were unmasked and five unsatisfactory masked. By comparing these three studies, Kounadi and Resch (2018) stress that disclosure of confidential information, where individuals were published as point data, takes place in different fields such as health geography, geography and spatial crime analysis. In fact, Kounadi and Leitner (2014) find 61 percent of the articles were related to health topics. Brownstein et al. (2006a) investigated exclusively in five different major journals regarding medicine. Haley et al. (2016) applied their investigations on readings concerning sexual and reproductive health which was distributed in PubMed.

Nevertheless, Leitner and Curtis (2006) describe in their example for spatial confidentiality that public bureaus would take a conservative approach to not breach confidentiality of mapped individuals and publish the data on an aggregated level. Kwan et al. (2004) state that public agencies are legally obligated to withhold private or sensitive information they gathered of individuals as a data set prior to publishing the data to the general public, and therefore the geographic data is usually aggregated to secure the geoprivacy, which on the other hand dissolves the spatial granularity level of the data for further investigation. Moreover, Kounadi and Resch (2018) state that the data holders such as researchers or research institutions can be in control of the data. Scholars, such as Leitner and

Curtis (2004, 2006) made some outline available for representing confidential data of individuals on maps. Although, as Wang, Kim and Kwan (2022) argue, data managers have no expertise in implementing the appropriate masking technique.

Zandbergen (2014) detected three deliberations which need to be taken into account when publishing spatial data to avoid re-identification and maintain certain attributes. First, the safekeeping of a person's confidential data is crucial in terms of privacy and is premised in the data compilation. Another deliberation is to inherent the spatial structure for application. The last is publishing data with the scientific community and the general public in the spirit of open data.

Researchers use different narratives of privacy in the context of spatial data and even adapt their description depending on their research data (Kounadi and Resch 2018). Kwan et al. (2004) refer to geoprivacy as the following:

"[…] geoprivacy […] refers to individual corrects to prevent disclosure of the location of one's home, workplace, daily activities, or trips. The purpose of protecting geoprivacy is to prevent individuals from being identified through locational information." (Kwan et al. 2004: 15)

The definition can be used universally for discrete location data and spatiotemporal trajectories of individuals (Resch and Kounadi 2018). Beresford and Stajano (2003) define the term location privacy as:

"[…] the ability to prevent other parties from learning one's current or past location." (Beresford and Stajano 2003:46)

Although, the research of Beresford and Stajano (2003) is about spatial trajectories, Kounadi and Leitner (2014) argue that this definition is also used for the term geoprivacy, while differently to other categories of privacy, this term adds a space-time component to a person which includes denoted movement through geographical areas and stationary positions such as residential locations. Kounadi and Resch (2018) argue, depending on the data not all geoprivacy definitions can secure the disclosure of privacy or even confidential data, and further, for a terminology description of spatial confidentiality when data is collected via e.g., a survey. Leitner and Curtis (2006) differentiate between two kinds of confidentiality regarding map presentation. On the one hand there is statistical (attribute) confidentiality, which is considered the be a statistical or attribute information about a person, such as age, gender or even health status and criminal records. On the other hand, the researchers define spatial or location confidentiality as the mapped presentation of a person, additionally linked with the statistical property. In the example a personal residency in a city area can be accurate simply mapped by a concrete data point. The map could show a statistical information and the map user could read the mapped residency as the location of people with the statistical information, exposing the person for identification.

Kounadi and Resch (2018) distinguish between nine spatial data types, where different approaches of geoprivacy are used. One of these data types is confidential discrete location data on individuals, which is the main theme in health and crime research concerning spatial data with a discrete geographical reference such as crime location in residential locations or a sick person's home. In their search for mapped confidential information, Kounadi and Leitner (2014) also find discrete data points illustrating the patient's residential locations.

Armstrong, et al. (1999) introduced an alternative approach addressing the disclosure problem of publishing geographic health data by using methods that maintain the geographical integrity for data analysis and ease concerns on the possibility to trace the private information to the individuum itself. Such methods are called geographical masks.

Zandbergen (2014) argues that minimizing the re-identification risk is crucial, as it can be used to describe the performance of a geomasking method.

"Spatial re-identification risk is the probability of identifying the real location of an individual." (Kounadi and Leitner 2014: 35)

Sweeney (2002) has introduced the privacy measure of *k*-anonymity, which can be applied to data releases to reduce the risk of re-identification by premising k-anonymity for disclosed data. k-anonymity in a data set implies that there is at least one other person for each randomly selected person in the data set (k-1) with the same attributes such as race, year of birth, or gender. For the example described above k = 2.

The derived idea from k-anonymity called spatial k-anonymity got attention in the last couple of years (Zandbergen 2014). For spatial k-anonymity to work, there must be at least k-1 sites that are indiscriminable from any other location (Kounadi and Leitner 2016).

Zandbergen (2014) further describes that shifting an original point to a random location within its vicinity is a typical practice of geomasking techniques. Moreover, when applying these geomasking techniques, it has been recommended to define a spatial k-anonymity level rather than arbitrarily defined distances between the original and the masked points. This is especially true when considering rural areas with low population density, where high distances can be achieved through geomasking, but may still yield a low spatial k-anonymity value, which indicates a higher re-identification risk.

Seidl et al. (2018) pointed out that no scientist investigated the influence of masked points, specifically points that fall onto new sites or onto another household. McLafferty (2004) notes that geographical masks transfer private information from the original site to a new one, which potentially cause new, yet unconsidered ethical implications. Seidl et al. (2018) proposed a topological framework which considers correct re-identification and false re-identification. Correct and false re-identification can be described as the following:

"*Correct identification means that a data point is linked to a correct person or household by the end user. False identification is an incorrect linkage of a data point to a different person or household*" (Seidl et al. 2018: 281)

Due to the nature of the data, which consists of discrete points, distances can be used to determine the risk locations or risk households for calculating spatial k-anonymity for these methods. This gives rise to one of the shortcomings of the methods, which use privacy metrics such as spatial k-anonymity and leave the considerations of the topology out (Seidl et al. 2018). Seidl et al. (2019) fill the need of an empirical study in regard of re-identification.

1.1. Problem Statement

Seidl et al (2019) argue that a large number of approaches to quantify the privacy risk of geomasked data neglect that additional geographical data is freely available online. This data sets can be downloaded in machine readable formats so that they are easily importable into various geoinformation systems (GIS). In order to gain additional information to facilitate re-identification by the end user, the masked data points can be projected onto the different, freely available map layers and basemaps. These layers and basemaps can contain different information such as parcel boundaries or urban structures.

Secondly, as described by Seidl et al. 2018, it is not just the quantity of geographical information, but the way geographical information is used for re-identification. Rather than having a general point household data set for a point-to-point re-identification, parcel polygons of residences and aerial imagery of residential buildings can be used for tracing a masked point to its original location.

Moreover, until now there is hardly any literature on testing geomasking techniques by re-identifying masked points to their original location with human cognition. Seidl et al. (2019) are pioneering authors who approached the re-identification of mock data points, which are assumed to be masked points, based on human perception with an empirical study. Up until now it is well established, that the ability of a human to process information for re-identification has been neglected, when applying geomasking techniques to geographical datasets.

This thesis is arguing that freely available geodata (such as orthophotos, or building parcels) contain visual information, which could lead to disclosure of a masked point. For example, the city of Vienna provides a rich online database of geographical data on *data.gv.at.* The data base contains a basemap of the city of Vienna, with street names, buildings on a parcel level, green spaces and traffic zones. Additionally, it is possible to download these thematic categories of the basemap as a single geographical data product. Orthophotos are provided by the city as well. All this information is accessible through a web map service (WMS) and a web feature service (WFS). Many GIS are able to connect directly to these services through an integrated user interface. Until now there is no literature or study on re-

identification risk (correct and false) by human cognition influenced by different cartographic information such as different types of layers shown on the map. Interestingly, scholars such as Seidl et al. 2018 and Seidl et al. 2019 mentioned in literature that freely available geo data could influence human perception to help the re-identification.

1.2. Aim of research

The aim of this study is to contribute to the research of re-identification of masked geodata. Other researchers such as Seidl et al. 2018, Seidl et al. 2019 already examined the question if the point topology and the frequency on which information of masked data on the map influence map users in their confidence of re-identifying masked data points to their original household location. Furthermore, this study is arguing that map elements with visual geographical information, such as the different kinds of layers, and disclaimers with different levels of information on geomasking techniques have an impact on human perception and the disclosure for re-identification. To examine the risk of correct re-identification and false re-identification with different geographical information, the next step is to test geomasking results on participants using human cognition. The purpose of this study is to identify possible information parameters for their impact on the risk of re-identification both correct and false with the level of the participants confidence.

1.3. Objective and Research Questions

1.3.1. Objective 1 and research question

The first objective is to investigate the effect of different map elements on human perception to re-identification. As every map element contains different information and every piece of geographical information could help to re-identify the original household. For this study, it is assumed that the perception of how much one thinks their re-identification is correct will be mirrored in confidence. Therefore, it is of interest to see how it influences the confidence of a potential re-identification task. How does confidence in re-identification vary by the use of different map elements, such as orthophotos, multi-purpose areas, multimodal reference network, building parcels and different levels of information on geomasking techniques?

To answer this research question, a survey will be conducted. Participants are presented with different sections of a map with masked points and different map elements. They will be asked, given the shown information, how certain one is that their re-identification of the masked points to their original location is correct. Thus, a low confidence is perceived as low certainty in a participant when re-identifying and vice versa for high confidence.

1.3.2. Objective 2 and research question

For the second objective, the effect of different map elements on human ability in assigning a point to its original location or false location needs to be investigated. Different information such as basemaps and disclaimers on the metadata of the geomasking techniques could help to improve spatial cognition and may or may not lead to correct or false re-identification. How do these map elements such as orthophotos, multi-purpose areas, intermodal reference network, building parcels and different levels of information on geomasking techniques enhance (or not) the spatial cognition when assigning masked points to their assumed original places?

The survey mentioned in research question one must include this research question as well. Participants are shown different sections of a map with masked points. They will then be asked to mark the point of origin of the given masked points.

1.3.3. Objective 3 and research question

For the third objective, the performance of different geomasking methods in regard of reducing the risk of false and correct re-identification and the confidence of re-identification needs to be investigated, including all the different map elements such as orthophotos, multi-purpose areas, intermodal reference network, building parcels, and different levels of information on geomasking techniques. How effective are the proposed geomasking techniques in minimizing the risk of false and correct re-identification and confidence in the re-identification task?

This research question is incorporated into the survey as well. The point data sets representing residents with a health condition will be masked with two different geomasking techniques and will be shown to the participants. A possible hypothesis can be, that certain geomasking techniques perform better in preventing re-identification than others.

2. Spatial re-identification and geoprivacy

Spatial information, which is associated to a person, such as a ZIP-code, a street address or coordinates, are considered solid geographic identifiers (Hampton et al. 2010; Zandbergen 2014; Charleux and Schottfield 2020). As mentioned before, researchers such as Brownstein et al. (2006a), Kounadi and Leitner (2014), Haley et al. (2016) found unmasked and masked confidential information on maps published in journals and argue that unmasked confidential discrete data is in the risk of re-identification.

*"*Spatial *re-identification risk is the probability of identifying the real location of an individual. Research on spatial re-identification has focused on the risk (a) when re-identifying locations from published maps that contain unmasked confidential locations and (b) when estimating the actual locations from published maps that contain masked confidential locations.*" (Kounadi and Leitner 2014: 35)

The following chapter is going to examine literature on influences of re-identification of unmasked locations, represented as discrete location points. It is assumed that influences of re-identification of unmasked data can be translated into re-identification based on human cognition as well.

2.1. Spatial re-identification of unmasked geodata

Curtis et al. (2006) claim that certain information, such as scale, dimension, quality, the kind of projection and the accuracy with which data points are presented on a map, have an influence on location re-engineering. Arguably, this can also be true for re-identification by human cognition.

Armstong (2002) showed in his experiment how easy it was to determine addresses of mapped points through reverse address matching with GIS-applications and the TIGER files, which included street names and census areas, and recommended to mask mapped points representing confidential information, especially in the field of health.

Curtis et al. (2006) show in their research the re-identification by re-engineering locations of mapped, discrete point data representing cases of death caused by Hurricane Kathrina. For the re-engineering, they first scanned the analog map, and rectified the image with ArcMap. Then they digitized the points representing the casualties and calculated the spatial center of the points.

Brownstein et al. (2005) presented their first results using a simulated data set, where they re-identified 26 percent from a low-resolution map and 79 percent from a high resolution map out of 550 addresses. In the paper of Brownstein et al. (2006b), the authors specify their actions on reverse identifying. Interestingly, they included a step of scanning and georeferencing the maps, equally as described in Curtis et al. (2006). Nevertheless, in contrast to Curtis et al. (2006) who used human cognition and their visual perception, Brownstein et

al. (2006b) used image analyzing software for the point detection and spatial assignment. This is a fully automized workflow where no human cognition was used.

Leitner et al. (2007) showed that indeed beginners to geoinformatics technologies where able to re-identify addresses through reverse address-matching with test maps. The authors used a choropleth map as a basemap and three different sizes of circles representing addresses to find its influence. Additionally, they printed the map on three different scales (7 maps in 1:130,000,7 maps in 1:190,000, and another 7 maps in 1:300,000). An example of the maps is shown in Figure 3. Each of the 21 participants got one map and performed the task in scanning the maps, geo-rectifying the scanned maps, digitizing the contour of the circle symbols, computing the centroids of their digitization, and identifying the address by overlaying these centroids on the U.S. Census Street network. The steps were similar to the steps described by Curtis et al. (2006), and partially included human cognition. Leitner et al. (2007) calculated the error by determining the average distance between the re-engineered points and their respective GPS position, for the points in three maps of their study. The distances were between 41 and 74 meters.

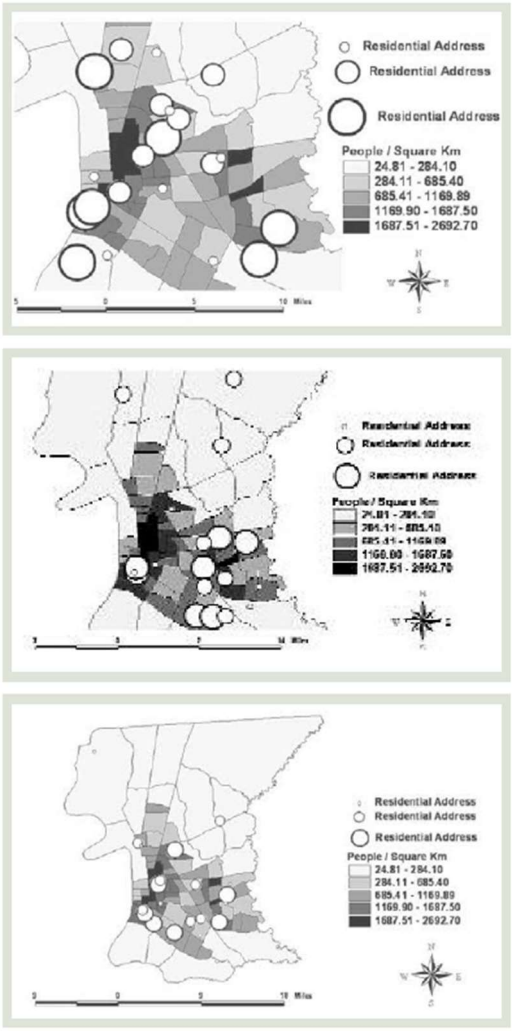


Figure 3 Test maps (three different scales from top to bottom) by Leitner et al. 2007

Armstrong (2002), Curtis et al. (2006), Brownstein et al. (2006b) recommended spatial masking of the data based on their findings. Furthermore, Kounadi et al. (2013) focused their study on reverse geocoding, which is built around withdrawing a written spatial content, e.g. an address, and summed up that Armstrong (2002), Curtis et al. (2006), Brownstein et al. (2006b) focused on their method of recording coordinates. This is especially interesting since in the scope of this work it will be also relied on coordinates rather than addresses.

In their study on confidential data on maps in published articles, Kounadi and Leitner (2014) find that maps with masked data have a larger scale (between 1:42 and 1:29,000) than the maps with unmasked data (between 1:56 to 1:153,846) and assume an easier re-identification by larger scales. By comparing the results between the two studies of Brownstein et al. (2006b) and of Leitner et al. (2007), it is also pointed out by Kounadi and Leitner (2014) that re-identification of unmasked data is indeed possible by a scale of 1:100,000. This comparison is especially interesting because it provides a comparison between re-identification which is partially using human cognition (Leitner et al. 2007), and technology aided re-identification (Brownstein et al. 2006b). For further research in this thesis, based on the previous assumption, it will be assumed that large scales facilitate the re-identification.

In the study of Kounadi and Leitner (2014) the articles with masked and unmasked data have likewise population densities. The masked data has 21 p.p km² to 12,507 p.p. km² and the unmasked data has 21 p.p. km² to 13,000 p.p. km². The authors postulate a simpler re-identification in areas with low population density than in areas with high population density. The best results in the study by Curtis et al. (2006), where 22 percent of the addresses were within 10 meters in a homogenic and lose urban structure such as in New Orleans East, suggest that urban structure has an influence on their results.

Kounadi et al. (2013) also pointed out that scholars use different spatial terminologies for correct re-identification. When Armstong (2002) used a match on the same road section as a condition for a correctly matched address, his results showed 97 percent positive matches. When using a stricter condition, where the match had to be in a directly neighboring address the success rate was 85 percent, and matching the exact same address it was 63 percent. Curtis et al. (2006) published rates of different distances to determine their degree of success. For this thesis, the degree of successful re-identification is quantified in percent of estimated locations which lie inside the original building parcels.

2.2. Geographical masking linked to correct and false re-identification

Kounadi and Leitner (2016) summed up that geographical masking relied on the principle of vagueness, which is the absence of precise specifics in information and can be seen in spatial aggregation. This involves merging confidential locations into administrative zones. The second principle is error, which can be described as the discrepancy between the given information on a map and reality. Such geographical masking methods are called isomasks with the original developments by Kwan et al. (2004), Leitner and Curtis (2004, 2006). As McLafferty (2004, p.52) is agreeing with the necessity of obfuscation confidential data, but he argues that associating this kind of information by geographical masking to a different geographical place rises a moral conundrum. The definition of Kwan et al. (2004) on geographical masking would be applicable to the two principles mentioned above:

> "*A geographical mask is a method of hiding or modifying the original location of a data point.*" (Kwan et al. 2004: 17)

The factor within geographical masks that controls the extent of spatial error inserted to the data is referred to as the masking degree (Kounadi and Leitner 2015a). The definition of geographic masking by Zandbergen (2014) would only cover the inclusion of error, as he states:

> "*Geographic masking is the process of altering the coordinates of point location data to limit the risk of re-identification upon release of the data*." (Zandbergen 2014: 4)

Armstrong et al. (1999) proposed several creative geographical masking techniques, amongst others affine transformation, where the coordinates are modified by translation, scaling, rotation, or an interplay of all of these tactics. Additionally, the authors proposed point aggregation, where one dot represents several individuals, and areal aggregation. Another technique is random perturbation, where each point is shifted arbitrarily but inside a predefined maximum radius. The authors proposed to use population density for calculating the distance inverse to the population density to uphold a higher distance in low population density areas and to allow lower distances vice versa. Kwan et al. (2004) established a weighted perturbation technique by weighting the maximum distance with population density. Hampton et al. (2010) studied the donut masking technique in the field of disease mapping. As the authors state, in this technique the masked dot lands inside a surrounding donut shaped area, giving an additional minimum distance to achieve a minimum distance to the original location and a maximum distance. As stated by Kounadi and Leitner (2016) the safeguarding of privacy takes place by relocating original points within regions of uncertainty generated by the masks, which represents the zone where a masked point could potentially be located, such as the torus by the donut masking technique by Hampton et al. (2010). The authors argue that the method provides a random perturbation of the point but inside a donut

shaped area. Further, the donut masking technique can be described as an adaptable mask as it adjusts the mask's extent at each point based on the population density underneath to a user-defined minimum and maximum k-anonymity criteria. Zandbergen (2014) argues that methods, such as donut masking, with a minimum distance of displacement reduce the risk of reverse geocoding, which can be argued also for correct re-identification for human cognition.

Cassa et al. (2006) introduced the population-density-based gaussian spatial blurring, which provides, due to the bivariate Gaussian approach, an incidental pick of length and orientation with respect to the center. Seidl et al. (2015) pointed out that it is in the nature of all these masking techniques mentioned before that they are not taking the risk of false re-identification into account, but they can be customizable to do so by excluding the affected area with specific utilization such as residential use. The authors introduced the Voronoi masking technique by creating Voronoi polygons, which refer to the residential unmasked points and move the unmasked point to the nearest situation of the polygons borders (see Figure 4). The denser the points are the more concise the spatial movements are. This masking technique can provide reducing the risk of false re-identification by including all potential locations in the creation of the Voronoi polygon, such as the residential locations of the study area, so that no dot is going to be moved onto another residential location. To be considered, two or more offsite located residential locations may result in little protection in regard of re-identification because of the concise movements in the masking procedure.

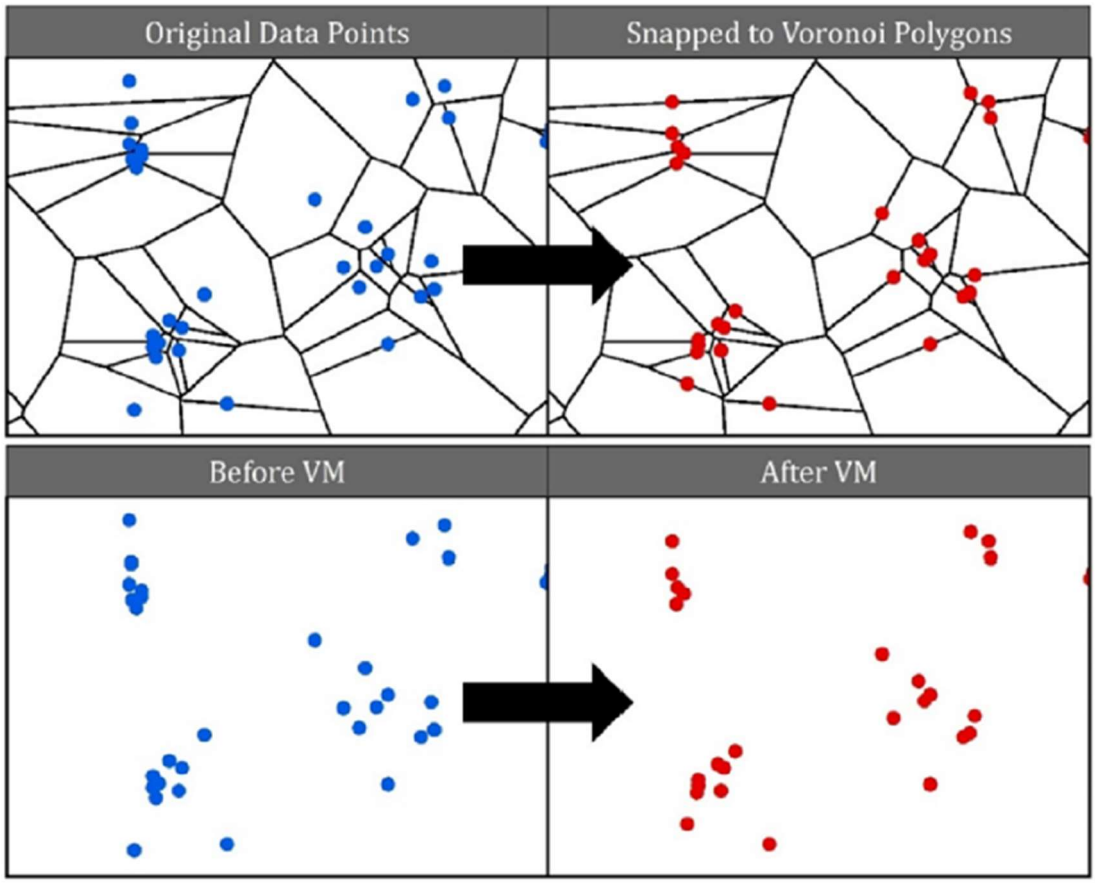


Figure 4 Difference of unmasked and Voronoi masked points by Seidl et al. 2015

Kounadi and Leitner (2016) tackled the issue of reducing the spatial error but simultaneously providing the actual k-anonymity by developing the masking technique of adapting areal elimination. By the means of a street network the street blocks were calculated, on which the address point data set was referred. Further, by dissolving these blocks, a minimum value of spatial k-anonymity referring to an area can be modeled by using a given disclosure value. As a last process, the confidential points can be aggregated or randomly moved inside this area. Nevertheless, this masking technique was modified by Charleux and Schotfield (2020) to the adaptive areal masking technique. Moreover, Polzin and Kounadi (2021) devised a combination of the masking techniques, adapting areal elimination of Kounadi and Leitner (2016) and Voronoi masking by Seidl et al. (2015), to additionally lessen the danger of false re-identification. On the one hand Polzin and Kounadi (2021) created the dissolved blocks with a minimum value of spatial k-anonymity, as described for the adaptive areal elimination by Kounadi and Leitner (2016). In addition, Polzin and Kounadi (2021) created Voronoi polygons and moved the confidential points to the nearest situation of the polygon borders, likewise as Seidl et al. (2015). In the final step Polzin and Kounadi (2021) shifted the points to the nearest crossroads (see Figure 5). The key purpose of Polzin and Kounadi (2021) was to include the topographical conditions of the masking area in regard to false re-identification and a pre-set spatial k-anonymity, as they argue that the available approaches are not able to provide both.

As topographical geographical masking techniques, Polzin and Kounadi (2021) consider the street aggregation by Leitner and Curtis (2004). The authors applied the street aggregation as a local geographical masking method on the center of a road segment and on the nearest road junction.

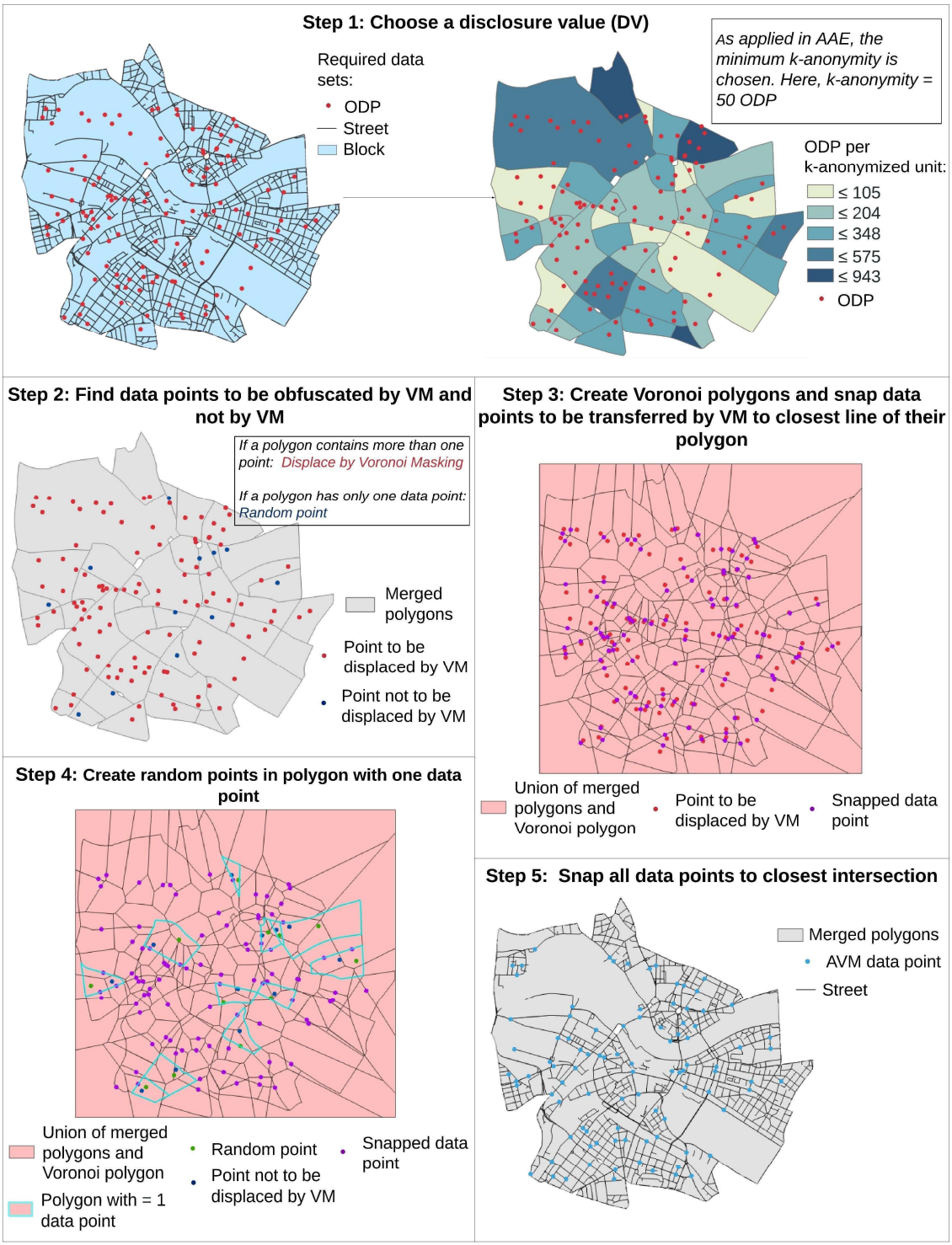


Figure 5 Steps for the adaptive Voronoi masking technique by Polzin and Kounadi (2021)

According to Seidl et al. (2018), the geographical masking technique of Zhang et al. (2017) called location swapping (see Figure 6) stands out in regard of false re-identification by forcing it. As the name already revealed, for masking purposes the point is shifted to another point representing another household inside a circle with a given radius. Seidl et al. (2018) did not include the technique of location-swapping-with-donut by Zhang et al. (2017), but might have to as the technique does not differ substantially from location swapping. Instead of a circle a location is chosen from a donut shaped area for the swapping. The benefit of the two swapping methods is the consideration of the geography, such as land cover and road proximity. Therefore, authors like Polzin and Kounadi (2021) argue that the swapping technique also include the topographical aspect.

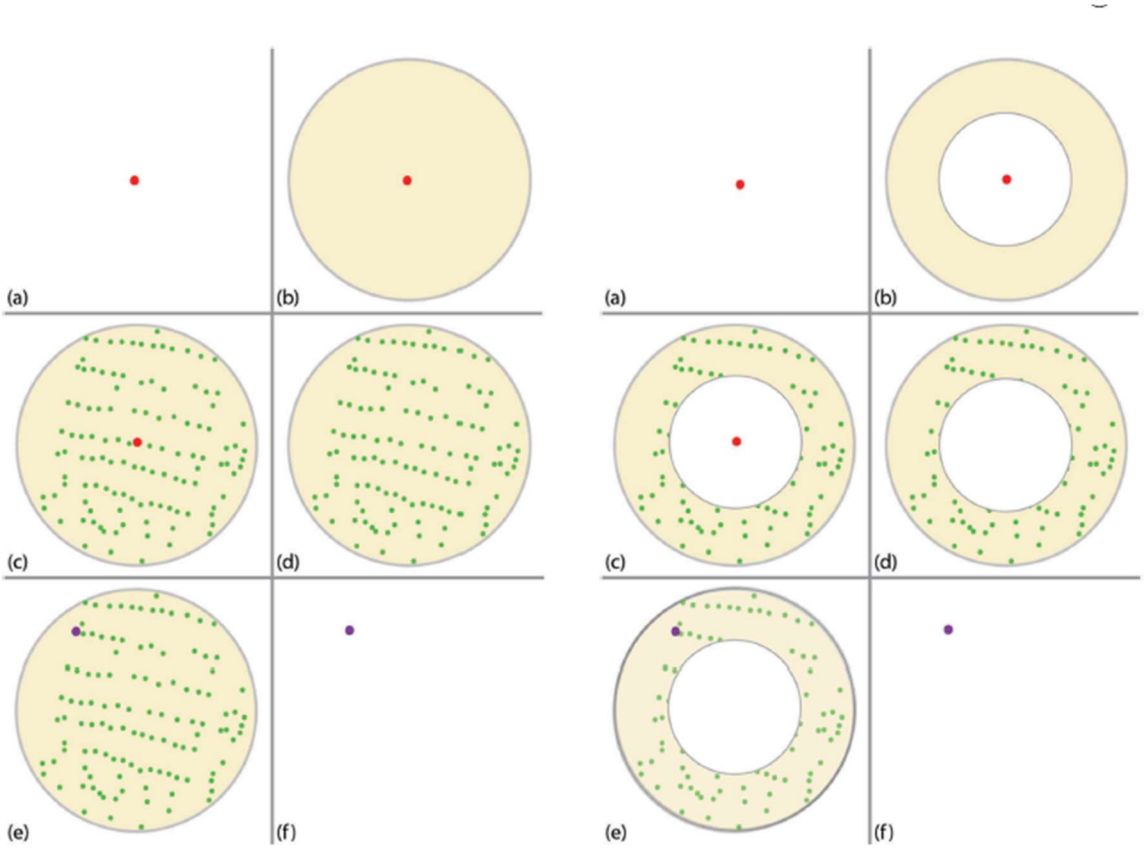


Figure 6 Location-swapping and location-swapping-with-donut by Zhang et al. (2017)

Swanlund et al. (2020) also considers the advantage of the masking technique of Zhang et al. (2017) called swapping method and Richter (2018) called verified neighbor approach, which shift the masked point to an actual address.

Swanlund et al. (2020) argue that the motivation of reducing the information loss caused by geographical masking leads to the evolution of geographical masking methods which include additional information, such as population data in administrative areas, are used in donut masking (Hampton et al. 2010) and weighted random perturbation (Kwan et al.

2004), as well as address points in location swapping (Zhang et al. 2017) and verified neighbor (Richter 2018). Nevertheless, the verified neighbor approach by Richter (2018) is also enhancing the risk of false re-identification. Finally, another technique is the street masking technique by Swanlund et al. (2020), which take the risk of false re-identification into consideration during the masking by shifting the point to the intersection of a street network, which as a result may shifts several points to one intersection. According to Seidl et al. (2015) the masking techniques which lean on adding two or more points to one position intensify the obfuscation, which increases the k-anonymity.

Authors such as Zandbergen (2014), and Seidl et al. (2018) summed that in literature it is distinguished between the area associated with the original point and the masked point, where the individuals or locations are used to calculate k. Allshouse et al. (2010) define the estimated k-anonymity as corresponding to the number of homes in an administrative unit, predefined by the original point. They used the following equation, where D is the Euclidean distance between the initial and shifted position, and the population density as number of households (N) in an area of an administrative unit (A) under the premise of a homogeneous distribution of households:

$$k_{est} = \pi * D_i^2 * (N_i/A_i)$$

Also, Allshouse et al. (2010) define the actual k-anonymity in their research as the number of households, that are nearer to the original point than the masked point through their data file E911 file. Nevertheless, the authors also point out that the assumption of a homogeneously distributed population may lead to an overestimation of k-anonymity. Further, Seidl et al. (2015) points out that the approach of estimated k-anonymity has a shortcoming regarding false re-identification. Even if the calculated k has a high value for the original point, it may be shifted to a site with low population density. Thus, it is suggested to include k-anonymity for the masked location as well.

Seidl et al. (2018) recap that the concept of spatial k-anonymity cannot just be used to measure the performance of the masking technique in terms of privacy, but also can be used to provide weighted distances during the masking procedure. Hampton et al. (2010) calculated the maximum distance of the donut shaped area using the underlying population density inversely. To that end the authors first determined a target k-anonymity value and reverse-engineered the maximum displacement distance, that is the outer circle of the donut. This leads toward a higher displacement distance in low population density areas and shorter displacement distances in high population density areas. Kounadi and Leitner (2016) aggregated the address point data to create a spatial database with a minimum k-anonymity to reduce the risk of re-identification for the masking method adaptive areal elimination.

Another influence on the calculation of k in spatial k-anonymity is the type of data available for the area of interest. Lu et al. (2012) argue that when relying on population density it may result in too small areas, since it is expected that there is an imbalance towards the

higher population rather than number of residential addresses due to larger households such as families. It is therefore suggested to use locations such as addresses over population density when computing k. To determine the spatial k-anonymity access to spatially highly resolved data, such as residential locations is necessary (Zandbergen 2014; Wang et al. 2022). Nevertheless, it can also work in a restricted way with aggregated reference data (Wang et al. 2022).

Seidl et al. (2018), which argue that the masking methods, which include spatial k-anonymity with distances in their computation, are constrained, because they do not include topological aspects. Moreover, a lot of basemaps include building parcels and orthophotos and therefor re-identification is not exclusively based on a point to point-paring task but rather a point to polygon or image paring task.

2.3. Human cognition and geoprivacy

There is little literature on human cognition and geoprivacy in general. On one hand it is pointed out by Seidl et al. (2019) that point pattern comparison has been investigated regarding human perception such as by Leitner and Curtis (2004, 2006) or Kounadi and Leitner (2015a). Other aspects are the perception of sensitive data on online maps by Kim et al. (2021), Kounadi, Bowers and Leitner (2014) or Ray et al. (2012) and reducing the risk of false re-identification of geographical masked data by Seidl et al. (2018) and Seidl et al. (2019).

2.3.1. Human cognition and point pattern comparison

Leitner and Curtis (2004) compared the spatial distribution of masked and unmasked points plus their hot spots in an empirical study by human cognition. 82 participants compared 34 map pairs for ranking the precepted affinity of the spatial distribution and marking the precepted hotspots of the masked and unmasked data. The authors also used different basemaps to examine the possible influence on the perception of the affinity and constructed their maps with an empty basemap with no information, a basemap showing the borders of a census unit and a road system (see Figure 7). Additionally for examining another influence, the authors used ten different masking methods which were differentiated between global and local techniques. Their analysis indicates that the different basemaps had an influence on how the participants ranked the perceived similarities only when points were geographical masked by rotating by some random angle around the central point of each grid cell.

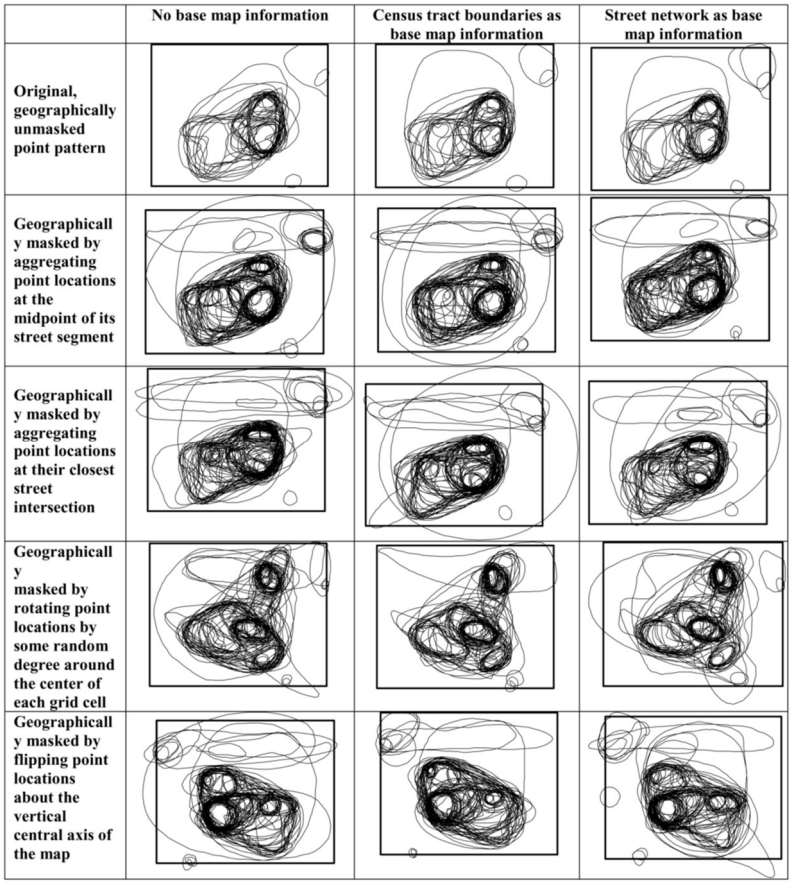


Figure 7 Marked hot spots perceived by humans (Leitner and Curtis 2004)

Leitner and Curtis (2006) also tested human perception in another empirical study with 42 map pairs and 62 participants. The applicants ranked the visually perceived similarity of the patterns on the map pairs and were asked to mark the hot spots as well. All maps had the same scale (1:31,800), using three related masking techniques which applied three different sizes of cells in their masking process. Also, three different base maps were applied, as mentioned before in Leitner and Curtis (2004). Nevertheless, differently to Leitner and Curtis (2004), Leitner and Curtis (2006) did not conclude an influence on the ranking of the perceived similarities. Yet, the masking methods and the cell sizes showed to have an influence on the perception. The smaller the cell sizes the more affine the patterns were perceived.

Another study conducted by Kounadi and Leitner (2015a) also included a ranking of the visually perceived affinity of 30 map pairs by participants. Their map pairs differed in the placement of one masking technique and two scales. The authors did not include any more variation on any of the map elements, as they wanted to avoid any additional effects on their results. Nevertheless, their study showed a strong link between the degree of inaccuracy in the hot spots induced through the masking methods and the perceived affinity.

These three studies mentioned above assumed and examined that map elements and different masking techniques have an influence on human perception in affinity by comparing masked and unmasked data. Nevertheless, as Seidl et al. (2019) pointed out that human perception has not been studied in the context of evaluating the privacy risk.

However, Kim et al. (2021) conducted a survey on the subjective perceived disclosure risk of a map with 865 participants. Their main objectives included the influence on the perceived disclosure risk by different map characteristics, two different masking methods and mapping socially vulnerable individuals. The participants were asked to rank their personal comfort while showing them their simulated mapped homes and trajectories. The authors saw the personal comfort as equal to the perceived disclosure risk. One of their main results regarding the masking methods, which included on one hand aggregation at different stages and on other hand point shifting, showed that a higher aggregation level and a larger shifting distance reduced the discomfort. Yet the participants expressed some doubt regarding the masking method which shifts points. A point moved by a small distance within its vicinity displeased the participants as it would be better to move it a bigger stance. Also, a wrongly association of a moved point to another, inculpable person was in the apprehension of the participants. The authors already linked their insight to the issue of false identification mentioned by McLafferty (2004) and employed by Seidl et al. (2018).

In the study of Ray et al. (2012) participants expressed favorable views regarding the choice of visualization of crime locations on an online map, in which the locations were moved to the streets to maintain privacy whilst spatially depicting the location.

2.3.2. Disclaimers of geographical masked data

While Armstong et al. (1999) published the geographical masking methods, the authors argued that revealing the information of the masking method and its parameters might be useful for the data analyst but recovers an increasing risk of re-identification especially if this information is exposed to a data infiltrator. Armstrong et al (1999) and Zimmerman and Pavlik (2008) query whether safekeepers of geographical data should reveal the mask metadata additional to the masked data. According to the authors, mask metadata refers to the applied detailed information during the geographical masking process. The authors debate that the revelation of mask metadata could be beneficial for researchers, as it would potentially allow them to assess the robustness of their findings or determine the feasibility of the analysis in the first place. On other hand there's a significant concern that disclosing mask metadata might also assist a malicious entity attempting to identify individuals within the dataset, according to the authors. As the results of the statistical model of Zimmerman and Pavlik (2008) showed, the release of several masked variations of the data, including mask metadata, rises the probability of revealing the data. This research was not conducted on humans' perception.

Researchers such as Kounadi and Resch (2018) recommend using disclaimers on online publications for sensitive data collected by participatory sensing on maps to reduce any possible misapprehension regarding the map reader. Simultaneously, the authors suggest not to include any further information about the meta data of the anonymization method to avoid exposition of the confidentiality of participatory sensing data. Specifically, the authors state that there is no universal framework for disclaimers, rather their content is influenced by the

publication and visualized information, which are open to the perception and understanding of each individual.

An interesting example for the influence of disclaimers of geographical masked maps was the study of Kounadi et al. (2014) which was conducted using human perception and showed that 51 percent of the study participants construed the masked crime positions published on the police.uk as the genuine position of the crime. Seidl et al. (2019) examined the frequency of masking disclaimers on confidentiality in re-identification, using visual perception as well. They differ between disclaimers which show up on each map, a disclaimer which is shown once and no disclaimers at all. Their studies suggested that displaying once the disclaimer of masked data lowered the confidence of the participants, while showing it on every map reinforced that effect.

3. Methodology

To answer the research questions, a survey was conducted. Therefore, data had to be produced, areas of interests had to be chosen, the survey including 10 maps in one form had to be produced. There were eight different forms in total, which differed in terms of the displayed information.

3.1. Simulated dataset at household level

A point data set representing the geographical location such as coordinates of individuals is needed. For Vienna there is no publicly available point data set representing health data or individuals in a household. Therefore, the data needed to be simulated by the author herself. This includes modelling data, which is mainly publicly available on *data.gv.at*, and could represent a person in a household with a certain health condition. For the modelling, the point data set of information on buildings (City of Vienna - https://data.wien.gv.at (a)) and the polygon data set with buildings (City of Vienna - https://data.wien.gv.at (b)) was used to get just the polygon buildings with residential utilization. The points, representing an individual with a health condition living in a household, where spatially randomly distributed within the predefined polygon buildings with residential utilization, but the number of points was dependent on the population count from 2012 within the registration district. The data set with the registration districts and the population data has was provided by City of Vienna - https://data.wien.gv.at (c) as well. This concept of modelling the data was chosen to make sure that several points can fall in one building or in none.

In order to distinguish between registration districts with low and high population, a threshold is defined, which is used to generate six different data sets (see Table 1). On the one hand the mean value of the population counts of the registration districts in 2012 is used as a threshold, which was 7,999. On the other hand, the standard deviation 4,539 of the population of these registration districts is used as a threshold. For each threshold three distinct data sets were created. A registration district with a lower population count has a higher probability for fewer sick residents than a registration district with a high population count. If the population count in the registration district of 2012 is smaller than a certain threshold, then there were 10, 20 or 30 points spread, otherwise 50, 100 or 150 points. The procedure made sure that the points were only spread inside the buildings which lie inside the population district.

Nr. of dataset	threshold	points spread, if the population count was smaller then the threshold	points spread, if the population count is greater or equal to the threshold	number of generated points
dataset 1	7,999	30	50	7,090 points
dataset 2	7,999	20	100	10,460 points
dataset 3	7,999	10	150	13,830 points
dataset 4	4,539	30	50	8,010 points
dataset 5	4,539	20	100	14,140 points
dataset 6	4,539	10	150	20,270 points

Table 1 Created data sets

3.2. Areas of interest/Study area

The study area will be the city of Vienna, which is the capital of Austria. As areas of interest five different neighborhoods with different types of residential building structures were chosen (see Figure 8).

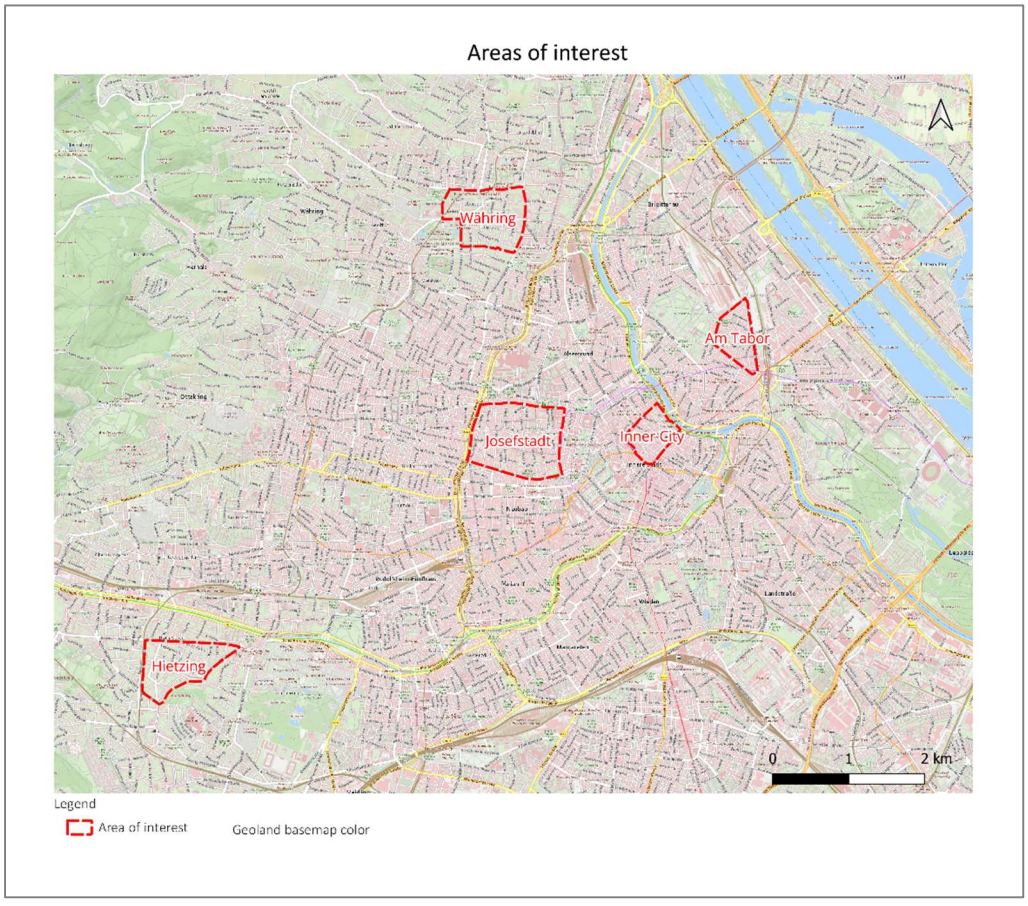


Figure 8 Areas of interest

3.2.1. AoI – Hietzing

Area one is located in the 13th district, Hietzing. It lies between Hietzinger Straße, Lainzer Straße and Hummelgasse. The northwest of Hietzinger Straße is a shopping street which serves as a town center in the district and borders outside to the area of interest. The building structures become more and more loose and rural because of the single-family houses with gardens. Inside of the area there is the small green area, Hügelpark. Nevertheless, Schloss Schönbrunn with its gardens is very close. In the south borders the Küniglberg.

3.2.2. AoI – Währing

Area two is situated in Währing, the 18th district of Vienna. It lies between Krottenbachstraße, Gymnasiumstraße, Haizingerstraße and Max-Emanuel-Straße. In the western part of the area lies the Türkenschanzpark, which is a greening area, and university buildings such as the Department of Astrophysics in the Sternwartepark and the main building of the University of Natural Resources and Life Sciences. The residential buildings of this area consist mainly of houses and villas surrounded by fenced in gardens.

3.2.3. AoI – Inner city

The fifth area is located in the first district, between Franz-Josefs-Kai, Rotenturmstraße, Graben and Tiefer Graben. This part of the city is famous for its dense building structure and narrow streets. The area is close to Graben and Stephansplatz, where touristic shops, restaurants and high end retail dominate the cityscape.

3.2.4. AoI – Josefstadt

Area four is the 8th district named Josefstadt. It is located between Alser Straße, Hernalser Gürtel, Lerchenfelder Straße and Landesgerichtsstraße. It has typical old courtyard residential buildings and a very dense building structure. There are hardly any greening areas such as parks.

3.2.5. AoI – Am Tabor

The fifth area of interest lies in the second district between Taborstraße, Nordbahnstraße and Heinestraße. It has old residential buildings. The Taborstraße is a shopping street. The southern part of this area leads to Prater Stern, a hub for motorized and public transit in Vienna. This part is pretty dense and has no green space. In the center lies the Volkertplatz, with market stands.

3.3. Chosen basemap

For the survey maps will be printed for the participants. Nevertheless, available spatial online data was downloaded or used for the survey. The maps contain the masked data sets with the following information layers:

- Orthophoto as from the WMS- https://www.wien.gv.at/viennagis/ (a)
- General map as from the WMS- https://www.wien.gv.at/viennagis/ (b)
- Building parcels (City of Vienna – https://data.wien.gv.at (b))
- Streets from the intermodal transport reference system of Austria (GIP.at) (City of Vienna - https://data.wien.gv.at (d))

The orthophotos cover the city of Vienna as aerial images, which have a pixel scale of 15 centimeters per pixel and were taken on 25th and 26th of March 2021 (City of Vienna - https://data.wien.gv.at (f)).

The multi-purpose area map is a general basemap and was offered over WMS by the City of Vienna (WMS- https://www.wien.gv.at/viennagis/ (b)) as a raster data and includes a pixel scale of 25 centimeter. As in the data set of the multi-purpose areas described (City of Vienna – https://data.wien.gv.at (b)), it includes different land use categories such as buildings, all built-up areas such as traffic areas, green areas such as parks, gardens and yards. Additionally, it shows street names and house numbers to the according address (City of Vienna - https://data.wien.gv.at (g)).

The building parcels are from the data of the City of Vienna – https://data.wien.gv.at (b) and was downloaded from the vector data multi-purpose map web feature service. The purpose to include this kind of feature layer as a basemap was to provide a layer with only the level of information that let the participants are able to refer to an existing building structure.

The last basemap was the traffic route from the Intermodal transport reference system of Austria (GIP.at) (Streets from the intermodal transport reference system of Austria (GIP.at) (City of Vienna - https://data.wien.gv.at (d)). It is a graph which consists of edges and nodes and includes street network, rail network, bicycle tracks and footpaths. For the re-identification task only the edges were visualized to provide as little information as possible but enough to infer the given topology such as building blogs, parks or plazas.

3.4. Choice of geomasking technique

To examine the two geomasking techniques for objective three, one of the geomasking methods will be chosen by the criteria of including the reduction of false re-identification. According to Swanlund et al. (2020) their developed street masking technique provides the reduction of false re-identification by shifting the points to the nodes of the street. The steps for the street network are displayed in Figure 9. Firstly, the point is snapped to the nearest street intersection of the downloaded street network provided by Open Street Map. Then a depth value is given by the user, which corresponds to the number of neighboring intersections or dead ends. In the next step the mean distance to all these nodes to the starting point is computed. Finally, the point is shifted to the node that is closest to the mean distance computed in the previous step. The underlying idea of this method is that the street network reflects the urban structure. A neatly meshed street network results in a smaller displacement distance than a loosely meshed one. This ensures that urban areas, which are assumed to have more neatly meshed street network, points are not shifted too far away from their origin, as they would be in rural areas, with an assumed loosely meshed grid.

The parameter, which was chosen for the maps in this survey, was a search dept of three, to provide a simple re-identification, as there was no experience or literature to rely on for chosen parameters for re-identification by human cognition.

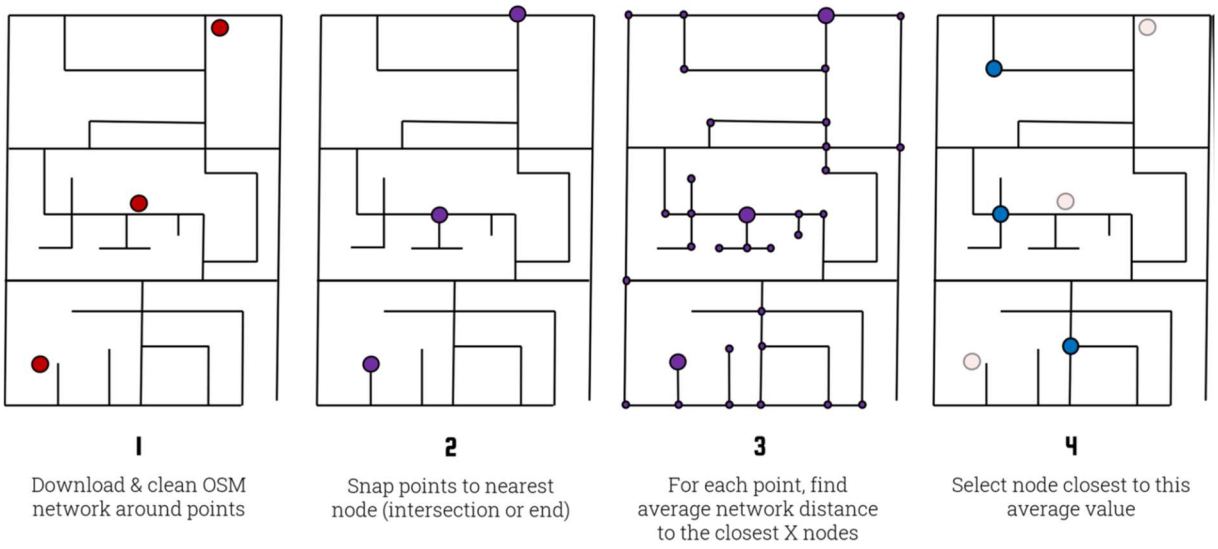


Figure 9 Steps of street masking by Swanlund et al. (2020)

The other masking method is the donut masking technique by Hampton et al. (2010), as it does not consider as Seidl et al. states (2018) false identification. Because of the randomness interwoven into the donut masking technique (Hampton et al: 2010), the strategy to provide simple re-identifiable masked points had to be changed into reducing the

displacement distance. Therefore, the spatial k-anonymity value for calculation of the outer radius was chosen to be 15, which can be interpreted as such that the masked residential point, representing a person, is indistinguishable from 14 other persons in the vicinity inside a circle. The radius of the circle is the distance between the masked point and the original location (Hampton at al: 2010). For the inner radius 20 percent of the maximum radius was used.

To create the masked points for the survey the open source Python package *maskmypy* was used on the created residential points which represent individuals (see Figure 10). To mask the points a python script was written based on the example code given in Swanlund et al. (2020). For the street masking the following code example was used for one map:

```
import geopandas as gpd # for loading/saving data

from maskmypy import Street

print("ladedaten")

input_points =
gpd.read_file(r"C:\Users\isabe\Documents\kartenproduktion\karten\Karte3_SM\Karte3_7999_30_50_ausschnitt.shp") #load points

street_mask = Street(input_points, depth=3) #Create the mask

print("maskieredaten")

masked_data = street_mask.execute() # Execute the mask

street_mask.displacement_distance() #calculate distance

masked_data.to_file(r"C:\Users\isabe\Documents\kartenproduktion\karten\Karte3_SM\K3_7999_30_50_D3.shp") #save results

print("fertig")
```

Figure 10 Python script for street masking

A similar script was created for donut masking also using the *maskmypy* Python package. It was as the following example for one map (see Figure 11):

```
import geopandas as gpd # for loading/saving data

from maskmypy import Donut_MaxK

print("ladedaten")

input_points =
gpd.read_file(r"C:\Users\isabe\Documents\kartenproduktion\karten\Karte9_DM\Karte9_4539_30_50_ausschnitt.shp") #load points

population=gpd.read_file(r"C:\Users\isabe\Documents\py\population_regDistririct\population_regDistrict.shp") #load population

donutmask = Donut_MaxK(

   input_points, # Name of the sensitive geodataframe

   population, # Name of the census geodataframe

   population_column='pop2012', # Name of the column containing the population field

   max_k_anonymity=15, # The maximum possible k-anonymity value

   donut_ratio=0.2, # The ratio used to define the minimum possible k-anonymity value.

   distribution='uniform')#, # The distribution to use when displacing points. Other options include 'gaussian' and 'areal'. 'Areal' distribution means points are more likely to be displaced further within the range.

   #container=container) # Optional, a geodataframe used to ensure that points do not leave a particular area.

print("maskieredaten")

masked_data = donutmask.execute() # Execute the mask

donutmask.displacement_distance() #calculate distance

masked_data.to_file(r"C:\Users\isabe\Documents\kartenproduktion\karten\Karte8_DM\K9_4539_30_50_k15.shp") #save results

print("fertig")
```

Figure 11 Python script for donut masking

3.5. Choice of map section

Each of the six datasets were geographically masked by donut masking and street masking. Finally, two sections from each of the five areas of interest was chosen under the considerations of the masked points. The maps A to E included sections with street masked points, the maps F to J included donut masked points (see Table 1). As there is no study or publicized survey on re-identification by human perception of actual geographical masked data and there was little literature or experience to rely on, sections with simple re-identifiable masked points were included. For the maps A to E, some of the street masked points shifted into a dead-end street. Due to the selected parameters, the locations of the unmasked points were in the nearby street section where the dead-end-street leads into.

Map A, B, F and G show areas of interests with loose urban structure, which is why a number of points with a single figure were represented in these maps (see Table 2). In the other maps C, D, E, H, I and J 13 to 31 masked points lie inside the maps. Finally, the scales of all maps were between 1:1,000 and 1:1,339. As claimed by Kounadi and Leitner (2014), a bigger scale should facilitate the task of re-identification. This thought was incorporated into the map visualization for the survey.

Map section and masked points					
Map section	areas of interest (=aoi)	masking method	Masked data set	scale	number of points
A	aoi 1	street masking	dataset 6	1: 1,200	6
B	aoi 2	street masking	dataset 1	1: 1,339	2
C	aoi 3	street masking	dataset 1	1: 1,200	13
D	aoi 4	street masking	dataset 2	1: 1,300	14
E	aoi 5	street masking	dataset 5	1: 1,300	17
F	aoi 2	donut masking	dataset 1	1: 1,000	4
G	aoi 1	donut masking	dataset 2	1: 1,100	3
H	aoi 3	donut masking	dataset 2	1: 1,339	14
I	aoi 4	donut masking	dataset 4	1: 1,200	15
J	aoi 5	donut masking	dataset 6	1: 1,299	31

Table 2 Map section and masked points

3.6. Chosen disclaimer

To examine the quality of the disclaimer on the masking techniques, two levels of information were developed. One level of information had the general description of the masking technique in writing and in figures. The other disclaimer of the more detailed information level had additional information of the parameters.

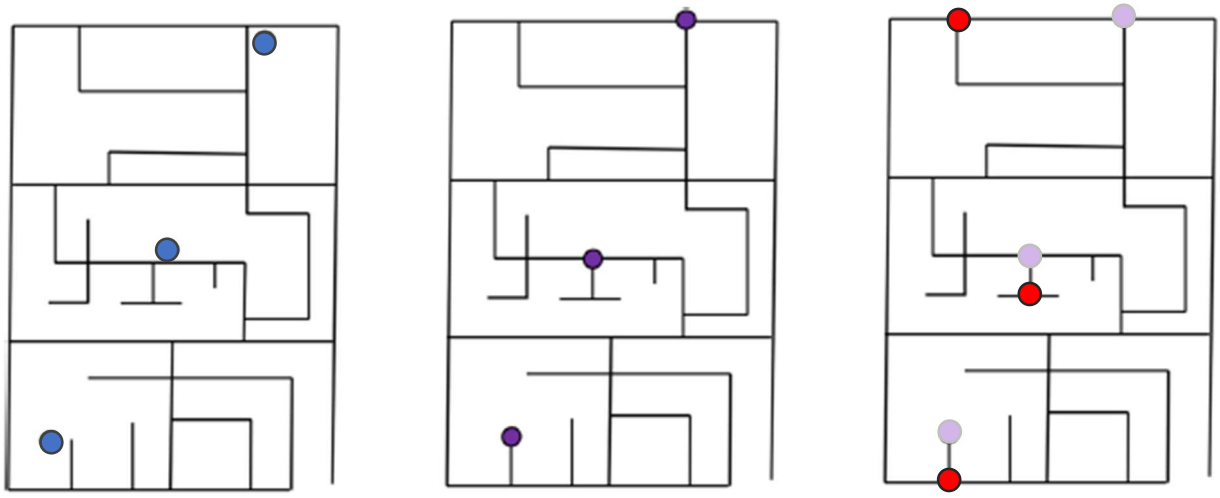
Additionally, the information of the applied masking techniques was described in the map title and description of the data sources to provide an allocation of the corresponding technique. Although, the results of the empirical study by Seidl et al. (2019) showed that a frequent display of the masking method reduced the confidence by the participants, in this survey it was necessary to supply the participants with this information on every map to avoid any confusion on the masking method and the parameters. This consideration has its origin in the investigation of Kounadi et al. (2014), where participants misunderstood the masked points as the original location, even though a disclaimer was included on the map.

3.6.1. Information level 1

The street masking technique of level one was described as the following:

Disclaimer of the masking method - Street masking

The method used to mask the points of the following five maps is known as street masking (Swanlund et al. 2020). As the name describes, the points are shifted from the residential place to the closest street intersection or closest end of a dead-end street in the street network of Open Street Map. From there the points are moved to a different street intersection or to an end of a dead-end street.



Figures modified from Swanlund et al., 2020

The unmasked points (blue) in their original location.

In the next step they are moved to the closest street intersection or end of a dead-end.

Finally, they are moved to a close street intersection or end of a dead-end. The masked points are shown in red.

Figure 12 Disclaimer of street masking information level 1

For the donut masking disclaimer of information level one also described the general masking method and included a figure:

Disclaimer of the masking method - Donut masking

The method, which is used to mask the points in the following five maps, is called donut masking (Hampton et al. 2010). As the name describes the masked points are inside a donut shaped area.

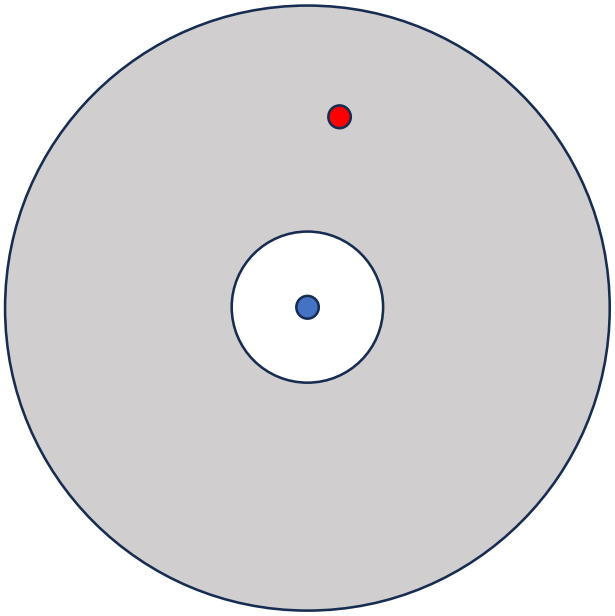


Figure based on Hampton et al., 2010

The blue, unmasked point is in the center of the donut. During the masking procedure it is moved into the shaded area (=red point).

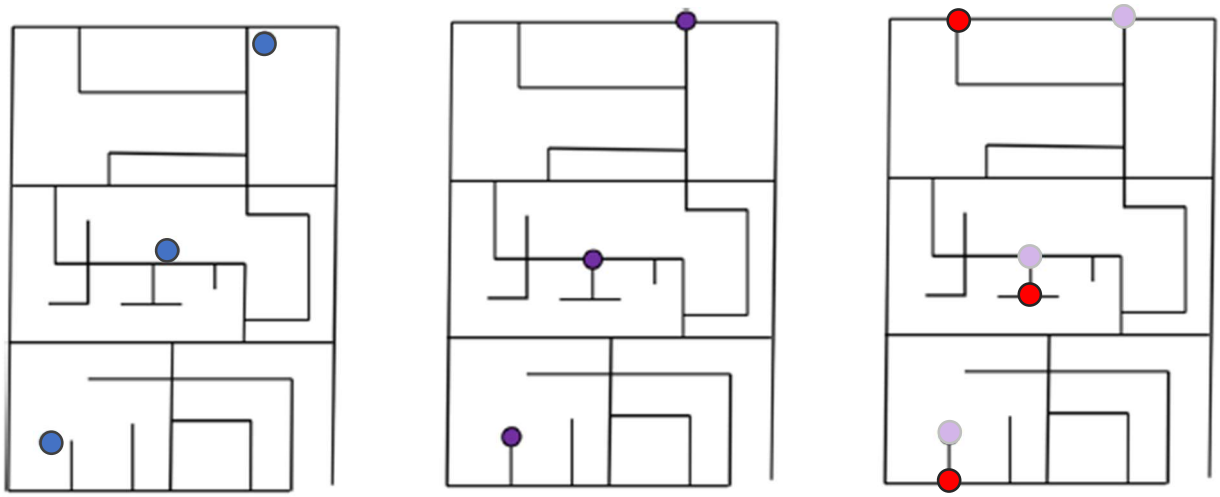
Figure 13 Disclaimer of donut masking information level 1

3.6.2. Information level 2

For the street masking method, the general description was included as well as the parameters which was a threshold of tree:

Disclaimer of the masking method - Street masking

The method used to mask the points of the following five maps is known as street masking (Swanlund et al. 2020). As the name describes, the points are shifted from the residential place to the next intersection or the next end of a dead-end street of the street network (=nodes). This will be the starting node for the next step, in which two close nodes of the network are chosen. The distances from each of the two close nodes to the starting node are calculated. These two distances are summed up and divided by three, to calculate the average distance of the 3 nodes (starting node and the two close nodes). The point is masked by shifting it to one of the two close nodes that is closest to the average distance.



Figures modified from Swanlund et al., 2020

The unmasked points (blue) in their original location.

In the next step they are moved to the closest node (=street intersection or dead-end).

Finally, they are moved to another node depending on the average distance of two close nodes and the closest node from the image to the left. The masked points are shown in red.

Figure 14 Disclaimer of street masking information level 2

For the disclaimer regarding the donut masking technique the general information is given additional to the parameters ok k=14 to anticipate the maximum and minimum radius of the donut shaped area, where the point moved. The description was as follows:

Disclaimer of the masking method - Donut masking

The method, which is used to mask the points in the following five maps, is called donut masking (Hampton et al. 2010). Donut masking moves the unmasked point from its original location to a random location within a donut-shaped area around the original position. The outer radius of the donut is the maximum shifting distance and is chosen so that only 14 other people are living inside. The inner radius is the minimum shifting distance and is 20 percent of the outer radius.

Our masked point is randomly placed somewhere in the donut shaped area around the original point, so that it cannot be distinguished from the other individuals inside the area.

Therefore, in a low population density area the point is shifted farther away than in a high population density area.

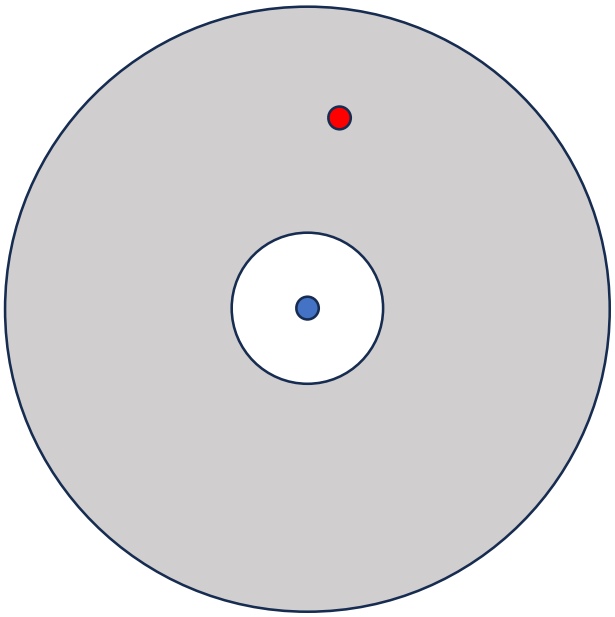


Figure based on Hampton et al., 2010

The blue, unmasked point is in the center of the donut. During the masking procedure it is moved into the shaded area (=red point).

Figure 15 Disclaimer of donut masking information level 2

3.7. Preparation of the forms

By including four different backgrounds and two different disclaimer information levels, there were a total of eight combinations of the different map element possible (see Table 3). There was only one of these eight combinations of map elements in one form, in order to avoid biasing the participant towards one or the other map background. Further, to keep a uniform testing ground within one form there was only one level of information regarding the disclaimer of masking technique involved.

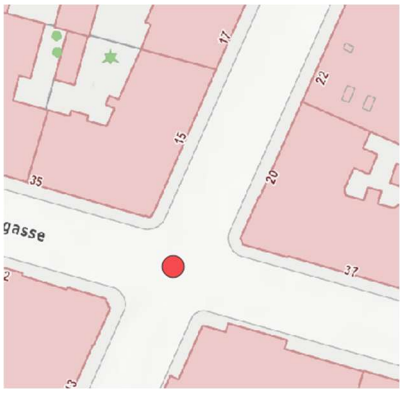
	Form 1	Form 2	Form 3	Form 4	Form 5	Form 6	Form 7	Form 8
basemap	multi-pur-pose area map	orthophoto	building parcel	intermodal transport reference system	multi-pur-pose area map	ortho-photo	building parcel	intermodal transport reference system
disclaimer infor-mation level	disclaimer 1	disclaimer 1	disclaimer 1	disclaimer 1	disclaimer 2	dis-claimer 2	disclaimer 2	disclaimer 2

Table 3 Combination of basemap and disclaimer

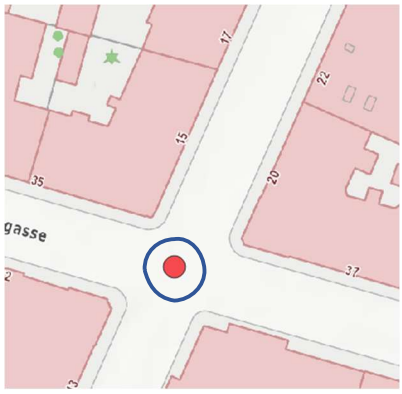
The general structure of the form included a consent and confidentiality part, a description of the theme and the explanation of the task. The participants were asked to follow the instructions:

Your task

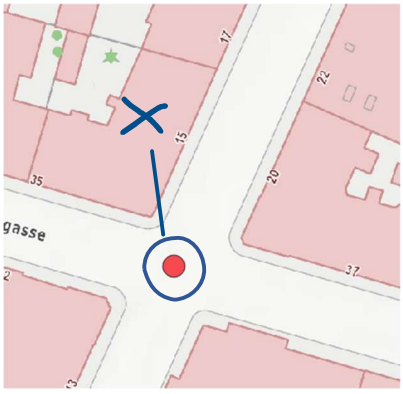
To complete the task, please follow the instructions in the images below.



Look at this section of a map. The red points are masked points representing an individual.



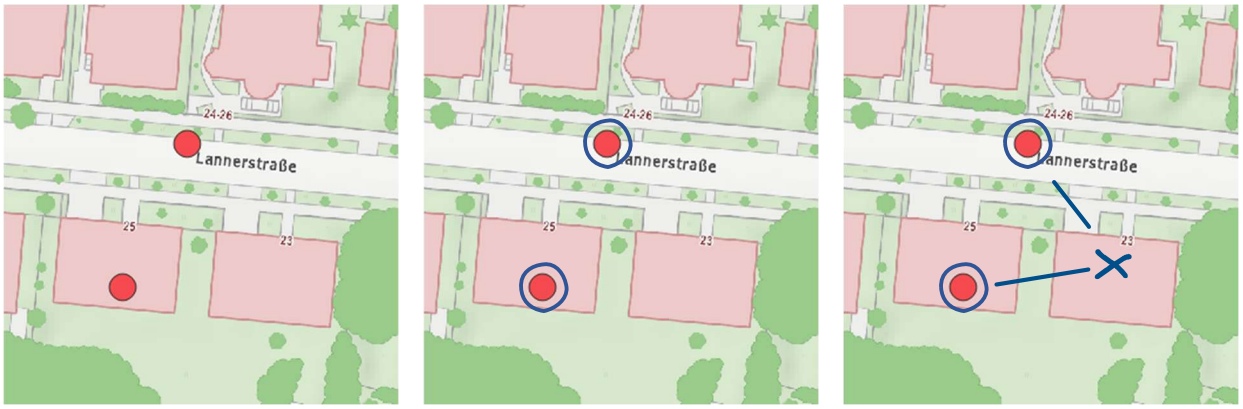
Do you think you can find the original location and hack the place? Choose one or more of the red points and mark it/them.



Make a cross where you think the original point is situated. Connect your chosen point with the cross.

Figure 16 Description of the task I

Please keep in mind that the original points are just on locations, where there is actual residential use. There may be more than one individual in one housing. If you think that this is the case, please mark the masked points and estimate the location of origin by marking it on the map as well as shown below.



The points with a number represent several individuals, which are moved to the same location. Please pick a point by marking it. Try to figure out the location of origin by marking it on the map as well. If the point represents more than one individuum, please feel free to mark as many locations as are shown by the number on the point. For example, if the point has the number two, then you may mark a maximum of two locations of origin on your map as shown below.

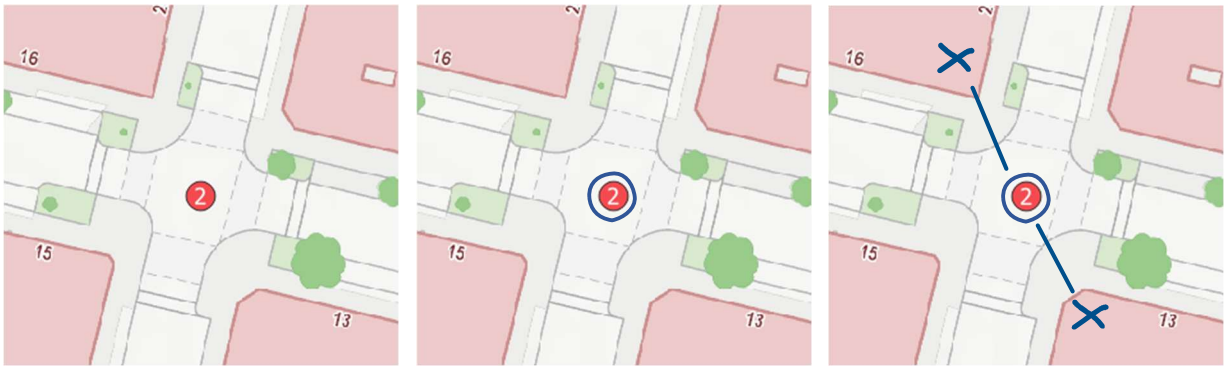


Figure 17 Description of the task II

This was followed by the disclaimer (information level one or two) with a description of the street masking technique by Swanlund et al. (2020). The next part of the form were five maps (A to E) with street masked points. The participants were asked to find the original position of at least one masked point according to the instructions for every map. Each map was accompanied by the question of "How confident do you feel in figuring out the original position?" (see Figure 18), as it is assumed that their level of confidence reflects the degree to which someone believes their re-identification is accurate. Therefore, the Likert scale by Likert (1932) extending from "completely confident", "fairly confident", "somewhat confident", slightly confident" and "not at all confident" was used. This kind of assessment by participants

was similarly used by Seidl et al. (2019) when participants were asked to assess their confidence in their assignment as well. Differently, the authors argued that confidence is indicative of the comprehension on geomasking.

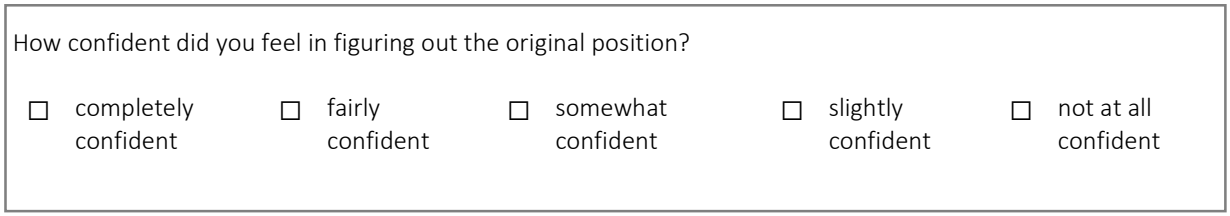


Figure 18 Confidence question on every map

Every map showed a section of one of the five study areas, so that every study area was included. After the first five maps, another disclaimer (information level one or two) with a description of the donut masking technique by Hampton et al. (2010) was illustrated. And another five maps (F to J) with the donut masked points were shown. In the last part same basic demographics like age, gender, school leaving qualifications, work in the field a related field of geography or with spatial data was asked. Additionally, a control question regarding the instructions of the masking methods "Did you find the instructions of the masking methods helpful?" was the final question to figure out if the participants have read the instructions.

3.8. Survey information

The collection of the data through the survey was conducted within the immediate social sphere. Therefore, the participants which took part on the survey were not representative of a general population. To collect as much data as possible a kick-off was organized in the department of geography in a classroom on 31st of July 2023. 18 participants completed one of the eight different forms on this day. The forms were printed on A3 format and different drawing tools were provided. During the task the participants were reassured that the task was not a reflection on their IQ but was about the ability on humans on re-identification with the given information and their perception. Furthermore, it was pointed out that the geographical masking methods are supposed to obscure confidential locations to make somebody insecure to reduce the risk of re-identification. But the participants should not be put off in accomplishing the task, because it is possible to find the correct identification.

After that the survey was conducted until 1st of September. Other participants were given a form as PDF to print it at home or apply the task on the computer with a PDF-program. Some participants got a printed version to fill out at home. In total 49 participants conducted the survey.

The survey forms were given to the participants so that the forms were equally distributed among the participants. With a total number of 49 participants this results in 12 participants filling out form for orthophoto, building parcels and intermodal reference system.

4. Results and Discussion

Each of the eight forms was filled out by six people, except for form 5. Form number five was filled out by seven people, which totals in 49 people taking part in the survey. There are no participants with the highest education level of elementary school. The most participants were university graduates. Some participants even had an expertise with spatial data. Most participants were in the age class of 20 to 34 and 35 to 49. In total there was a gender ratio of approximately two to one in favor to males (see Table 4).

highest education level	Form 1	Form 2	Form 3	Form 4	Form 5	Form 6	Form 7	Form 8
elementary school graduate	0	0	0	0	0	0	0	0
secondary school graduate	1	0	0	0	1	1	0	1
university entrance level	0	1	1	1	0	0	0	0
university graduate	5	5	5	5	6	5	6	5
expertise with spatial data	3	2	1	3	2	1	3	2
age class								
0-14	0	0	0	0	0	0	0	0
15-19	0	0	0	1	0	0	0	0
20-34	4	1	4	3	3	1	3	5
35-49	2	5	2	2	3	3	2	0
50-64	0	0	0	0	1	1	0	1
65	0	0	0	0	0	1	1	0
gender								
male	4	3	3	4	5	4	5	4
female	2	3	3	2	2	2	1	2

Table 4 Demographics of the survey participants

For every map the average displacing value, as well as the estimated k-anonymity by Hampton et al. (2010) with census data (City of Vienna - https://data.wien.gv.at (c)) and true k-anonymity by Allshouse et al. (2010) with an address data set (City of Vienna - https://data.wien.gv.at (e)) was calculated (see Table 5). Comparing the street masked maps A to E, the average displacing distance per map was in the range between 59.45 and 113.18 meters. The donut masked maps F to J have an average displacing distance between 9.64 and 24.89 meters, which is lower compared to the street masked maps. This is also reflected in both k-anonymity values. For the estimated k-anonymity the values range between map A and E from 113.23 to 1068.90, whereas it ranges in the maps F to J between 8.23 to 9.24. The true k-anonymity values are clearly lower, than the estimated k-anonymity. As already argued by

Allshouse et al. (2010), the estimated k-anonymity seems to be inaccurate and overestimated. Nevertheless, Swanlund et al. (2020) also share the concern that non-residential addresses are included in the measurement of k-anonymity.

map	Average distance (m)	est k-anonymity	true k-anonymity
A	86.51	113.23	29.78
B	81.51	154.96	24.50
C	59.45	122.70	28.08
D	113.18	1,068.90	87.33
E	65.99	482.41	22.59
F	19.76	8.23	1.50
G	24.89	8.82	0.00
H	15.18	8.54	1.50
I	10.36	8.31	0.67
J	9.64	9.24	0.55

Table 5 Average distance and k-anonymity values by map

Firstly, the varying confidence for the different maps was examined. Figure 19 displays the frequency of different levels of confidence the participants experienced during the re-identification task by map (A to J). Every form included ten questions regarding the confidence of the re-identified residential points, which with 49 participants results in 490 answers (see Table 6).

Maps A to E show street masked points and maps F to J include donut masked points. Maps with small population densities, such as maps A and B, have low values in regard of frequency of the confidence level not at all confident, compared with all the street masked maps (A to E). This relation can also be seen for the maps F and G compared to all the donut masked maps (F to J). Simultaneously, the frequencies for the levels completely and fairly confident are highest in map A and B. On the other hand, this kind of relation with complete and fairly confident and population density is not reflected in the street masked points.

Kounadi and Leitner (2014) claimed that re-identification becomes easier with low population densities. Considering the underlying results of the survey it can be observed that the confidence in re-identification is somewhat higher for maps with low population density. This can be interpreted as such that the participants may perceive the re-identification task as easier in maps with low population density. In this context, it should be mentioned that in this survey the maps with lower population density have significantly fewer residential points than maps with higher population density. Under this premise one could argue that the low number of masked points influences the confidence in re-identification towards higher levels.

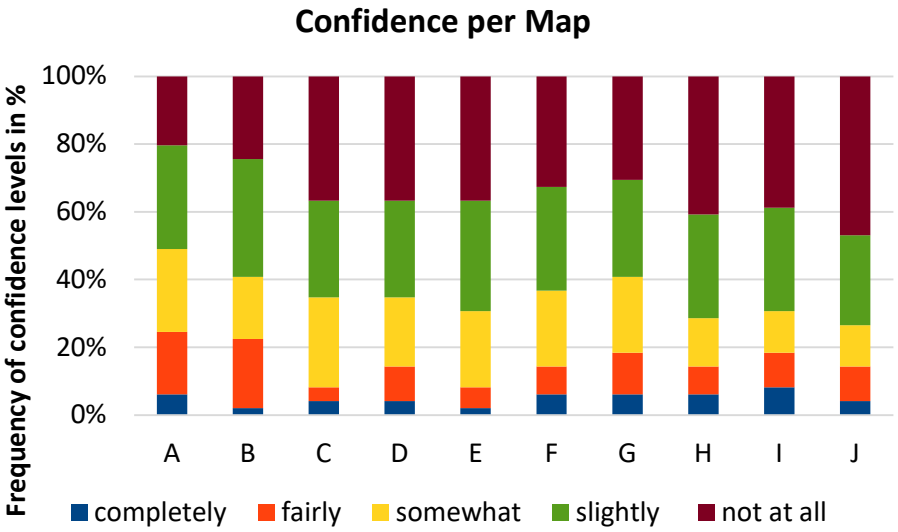


Figure 19 Frequency of confidence per map

re-identification	A	B	C	D	E	F	G	H	I	J
completely	6.12%	2.04%	4.08%	4.08%	2.04%	6.12%	6.12%	6.12%	8.16%	4.08%
fairly	18.37%	20.41%	4.08%	10.20%	6.12%	8.16%	12.24%	8.16%	10.20%	10.20%
somewhat	24.49%	18.37%	26.53%	20.41%	22.45%	22.45%	22.45%	14.29%	12.24%	12.24%
slightly	30.61%	34.69%	28.57%	28.57%	32.65%	30.61%	28.57%	30.61%	30.61%	26.53%
not at all	20.41%	24.49%	36.73%	36.73%	36.73%	32.65%	30.61%	40.82%	38.78%	46.94%
total numbers										
completely	3	1	2	2	1	3	3	3	4	2
fairly	9	10	2	5	3	4	6	4	5	5
somewhat	12	9	13	10	11	11	11	7	6	6
slightly	15	17	14	14	16	15	14	15	15	13
not at all	10	12	18	18	18	16	15	20	19	23
sum	49	49	49	49	49	49	49	49	49	49

Table 6 Frequency of confidence per map

Contrary to Kounadi and Leitner (2014), regarding the low population density maps it is not clearly reflected in the correct re-identification rates that re-identification was correct more often. Taking a look at Figure 20 and Table 7, it can be argued that for the street masked points on map A and B correct re-identification was higher compared to the other street masked points on the other maps. However, this cannot be said about the donut masked points where the correct re-identification was at a similar level except for map G, where it was the lowest. Map G also is one of the maps with low population density.

Moreover, it was not investigated if there is an influence on the confidence depending on whether the street masked or the donut masked maps were shown in the survey first.

Furthermore, Curtis et al. (2006) found that population density had in influence on their results of correct re-engineered points, the results of this study do not support their finding.

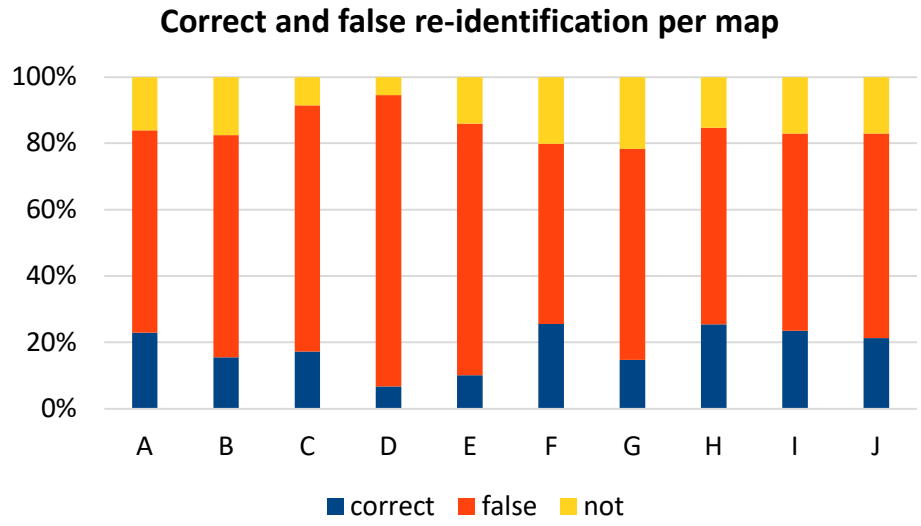


Figure 20 Correct and false re-identification per map

re-identification	A	B	C	D	E	F	G	H	I	J
correct	22.85%	15.46%	17.16%	6.72%	10.04%	25.54%	14.69%	25.37%	23.48%	21.23%
false	61.05%	67.01%	74.25%	87.81%	75.90%	54.35%	63.64%	59.31%	59.45%	61.67%
not	16.10%	17.53%	8.58%	5.47%	14.06%	20.11%	21.68%	15.32%	17.07%	17.10%
total numbers										
correct	61	15	92	43	70	47	21	154	154	283
false	163	65	398	562	529	100	91	360	390	822
not	43	17	46	35	98	37	31	93	112	228
sum	267	97	536	640	697	184	143	607	656	1,333

Table 7 Correct and false re-identification per map

4.1. Varying confidence per basemaps

Firstly, the varying confidence with regard to the different basemaps was examined. Figure 21 displays the frequency of different confidence levels the participants experienced during the re-identification task by basemaps. Each form contained ten confidence questions in addition to the maps, which results in 120 confidence statements per basemap for orthophotos, building parcels and intermodal ref. system and 130 for the basemaps with multi-purpose areas (see Table 8), which totals in 490 answers covering all forms.

The results exhibit that 33.33 % felt slightly confident and 34.17 % not at all confident concerning the re-identification with the multi-purpose areas as a basemap. The maps with building parcels show similar results regarding these two levels of confidence. The frequency of the confidence levels with basemaps displaying the multi-purpose areas and building parcels differ slightly at the level of completely confident, fairly confident and somewhat confident. Interestingly, maps with orthophotos distinguished the most in the frequency of confidence compared to other basemaps. It shows the highest level of confidence, with a rate of completely confident of 16.67 % and the lowest in the level slightly confident of 22.50% and not at all confident of 19.17%. This is especially interesting, because this basemap is the only one which does not include geometries, but remote sensing data for the re-identification task. Not surprisingly, the basemap with the intermodal reference system, which displays no buildings, had the least confidence. None of the participants felt completely confident during the re-identification task when provided with this kind of basemap and only 4,17 % felt fairly confident. It also had the highest rate of the level not at all confident with 50.00 % percent (=60 people).

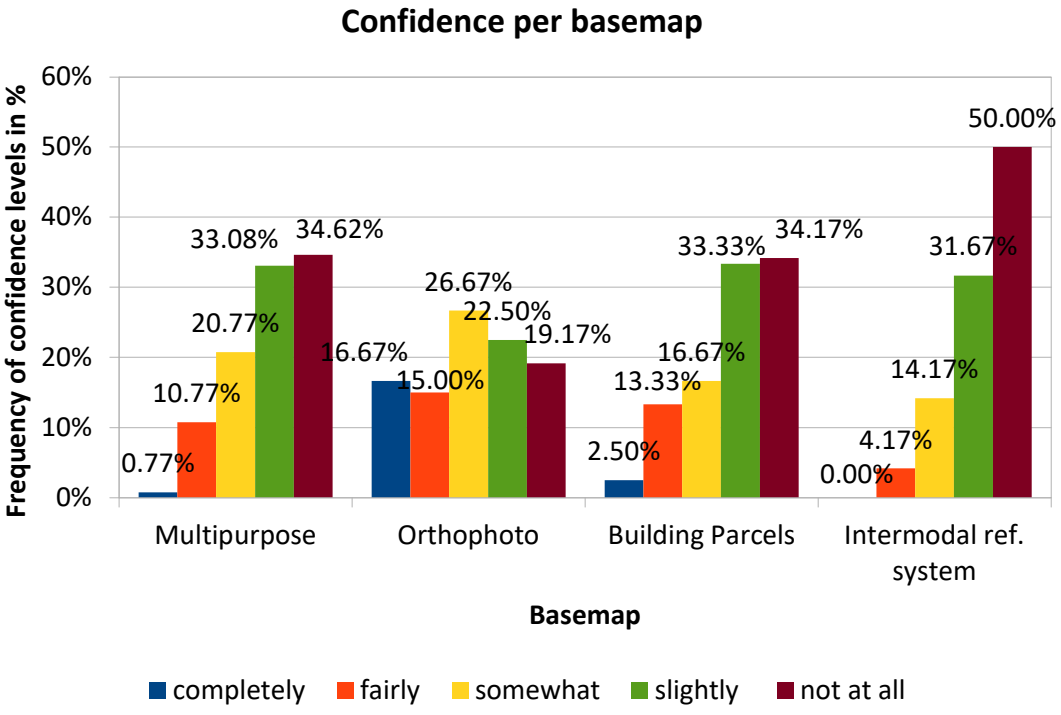


Figure 21 Frequency of confidence levels by basemap

Basemaps	Multi-purpose area	Ortho-photo	Building Parcels	Intermodal ref. system
confidence level				
completely	0,77%	16,67%	2,50%	0,00%
fairly	10,77%	15,00%	13,33%	4,17%
somewhat	20,77%	26,67%	16,67%	14,17%
slightly	33,08%	22,50%	33,33%	31,67%
not at all	34,62%	19,17%	34,17%	50,00%
Total numbers of counts				
completely	1	20	3	0
fairly	14	18	16	5
somewhat	27	32	20	17
slightly	43	27	40	38
not at all	45	23	41	60
Sum	130	120	120	120

Table 8 Frequency of confidence levels by basemap

4.2. Varying confidence by masking disclaimer

To answer the second part of the RQ1, which includes how confidence varies with the two different levels of information on masking techniques given in the disclaimers. Figure 22 illustrates the different levels of confidence by the two disclaimers. Due to the odd number of participants, disclaimer two had one more form filled out (see Table 9).

Clearly, disclaimer one, which has no information on the masking parameters, performs with a higher level of not at all confident (39.17%) than the disclaimer two where 30.00 % of the answers were not at all confident. Further, the frequency of the two levels of confidence for fairly confident and somewhat confident in disclaimer two was at 13.60 % and 21.60 %, which is higher than in disclaimer one with values of 7.92 % and 17.50 %. The levels completely confident and slightly confident are very similar within both disclaimers.

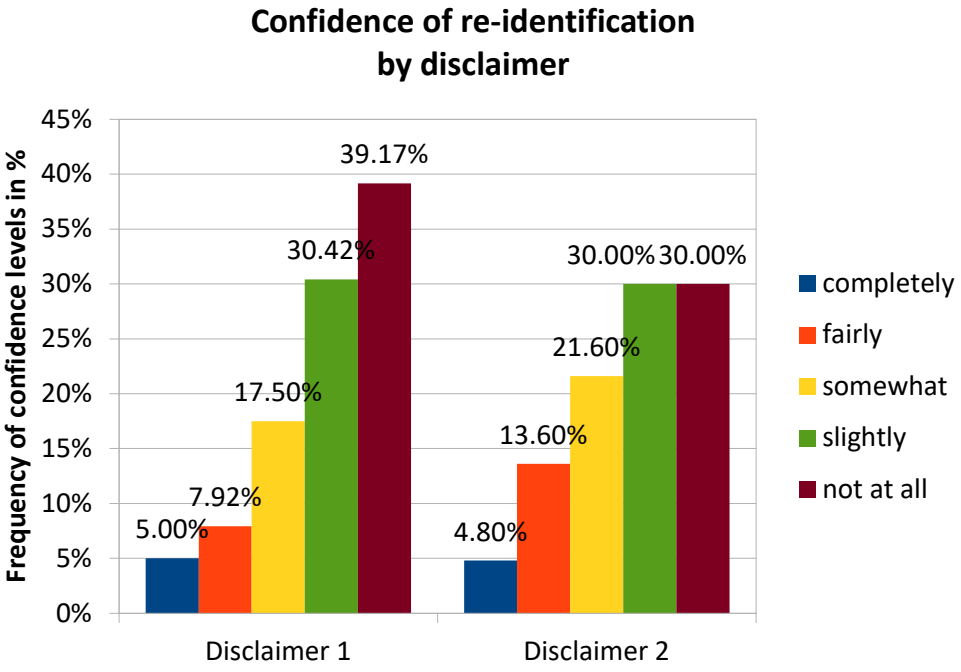


Figure 22 Frequency of confidence level by disclaimer

confidence level	Disclaimer 1	Disclaimer 2
completely	5,00%	4,80%
fairly	7,92%	13,60%
somewhat	17,50%	21,60%
slightly	30,42%	30,00%
not at all	39,17%	30,00%
total numbers		
completely	12	12
fairly	19	34
somewhat	42	54
slightly	73	75
not at all	94	75
Sum	240	250

Table 9 Frequency of confidence levels by disclaimer

Figure 23 and Table 10 show a boxplot of confidence by disclaimer, as well as the data. The center line of the box plots represents the median cumulative level of confidence as collected from the survey. The lower and upper boundaries on the box show the 25th and 75th percentile, respectively. The error bars depict the 5th and 95th percentile. The cumulative confidence is the sum of all confidence levels all one participant gave in the entire form. To this end numeric values have been assigned to the confidence levels according the Likert scale, with one corresponding to not at all confident, and five corresponding to completely

confident. In the analysis by disclaimer the minimum cumulative confidence is ten and the maximum cumulative confidence is 50, because there are only there are ten confidence questions on the form. A similar analysis has been conducted by Seidl et al. (2019) and provided the basis for this work.

It can be observed that the box of disclaimer two is higher than for disclaimer one. This can relate to a higher cumulative confidence for disclaimer two than for disclaimer one, which is reasonable since disclaimer two contained more information that could potentially have helped the participants in the re-identification task.

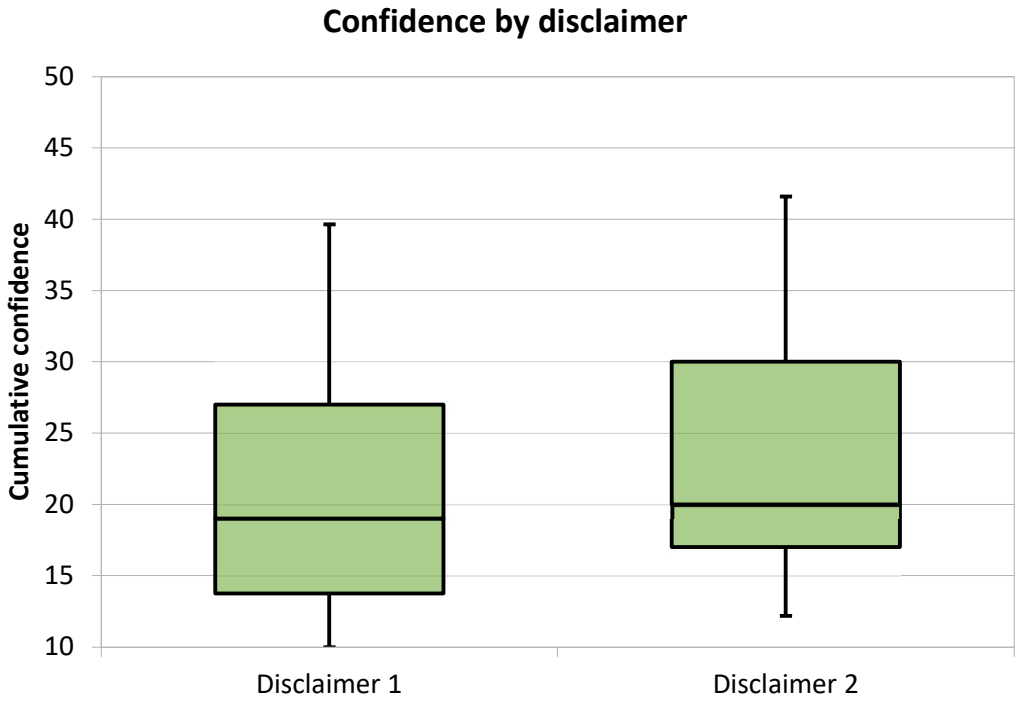


Figure 23 Boxplot – Cumulative confidence by disclaimer

	P(5%)	P(25%)	P(50%)	P(75%)	P(95%)
Disclaimer 1	10	13.75	19	27	39.65
Disclaimer 2	12.2	17	20	30	41.6

Table 10 Cumulative confidence by disclaimer

4.2.1. Varying confidence by basemap for disclaimer 1 and 2

To investigate how the confidence varied for the different basemaps and each disclaimer, data was analyzed separately for disclaimer one and disclaimer two. Figure 24 shows the histogram of the confidence by basemap for all participants who filled out the forms with disclaimer one (=form 1, 2, 3, 4). It is clearly visible that the basemaps of building parcels

and the intermodal reference system show the lowest confidence, with no participants giving the levels completely confident and fairly confident as compared to the other two basemaps (see Table 11). Furthermore, the frequency of the level of confidence not at all confident is the highest with 53.33 % for the building parcels basemap and 48.33 % for the intermodal reference system basemap.

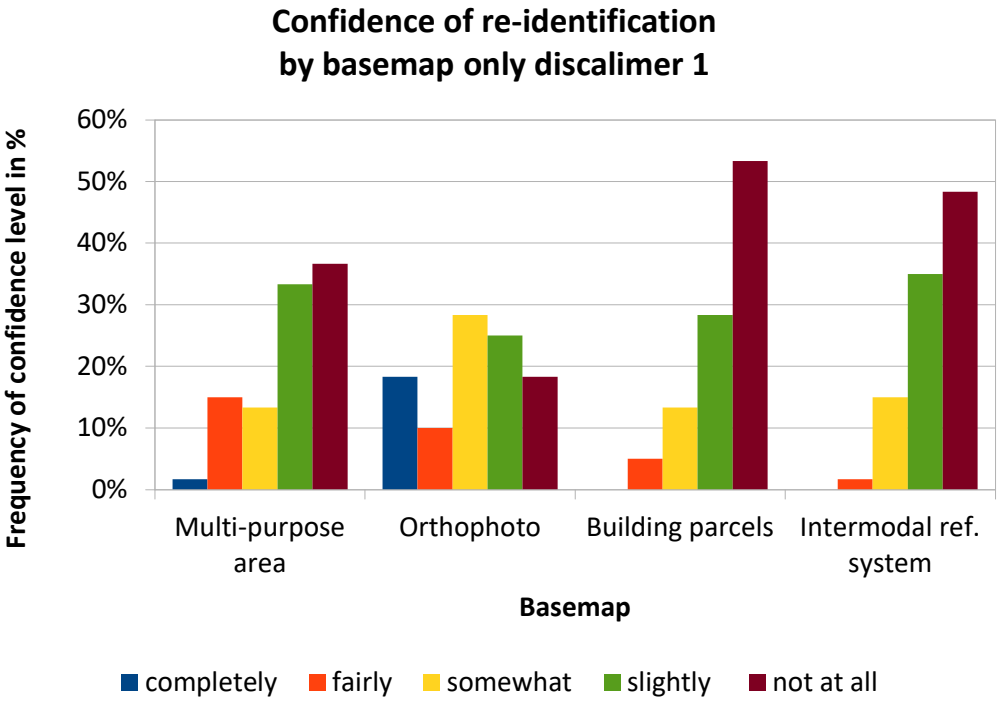


Figure 24 Frequency of confidence level by basemap for disclaimer 1

confidence level	Multi-purpose area	Orthophoto	Building parcels	Intermodal system
completely	1.67%	18.33%	0.00%	0.00%
fairly	15.00%	10.00%	5.00%	1.67%
somewhat	13.33%	28.33%	13.33%	15.00%
slightly	33.33%	25.00%	28.33%	35.00%
not at all	36.67%	18.33%	53.33%	48.33%
Total numbers				
completely	1	11	0	0
fairly	9	6	3	1
somewhat	8	17	8	9
slightly	20	15	17	21
not at all	22	11	32	29
Sum	60	60	60	60

Table 11 Frequency of confidence level by basemap for disclaimer 1

The boxplot for the cumulated confidence in maps with disclaimer one by basemap was created using the same percentiles as in Figure 25. It shows that the maps with orthophotos have a higher confidence than the others. Furthermore, the median cumulative confidence for the orthophoto basemaps is at 27, where the other basemaps have a lower median cumulative confidence of 20, 15, and 16.5, for multi-purpose area, building parcels and intermodal reference system (see Table 12). Additionally, the basemap with the intermodal reference system exhibits the lowest spread in confidence paired with a very low confidence in general. Furthermore, basemaps with building parcels have the lowest median of cumulative confidence, however the highest level of confidence can reach up to the median value for orthophotos.

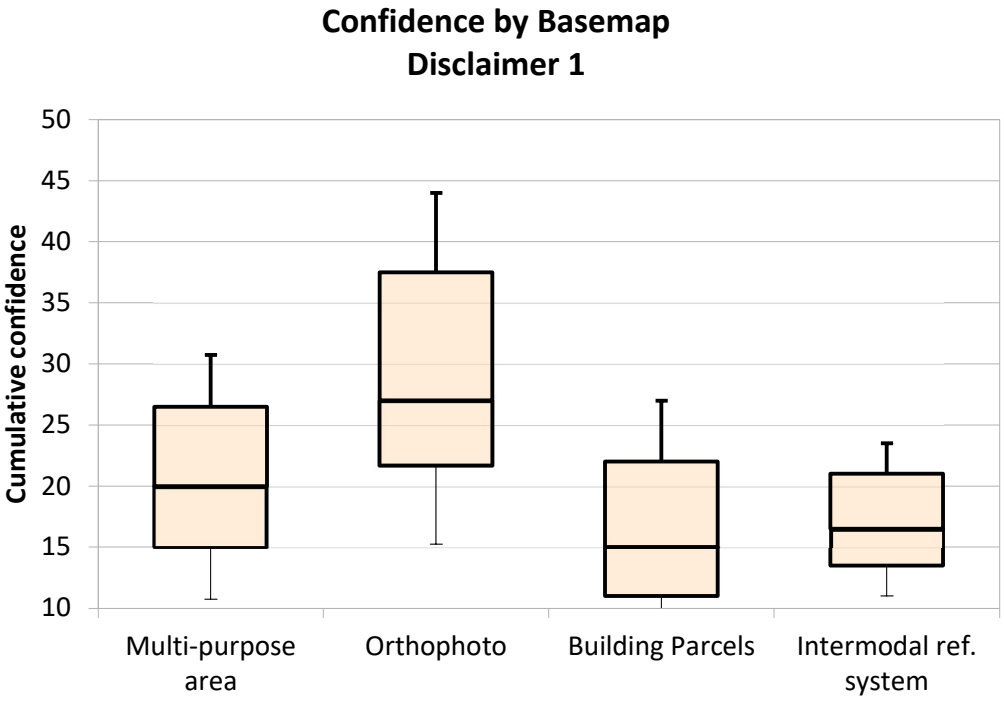


Figure 25 Boxplot - Cumulative confidence by basemap for disclaimer 1

	P(5%)	P(25%)	P(50%)	P(75%)	P(95%)
Multi-purpose area	13.5	15	20	26.5	30.75
Orthophoto	13.25	21.75	27	37.5	44
Building parcels	10	11	15	22	27
Intermodal ref. system	10.75	13.5	16.5	21	23.5

Table 12 Cumulative confidence by basemap for disclaimer 1

Similar as for disclaimer one, Figure 26 and Table 13 show the frequency of confidence levels by basemap but only for the participants who filled out the forms with disclaimer two (=form 5, 6, 7, 8). The frequency of the confidence level not at all confident was very high (51.67 %) regarding the intermodal reference system as a basemap. However, the level of confidence with disclaimer two per basemap appeared to be somewhat higher for the other three basemaps when compared to disclaimer one. Comparing the boxplots for the confidence by basemap for disclaimer one (Figure 25 and Table 12) and disclaimer two (Figure 27 and Table 14) the median cumulative confidence values are very similar except for the basemap with building parcels (15 for disclaimer one and 25 for disclaimer 2). This can hint towards a connection of additional information on masking techniques paired with basemaps of building parcels toward a higher confidence.

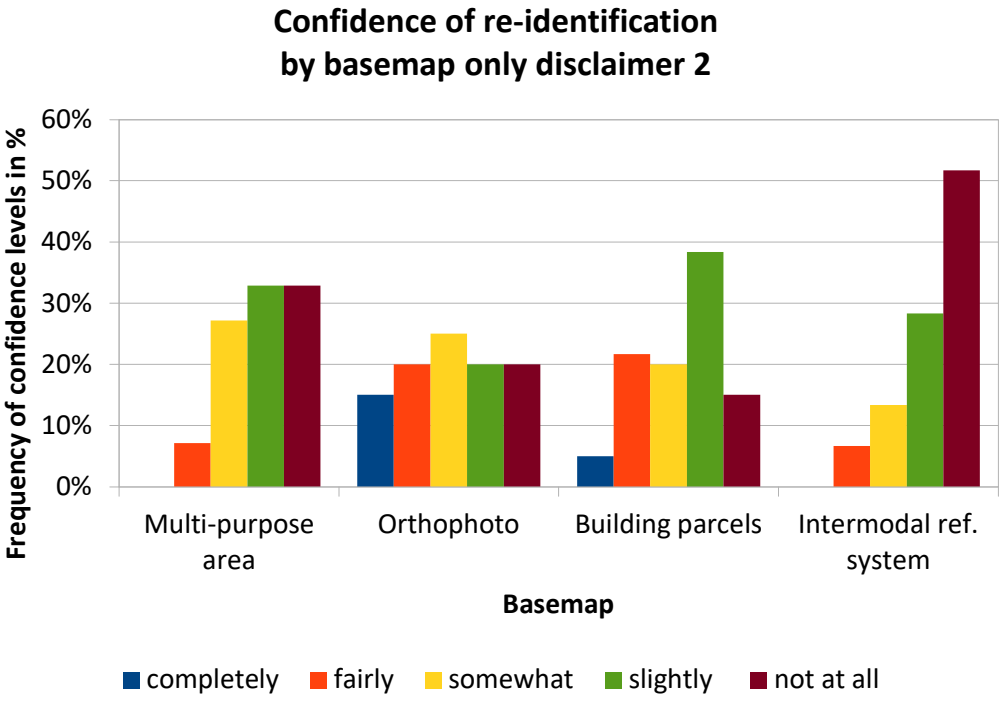


Figure 26 Frequency of confidence level by basemap for disclaimer 2

confidence level	Multi-purpose area	Orthophoto	Building parcels	Intermodal sys-tem
completely	0,00%	15,00%	5,00%	0,00%
fairly	7,14%	20,00%	21,67%	6,67%
somewhat	27,14%	25,00%	20,00%	13,33%
slightly	32,86%	20,00%	38,33%	28,33%
not at all	32,86%	20,00%	15,00%	51,67%
Total numbers				
completely	0	9	3	0
fairly	5	12	13	4
somewhat	19	15	12	8
slightly	23	12	23	17
not at all	23	12	9	31
Sum	70	60	60	60

Table 13 Frequency of confidence levels by basemap for disclaimer 2

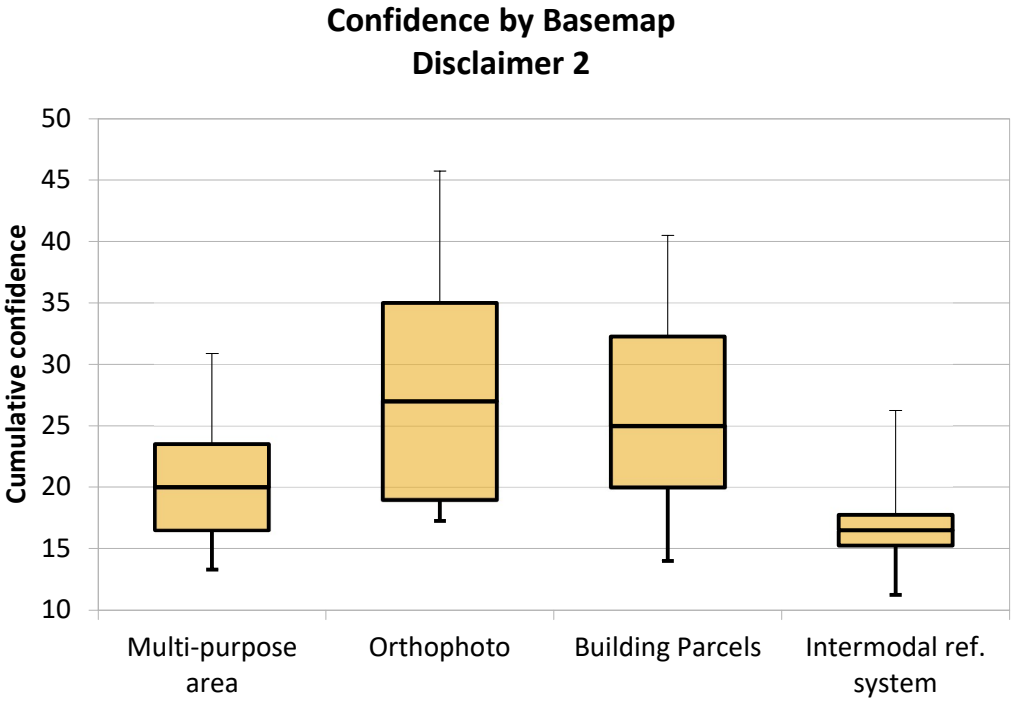


Figure 27 Boxplot - Cumulative confidence by basemap for disclaimer 2

	P(5%)	P(25%)	P(50%)	P(75%)	P(95%)
Multi-purpose area	13.3	16.5	20	23.5	30.9
Orthophoto	17.25	19	27	35	45.75
Building Parcels	14	20	25	32.25	40.5
Intermodal ref. system	11.25	15.25	16.5	17.75	26.25

Table 14 Cumulative confidence by basemap for disclaimer 2

4.3. Varying re-identification by basemaps

The task of the re-identification gave three different categories, for correct, false and not re-identified. A point was considered correctly re-identified if a participant placed the estimated point into the same building parcel. A point was considered false re-identified if the participant placed it into another building parcel. However, if for some reason the point was placed not in the building parcel at all, the point is considered not re-identified. In general, of a total of 5,160 residential points, which were selected for re-identification from the participants, 940 were correctly re-identified, 3,480 residential points were false re-identified and 740 residential points were not re-identified (see Table 15). Note that it was possible to re-identify 5,831 residential points. Each of the 49 participants had 119 points in one form. However, as the task was explained in the forms, the participants were instructed to re-identify at least one residential point and some of them choose to do one per map. However, most of the participants selected all the points to accomplish the task.

In order to examine research question two how different basemaps enhance spatial re-identification with human cognition, the distribution of the correct, false and not re-identified points by basemaps is shown in Figure 28. As it was clearly revealed, the participants were able to correct and false re-identify the residential points with all the different basemaps. But it showed some differences in the distribution as well. While the basemaps of the multi-purpose area, orthophoto and building parcels have a rate of 19.24 %, 23.05 % and 21.28 % for correct re-identification and a rate of 79.99 %, 75.46 % and 76.52 % for false re-identification, which are all similar values, the basemap of intermodal reference system performs differently. The correct re-identification rate was 9.44 %, the false re-identification rate was 36.82 % and the rate of the not at all re-identified points was 53.74 %. Due to the nature of the intermodal reference system, the participants were only able to deduce the building blocks and could vaguely percept the building parcels. Also, the not re-identified points are represented in every basemap. However, the false re-identified residential points (count of 3,480) always outnumbered the correct re-identified residential points (count of 940).

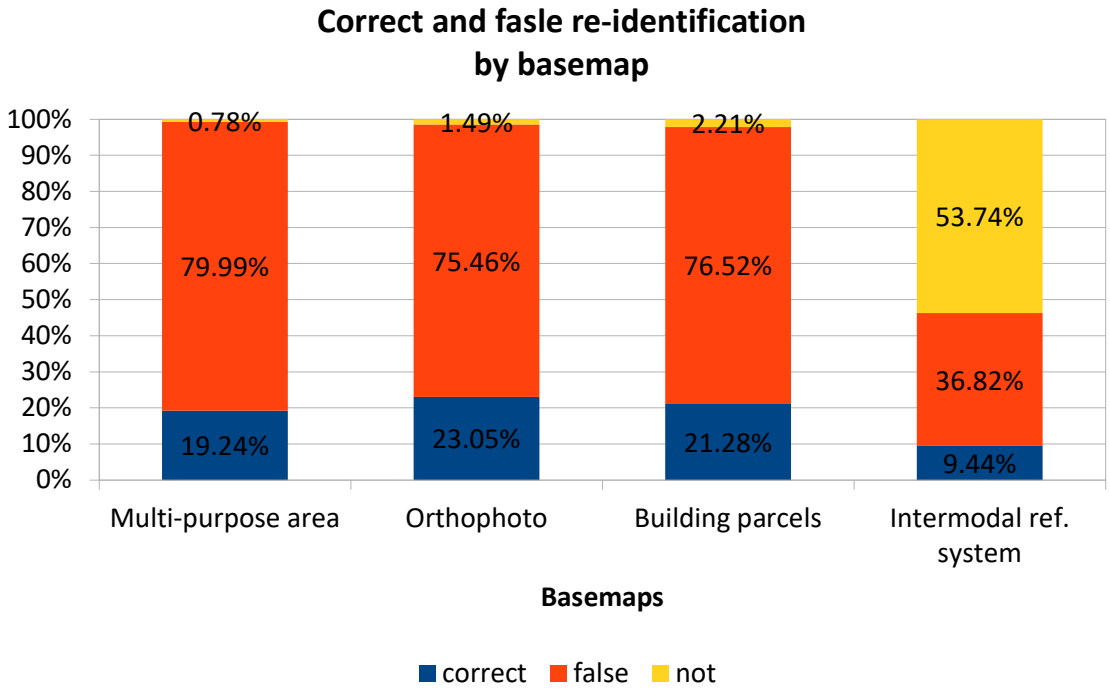


Figure 28 Re-identification by basemap

re-identification	Multi-pur-pose area	Ortho-photo	Building parcels	Intermodal ref. system
correct	19.24%	23.05%	21.28%	9.44%
false	79.99%	75.46%	76.52%	36.82%
not	0.78%	1.49%	2.21%	53.74%
Total numbers				
correct	272	278	270	120
false	1,131	910	971	468
not	11	18	28	683
sum	1,414	1,206	1,269	1,271

Table 15 Re-identification by basemap

4.4. Varying re-identification by masking disclaimer

To investigate the effect of map disclaimers with two different levels of information on the enhancement of correct and false re-identification, Figure 29 and Table 16 exhibited the distribution of correct and false re-identification per disclaimer. The ratio of correct re-identified residential points in maps with disclaimer one was lower (13.73 %) than with disclaimer two (22.67 %). This can be indicative towards disclaimer two, which is the one with more information, being more helpful in the re-identification task than disclaimer one. With the ratios of not re-identified residential points being almost the same (15.56 % for disclaimer

one and 13.17 % for disclaimer two), it remains that the ratio of false re-identified residential points is larger for disclaimer one (70.71 %) than for disclaimer two (64.16 %).

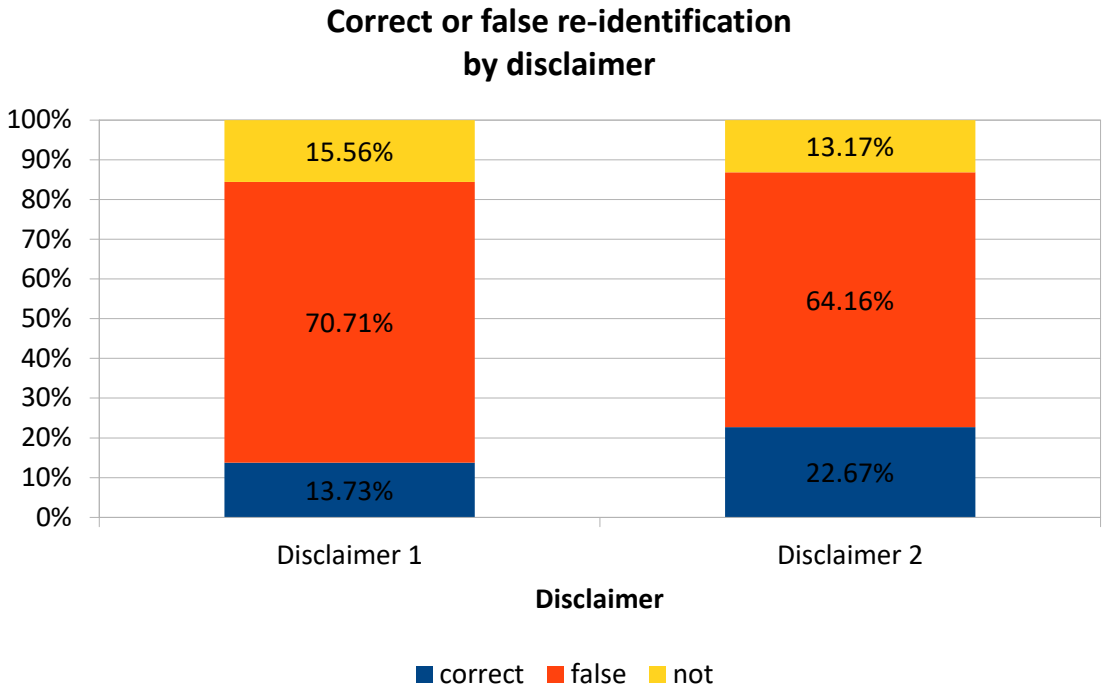


Figure 29 Re-identification by disclaimers

re-identification	Disclaimer 1	Disclaimer 2
correct	13.73%	22.67%
false	70.71%	64.16%
not	15.56%	13.17%
total numbers		
correct	353	587
false	1,818	1,661
not	400	341

Table 16 Re-identification by disclaimers

4.4.1. Varying re-identification by basemap and disclaimer 1 and 2

For further detailed investigation, correct and false re-identification was represented in two separate figures for disclaimer one (see Figure 30, also Table 17) and disclaimer two (see Figure 31, also Table 18). They show that the correct re-identification rates in all the basemaps are higher for disclaimer two than for disclaimer one. Again, this indicates that the additional information of parameters improved the ability of human cognition for re-identifying the correct residential points. Firstly, independent of the disclaimers, participants were able to correctly re-identify points in each category of basemap. Secondly, the rates of correct re-identification varied between the different disclaimers and basemaps. With disclaimer one, the highest rate of correct re-identified residential points, where participants had little information on the masking technique, was in the basemap with orthophotos (20.67 %). The highest rate of correct re-identifications based on disclaimer two was with the basemap of building parcels (27.80 %), followed by multi-purpose area (25.27 %) and orthophoto (25.16 %).

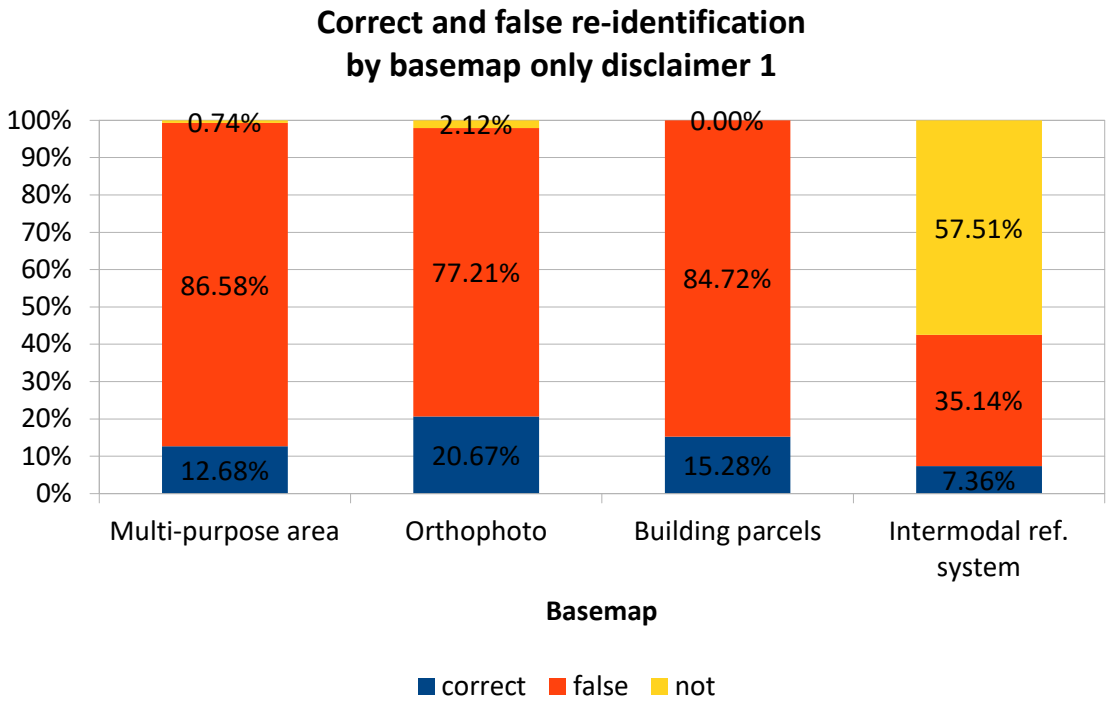


Figure 30 Re-identification by basemap and disclaimer 1

re-identification	Multi-purpose area	Orthophoto	Building parcels	Intermodal ref. system
correct	12.68%	20.67%	15.28%	7.36%
false	86.58%	77.21%	84.72%	35.14%
not	0.74%	2.12%	0.00%	57.51%
total numbers				
correct	86	117	101	49
false	587	437	560	234
not	5	12	0	383
sum	678	566	661	666

Table 17 Re-identification by basemap and disclaimer 1

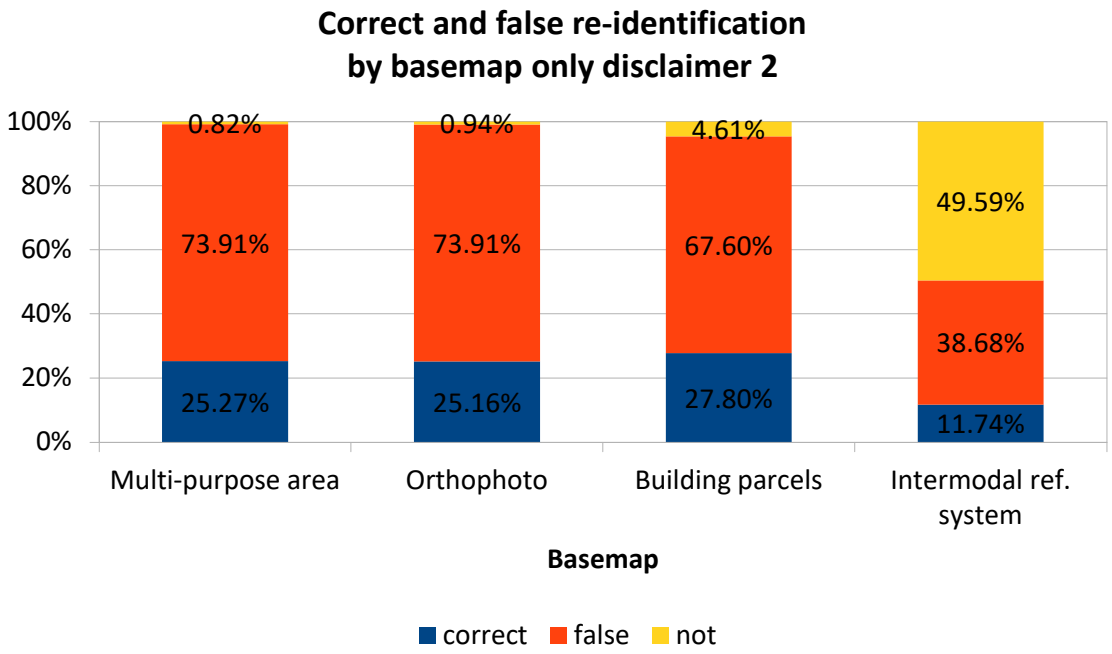


Figure 31 Re-identification by basemap for disclaimer 2

re-identification	Multi-purpose area	Orthophoto	Building parcels	Intermodal ref. system
correct	25.27%	25.16%	27.80%	11.74%
false	73.91%	73.91%	67.60%	38.68%
not	0.82%	0.94%	4.61%	49.59%
total numbers				
correct	186	161	169	71
false	544	473	411	234
not	6	6	28	300
sum	736	640	608	605

Table 18 Re-identification by basemap for disclaimer 2

4.5. Varying confidence and re-identification for masking methods

In terms of confidence by masking method, there are an equal number of 245 confidence declarations for each masking method. The participants were asked to give their confidence on a scale from one to five for each of the five maps per masking method. This results in 10 confidence statement per form, and totals in 490 statements over the whole survey.

In Figure 32 (also see Table 19) the confidence levels at each masking method were compared to each other. The levels seem to be very similar to each other, although the frequencies of the level not at all confident and completely confident were higher for donut masking (6.12 % and 37.96 %) than for street masking (3.67 % and 31.02 %). This could be an indicator that donut masking may lead to a polarization towards the extreme ends of the confidence scale compared to street masking (see Table 19).

Interestingly, in Figure 33 and Table 20 it is shown that maps with street masked residential points have a lower correct re-identification rate (12.56 %) and not re-identified points (10.68 %), than residential points, which were donut masked with a rate of 22.55 % of correct re-identifications and 17.17 % non-re-identifications. In return the false re-identifications were higher for street masked points (76.75 %) and lower for donut masked points (60.28 %).

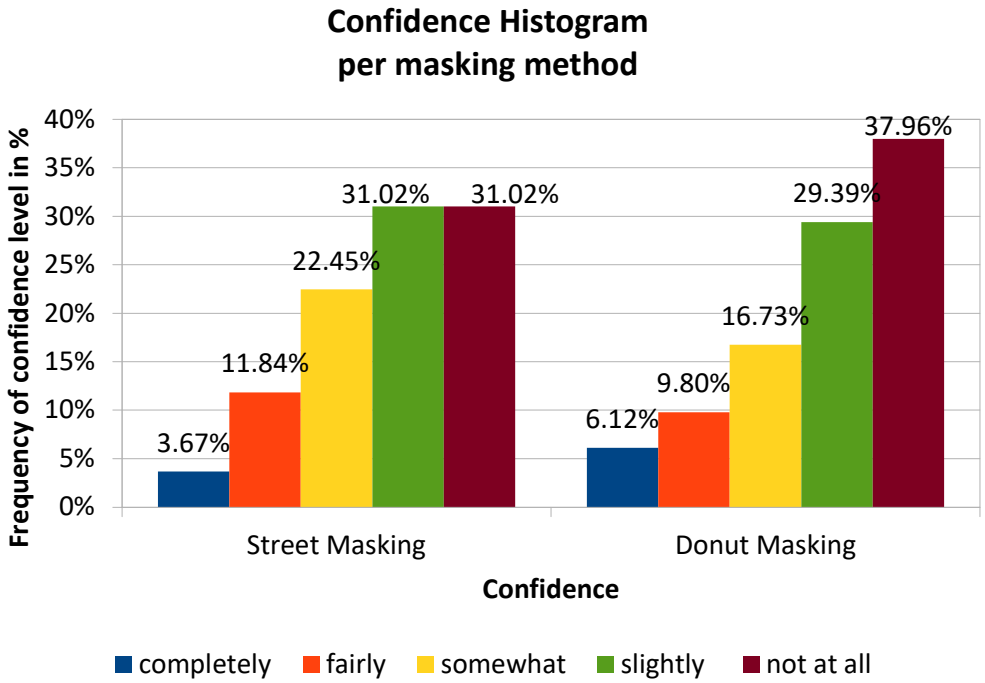


Figure 32 Frequency of confidence level by masking method

confidence level	Street Masking	Donut Masking
completely	3.67%	6.12%
fairly	11.84%	9.80%
somewhat	22.45%	16.73%
slightly	31.02%	29.39%
not at all	31.02%	37.96%
Total numbers		
completely	9	15
fairly	29	24
somewhat	55	41
slightly	74	72
not at all	76	93
Sum	245	245

Table 19 Frequency of confidence levels by masking method

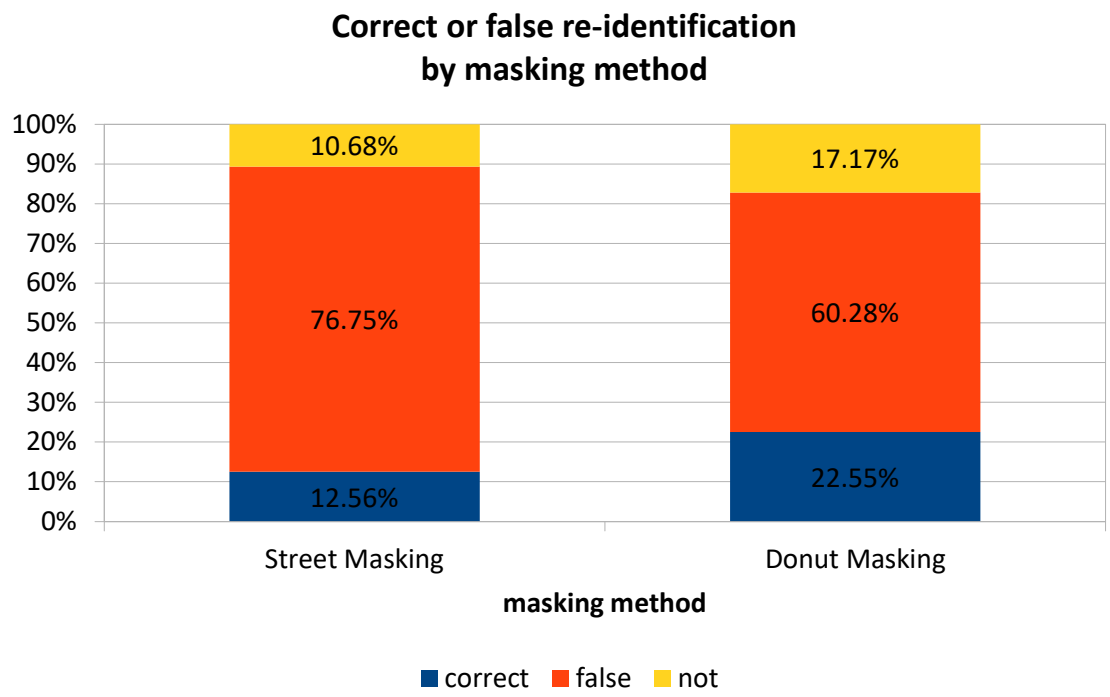


Figure 33 Re-identification by masking method

Re-identifications	Street Masking	Donut Masking
correct	12.56%	22.55%
false	76.75%	60.28%
not	10.68%	17.17%
total numbers		
correct	281	771
false	1,717	2,045
not	239	892
sum	2,237	3,708

Table 20 Re-identification by masking method

In Figure 34, the boxplot shows the confidence on re-identification by masking method. The y axis represents the cumulative confidence over all maps. Since 5 maps were used for each masking method and participants gave a confidence level for each map (1=not at all confident, 5 = completely confident), the total cumulative confidence can reach a maximum of 25 (which represents complete confidence for every map) and a minimum of 5 (least confidence in every map). The lower the value, the lower is the level of overall confidence.

The median cumulative level of the confidence on maps with street masked points is at 11, which indicates a slightly higher confidence as compared to maps with donut masked points, where the median lies at 10 (also see Table 21). A similar trend can be seen for the 75th

percentile which lies at 14 for street masking and 13 for donut masking. However, it can be seen for donut masking that the 95th has a higher value with 22.6 versus 20 for street masking. This indicates that individual participants may have experienced a higher confidence for donut masking.

Higher confidence in re-identification for street masked points is linked to a higher false re-identification (see Table 20). This result stands in contrast to results of the empirical study by Seidl et al. (2019), where they found a lower confidence in assigning points to a building parcel with the point lying outside the building. Their study showed that points at equal distance to other buildings had low confidence. However, the street masked points used in this survey were not necessarily placed at an equal distance between buldings.

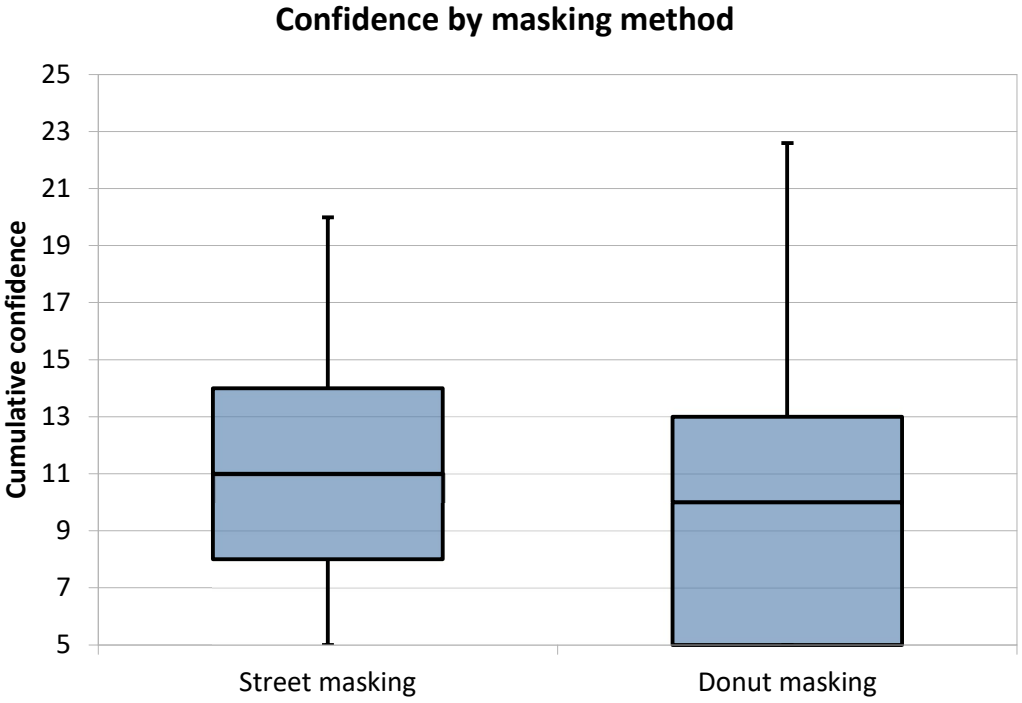


Figure 34 Boxplot - Cumulative confidence by masking method

	P(5%)	P(25%)	P(50%)	P(75%)	P(95%)
Street masking	5	8	11	14	20
Donut masking	5	5	10	13	22.6

Table 21 Cumulative confidence by masking method

Swanlund et al. (2020) developed the street masking method with the motivation to reduce the risk of false re-identification by shifting the masked points to the street intersections and automatically retrieve additional specialized data such as open street maps. Swanlund et al. (2020) argues that this stands in contrast to manually adding data such as the

population density, which was used for donut masking by Hampton et al. (2010). Nevertheless, the street masking method by Swanlund et al. (2020) disregards the effect that points can be closer to the neighboring building. This is especially true when taking into account the results of Seidl et al. (2019), as mentioned above.

At this point it should be noted that the spatial error for the street masked points was higher than for the donut masked points. This is attributed to the design of the survey, where the parameters for donut masking were chosen as such that it is not too hard to re-identify the masked points. This is also reflected in the values for estimated and true k-anonymity (see Table 5). Also, Kounadi and Leitner (2015a) were able to observe an influence of the displacement distance and the perceived affinity of patterns of masked points, which can be translated to the confidence on their re-identification on this study.

4.5.1. Varying confidence and re-identification by basemap for street masking

To get a deeper insight into the performance of the masking methods with regard to the different basemaps, the confidence level of re-identification for each masking method are put into relation for the different basemaps. The number of confidence statements per masking method results from five related map questions in one form. Figure 35 shows the frequency of confidence levels acquired from the survey considering only street masked points. It is easily seen that the basemap of the intermodal reference system induced the lowest confidence in the participants and further has the highest number of not re-identified points with 41.67 % (see Figure 36 and Table 23). Moreover, the frequency of confidences per basemap already hints towards the highest confidence being achieved by the orthophoto basemap. Especially when considering the absolute numbers of the level completely confident (see Table 22), the basemap with multi-purpose areas, building parcels and the intermodal reference system have zero or one count, while the basemap with orthophoto has seven counts.

For the correct re-identification the basemaps with the multi-purpose areas and orthophotos show the highest performance with 16.94 % and 15.36 %, respectively. The maps with the intermodal reference system show the lowest re-identification with 6.85 %.

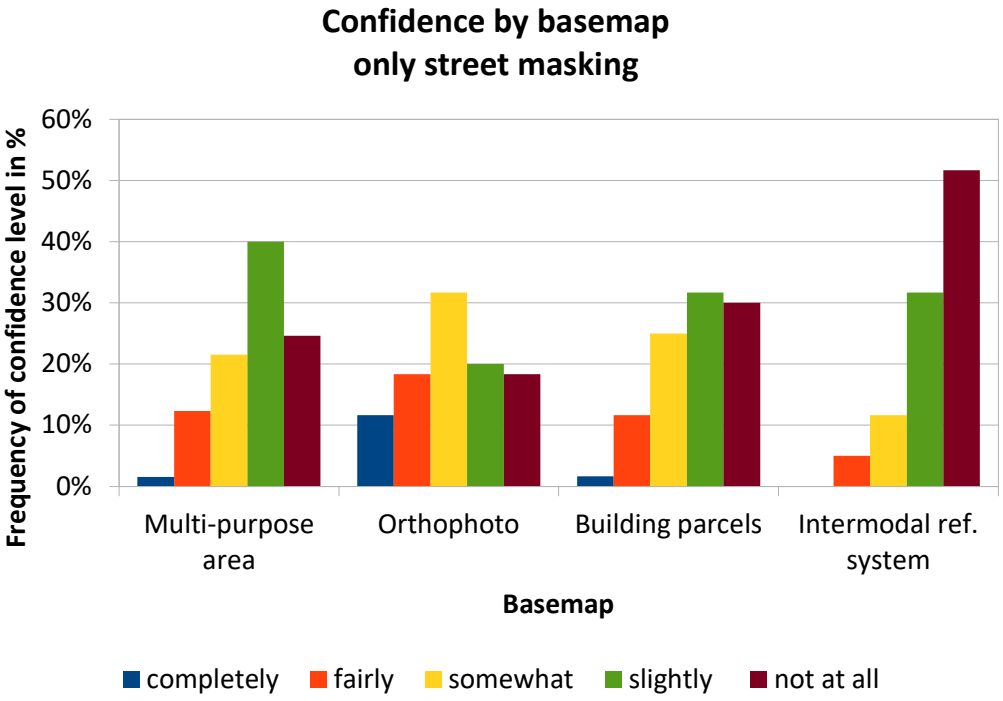


Figure 35 Frequency of confidence levels by basemap for street masking

confidence level	Multi-purpose area	Ortho-photo	Building parcels	Intermodal ref. system
completely	1.54%	11.67%	1.67%	0.00%
fairly	12.31%	18.33%	11.67%	5.00%
somewhat	21.54%	31.67%	25.00%	11.67%
slightly	40.00%	20.00%	31.67%	31.67%
not at all	24.62%	18.33%	30.00%	51.67%
total numbers				
completely	1	7	1	0
fairly	8	11	7	3
somewhat	14	19	15	7
slightly	26	12	19	19
not at all	16	11	18	31
sum	65	60	60	60

Table 22 Frequency of confidence levels by basemap for street masking

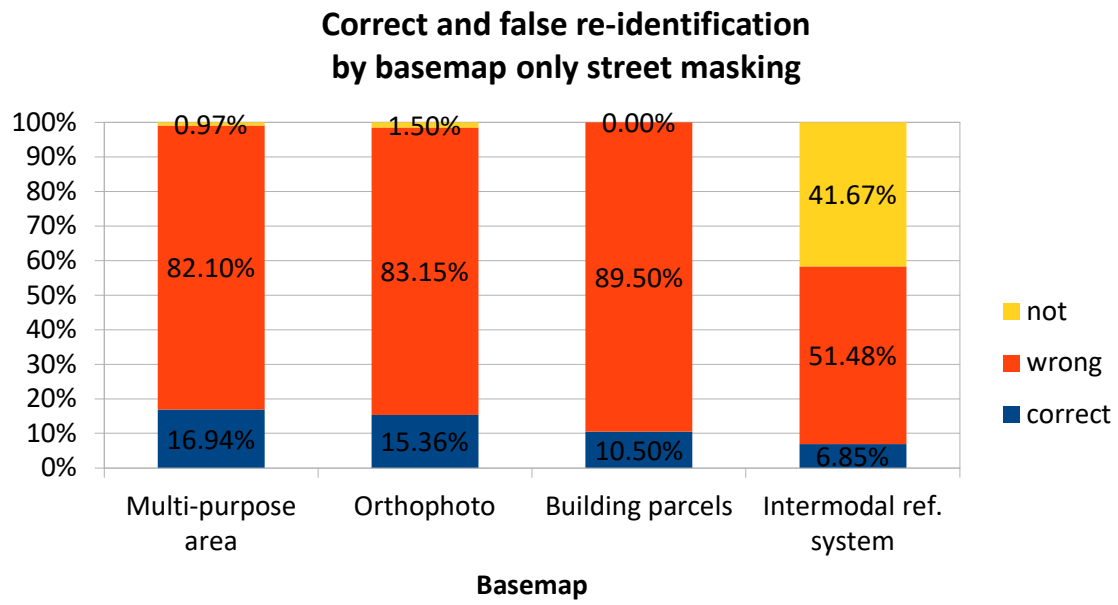


Figure 36 Re-identification by basemap for street masking

re-identification	Multi-purpose area	Orthophoto	Building parcels	Intermodal ref. system
correct	16.94%	15.36%	10.50%	6.85%
wrong	82.10%	83.15%	89.50%	51.48%
not	0.97%	1.50%	0.00%	41.67%
total numbers				
correct	105	82	57	37
wrong	509	444	486	278
not	6	8	0	225

Table 23 Re-identification by basemap for street masking

Improved visualization for the overall confidence per basemap is given by the boxplots in Figure 37 (also see Table 24). The same percentiles as in Table 21 were used. It can now be clearly seen that the participants felt most confident when presented with the orthophoto basemap with a median cumulative confidence value of 15. Again, it is shown that the intermodal reference system has the lowest median confidence value of 9, whereas for the basemap of multi-purpose area and building parcels the median is the same at 11. However, it seems that the overall confidence is lower for the building parcel basemap.

The high performance of maps with orthophotos in terms of confidence and re-identification immediately raises concerns with regard to confidentiality and privacy when sensitive data is published on maps with orthophotos. This also takes into consideration the high rates of false re-identification.

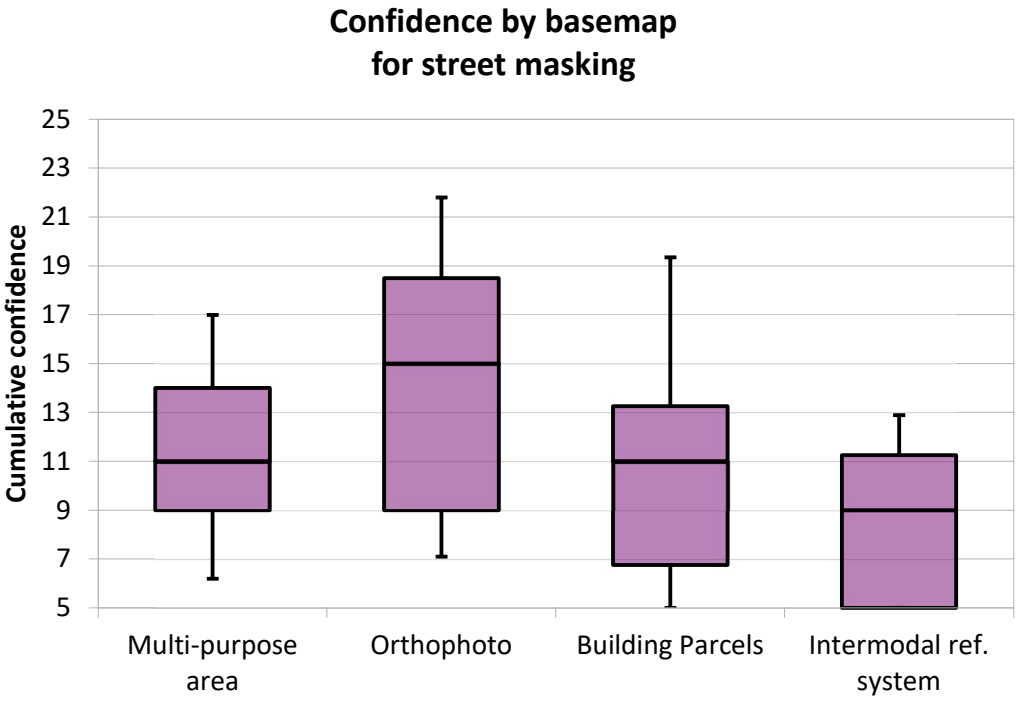


Figure 37 Boxplot - Cumulative confidence by basemap for street masking

	P(5%)	P(25%)	P(50%)	P(75%)	P(95%)
Multi-purpose	6.2	9	11	14	17
Orthophoto	7.1	9	15	18.5	21.8
Building Parcels	5	6.75	11	13.25	19.35
Intermodal ref. system	5	5	9	11.25	12.9

Table 24 Cumulative confidence by basemap for street masking

4.5.2. Varying confidence and re-identification by basemap for donut masking

For the second masking method included in the survey, the frequency of the levels of confidence for re-identified donut masked residential points are displayed as well (see Figure 38 and Table 25). It is clearly seen that the confidence is extremely low for most of the backgrounds. Only with the case of orthophoto the confidence is quite good, as the level completely confident is at 21.67 %. This means that 13 times the re-identification in the forms with orthophoto basemaps were judged with the level completely confident, whereas the other basemaps had zero or two counts in that category.

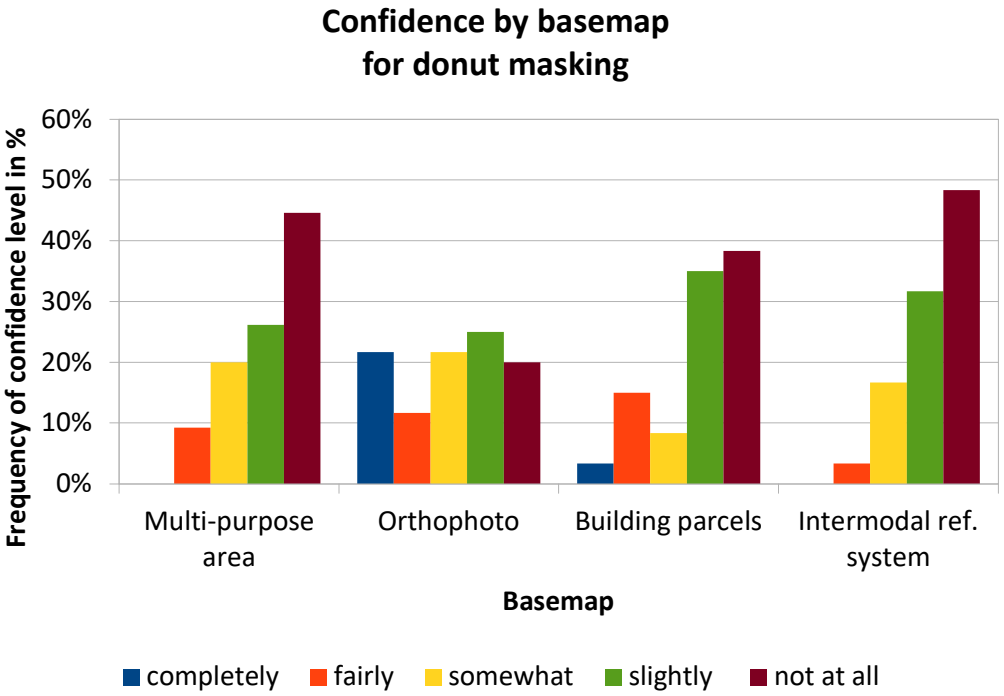


Figure 38 Frequency of confidence levels by basemap for donut masking

confidence level	Multi-purpose area	Orthophoto	Building parcels	Intermodal ref. system
completely	0.00%	21.67%	3.33%	0.00%
fairly	9.23%	11.67%	15.00%	3.33%
somewhat	20.00%	21.67%	8.33%	16.67%
slightly	26.15%	25.00%	35.00%	31.67%
not at all	44.62%	20.00%	38.33%	48.33%
Total numbers				
completely	0	13	2	0
fairly	6	7	9	2
somewhat	13	13	5	10
slightly	17	15	21	19
not at all	29	12	23	29
sum	65	60	60	60

Table 25 Frequency of confidence levels by basemap for donut masking

In general, one can observe higher correct re-identification rates with donut masking (see Figure 36) when compared to street masking (see Figure 39).However as mentioned above the donut masked method has lower k-anonymity levels and a smaller spatial error (see Table 5). This is reflected in all four basemaps. Nevertheless, as seen in Figure 39 (also Table 26) the rate of correct re-identification for orthophoto and building parcel basemaps is higher with 29.17 % and 29.34 % than for the other two. The basemap of the intermodal reference system has the lowest correct re-identification with 11.35 %.

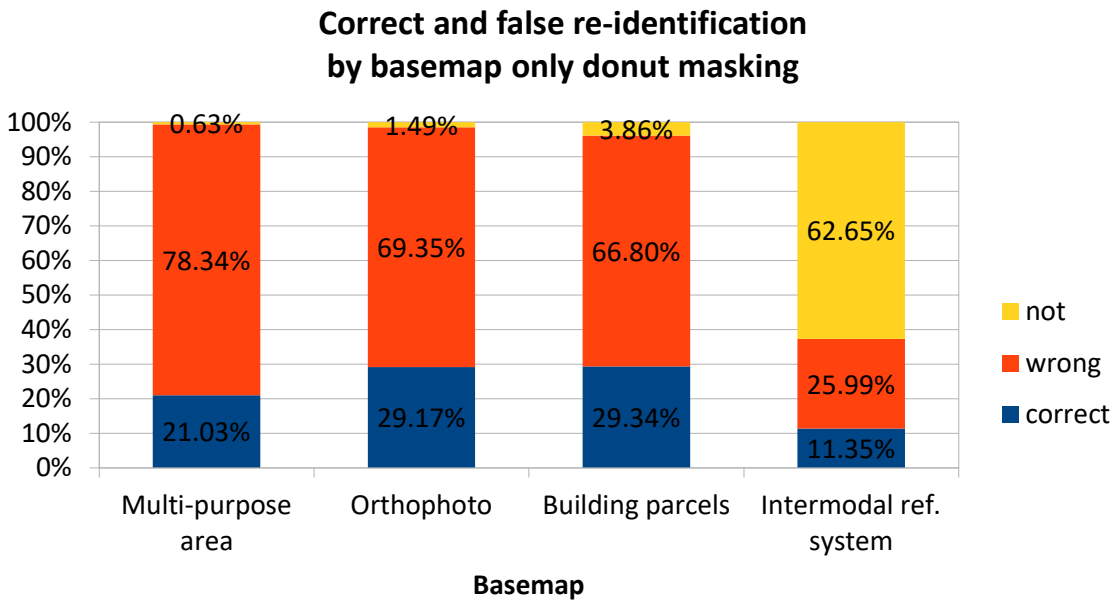


Figure 39 Re-identification by basemap for donut masking

re-identification	Multi-purpose area	Orthophoto	Building parcels	Intermodal ref. system
correct	21.03%	29.17%	29.34%	11.35%
wrong	78.34%	69.35%	66.80%	25.99%
not	0.63%	1.49%	3.86%	62.65%
total numbers				
correct	167	196	213	83
wrong	622	466	485	190
not	5	10	28	458

Table 26 Re-identification by basemap for donut masking

In Figure 40 (also see Table 27) it is shown in the boxplots that maps with orthophotos have the highest median cumulative confidence with value 12. The median cumulative confidence for the other basemaps is significantly lower and lies between 8.5 and 10, which is under the 25th percentile of the orthophotos. Further the 75th percentile ranges up to 26.5, which could indicate a larger scatter towards higher confidences.

Like street masking, the confidence and correct re-identification for maps with orthophotos is high. Both masking methods with orthophoto maps follow this trend, where re-identification rate is one of the highest of alle the other basemaps and show the highest confidence. This bears a possible danger of masked points which are published on orthophotos on correct and false re-identification. One possible explanation for the high rates of confidentiality on orthophotos may be that the participants or people in general are more

familiar with photos of the reality than with other kinds of basemaps. Considering this connection, the intermodal reference system was the one with the highest not re-identified points and the lowest confidence.

Leitner and Curits (2004) were able to observe some differences on the perceived hotspots on masked maps by different basemaps in one of the diverse geomasking techniques. Differently, in this survey a different perception regarding the confidence on re-identification on every basemap can be observed by every masking method.

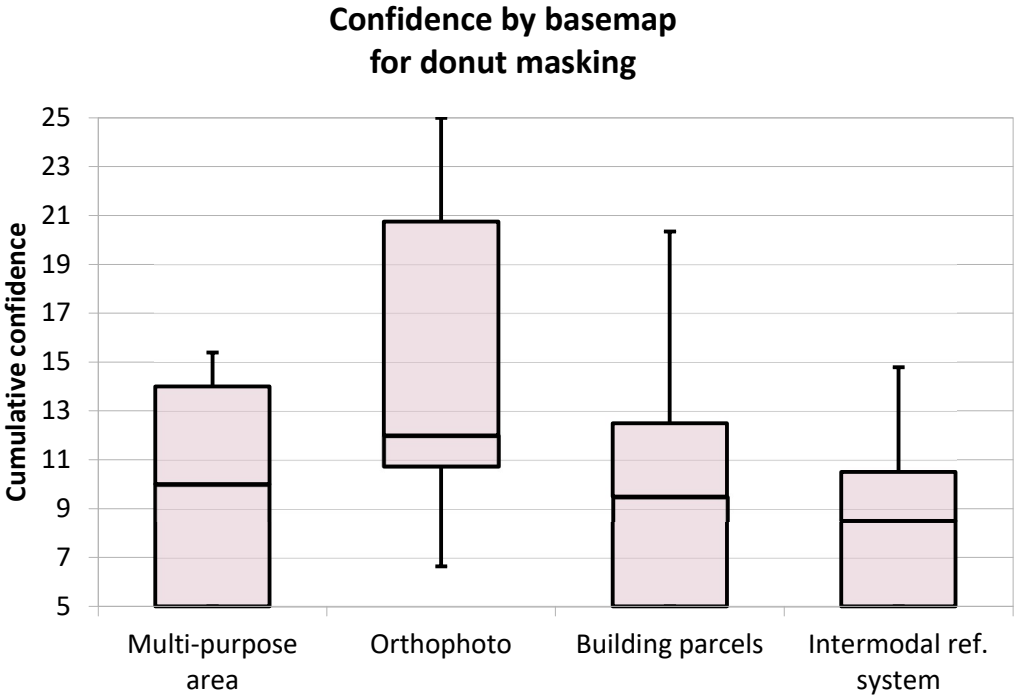


Figure 40 Boxplot - Cumulative confidence by basemap for donut masking

	P(5%)	P(25%)	P(50%)	P(75%)	P(95%)
Multi-purpose area	5	5	10	14	15.4
Orthophoto	6.65	10.75	12	20.75	25
Building parcels	5	5	9.5	12.5	20.35
Intermodal ref. system	5	5	8.5	10.5	14.8

Table 27 Cumulative confidence by basemap for donut masking

4.5.3. Confidence and re-identification by disclaimer for masking methods

When comparing the two disclaimers per masking technique, there are some differences regarding the observed confidence. The levels of confidence varied slightly regarding the maps with street masked points. Figure 41 and Table 28 show similar frequencies of the confidence levels between disclaimer one and two. In this context, one could argue that

for the street masking method the level of information provided in the disclaimers have hardly any influence on the confidence.

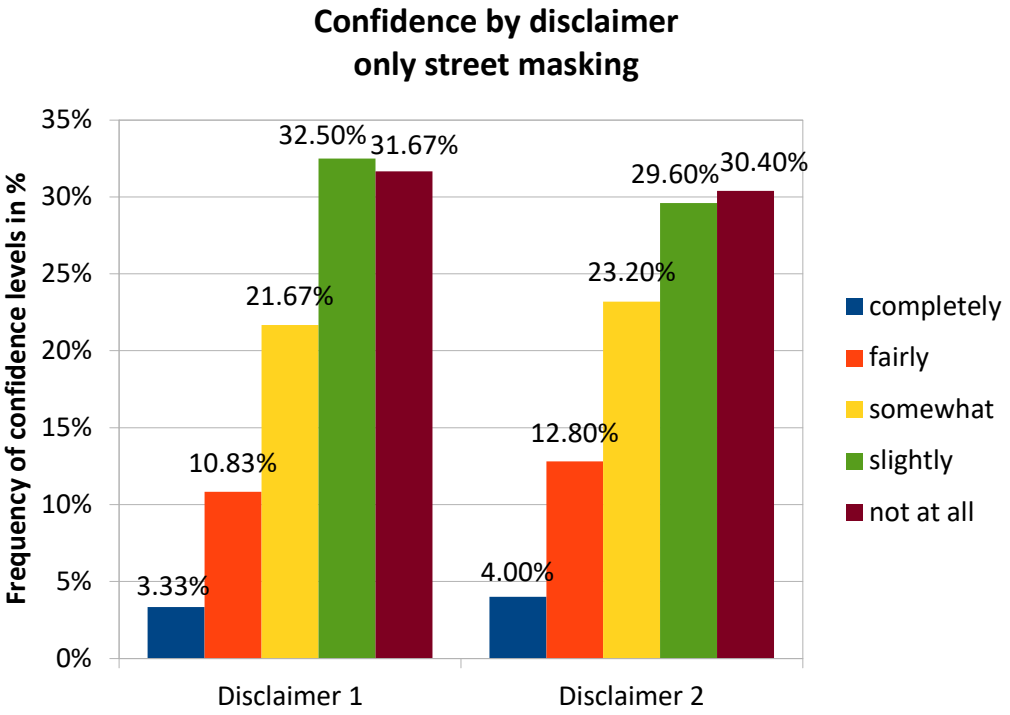


Figure 41 Confidence levels by disclaimer for street masking

confidence level	Disclaimer 1	Disclaimer 2
completely	3.33%	4.00%
fairly	10.83%	12.80%
somewhat	21.67%	23.20%
slightly	32.50%	29.60%
not at all	31.67%	30.40%
total numbers		
completely	4	5
fairly	13	16
somewhat	26	29
slightly	39	37
not at all	38	38

Table 28 Confidence levels by disclaimer for street masking

When considering donut masking, a different picture occurs (see Figure 42 and Table 29). It can clearly be seen that in this case the lack of information in disclaimer one leads to lower confidence of 46.67 % of all confidence levels being not at all confident. Differently, disclaimer two provided more information, which is mirrored in higher confidence levels than for disclaimer one.

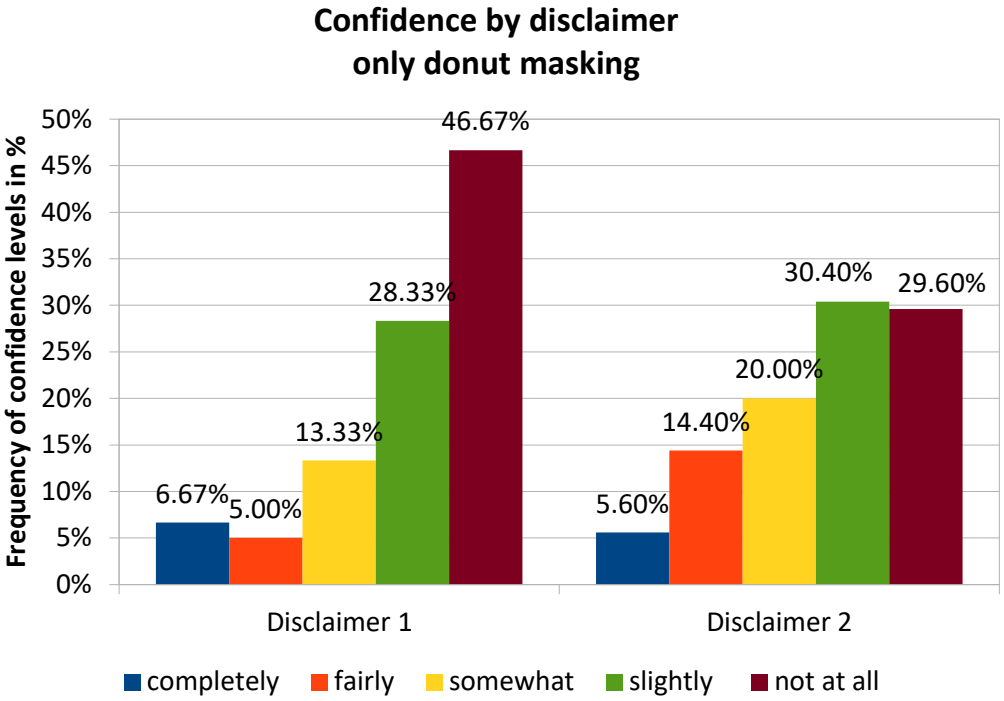


Figure 42 Confidence levels by disclaimer for donut masking

confidence level	Disclaimer 1	Disclaimer 2
completely	6.67%	5.60%
fairly	5.00%	14.40%
somewhat	13.33%	20.00%
slightly	28.33%	30.40%
not at all	46.67%	29.60%
total numbers		
completely	8	7
fairly	6	18
somewhat	16	25
slightly	34	38
not at all	56	37

Table 29 Confidence levels by disclaimer for donut masking

Similar relations between correct and false re-identified points can be observed for the two disclaimers per masking method. When only considering street masking, the ratio of correct and false re-identifications only differs slightly between the two disclaimers. This is shown in (Figure 43 and Table 30).

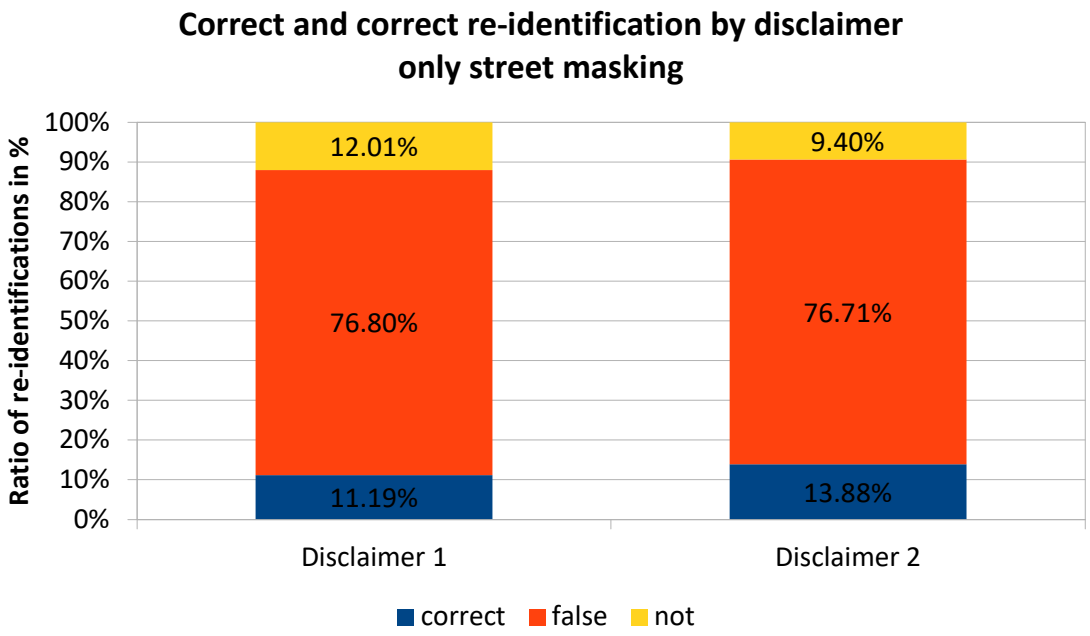


Figure 43 Re-identification by disclaimer for street masking

re-identifications	Disclaimer 1	Disclaimer 2
correct	11.19%	13.88%
false	76.80%	76.71%
not	12.01%	9.40%
Total numbers		
correct	123	158
false	844	873
not	132	107

Table 30 Re-identification by disclaimer for street masking

When looking at the influence of the disclaimers on the donut masked points, interestingly, one can see that there are far more correct re-identified points when more information is provided with disclaimer two. In combination with the higher confidence (see Figure 42) this raises concerns regarding the disclosure risk of personal information for donut masked points combined with a high level of information provided. It is worth mentioning that the general correct re-identification rate may be higher for donut masking, which is indicated by a slight increase of correct re-identified residential points with disclaimer one, when comparing street masking to donut masking (left side of Figure 43 and Figure 44). However, the higher re-identification rate may also be attributed to the small spatial error which was generally introduced in donut masked points.

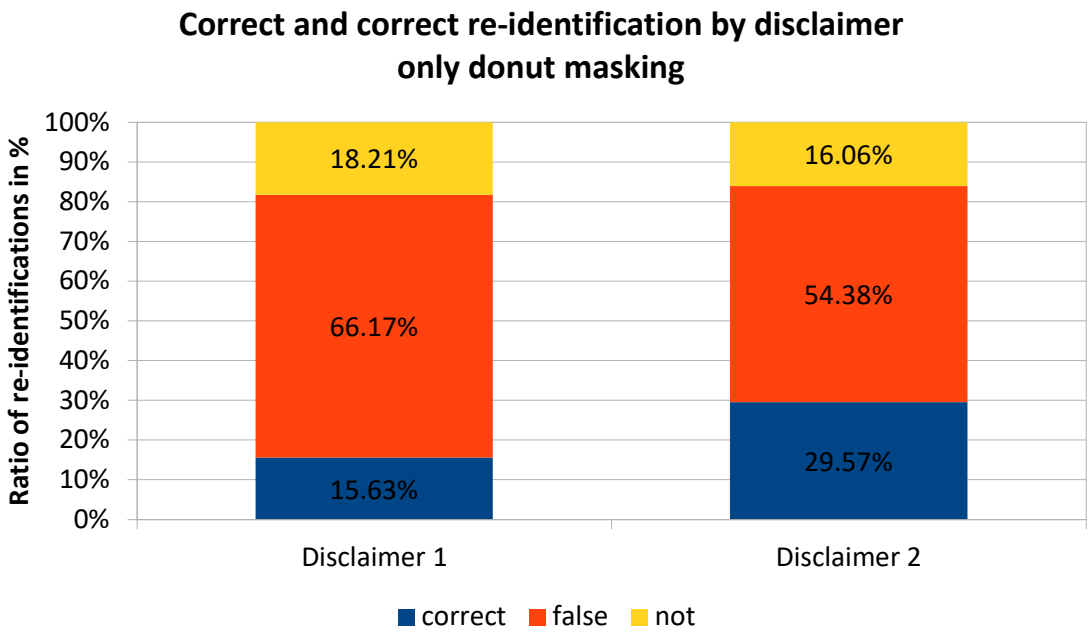


Figure 44 Re-identification by disclaimer for donut masking

re-identifications	Disclaimer 1	Disclaimer 2
correct	15.63%	29.57%
false	66.17%	54.38%
not	18.21%	16.06%
total number		
correct	230	429
false	974	789
not	268	233

Table 31 Re-identification by disclaimer for donut masking

5. Conclusion

As this survey is one of the first of its kind, it provides insight into new fields of research. The survey demonstrated that correct re-identification of geographical masked data points to different basemaps with building parcels or orthophotos by human cognition, as described by Seidl et al (2018) was indeed possible but varied in the confidence.

The confidence of the participants varied regarding the different map elements in this survey (see research question one) and showed an influence on the ability to re-identify as well (see research question two). Altogether the different basemaps had a high influence on the confidence of participants, while the disclaimers information had an influence on the correct re-identification rates.

The basemap of orthophotos had an effect on the participants stating higher confidence levels than the other basemaps, which can be seen as problematic for false and correct re-identified residential points. The high confidence level can lead to linking the confidential information to the correct resident, unearthing confidential information, or linking the confidential information to the wrong residence or household. It can be argued that the ethical concern of McLafferty (2004) is by the results of this survey reinforced.

Additionally, two disclaimers, which differ in level of information, were tested. The results show little effect on the confidence levels by the participants, where disclaimer two with information on the parameters of the masking technique led to slightly higher confidence than disclaimer one with no information on the parameters. Even though, more points were correctly re-identified on maps provided with disclaimer two. Therefore, to prevent the revelation of confident information with a geographical link through correct re-identification with high confidence, it can be suggested from this survey that the combination of the orthophoto basemap and disclaimers with parameters should be avoided for publications. The suggestion of Kounadi and Resch (2018) was to avoid any detailed information on the masking technique and is supported by the results of the survey.

On the other hand, the lowest confidence and lowest rate of correct re-identifications was with the basemap of the intermodal reference system. An effect occurred to be stronger than it was first assumed. Points were not re-identified, obviously because of the visually missing building parcels, which were used to determine correct or false re-identification. Nevertheless, one could argue that the applied concept of correct and false re-identification can be modified for further research. However, the low confidence may have originated from the task of re-identifying residential location and not displaying them. This possible discovery can be included for further publications on confident spatial residential locations on point level as the visualization of the intermodal reference system or a street network offers some spatial reference in general but no direct visual display to residential buildings.

Regarding the effectiveness of the masking techniques concerning the confidence of the re-identifications, the confidence levels for the street masked maps was a little higher than for the donut masked maps. Nevertheless, the correct re-identifications for the donut masked residential points were higher, probably because of the lower spatial error. The two different

disclaimers showed hardly any effect on the confidence and on the re-identifications of the street masked points. In contrast, providing the participants with information on the parameters of the donut masking technique led to small increase in confidence, and even doubled the rate of correct re-identifications compared the rate of correct re-identifications of disclaimer one. This could hint towards a higher risk of correct re-identification when metadata with information on the parameter of the donut masking method is disclosed. Although maps with donut masked points had the lower spatial error than maps with street masked points. Regardless of the information level on the disclaimer, the information was enough for some correct re-identifications. In this case one could agree with Kounadi and Resch (2018), which recommended not to reveal any detailed information of the metadata such as masking method.

This study also raises concerns towards the transgressor scenario described by Kounadi et al. (2013), although the authors only describe a scenario with unmasked points. In an adapted scenario a person with evil intentions uses a map with sensitive masked points, recovers the coordinates and uses the parameters shown on the map to re-identify the masked points to the corresponding building or household. Finally, in following the steps described by Kounadi et al. (2013), a transgressor recovers the coordinates of the estimated residential location, using reverse geocoding tools to fetch an address and combines the address with other information associated to the address.

6. Literature

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. *Geocarto International*, *25*(6), 443–452. https://doi.org/10.1080/10106049.2010.496496

Armstrong, M. (2002). Geographic information technologies and their potentially erosive effects on personal privacy. In *Studies in the Social Sciences: Vol. 37/1* (pp. 19–28).

Armstrong, M. P., Rushton, G., & Zimmerman, D. L. (1999). Geographically masking health data to preserve confidentiality. *Statistics in Medicine*, *18*(5), 497–525. https://doi.org/10.1002/(sici)1097-0258(19990315)18:5<497::aid-sim45>3.0.co;2-#

Beresford, A. R., & Stajano, F. (2003). Location privacy in pervasive computing. *IEEE Pervasive Computing*, *2*(1), 46–55. https://doi.org/10.1109/MPRV.2003.1186725

Brownstein, J. S., Cassa, C. A., Kohane, I. S., & Mandl, K. D. (2005). Reverse geocoding: concerns about patient confidentiality in the display of geospatial health data. *AMIA ... Annual Symposium Proceedings. AMIA Symposium*, *2005*, 905.

Brownstein, J. S., Cassa, C. A., Kohane, I. S., & Mandl, K. D. (2006b). An unsupervised classification method for inferring original case locations from low-resolutaion disease maps. *International Journal of Health Geographics*, *5*(1), 56. https://doi.org/10.1186/1476-072X-5-56

Brownstein, J. S., Cassa, C. A., & Mandl, K. D. (2006a). No Place to Hide — Reverse Identification of Patients from Published Maps. *New England Journal of Medicine*, *355*(16), 1741–1742. https://doi.org/10.1056/NEJMc061891

Burdon, M. (2010). Privacy Invasive Geo-Mashups: Privacy 2.0 and the Limits of First Generation Information Privacy Laws . *University of Illinois Journal of Law, Technology & Policy*, *1*, 1–50.

Cassa, C. A., Grannis, S. J., Overhage, J. M., & Mandl, K. D. (2006). A Context-sensitive Approach to Anonymizing Spatial Surveillance Data: Impact on Outbreak Detection. *Journal of the American Medical Informatics Association*, *13*(2), 160–165. https://doi.org/10.1197/jamia.M1920

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

Curtis, A. J., Mills, J. W., & Leitner, M. (2006). Spatial confidentiality and GIS: Re-engineering mortality locations from published maps about Hurricane Katrina. *International Journal of Health Geographics*, *5*(1), 44. https://doi.org/10.1186/1476-072X-5-44

City of Vienna – https://data.wien.gv.at (a): Gebäudeinformation Standorte Wien https://www.data.gv.at/katalog/en/dataset/gebaeudeinformation-wien ID: 7a8aae59-71a4-4500-b38b-bdf15c7f627f

City of Vienna – https://data.wien.gv.at (b): Vector data of the multi-purpose area map (MZK) https://www.data.gv.at/katalog/de/dataset/stadt-wien_flchenmehrzweck-kartevektordatenwien ID: 7cf0da04-1f77-4321-929e-78172c74aa0b

City of Vienna - https://data.wien.gv.at (c): Bevölkerungszahlen und -dichte in Wien nach Zählbezirken https://www.data.gv.at/application/bevolkerungszahlen-und-dichte-in-wien-nach-zahlbezirken/#screenshots

City of Vienna - https://data.wien.gv.at (d): Intermodal transport reference system of Austria (GIP.at) https://www.data.gv.at/katalog/en/dataset/intermodales-verkehrsreferenzsystem-osterreich-gip-at-beta ID: 3fefc838-791d-4dde-975b-a4131a54e7c5

City of Vienna - https://data.wien.gv.at (e): Adressen Standorte Wien https://www.data.gv.at/katalog/de/dataset/stadt-wien_adressdatenderstadtwien ID: 1d5c2411-9719-4c8f-b99d-57a5f4a4ae41

City of Vienna - https://data.wien.gv.at (f): Orthofoto 2022 Wien https://www.data.gv.at/katalog/de/dataset/orthofoto-2022-wien ID: bcba4f91-27b7-4950-90a7-cecee4617a8b

City of Vienna - https://data.wien.gv.at (g): Mehrzweckkarte Rasterdaten Wien https://www.data.gv.at/katalog/de/dataset/stadt-wien_mehrzweckkarterasterdaten-wien ID: 2e0129ab-c9bd-4679-a987-30fd5008472a

Haley, D. F., Matthews, S. A., Cooper, H. L. F., Haardörfer, R., Adimora, A. A., Wingood, G. M., & Kramer, M. R. (2016). Confidentiality considerations for use of social-spatial data on the social determinants of health: Sexual and reproductive health case study. *Social Science & Medicine*, *166*, 49–56. https://doi.org/10.1016/j.socscimed.2016.08.009

Hampton, K. H., Fitch, M. K., Allshouse, W. B., Doherty, I. A., Gesink, D. C., Leone, P. A., Serre, M. L., & Miller, W. C. (2010). Mapping Health Data: Improved Privacy Protection With Donut Method Geomasking. *American Journal of Epidemiology*, *172*(9), 1062–1069. https://doi.org/10.1093/aje/kwq248

Kim, J., Kwan, M.-P., Levenstein, M. C., & Richardson, D. B. (2021). How do people perceive the disclosure risk of maps? Examining the perceived disclosure risk of maps and its implications for geoprivacy protection. *Cartography and Geographic Information Science*, *48*(1), 2–20. https://doi.org/10.1080/15230406.2020.1794976

Kounadi, O., Bowers, K., & Leitner, M. (2015). Crime Mapping On-line: Public Perception of Privacy Issues. *European Journal on Criminal Policy and Research*, *21*(1), 167–190. https://doi.org/10.1007/s10610-014-9248-4

Kounadi, O., Lampoltshammer, T. J., Leitner, M., & Heistracher, T. (2013). Accuracy and privacy aspects in free online reverse geocoding services. *Cartography and Geographic Information Science*, *40*(2), 140–153. https://doi.org/10.1080/15230406.2013.777138

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), 34–45. https://doi.org/10.1177/1556264614544103

Kounadi, O., & Leitner, M. (2015a). Defining a Threshold Value for Maximum Spatial Information Loss of Masked Geo-Data. *ISPRS International Journal of Geo-Information*, *4*(2), 572–590. https://doi.org/10.3390/ijgi4020572

Kounadi, O., & Leitner, M. (2015b). Spatial Information Divergence: Using Global and Local Indices to Compare Geographical Masks Applied to Crime Data. *Transactions in GIS*, *19*(5), 737–757. https://doi.org/10.1111/tgis.12125

Kounadi, O., & Leitner, M. (2016). Adaptive areal elimination (AAE): A transparent way of disclosing protected spatial datasets. *Computers, Environment and Urban Systems*, *57*, 59–67. https://doi.org/10.1016/j.compenvurbsys.2016.01.004

Kounadi, O., & Resch, B. (2018). A Geoprivacy by Design Guideline for Research Campaigns That Use Participatory Sensing Data. *Journal of Empirical Research on Human Research Ethics*, *13*(3), 203–222. https://doi.org/10.1177/1556264618759877

Krumm, J. (2007). Inference Attacks on Location Tracks. In *Pervasive Computing* (pp. 127–143). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-72037-9_8

Kwan, M.-P., Casas, I., & Schmitz, B. (2004). Protection of Geoprivacy and Accuracy of Spatial Information: How Effective Are Geographical Masks? *Cartographica: The International Journal for Geographic Information and Geovisualization*, *39*(2), 15–28. https://doi.org/10.3138/X204-4223-57MK-8273

Leitner, M., & Curtis, A. (2004). Cartographic Guidelines for Geographically Masking the Locations of Confidential Point Data. *Cartographic Perspectives*, *49*, 22–39. https://doi.org/10.14714/CP49.439

Leitner, M., & Curtis, A. (2006). A first step towards a framework for presenting the location of confidential point data on maps—results of an empirical perceptual study. *International Journal of Geographical Information Science*, *20*(7), 813–822. https://doi.org/10.1080/13658810600711261

Leitner, M., Mills, J. W., & Curtis, A. (2007). Can Novices to Geospatial Technology Compromise Spatial Confidentiality? *KN - Journal of Cartography and Geographic Information*, *57*(2), 78–84. https://doi.org/10.1007/BF03544029

Likert, R. (1932). A technique for the measurement of attitudes. In *Archives of Psychology* (140th ed., Vol. 22, pp. 5–55).

Lu, Y., Yorke, C., & Zhan, F. B. (2012). Considering Risk Locations When Defining Perturbation Zones for Geomasking. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *47*(3), 168–178. https://doi.org/10.3138/carto.47.3.1112

McLafferty, S. (2004). The Socialization of GIS. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *39*(2), 51–53. https://doi.org/10.3138/F333-6V74-815U-4631

Polzin, F., & Kounadi, O. (2021). Adaptive Voronoi Masking: A Method to Protect Confidential Discrete Spatial Data. *11th International Conference on Geographic Information Science (GIScience 2021)*. https://doi.org/10.4230/LIPIcs.GIScience.2021.II.1

Ray, K., Davidson, R., Husain, F., Vegeris, S., Vowden, K., & Karn, J. (2012). Perceptions of the policing and crime mapping 'Trailblazers'. *Home Office, Policing research and analysis, Research Report, 67.*

Richter, W. (2018). The verified neighbor approach to geoprivacy: An improved method for geographic masking. *Journal of Exposure Science & Environmental Epidemiology*, *28*(2), 109–118. https://doi.org/10.1038/jes.2017.17

Seidl, D. E., Jankowski, P., & Clarke, K. C. (2018). Privacy and False Identification Risk in Geomasking Techniques. *Geographical Analysis*, *50*(3), 280–297. https://doi.org/10.1111/gean.12144

Seidl, D. E., Jankowski, P., & Nara, A. (2019). An empirical test of household identification risk in geomasked maps. *Cartography and Geographic Information Science*, *46*(6), 475–488. https://doi.org/10.1080/15230406.2018.1544932

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

Swanlund, D., Schuurman, N., Zandbergen, P., & Brussoni, M. (2020). Street masking: a network-based geographic mask for easily protecting geoprivacy. *International Journal of Health Geographics*, *19*(1), 26. https://doi.org/10.1186/s12942-020-00219-z

Sweeney, L. (2002). k-Anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *10*(05), 557–570. https://doi.org/10.1142/S0218488502001648

Wang, J., Kim, J., & Kwan, M.-P. (2022). An exploratory assessment of the effectiveness of geomasking methods on privacy protection and analytical accuracy for individual-level geospatial data. *Cartography and Geographic Information Science*, *49*(5), 385–406. https://doi.org/10.1080/15230406.2022.2056510

WMS- https://www.wien.gv.at/viennagis/ (a): Geoland Basemap Orthofoto https://maps.wien.gv.at/basemap/1.0.0/WMTSCapabilities_31256.xml

WMS- https://www.wien.gv.at/viennagis/ (b): Geoland Basemap Farbe https://maps.wien.gv.at/basemap/1.0.0/WMTSCapabilities_31256.xml

Zandbergen, P. A. (2014). Ensuring Confidentiality of Geocoded Health Data: Assessing Geographic Masking Strategies for Individual-Level Data. *Advances in Medicine*, *2014*, 1–14. https://doi.org/10.1155/2014/567049

Zhang, S., Freundschuh, S. M., Lenzer, K., & Zandbergen, P. A. (2017). The location swapping method for geomasking. *Cartography and Geographic Information Science*, *44*(1), 22–34. https://doi.org/10.1080/15230406.2015.1095655

Zimmerman, D. L., & Pavlik, C. (2007). Quantifying the Effects of Mask Metadata Disclosure and Multiple Releases on the Confidentiality of Geographically Masked Health Data. *Geographical Analysis*, *40*(1), 52–76. https://doi.org/10.1111/j.0016-7363.2007.00713.x

7. Appendix A

This is form 1, with disclaimer one showing the multi-purpose area basemaps.

Geomasking and human cognition

Date	31.07.2023
Author-Interviewer	Isabella Noll
MSc program	Geography
Department	Department for Geography and Regional Research

Consent and confidentiality

I consent to participate in this research survey and the following has been explained to me:

• The survey responses will be made publicly available for the master's thesis of Isabella Noll.

• The author will retain the responses for a period of 2 years for research verification purposes.

• All personal information of the participants will be held confidential and will not be made publicly available at any time.

• By signing you consent to the conditions given above.

Name:

Signature:

Reidentification of health geodata representing individuals

As part of my master's thesis, I am conducting an examination on confidentiality and reidentification of geographical health datasets, such as households or residential addresses of individuals, represented by discrete points on maps. The study area is Vienna.

To safeguard personal data of individuals with spatial reference, such as addresses and coordinates, which are represented as discrete points, I employed a technique known as "geomasking" in which the original locations are shifted to avoid knowing the actual position on the map.

In the upcoming exercise, you will be presented with a series of maps where points represent individuals of households or residential addresses. They have been masked for confidentiality purposes. The task is divided into two parts:

- 5 maps are masked by a method called street masking (Swanlund et al., 2020).
- 5 maps are masked by a method called donut masking (Hampton et al, 2010).

Each method will be described to you, once before the actual task. Note that points with a number have been masked with street masking, and points without a number have been masked with donut masking.

Points masked by donut

Points masked by street masking.

Your task

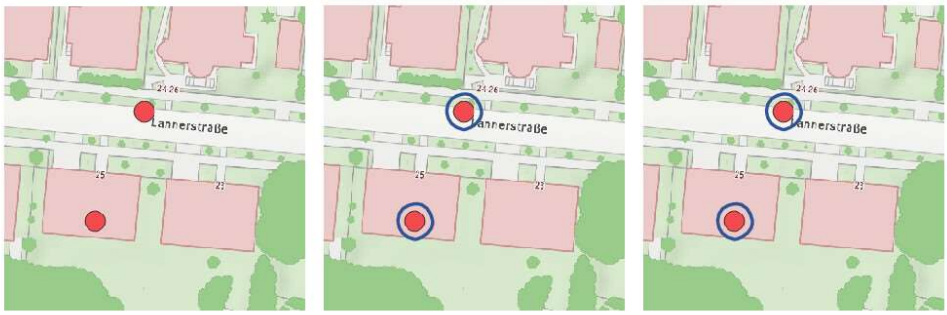
To complete the task, please follow the instructions in the images below.

Look at this section of a map. The red points are masked points representing an individual.

Do you think you can find the original location and hack the place? Choose one or more of the red points and mark it/them.

Make a cross where you think the original point is situated. Connect your chosen point with the cross.

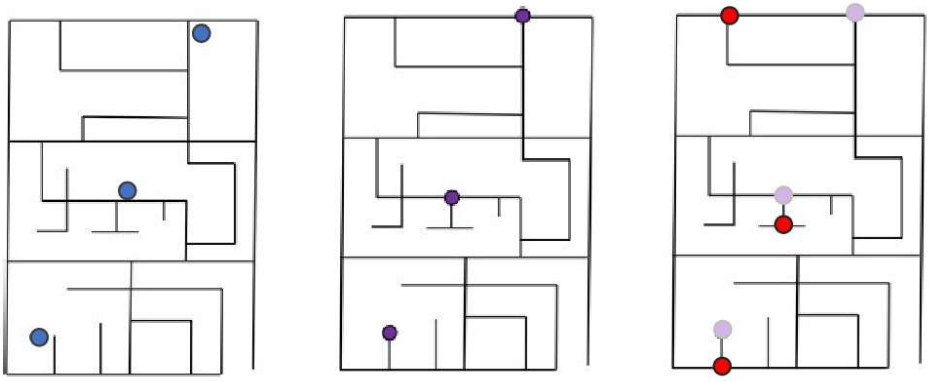
Please keep in mind that the original points are just on locations, where there is actual residential use. There may be more than one individual in one housing. If you think that this is the case, please mark the masked points and estimate the location of origin by marking it on the map as well as shown below.



The points with a number represent several individuals, which are moved to the same location. Please pick a point by marking it. Try to figure out the location of origin by marking it on the map as well. If the point represents more than one individuum, please feel free to mark as many locations as are shown by the number on the point. For example, if the point has the number two, then you may mark a maximum of two locations of origin on your map as shown below.

Disclaimer of the masking method - Street masking

The method used to mask the points of the following five maps is known as street masking (Swanlund et al., 2020). As the name describes, the points are shifted from the residential place to the closest street intersection or closest end of a dead-end street in the street network of Open Street Map. From there the points are moved to a different street intersection or to an end of a dead-end street.



Figures modified from Swanlund et al., 2020

The unmasked points (blue) in their original location.

In the next step they are moved to the closest street intersection or end of a dead-end.

Finally, they are moved to a close street intersection or end of a dead-end. The masked points are shown in red.

1. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map A - street masking

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

2. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map B - street masking

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

3. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map C - street masking

Legend

1 geomasked points by street masking

Geoland basemap color

Data Source

Raster data of the multi-purpose map of Vienna | Stadt Wien- https://data.wien.gv.at

Dataset of individuals | own work, masked by street masking (Swoniund et al., 2020)

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

4. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map D - street masking

Legend

1 geomasked points by street masking

Geoland basemap color

Data Source

Raster data of the multi-purpose map of Vienna | Stadt Wien- https://data.wien.gv.at

Dataset of individuals | own work, masked by street masking (Swanlund et al., 2020)

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

5. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map E - street masking

Legend

1 geomasked points by street masking

Geoland basemap color

Data Source

Raster data of the multi-purpose map of Vienna | Stadt Wien - https://data.wien.gv.at

Dataset of individuals | own work, masked by street masking (Swanlund et al., 2020)

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

Disclaimer of the masking method - Donut masking

The method, which is used to mask the points in the following five maps, is called donut masking (Hampton et al., 2010). As the name describes the masked points are inside a donut shaped area.

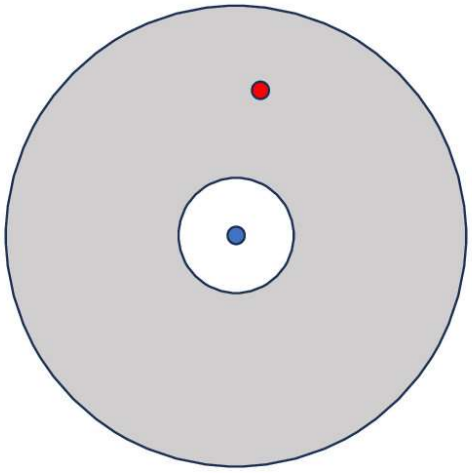


Figure based on Hampton et al., 2010

The blue, unmasked point is in the center of the donut. During the masking procedure it is moved into the shaded area (=red point).

With this method, there is always only one individual in the masked point. This means that there are no numbers on the masked points shown in the following maps.

6. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map F - donut masking

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

7. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map G - donut masking

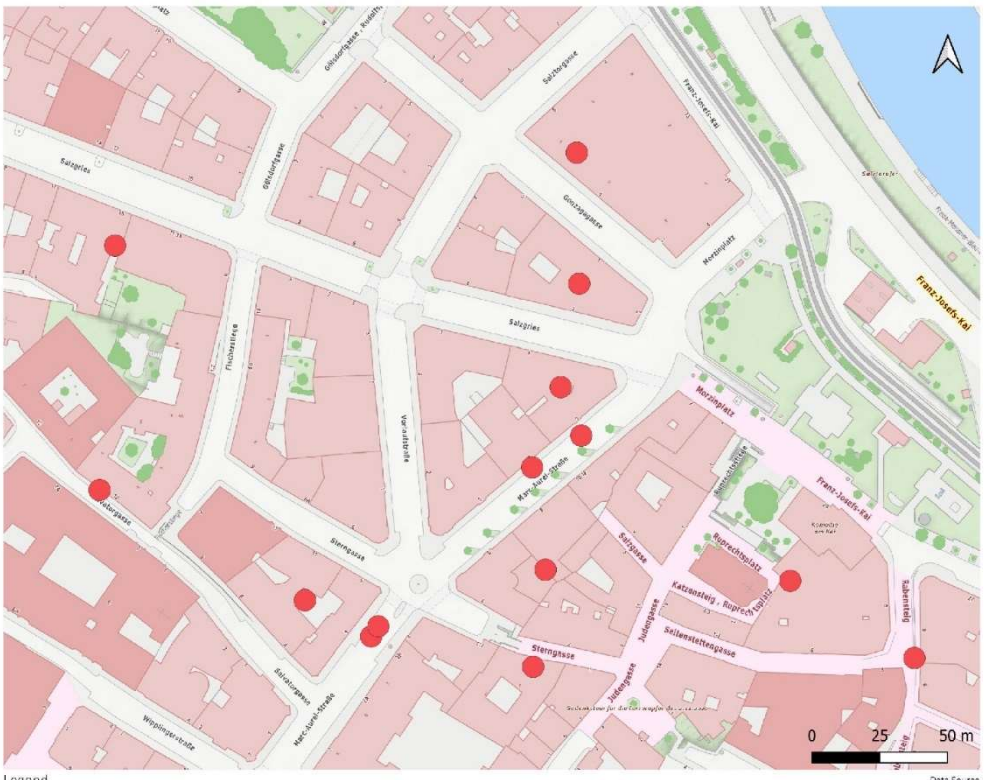
How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

8. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map H - donut masking



Legend

geomasked points by donut masking

Geoland basemap color

Data Source

Raster data of the multi-purpose map of Vienna | Stadt Wien- https://data.wien.gv.at

Dataset of individuals | own work, masked by donut masking (Hampton et al., 2010)

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

9. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map I - donut masking

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

10. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map J - donut masking

How confident did you feel in figuring out the original position?

☐ completely confident	☐ fairly confident	☐ somewhat confident	☐ slightly confident	☐ not at all confident

Demographics

Age:

☐ 0-14 ☐ 15-19 ☐ 20-34 ☐ 35-49 ☐ 50-64 ☐ 65 and older

Gender:

☐ male ☐ female ☐ other

School leaving qualification:

☐ elementary/ primary school ☐ secondary school without university entrance qualification ☐ secondary school with university entrance qualification ☐ University or/and Fachhochschule

Do you work in the field of Geography, in a related field or with spatial data (cartography, geoinformatics, spatial/landscape/city planning, forestry, astronomy, geology, geo science)?

☐ Yes. ☐ No.

Did you recognize any of the areas shown in the map section?

☐ Yes. ☐ No.

Did you find the instructions of the masking methods helpful?

☐ Yes. ☐ No.

8. Appendix B

This is form two showing disclaimer one with orthophotos as a basemap.

Geomasking and human cognition

Date	31.07.2023
Author-Interviewer	Isabella Noll
MSc program	Geography
Department	Department for Geography and Regional Research

Consent and confidentiality

I consent to participate in this research survey and the following has been explained to me:

- The survey responses will be made publicly available for the master's thesis of Isabella Noll.
- The author will retain the responses for a period of 2 years for research verification purposes.
- All personal information of the participants will be held confidential and will not be made publicly available at any time.
- By signing you consent to the conditions given above.

Name:

Signature:

Reidentification of health geodata representing individuals

As part of my master's thesis, I am conducting an examination on confidentiality and reidentification of geographical health datasets, such as households or residential addresses of individuals, represented by discrete points on maps. The study area is Vienna.

To safeguard personal data of individuals with spatial reference, such as addresses and coordinates, which are represented as discrete points, I employed a technique known as "geomasking" in which the original locations are shifted to avoid knowing the actual position on the map.

In the upcoming exercise, you will be presented with a series of maps where points represent individuals of households or residential addresses. They have been masked for confidentiality purposes. The task is divided into two parts:

- 5 maps are masked by a method called street masking (Swanlund et al., 2020).
- 5 maps are masked by a method called donut masking (Hampton et al, 2010).

Each method will be described to you, once before the actual task. Note that points with a number have been masked with street masking, and points without a number have been masked with donut masking.

Points masked by donut

Points masked by street masking.

Your task

To complete the task, please follow the instructions in the images below.

Look at this section of a map. The red points are masked points representing an individual.

Do you think you can find the original location and hack the place? Choose one or more of the red points and mark it/them.

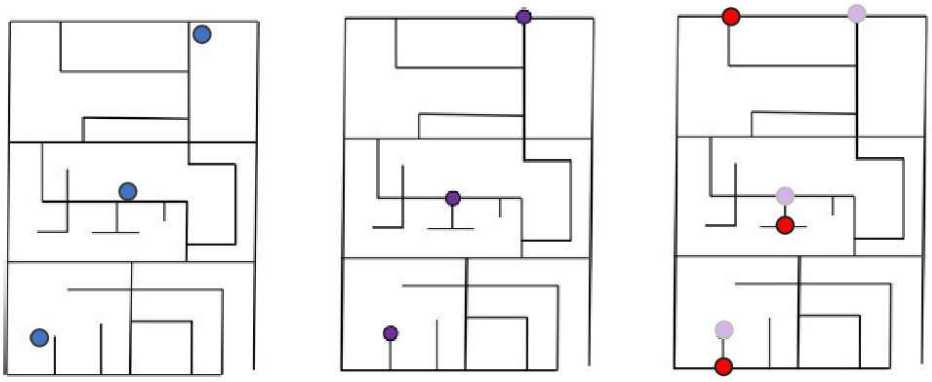
Make a cross where you think the original point is situated. Connect your chosen point with the cross.

Please keep in mind that the original points are just on locations, where there is actual residential use. There may be more than one individual in one housing. If you think that this is the case, please mark the masked points and estimate the location of origin by marking it on the map as well as shown below.

The points with a number represent several individuals, which are moved to the same location. Please pick a point by marking it. Try to figure out the location of origin by marking it on the map as well. If the point represents more than one individuum, please feel free to mark as many locations as are shown by the number on the point. For example, if the point has the number two, then you may mark a maximum of two locations of origin on your map as shown below.

Disclaimer of the masking method - Street masking

The method used to mask the points of the following five maps is known as street masking (Swanlund et al., 2020). As the name describes, the points are shifted from the residential place to the closest street intersection or closest end of a dead-end street in the street network of Open Street Map. From there the points are moved to a different street intersection or to an end of a dead-end street.



Figures modified from Swanlund et al., 2020

The unmasked points (blue) in their original location.

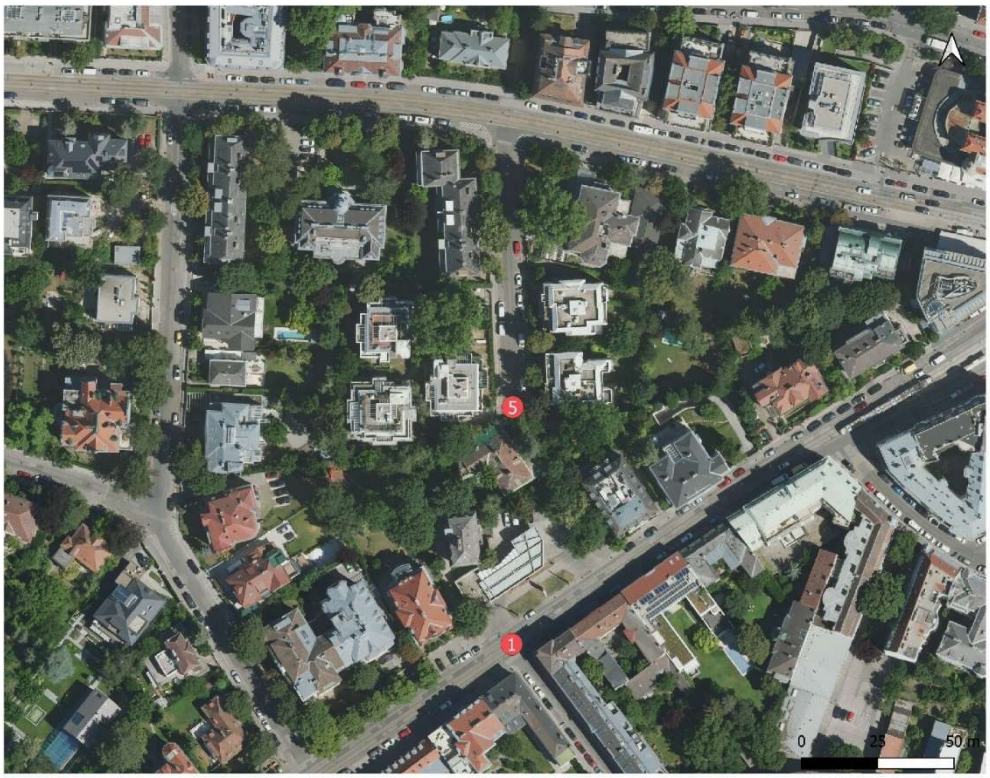
In the next step they are moved to the closest street intersection or end of a dead-end.

Finally, they are moved to a close street intersection or end of a dead-end. The masked points are shown in red.

1. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map A - street masking



Legend

1 geomasked points by street masking

Geoland basemap orthophoto

Data Source
Orthophoto of Vienna | Stadt Wien - https://data.wien.gv.at
Dataset of individuals | own work, masked by street masking (Swanlund et al., 2020)

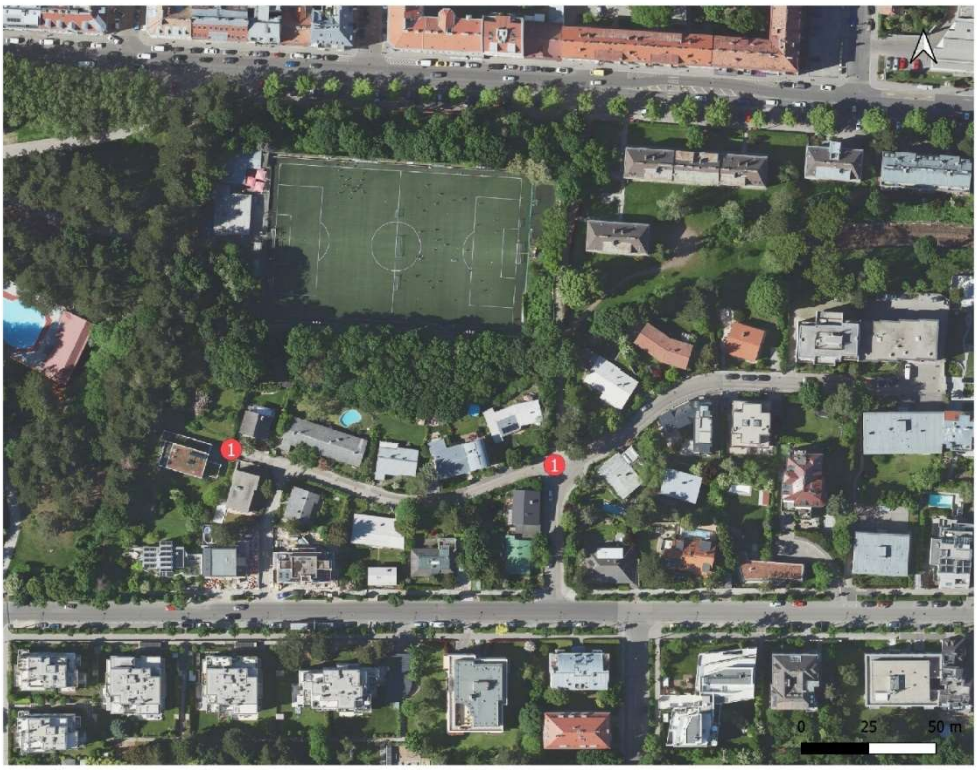
How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

2. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map B - street masking



Legend

1 geomasked points by street masking

Geoland basemap orthophoto

Data Source

Orthophoto map of Vienna | Stadt Wien- https://data.wien.gv.at

Dataset of individuals | own work, masked by street masking (Swanlund et al., 2020)

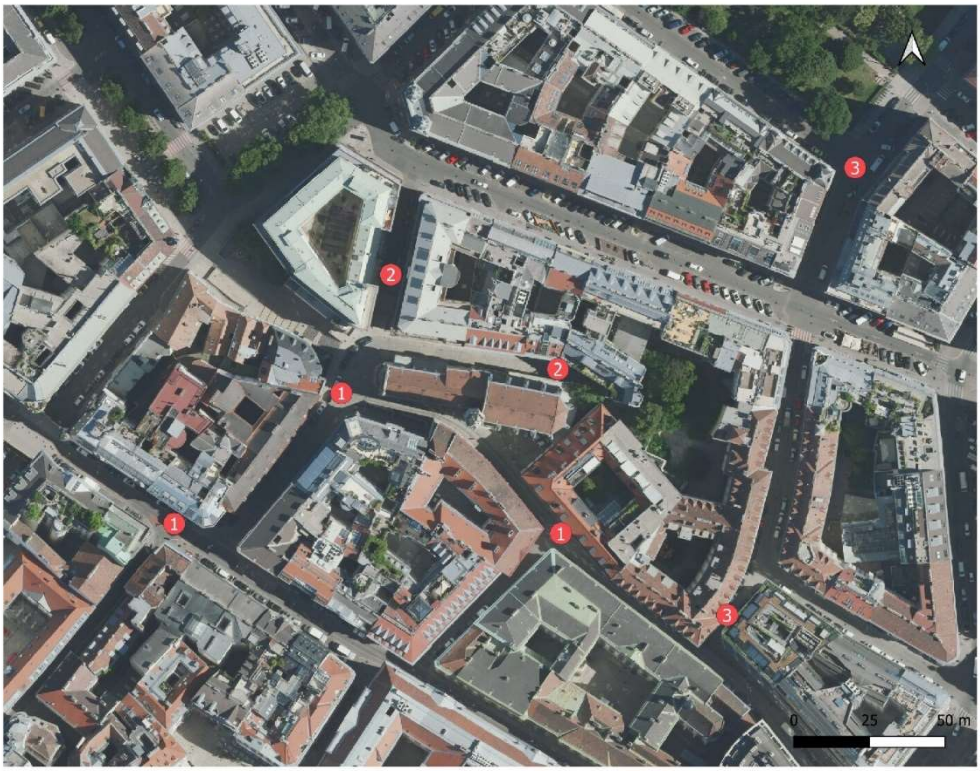
How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

3. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map C - street masking



Legend

1 geomasked points by street masking

Geoland basemap orthophoto

Data Source

Orthophoto of Vienna | Stadt Wien- https://data.wien.gv.at

Dataset of individuals | own work, masked by street masking (Swanlund et al., 2020)

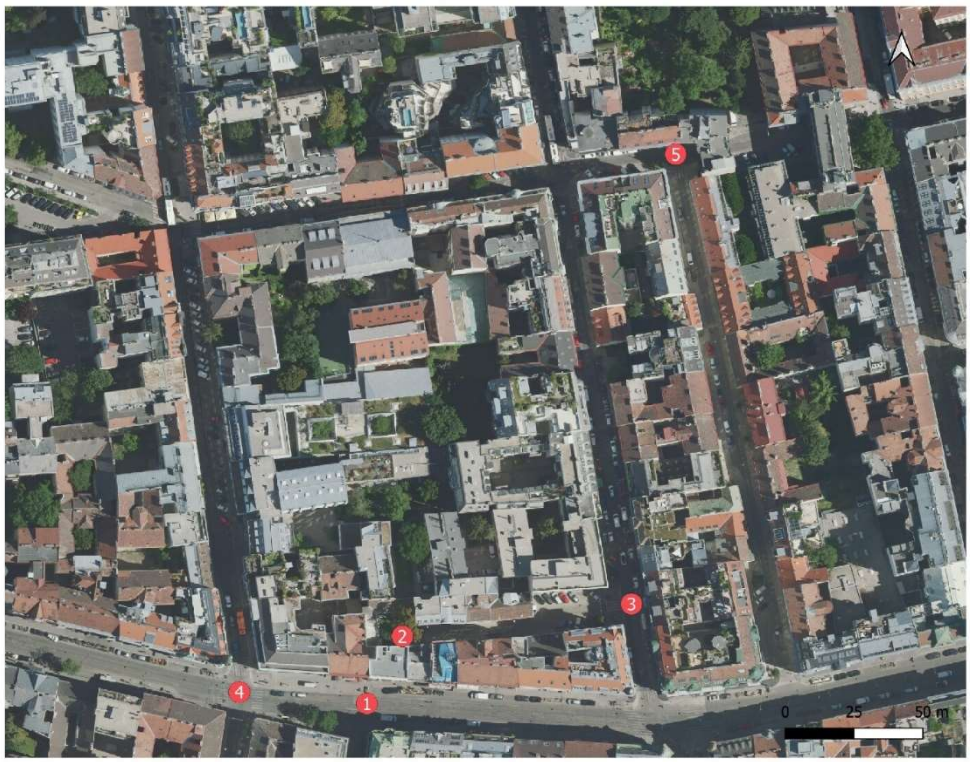
How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

4. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map D - street masking



How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

5. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map E - street masking

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

Disclaimer of the masking method - Donut masking

The method, which is used to mask the points in the following five maps, is called donut masking (Hampton et al., 2010). As the name describes the masked points are inside a donut shaped area.

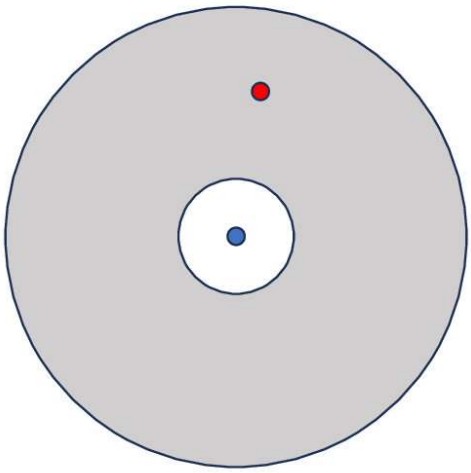


Figure based on Hampton et al., 2010

The blue, unmasked point is in the center of the donut. During the masking procedure it is moved into the shaded area (=red point).

With this method, there is always only one individual in the masked point. This means that there are no numbers on the masked points shown in the following maps.

6. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map F - donut masking

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

7. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map G - donut masking

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

8. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map H - donut masking

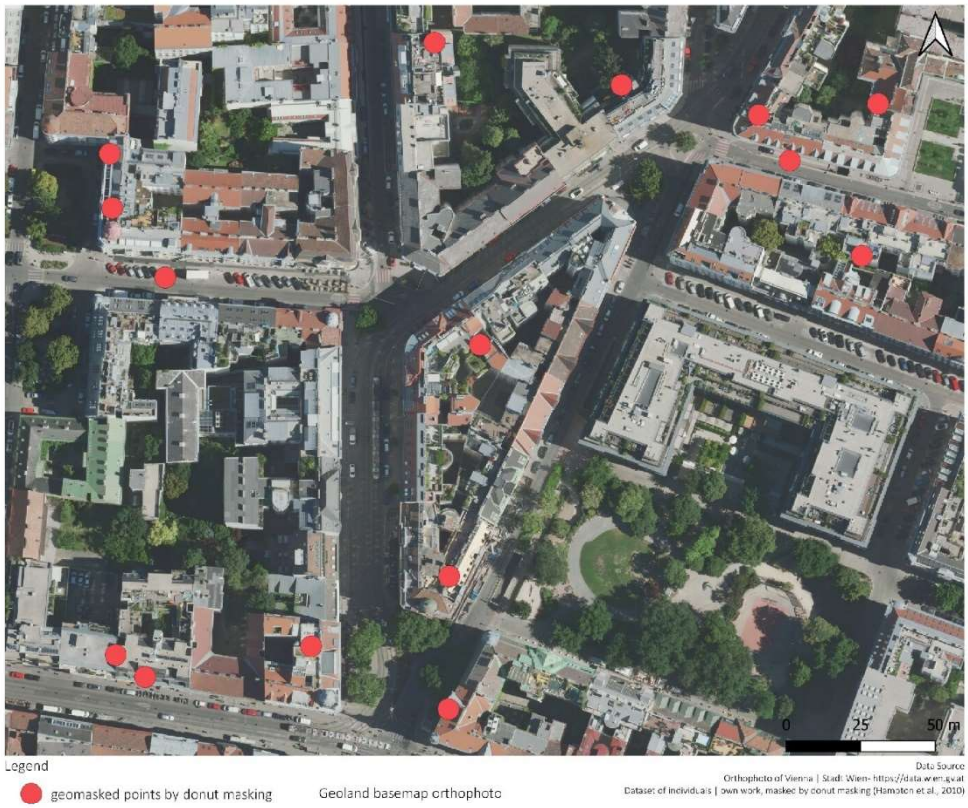
How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

9. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map I - donut masking



How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

10. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map J - donut masking

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

Demographics

Age:

☐ 0-14 ☐ 15-19 ☐ 20-34 ☐ 35-49 ☐ 50-64 ☐ 65 and older

Gender:

☐ male ☐ female ☐ other

School leaving qualification:

☐ elementary/ primary school ☐ secondary school without university entrance qualification ☐ secondary school with university entrance qualification ☐ University or/and Fachhochschule

Do you work in the field of Geography, in a related field or with spatial data (cartography, geoinformatics, spatial/landscape/city planning, forestry, astronomy, geology, geo science)?

☐ Yes. ☐ No.

Did you recognize any of the areas shown in the map section?

☐ Yes. ☐ No.

Did you find the instructions of the masking methods helpful?

☐ Yes. ☐ No.

9. Appendix C

This is form 3, which includes building parcels and disclaimer 1.

Geomasking and human cognition

Date	31.07.2023
Author-Interviewer	Isabella Noll
MSc program	Geography
Department	Department for Geography and Regional Research

Consent and confidentiality

I consent to participate in this research survey and the following has been explained to me:

- The survey responses will be made publicly available for the master's thesis of Isabella Noll.
- The author will retain the responses for a period of 2 years for research verification purposes.
- All personal information of the participants will be held confidential and will not be made publicly available at any time.
- By signing you consent to the conditions given above.

Name:

Signature:

Reidentification of health geodata representing individuals

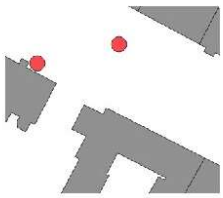
As part of my master's thesis, I am conducting an examination on confidentiality and reidentification of geographical health datasets, such as households or residential addresses of individuals, represented by discrete points on maps. The study area is Vienna.

To safeguard personal data of individuals with spatial reference, such as addresses and coordinates, which are represented as discrete points, I employed a technique known as "geomasking" in which the original locations are shifted to avoid knowing the actual position on the map.

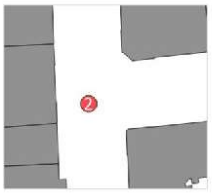
In the upcoming exercise, you will be presented with a series of maps where points represent individuals of households or residential addresses. They have been masked for confidentiality purposes. The task is divided into two parts:

- 5 maps are masked by a method called street masking (Swanlund et al., 2020).
- 5 maps are masked by a method called donut masking (Hampton et al, 2010).

Each method will be described to you, once before the actual task. Note that points with a number have been masked with street masking, and points without a number have been masked with donut masking.



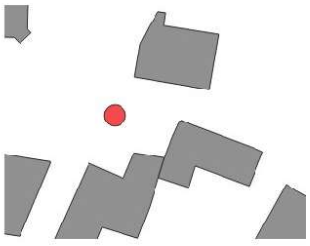
Points masked by donut



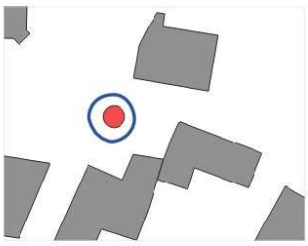
Points masked by street masking.

Your task

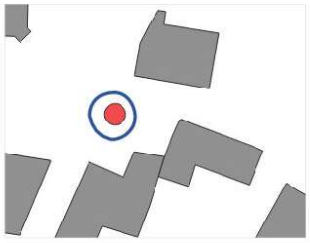
To complete the task, please follow the instructions in the images below.



Look at this section of a map. The red points are masked points representing an individual.

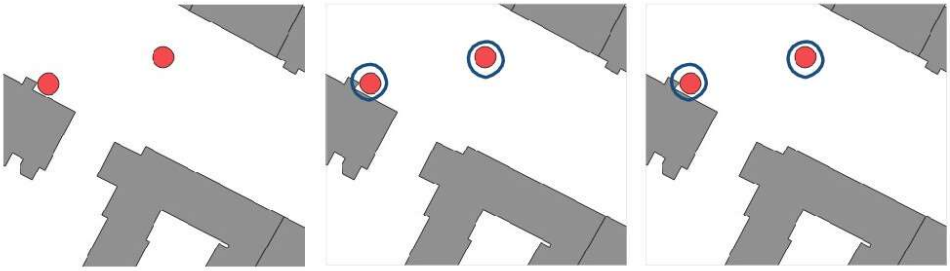


Do you think you can find the original location and hack the place? Choose one or more of the red points and mark it/them.

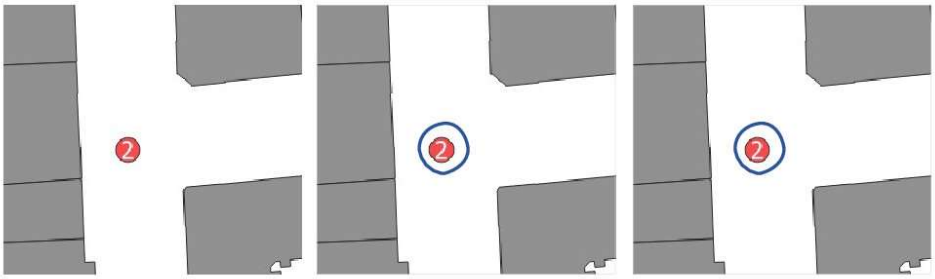


Make a cross where you think the original point is situated. Connect your chosen point with the cross.

Please keep in mind that the original points are just on locations, where there is actual residential use. There may be more than one individual in one housing. If you think that this is the case, please mark the masked points and estimate the location of origin by marking it on the map as well as shown below.

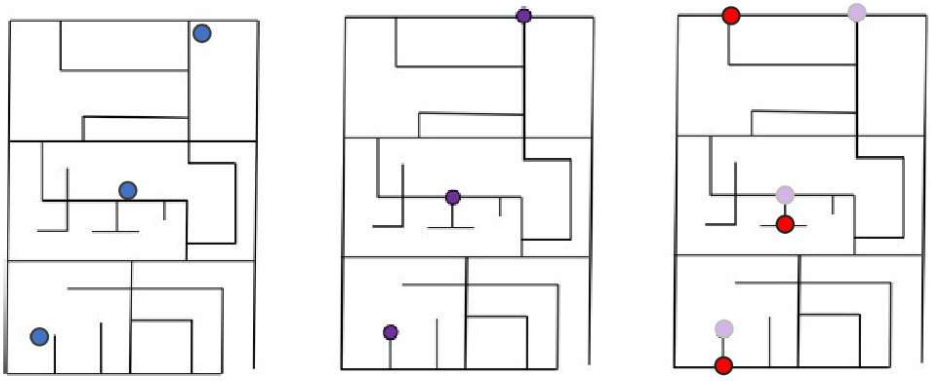


The points with a number represent several individuals, which are moved to the same location. Please pick a point by marking it. Try to figure out the location of origin by marking it on the map as well. If the point represents more than one individuum, please feel free to mark as many locations as are shown by the number on the point. For example, if the point has the number two, then you may mark a maximum of two locations of origin on your map as shown below.



Disclaimer of the masking method - Street masking

The method used to mask the points of the following five maps is known as street masking (Swanlund et al., 2020). As the name describes, the points are shifted from the residential place to the closest street intersection or closest end of a dead-end street in the street network of Open Street Map. From there the points are moved to a different street intersection or to an end of a dead-end street.



Figures modified from Swanlund et al., 2020

The unmasked points (blue) in their original location.

In the next step they are moved to the closest street intersection or end of a dead-end.

Finally, they are moved to a close street intersection or end of a dead-end. The masked points are shown in red.

1. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

2. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

3. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

4. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

5. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

Disclaimer of the masking method - Donut masking

The method, which is used to mask the points in the following five maps, is called donut masking (Hampton et al., 2010). As the name describes the masked points are inside a donut shaped area.

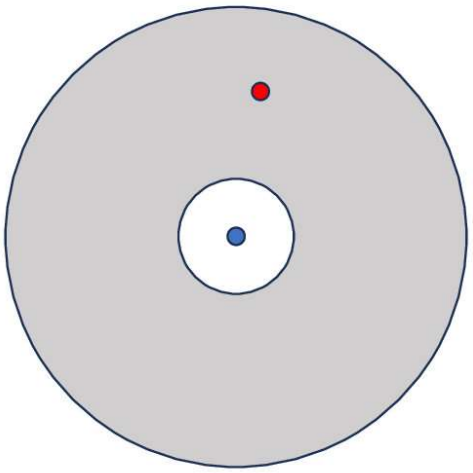


Figure based on Hampton et al., 2010

The blue, unmasked point is in the center of the donut. During the masking procedure it is moved into the shaded area (=red point).

With this method, there is always only one individual in the masked point. This means that there are no numbers on the masked points shown in the following maps.

6. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

7. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

8. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

9. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

10. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

Demographics

Age:

☐ 0-14 ☐ 15-19 ☐ 20-34 ☐ 35-49 ☐ 50-64 ☐ 65 and older

Gender:

☐ male ☐ female ☐ other

School leaving qualification:

☐ elementary/ primary school ☐ secondary school without university entrance qualification ☐ secondary school with university entrance qualification ☐ University or/and Fachhochschule

Do you work in the field of Geography, in a related field or with spatial data (cartography, geoinformatics, spatial/landscape/city planning, forestry, astronomy, geology, geo science)?

☐ Yes. ☐ No.

Did you recognize any of the areas shown in the map section?

☐ Yes. ☐ No.

Did you find the instructions of the masking methods helpful?

☐ Yes. ☐ No.

10. Appendix D

Form 4 includes disclaimer 1 and the intermodal transport reference system as a basemap.

Geomasking and human cognition

Date	31.07.2023
Author-Interviewer	Isabella Noll
MSc program	Geography
Department	Department for Geography and Regional Research

Consent and confidentiality

I consent to participate in this research survey and the following has been explained to me:

• The survey responses will be made publicly available for the master's thesis of Isabella Noll.

• The author will retain the responses for a period of 2 years for research verification purposes.

• All personal information of the participants will be held confidential and will not be made publicly available at any time.

• By signing you consent to the conditions given above.

Name:

Signature:

Reidentification of health geodata representing individuals

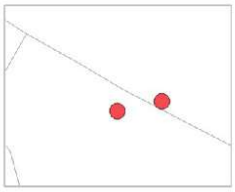
As part of my master's thesis, I am conducting an examination on confidentiality and reidentification of geographical health datasets, such as households or residential addresses of individuals, represented by discrete points on maps. The study area is Vienna.

To safeguard personal data of individuals with spatial reference, such as addresses and coordinates, which are represented as discrete points, I employed a technique known as "geomasking" in which the original locations are shifted to avoid knowing the actual position on the map.

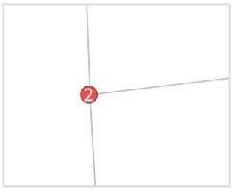
In the upcoming exercise, you will be presented with a series of maps where points represent individuals of households or residential addresses. They have been masked for confidentiality purposes. The task is divided into two parts:

- 5 maps are masked by a method called street masking (Swanlund et al., 2020).
- 5 maps are masked by a method called donut masking (Hampton et al, 2010).

Each method will be described to you, once before the actual task. Note that points with a number have been masked with street masking, and points without a number have been masked with donut masking.



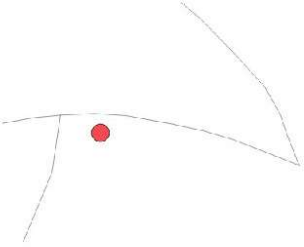
Points masked by donut



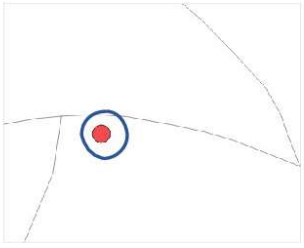
Points masked by street masking.

Your task

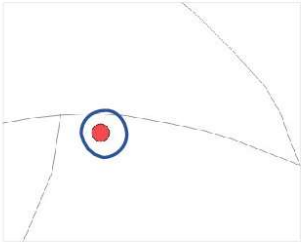
To complete the task, please follow the instructions in the images below.



Look at this section of a map. The red points are masked points representing an individual.

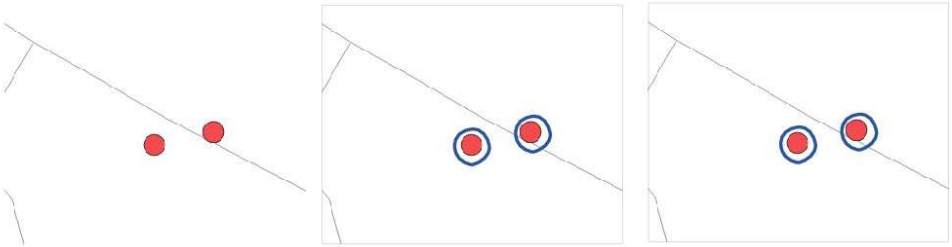


Do you think you can find the original location and hack the place? Choose one or more of the red points and mark it/them.

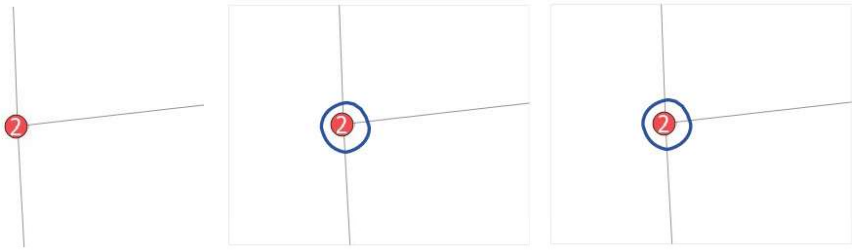


Make a cross where you think the original point is situated. Connect your chosen point with the cross.

Please keep in mind that the original points are just on locations, where there is actual residential use. There may be more than one individual in one housing. If you think that this is the case, please mark the masked points and estimate the location of origin by marking it on the map as well as shown below.

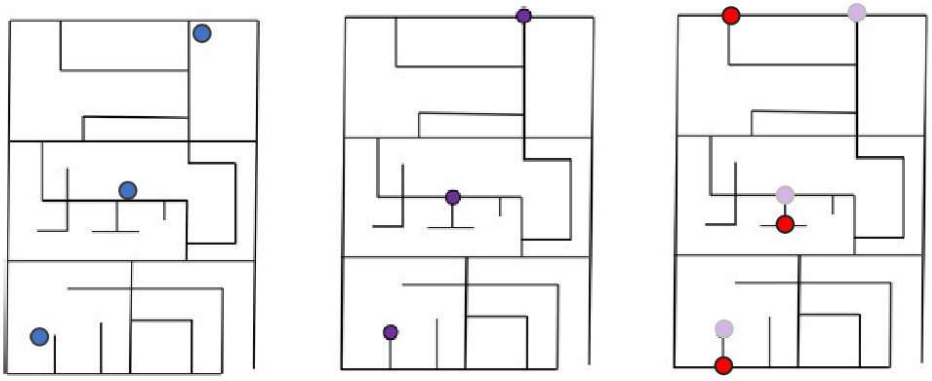


The points with a number represent several individuals, which are moved to the same location. Please pick a point by marking it. Try to figure out the location of origin by marking it on the map as well. If the point represents more than one individuum, please feel free to mark as many locations as are shown by the number on the point. For example, if the point has the number two, then you may mark a maximum of two locations of origin on your map as shown below.



Disclaimer of the masking method - Street masking

The method used to mask the points of the following five maps is known as street masking (Swanlund et al., 2020). As the name describes, the points are shifted from the residential place to the closest street intersection or closest end of a dead-end street in the street network of Open Street Map. From there the points are moved to a different street intersection or to an end of a dead-end street.



Figures modified from Swanlund et al., 2020

The unmasked points (blue) in their original location.

In the next step they are moved to the closest street intersection or end of a dead-end.

Finally, they are moved to a close street intersection or end of a dead-end. The masked points are shown in red.

1. Map

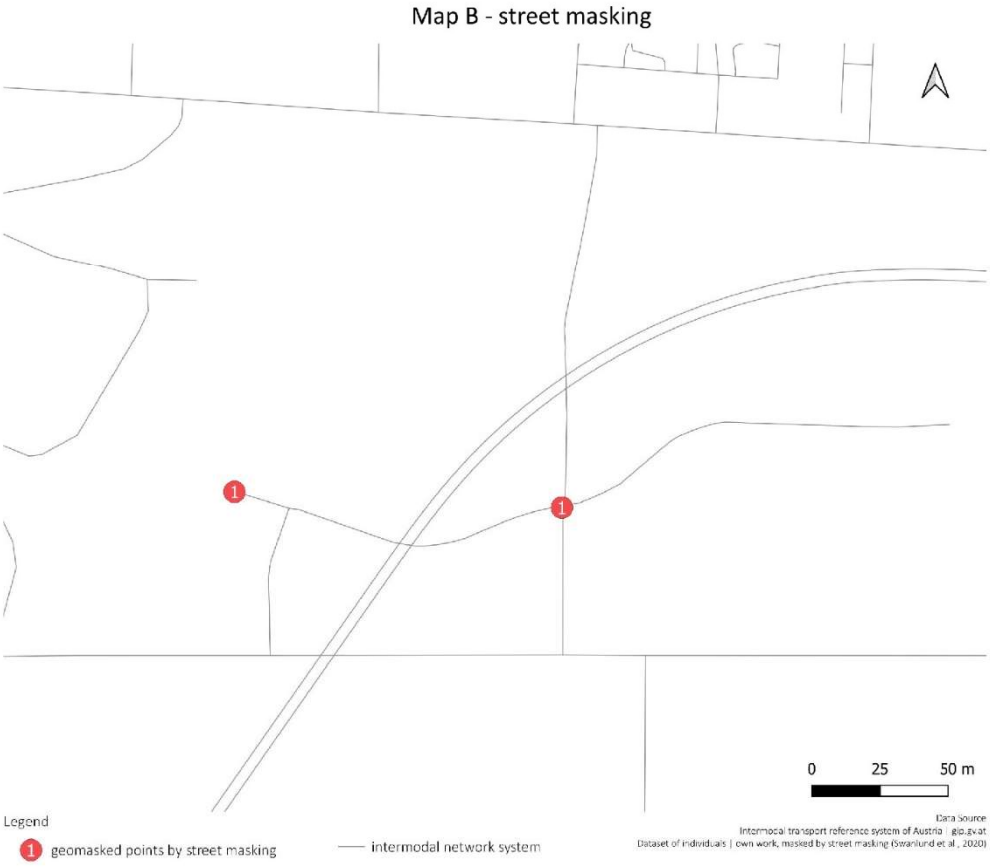
Please figure out the original position of the points you choose on the map below according to the instructions above:

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

2. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

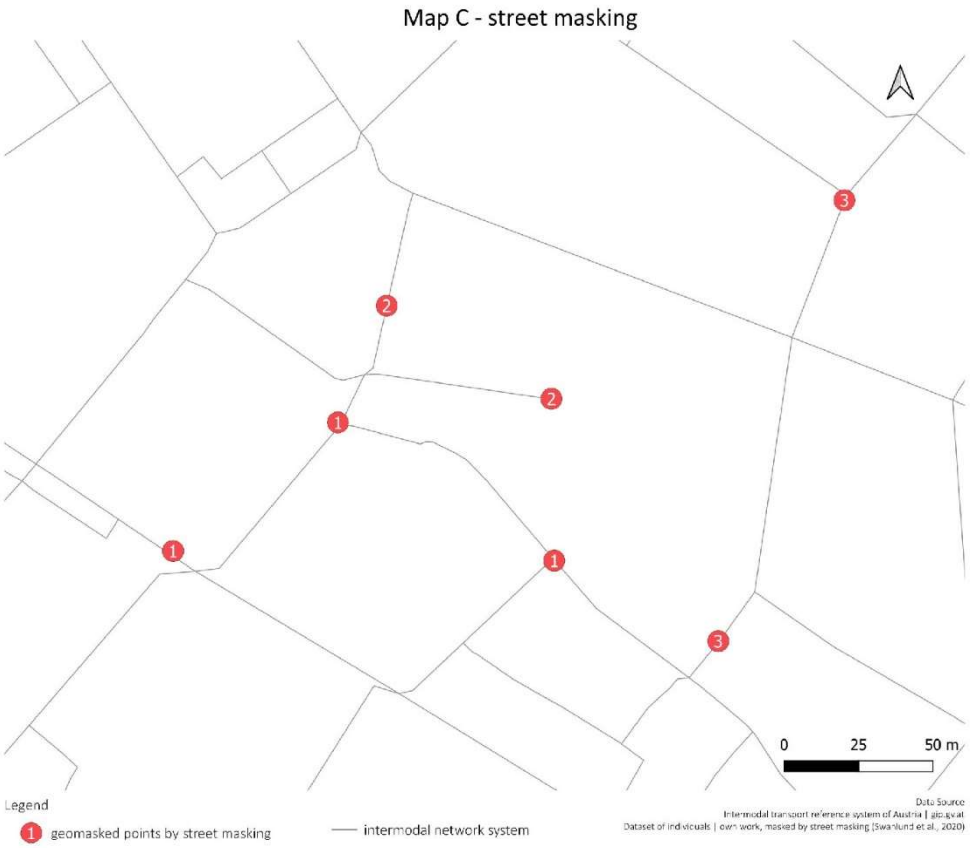


How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

3. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

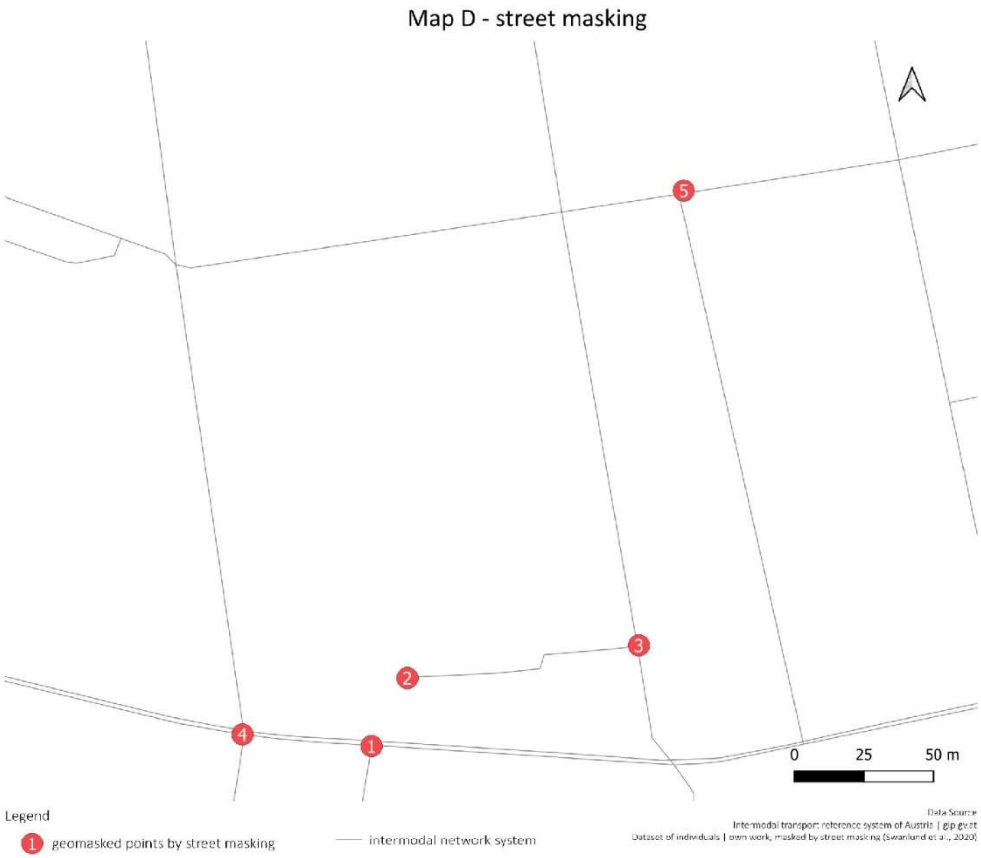


How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

4. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

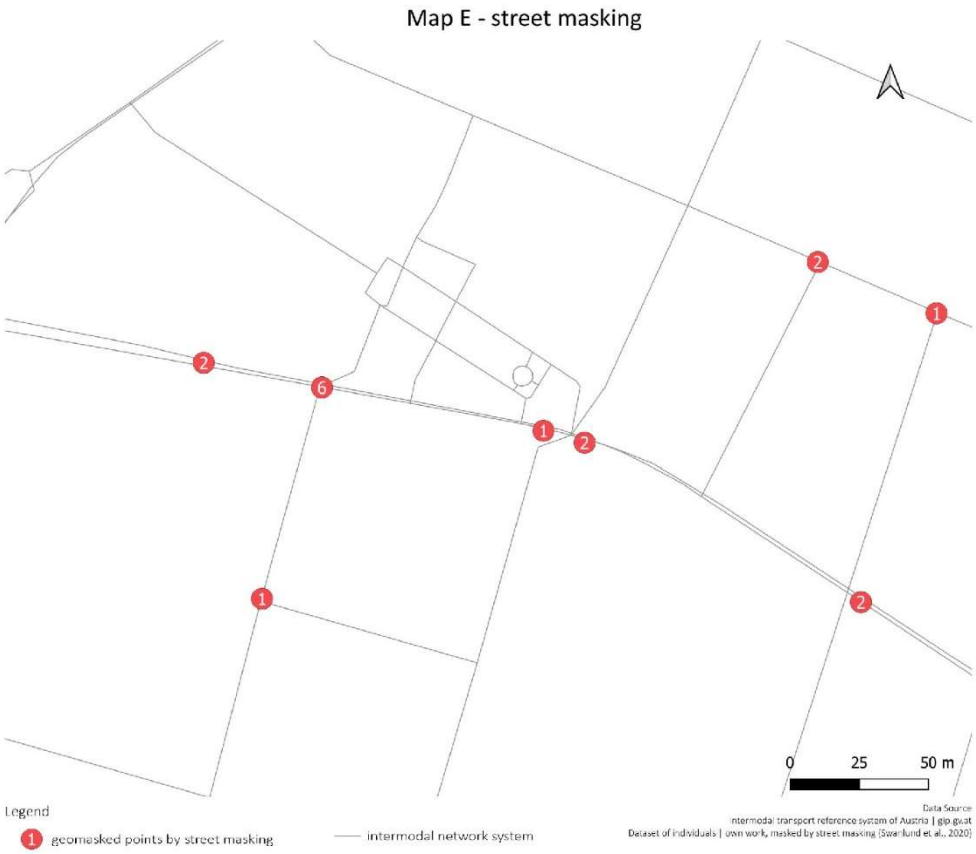


How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

5. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:



How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

Disclaimer of the masking method - Donut masking

The method, which is used to mask the points in the following five maps, is called donut masking (Hampton et al., 2010). As the name describes the masked points are inside a donut shaped area.

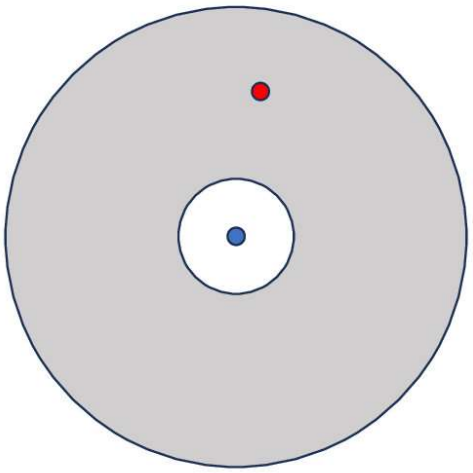


Figure based on Hampton et al., 2010

The blue, unmasked point is in the center of the donut. During the masking procedure it is moved into the shaded area (=red point).

With this method, there is always only one individual in the masked point. This means that there are no numbers on the masked points shown in the following maps.

6. Map

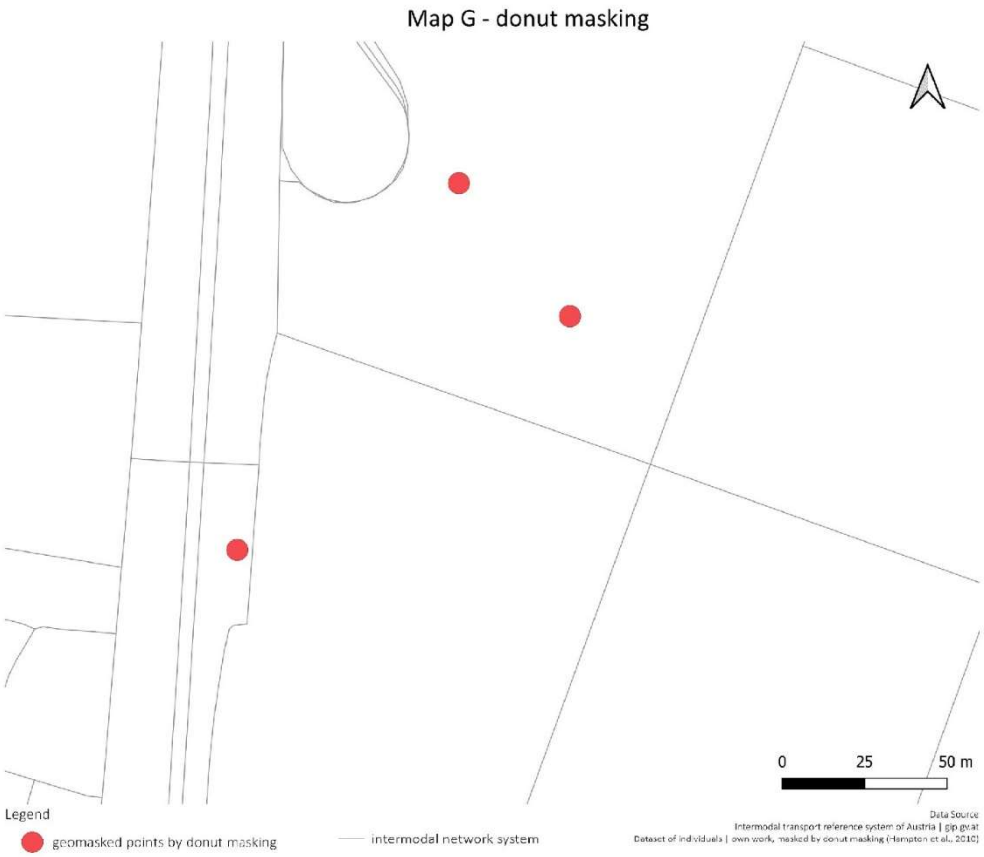
Please figure out the original position of the points you choose on the map below according to the instructions above:

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

7. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:



How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

8. Map

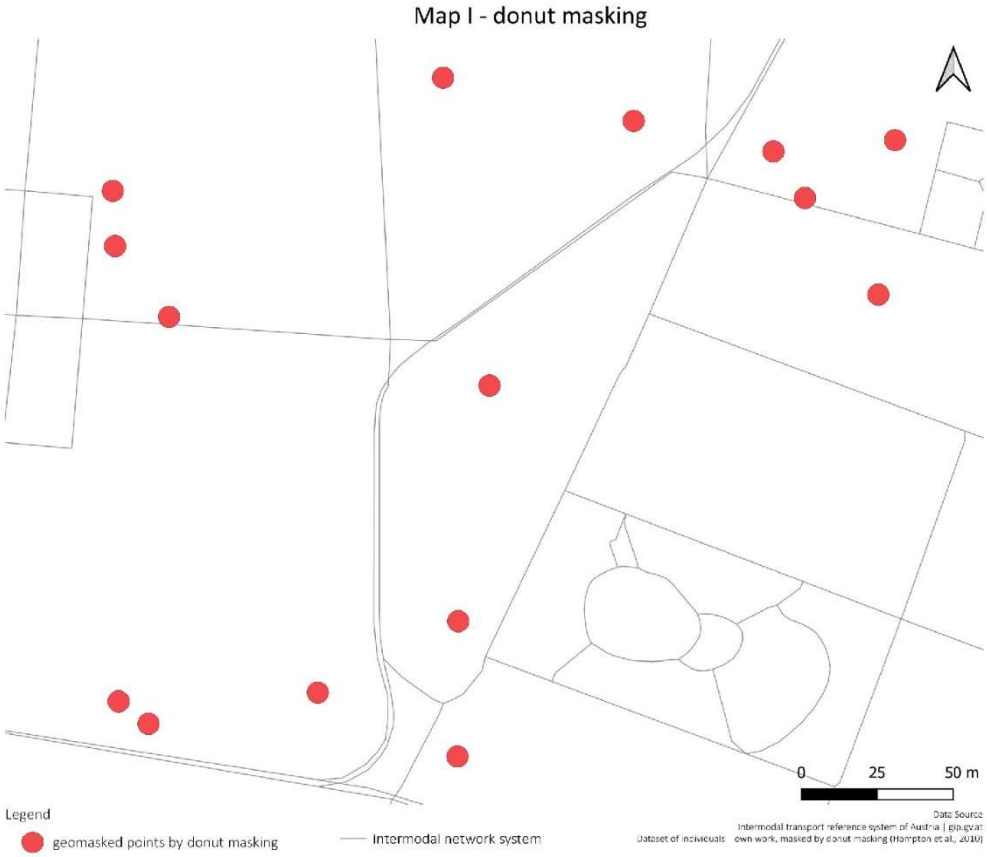
Please figure out the original position of the points you choose on the map below according to the instructions above:

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

9. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

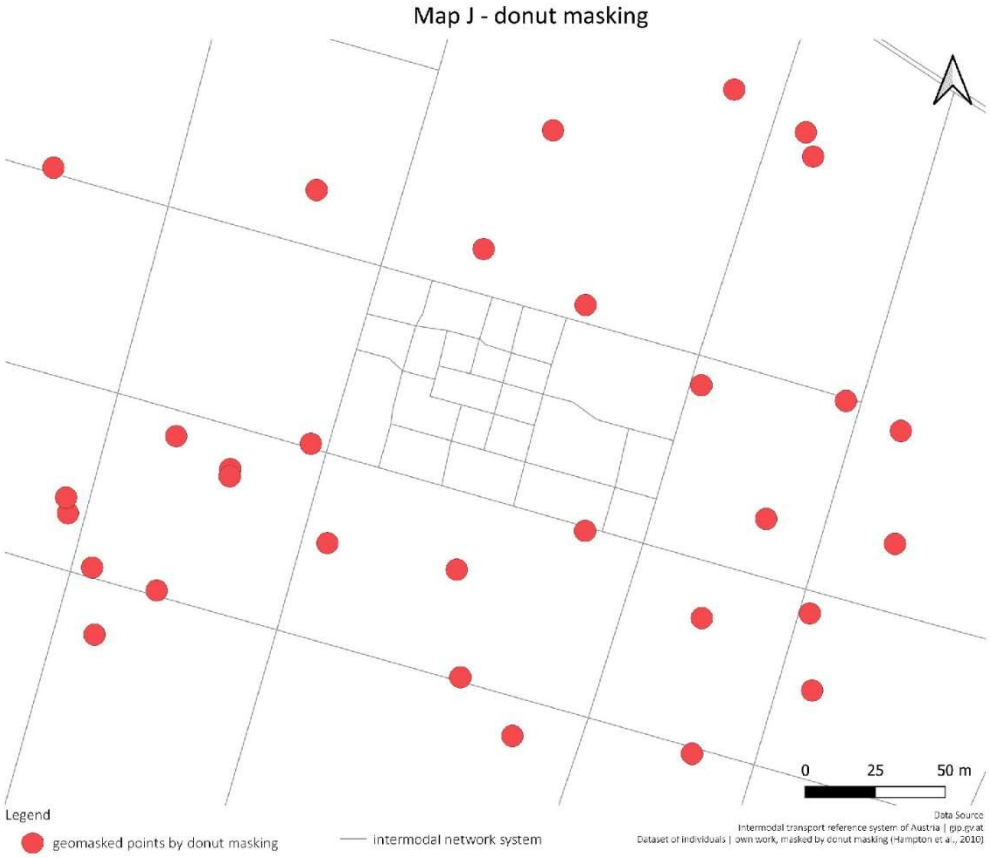


How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

10. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:



How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

Demographics

Age:

☐ 0-14 ☐ 15-19 ☐ 20-34 ☐ 35-49 ☐ 50-64 ☐ 65 and older

Gender:

☐ male ☐ female ☐ other

School leaving qualification:

☐ elementary/ primary school ☐ secondary school without university entrance qualification ☐ secondary school with university entrance qualification ☐ University or/and Fachhochschule

Do you work in the field of Geography, in a related field or with spatial data (cartography, geoinformatics, spatial/landscape/city planning, forestry, astronomy, geology, geo science)?

☐ Yes. ☐ No.

Did you recognize any of the areas shown in the map section?

☐ Yes. ☐ No.

Did you find the instructions of the masking methods helpful?

☐ Yes. ☐ No.

11. Appendix E

This shows form 5, which includes disclaimer 2 and basemap of multi-purpose areas.

Geomasking and human cognition

Date	31.07.2023
Author-Interviewer	Isabella Noll
MSc program	Geography
Department	Department for Geography and Regional Research

Consent and confidentiality

I consent to participate in this research survey and the following has been explained to me:

- The survey responses will be made publicly available for the master's thesis of Isabella Noll.
- The author will retain the responses for a period of 2 years for research verification purposes.
- All personal information of the participants will be held confidential and will not be made publicly available at any time.
- By signing you consent to the conditions given above.

Name:

Signature:

Reidentification of health geodata representing individuals

As part of my master's thesis, I am conducting an examination on confidentiality and reidentification of geographical health datasets, such as households or residential addresses of individuals, represented by discrete points on maps. The study area is Vienna.

To safeguard personal data of individuals with spatial reference, such as addresses and coordinates, which are represented as discrete points, I employed a technique known as "geomasking" in which the original locations are shifted to avoid knowing the actual position on the map.

In the upcoming exercise, you will be presented with a series of maps where points represent individuals of households or residential addresses. They have been masked for confidentiality purposes. The task is divided into two parts:

- 5 maps are masked by a method called street masking (Swanlund et al., 2020).
- 5 maps are masked by a method called donut masking (Hampton et al, 2010).

Each method will be described to you, once before the actual task. Note that points with a number have been masked with street masking, and points without a number have been masked with donut masking.

Points masked by donut

Points masked by street masking.

Your task

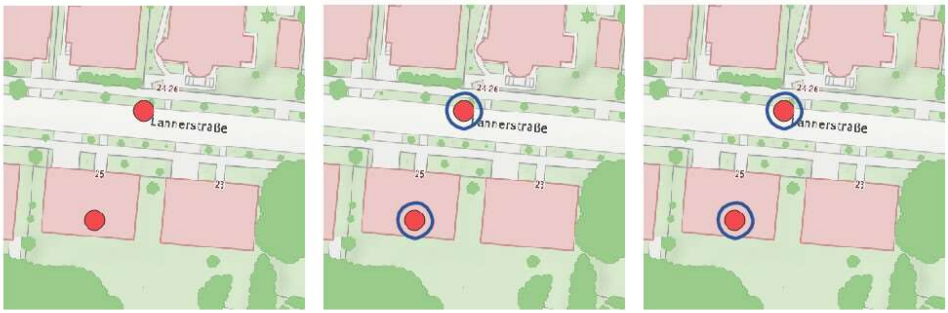
To complete the task, please follow the instructions in the images below.

Look at this section of a map. The red points are masked points representing an individual.

Do you think you can find the original location and hack the place? Choose one or more of the red points and mark it/them.

Make a cross where you think the original point is situated. Connect your chosen point with the cross.

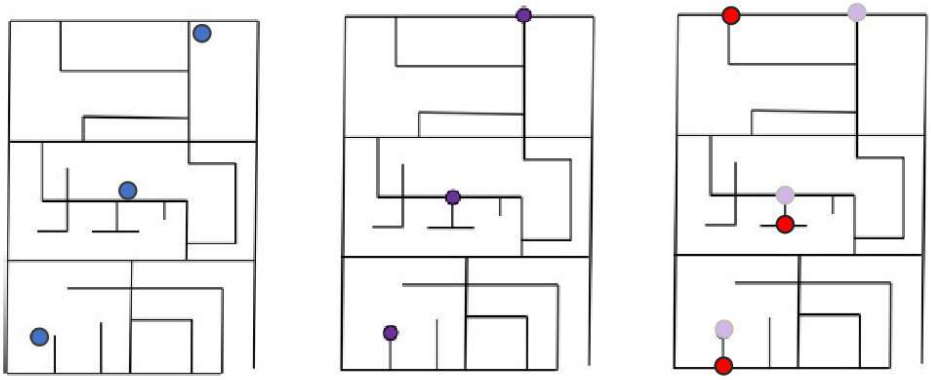
Please keep in mind that the original points are just on locations, where there is actual residential use. There may be more than one individual in one housing. If you think that this is the case, please mark the masked points and estimate the location of origin by marking it on the map as well as shown below.



The points with a number represent several individuals, which are moved to the same location. Please pick a point by marking it. Try to figure out the location of origin by marking it on the map as well. If the point represents more than one individuum, please feel free to mark as many locations as are shown by the number on the point. For example, if the point has the number two, then you may mark a maximum of two locations of origin on your map as shown below.

Disclaimer of the masking method - Street masking

The method used to mask the points of the following five maps is known as street masking (Swanlund et al., 2020). As the name describes, the points are shifted from the residential place to the next intersection or the next end of a dead-end street of the street network (=nodes). This will be the starting node for the next step, in which two close nodes of the network are chosen. The distances from each of the two close nodes to the starting node are calculated. These two distances are summed up and divided by three, to calculate the average distance of the 3 nodes (starting node and the two close nodes). The point is masked by shifting it to one of the two close nodes that is closest to the average distance.



Figures modified from Swanlund et al., 2020

The unmasked points (blue) in their original location.

In the next step they are moved to the closest node (=street intersection or dead-end).

Finally, they are moved to another node depending on the average distance of two close nodes and the closest node from the image to the left. The masked points are shown in red.

1. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map A - street masking

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

2. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map B - street masking

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

3. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map C - street masking

How confident did you feel in figuring out the original position?

☐ completely confident	☐ fairly confident	☐ somewhat confident	☐ slightly confident	☐ not at all confident

4. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map D - street masking

Legend

1 geomasked points by street masking

Geoland basemap color

Data Source

Raster data of the multi-purpose map of Vienna | Stadt Wien- https://data.wien.gv.at

Dataset of individuals | own work, masked by street masking (Swanlund et al., 2020)

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

5. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map E - street masking

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

Disclaimer of the masking method - Donut masking

The method, which is used to mask the points in the following five maps, is called donut masking (Hampton et al., 2010). Donut masking moves the unmasked point from its original location to a random location within a donut-shaped area around the original position. The outer radius of the donut is the maximum shifting distance and is chosen so that only 14 other people are living inside. The inner radius is the minimum shifting distance and is 20 percent of the outer radius.

Our masked point is randomly placed somewhere in the donut shaped area around the original point, so that it cannot be distinguished from the other individuals inside the area.

Therefore, in a low population density area the point is shifted farther away than in a high population density area.

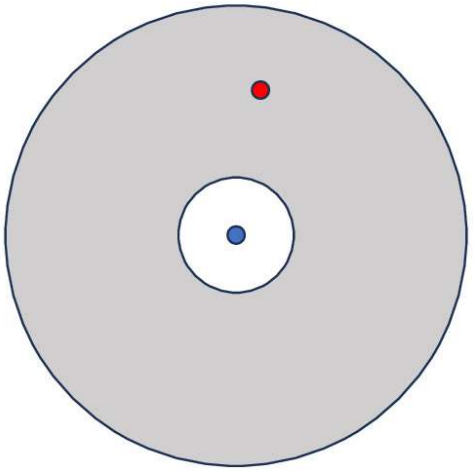


Figure based on Hampton et al., 2010

The blue, unmasked point is in the center of the donut. During the masking procedure it is moved into the shaded area (=red point).

With this method, there is always only one individual in the masked point. This means that there are no numbers on the masked points shown in the following maps.

6. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map F - donut masking

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

7. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map G - donut masking

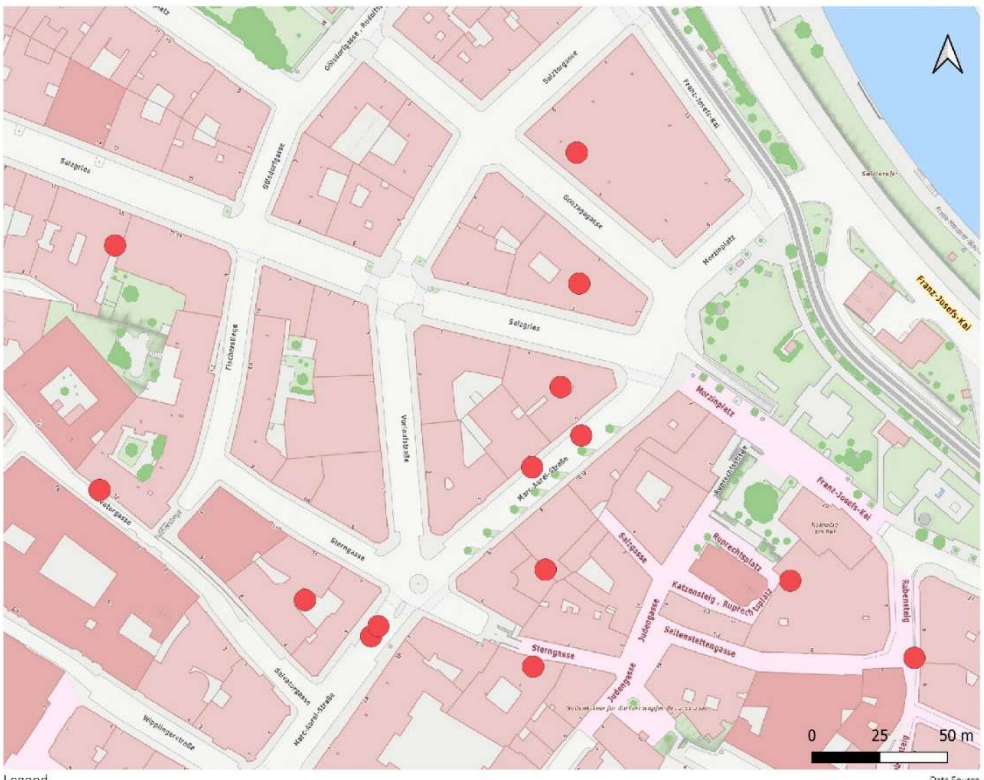
How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

8. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map H - donut masking



How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

9. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map I - donut masking

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

10. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map J - donut masking

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

Demographics

Age:

☐ 0-14 ☐ 15-19 ☐ 20-34 ☐ 35-49 ☐ 50-64 ☐ 65 and older

Gender:

☐ male ☐ female ☐ other

School leaving qualification:

☐ elementary/ primary school ☐ secondary school without university entrance qualification ☐ secondary school with university entrance qualification ☐ University or/and Fachhochschule

Do you work in the field of Geography, in a related field or with spatial data (cartography, geoinformatics, spatial/landscape/city planning, forestry, astronomy, geology, geo science)?

☐ Yes. ☐ No.

Did you recognize any of the areas shown in the map section?

☐ Yes. ☐ No.

Did you find the instructions of the masking methods helpful?

☐ Yes. ☐ No.

12. Appendix F

This appendix shows form 6 with disclaimer 2 and the basemap of orthophotos.

Geomasking and human cognition

Date	31.07.2023
Author-Interviewer	Isabella Noll
MSc program	Geography
Department	Department for Geography and Regional Research

Consent and confidentiality

I consent to participate in this research survey and the following has been explained to me:

• The survey responses will be made publicly available for the master's thesis of Isabella Noll.

• The author will retain the responses for a period of 2 years for research verification purposes.

• All personal information of the participants will be held confidential and will not be made publicly available at any time.

• By signing you consent to the conditions given above.

Name:

Signature:

Reidentification of health geodata representing individuals

As part of my master's thesis, I am conducting an examination on confidentiality and reidentification of geographical health datasets, such as households or residential addresses of individuals, represented by discrete points on maps. The study area is Vienna.

To safeguard personal data of individuals with spatial reference, such as addresses and coordinates, which are represented as discrete points, I employed a technique known as "geomasking" in which the original locations are shifted to avoid knowing the actual position on the map.

In the upcoming exercise, you will be presented with a series of maps where points represent individuals of households or residential addresses. They have been masked for confidentiality purposes. The task is divided into two parts:

- 5 maps are masked by a method called street masking (Swanlund et al., 2020).
- 5 maps are masked by a method called donut masking (Hampton et al, 2010).

Each method will be described to you, once before the actual task. Note that points with a number have been masked with street masking, and points without a number have been masked with donut masking.

Points masked by donut

Points masked by street masking.

Your task

To complete the task, please follow the instructions in the images below.

Look at this section of a map. The red points are masked points representing an individual.

Do you think you can find the original location and hack the place? Choose one or more of the red points and mark it/them.

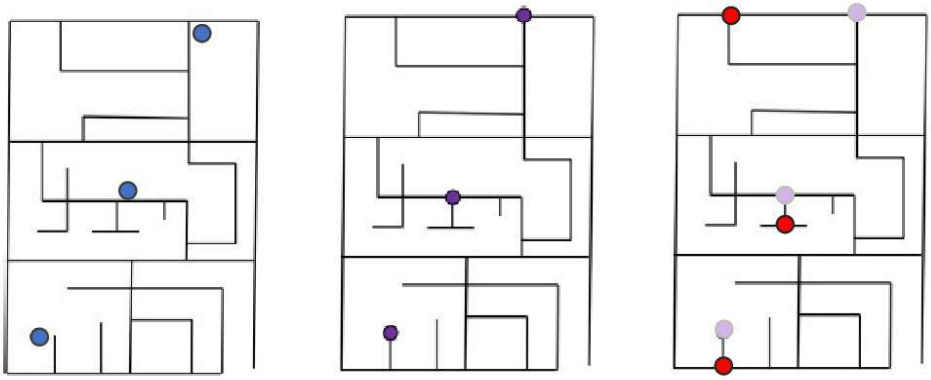
Make a cross where you think the original point is situated. Connect your chosen point with the cross.

Please keep in mind that the original points are just on locations, where there is actual residential use. There may be more than one individual in one housing. If you think that this is the case, please mark the masked points and estimate the location of origin by marking it on the map as well as shown below.

The points with a number represent several individuals, which are moved to the same location. Please pick a point by marking it. Try to figure out the location of origin by marking it on the map as well. If the point represents more than one individuum, please feel free to mark as many locations as are shown by the number on the point. For example, if the point has the number two, then you may mark a maximum of two locations of origin on your map as shown below.

Disclaimer of the masking method - Street masking

The method used to mask the points of the following five maps is known as street masking (Swanlund et al., 2020). As the name describes, the points are shifted from the residential place to the next intersection or the next end of a dead-end street of the street network (=nodes). This will be the starting node for the next step, in which two close nodes of the network are chosen. The distances from each of the two close nodes to the starting node are calculated. These two distances are summed up and divided by three, to calculate the average distance of the 3 nodes (starting node and the two close nodes). The point is masked by shifting it to one of the two close nodes that is closest to the average distance.



Figures modified from Swanlund et al., 2020

The unmasked points (blue) in their original location.

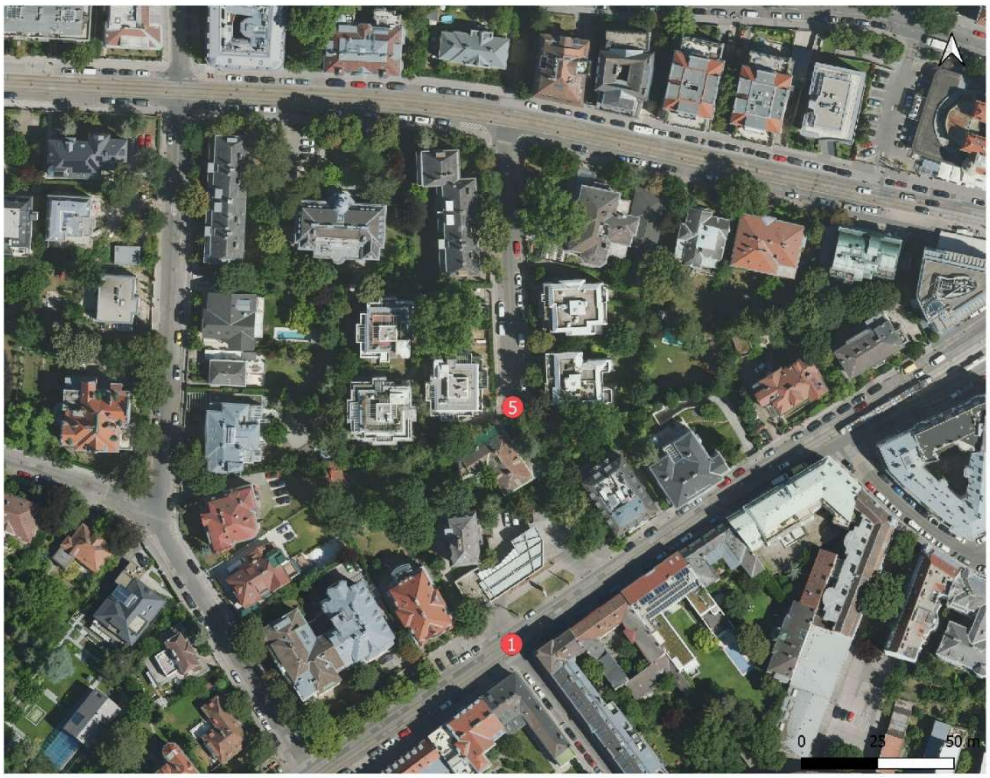
In the next step they are moved to the closest node (=street intersection or dead-end).

Finally, they are moved to another node depending on the average distance of two close nodes and the closest node from the image to the left. The masked points are shown in red.

1. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map A - street masking



Legend

1 geomasked points by street masking

Geoland basemap orthophoto

Data Source
Orthophoto of Vienna | Stadt Wien - https://data.wien.gv.at
Dataset of individuals | own work, masked by street masking (Swanlund et al., 2020)

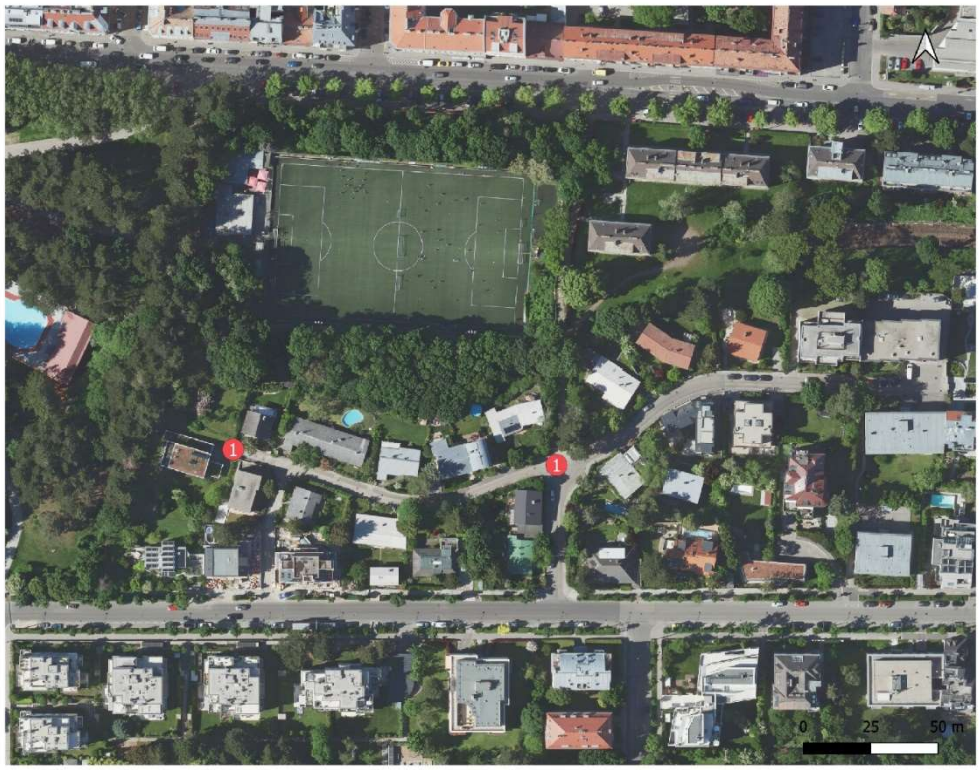
How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

2. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map B - street masking



Legend

1 geomasked points by street masking

Geoland basemap orthophoto

Data Source

Orthophoto map of Vienna | Stadt Wien- https://data.wien.gv.at

Dataset of individuals | own work, masked by street masking (Swanlund et al., 2020)

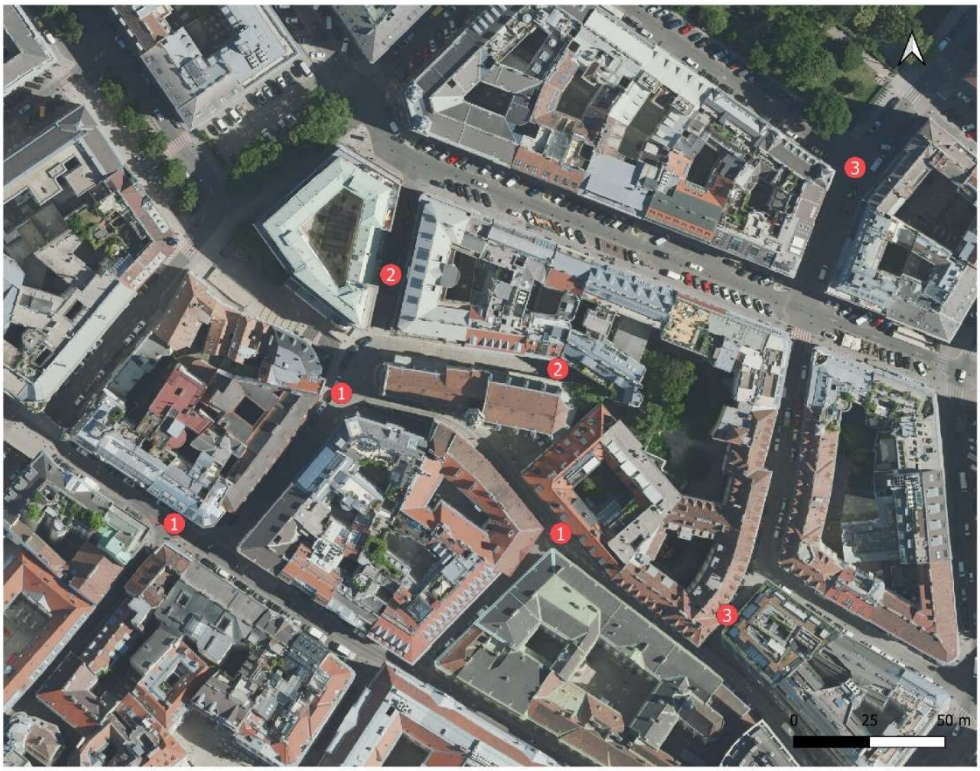
How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

3. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map C - street masking



Legend

1 geomasked points by street masking

Geoland basemap orthophoto

Data Source

Orthophoto of Vienna | Stadt Wien- https://data.wien.gv.at

Dataset of individuals | own work, masked by street masking (Swanlund et al., 2020)

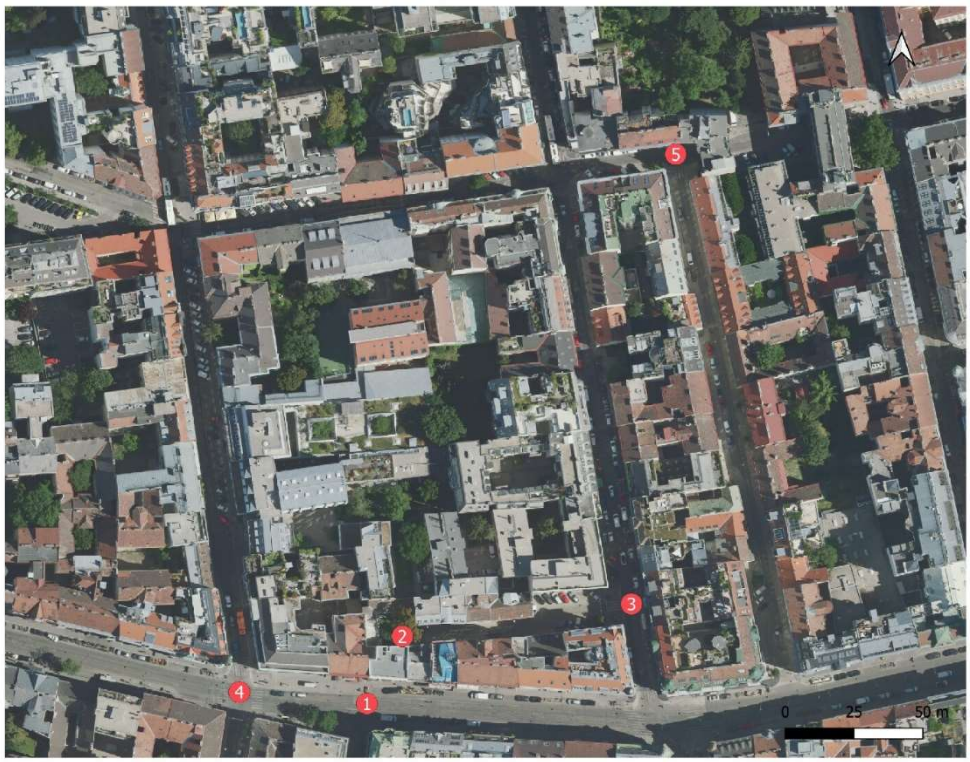
How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

4. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map D - street masking



How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

5. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map E - street masking

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

Disclaimer of the masking method - Donut masking

The method, which is used to mask the points in the following five maps, is called donut masking (Hampton et al., 2010). Donut masking moves the unmasked point from its original location to a random location within a donut-shaped area around the original position. The outer radius of the donut is the maximum shifting distance and is chosen so that only 14 other people are living inside. The inner radius is the minimum shifting distance and is 20 percent of the outer radius.

Our masked point is randomly placed somewhere in the donut shaped area around the original point, so that it cannot be distinguished from the other individuals inside the area.

Therefore, in a low population density area the point is shifted farther away than in a high population density area.

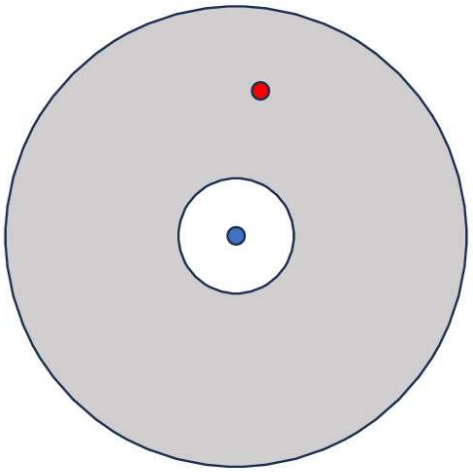


Figure based on Hampton et al., 2010

The blue, unmasked point is in the center of the donut. During the masking procedure it is moved into the shaded area (=red point).

With this method, there is always only one individual in the masked point. This means that there are no numbers on the masked points shown in the following maps.

6. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map F - donut masking

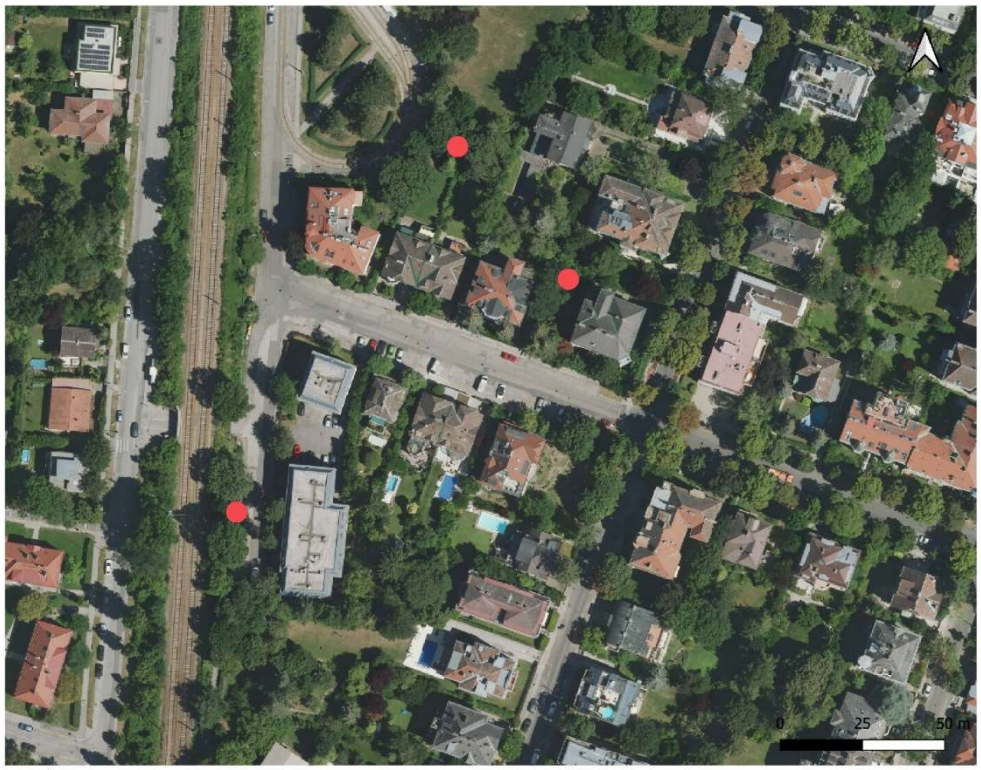
How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

7. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map G - donut masking



How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

8. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map H - donut masking

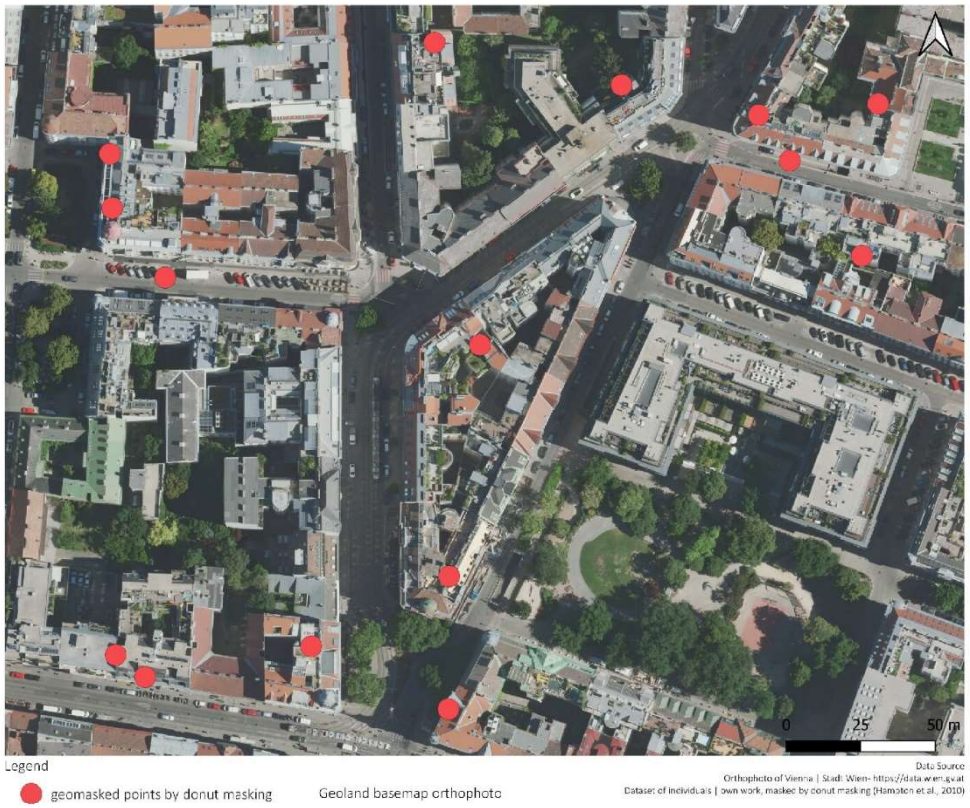
How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

9. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map I - donut masking



How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

10. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

Map J - donut masking

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

Demographics

Age:

☐ 0-14 ☐ 15-19 ☐ 20-34 ☐ 35-49 ☐ 50-64 ☐ 65 and older

Gender:

☐ male ☐ female ☐ other

School leaving qualification:

☐ elementary/ primary school ☐ secondary school without university entrance qualification ☐ secondary school with university entrance qualification ☐ University or/and Fachhochschule

Do you work in the field of Geography, in a related field or with spatial data (cartography, geoinformatics, spatial/landscape/city planning, forestry, astronomy, geology, geo science)?

☐ Yes. ☐ No.

Did you recognize any of the areas shown in the map section?

☐ Yes. ☐ No.

Did you find the instructions of the masking methods helpful?

☐ Yes. ☐ No.

13. Appendix G

Form 7 includes disclaimer 2 and building parcels as a basemap.

Geomasking and human cognition

Date	31.07.2023
Author-Interviewer	Isabella Noll
MSc program	Geography
Department	Department for Geography and Regional Research

Consent and confidentiality

I consent to participate in this research survey and the following has been explained to me:

- The survey responses will be made publicly available for the master's thesis of Isabella Noll.
- The author will retain the responses for a period of 2 years for research verification purposes.
- All personal information of the participants will be held confidential and will not be made publicly available at any time.
- By signing you consent to the conditions given above.

Name:

Signature:

Reidentification of health geodata representing individuals

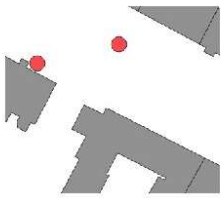
As part of my master's thesis, I am conducting an examination on confidentiality and reidentification of geographical health datasets, such as households or residential addresses of individuals, represented by discrete points on maps. The study area is Vienna.

To safeguard personal data of individuals with spatial reference, such as addresses and coordinates, which are represented as discrete points, I employed a technique known as "geomasking" in which the original locations are shifted to avoid knowing the actual position on the map.

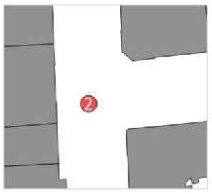
In the upcoming exercise, you will be presented with a series of maps where points represent individuals of households or residential addresses. They have been masked for confidentiality purposes. The task is divided into two parts:

- 5 maps are masked by a method called street masking (Swanlund et al., 2020).
- 5 maps are masked by a method called donut masking (Hampton et al, 2010).

Each method will be described to you, once before the actual task. Note that points with a number have been masked with street masking, and points without a number have been masked with donut masking.



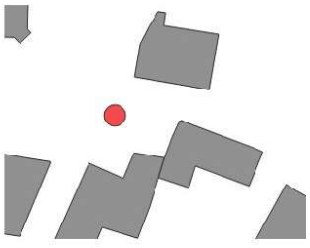
Points masked by donut



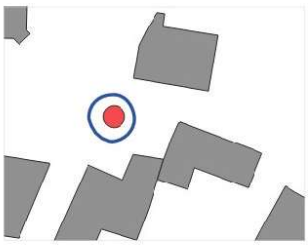
Points masked by street masking.

Your task

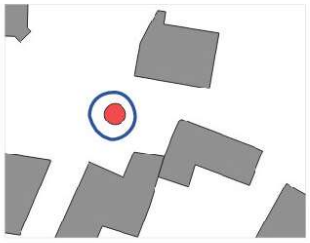
To complete the task, please follow the instructions in the images below.



Look at this section of a map. The red points are masked points representing an individual.

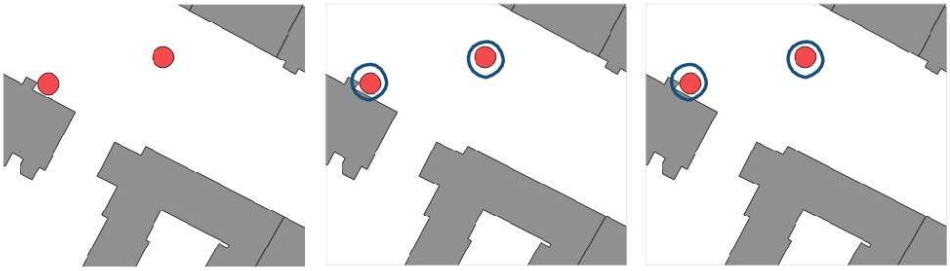


Do you think you can find the original location and hack the place? Choose one or more of the red points and mark it/them.

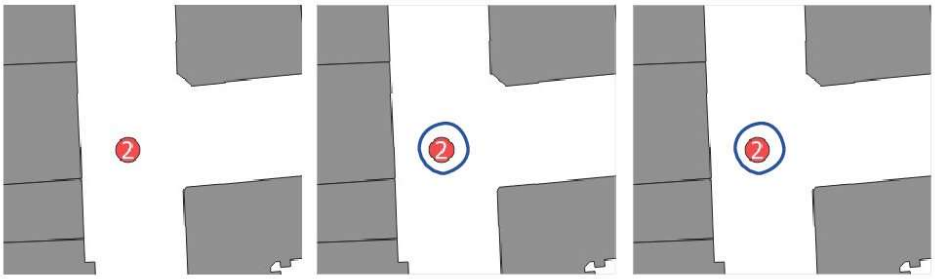


Make a cross where you think the original point is situated. Connect your chosen point with the cross.

Please keep in mind that the original points are just on locations, where there is actual residential use. There may be more than one individual in one housing. If you think that this is the case, please mark the masked points and estimate the location of origin by marking it on the map as well as shown below.

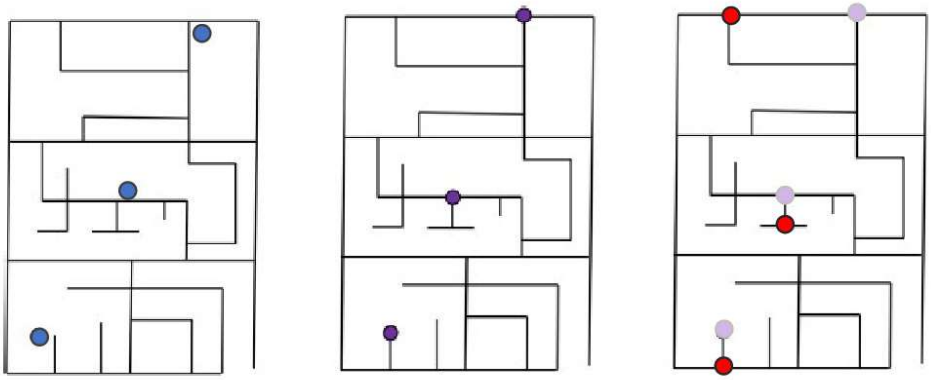


The points with a number represent several individuals, which are moved to the same location. Please pick a point by marking it. Try to figure out the location of origin by marking it on the map as well. If the point represents more than one individuum, please feel free to mark as many locations as are shown by the number on the point. For example, if the point has the number two, then you may mark a maximum of two locations of origin on your map as shown below.



Disclaimer of the masking method - Street masking

The method used to mask the points of the following five maps is known as street masking (Swanlund et al., 2020). As the name describes, the points are shifted from the residential place to the next intersection or the next end of a dead-end street of the street network (=nodes). This will be the starting node for the next step, in which two close nodes of the network are chosen. The distances from each of the two close nodes to the starting node are calculated. These two distances are summed up and divided by three, to calculate the average distance of the 3 nodes (starting node and the two close nodes). The point is masked by shifting it to one of the two close nodes that is closest to the average distance.



Figures modified from Swanlund et al., 2020

The unmasked points (blue) in their original location.

In the next step they are moved to the closest node (=street intersection or dead-end).

Finally, they are moved to another node depending on the average distance of two close nodes and the closest node from the image to the left. The masked points are shown in red.

1. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

2. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

3. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

4. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

5. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

Disclaimer of the masking method - Donut masking

The method, which is used to mask the points in the following five maps, is called donut masking (Hampton et al., 2010). Donut masking moves the unmasked point from its original location to a random location within a donut-shaped area around the original position. The outer radius of the donut is the maximum shifting distance and is chosen so that only 14 other people are living inside. The inner radius is the minimum shifting distance and is 20 percent of the outer radius.

Our masked point is randomly placed somewhere in the donut shaped area around the original point, so that it cannot be distinguished from the other individuals inside the area.

Therefore, in a low population density area the point is shifted farther away than in a high population density area.

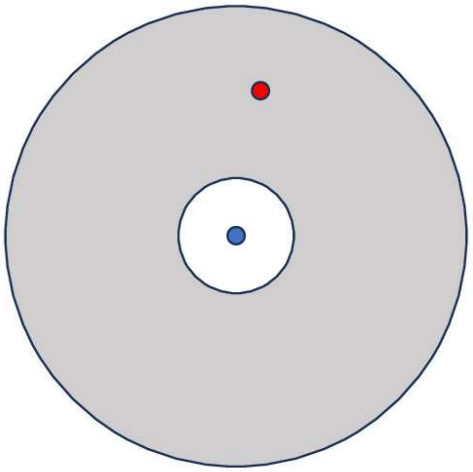


Figure based on Hampton et al., 2010

The blue, unmasked point is in the center of the donut. During the masking procedure it is moved into the shaded area (=red point).

With this method, there is always only one individual in the masked point. This means that there are no numbers on the masked points shown in the following maps.

6. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

7. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

8. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

9. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

10. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

Demographics

Age:

☐ 0-14 ☐ 15-19 ☐ 20-34 ☐ 35-49 ☐ 50-64 ☐ 65 and older

Gender:

☐ male ☐ female ☐ other

School leaving qualification:

☐ elementary/ primary school ☐ secondary school without university entrance qualification ☐ secondary school with university entrance qualification ☐ University or/and Fachhochschule

Do you work in the field of Geography, in a related field or with spatial data (cartography, geoinformatics, spatial/landscape/city planning, forestry, astronomy, geology, geo science)?

☐ Yes. ☐ No.

Did you recognize any of the areas shown in the map section?

☐ Yes. ☐ No.

Did you find the instructions of the masking methods helpful?

☐ Yes. ☐ No.

14. Appendix H

This is form 8 with disclaimer 2 and the intermodal transport reference system as a basemap.

Geomasking and human cognition

Date	31.07.2023
Author-Interviewer	Isabella Noll
MSc program	Geography
Department	Department for Geography and Regional Research

Consent and confidentiality

I consent to participate in this research survey and the following has been explained to me:

• The survey responses will be made publicly available for the master's thesis of Isabella Noll.

• The author will retain the responses for a period of 2 years for research verification purposes.

• All personal information of the participants will be held confidential and will not be made publicly available at any time.

• By signing you consent to the conditions given above.

Name:

Signature:

Reidentification of health geodata representing individuals

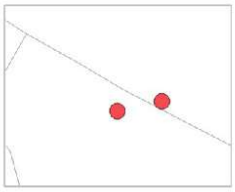
As part of my master's thesis, I am conducting an examination on confidentiality and reidentification of geographical health datasets, such as households or residential addresses of individuals, represented by discrete points on maps. The study area is Vienna.

To safeguard personal data of individuals with spatial reference, such as addresses and coordinates, which are represented as discrete points, I employed a technique known as "geomasking" in which the original locations are shifted to avoid knowing the actual position on the map.

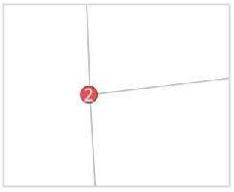
In the upcoming exercise, you will be presented with a series of maps where points represent individuals of households or residential addresses. They have been masked for confidentiality purposes. The task is divided into two parts:

- 5 maps are masked by a method called street masking (Swanlund et al., 2020).
- 5 maps are masked by a method called donut masking (Hampton et al, 2010).

Each method will be described to you, once before the actual task. Note that points with a number have been masked with street masking, and points without a number have been masked with donut masking.



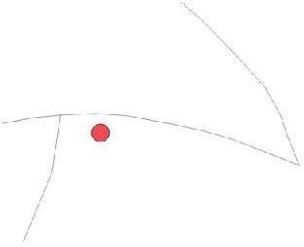
Points masked by donut



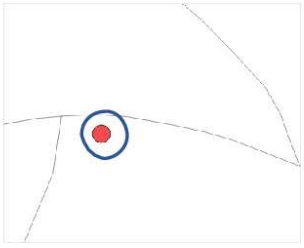
Points masked by street masking.

Your task

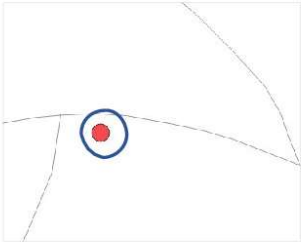
To complete the task, please follow the instructions in the images below.



Look at this section of a map. The red points are masked points representing an individual.

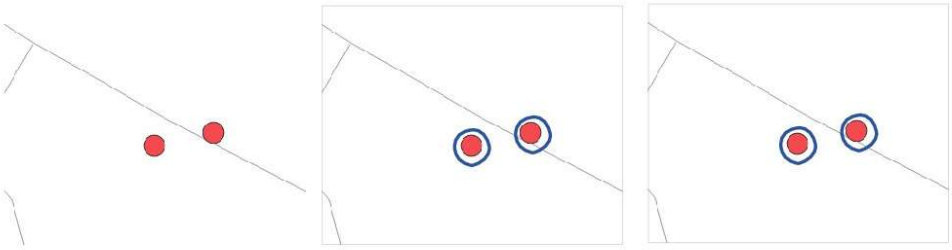


Do you think you can find the original location and hack the place? Choose one or more of the red points and mark it/them.

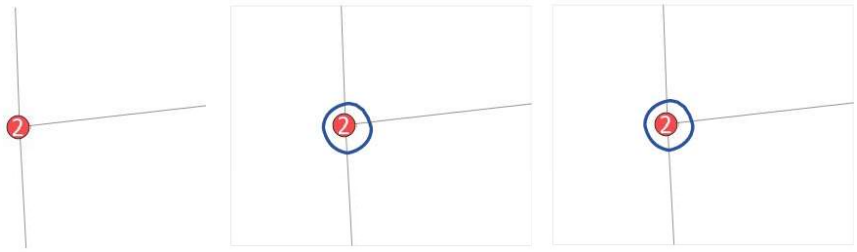


Make a cross where you think the original point is situated. Connect your chosen point with the cross.

Please keep in mind that the original points are just on locations, where there is actual residential use. There may be more than one individual in one housing. If you think that this is the case, please mark the masked points and estimate the location of origin by marking it on the map as well as shown below.

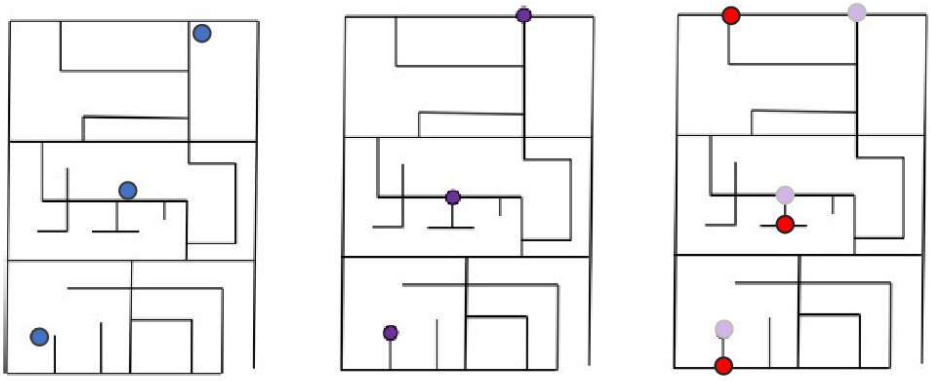


The points with a number represent several individuals, which are moved to the same location. Please pick a point by marking it. Try to figure out the location of origin by marking it on the map as well. If the point represents more than one individuum, please feel free to mark as many locations as are shown by the number on the point. For example, if the point has the number two, then you may mark a maximum of two locations of origin on your map as shown below.



Disclaimer of the masking method - Street masking

The method used to mask the points of the following five maps is known as street masking (Swanlund et al., 2020). As the name describes, the points are shifted from the residential place to the next intersection or the next end of a dead-end street of the street network (=nodes). This will be the starting node for the next step, in which two close nodes of the network are chosen. The distances from each of the two close nodes to the starting node are calculated. These two distances are summed up and divided by three, to calculate the average distance of the 3 nodes (starting node and the two close nodes). The point is masked by shifting it to one of the two close nodes that is closest to the average distance.



Figures modified from Swanlund et al., 2020

The unmasked points (blue) in their original location.

In the next step they are moved to the closest node (=street intersection or dead-end).

Finally, they are moved to another node depending on the average distance of two close nodes and the closest node from the image to the left. The masked points are shown in red.

1. Map

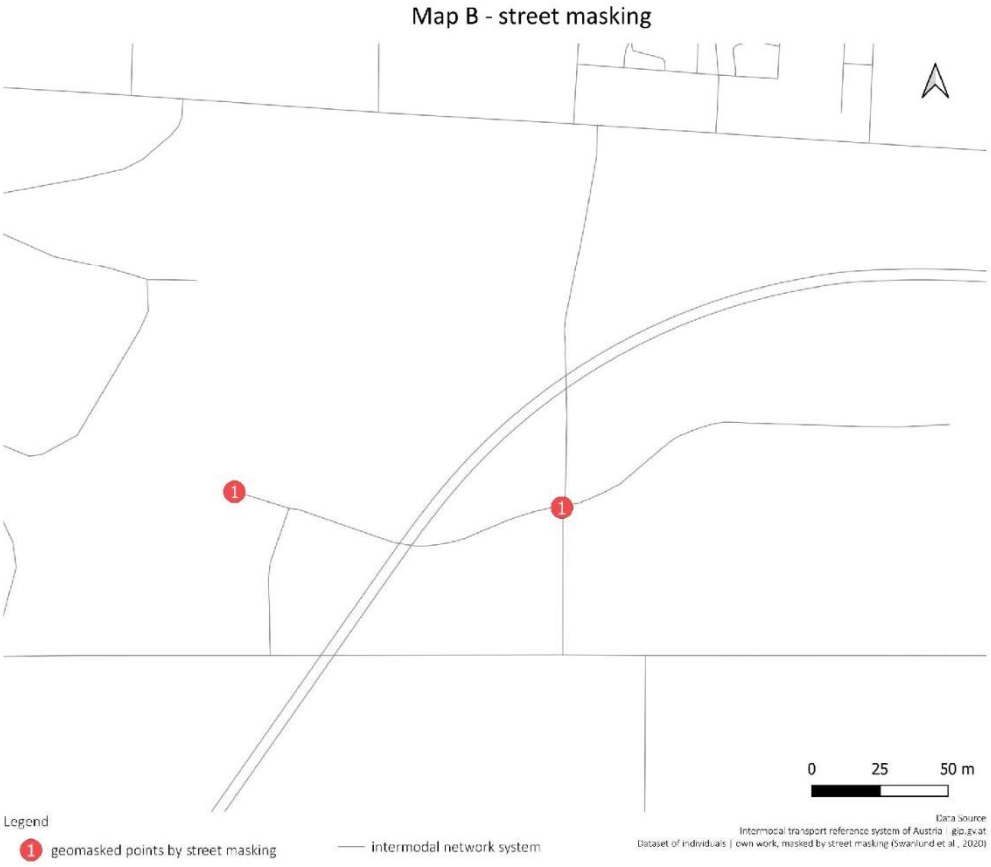
Please figure out the original position of the points you choose on the map below according to the instructions above:

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

2. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

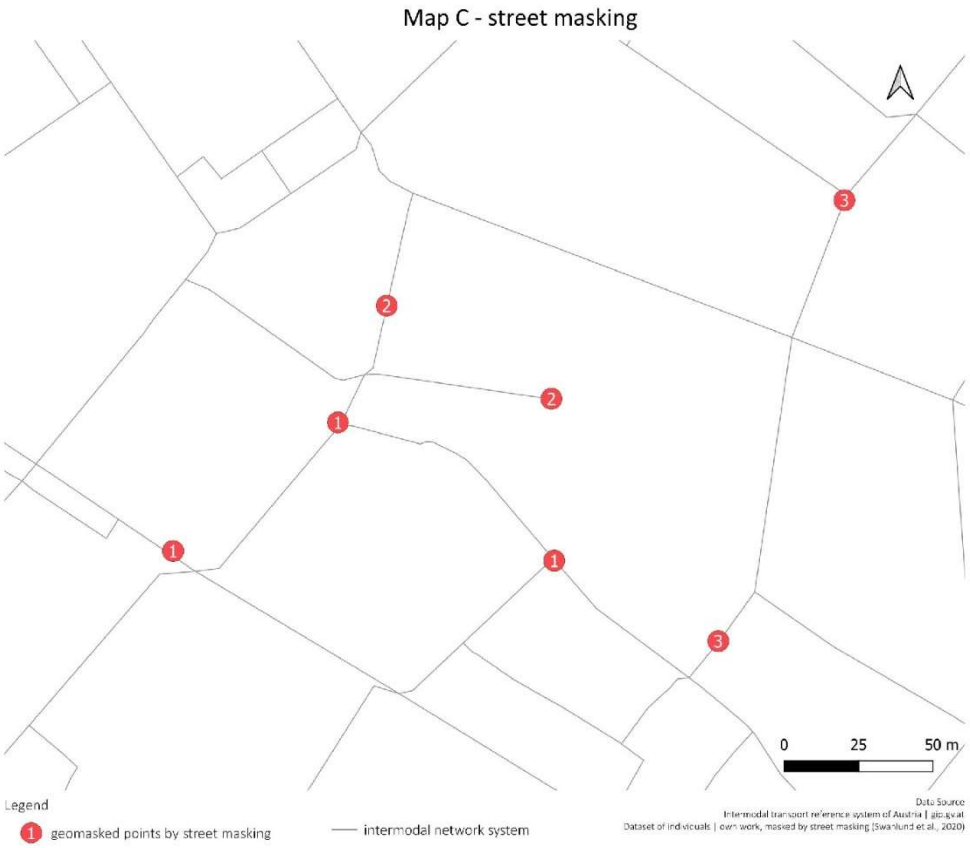


How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

3. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

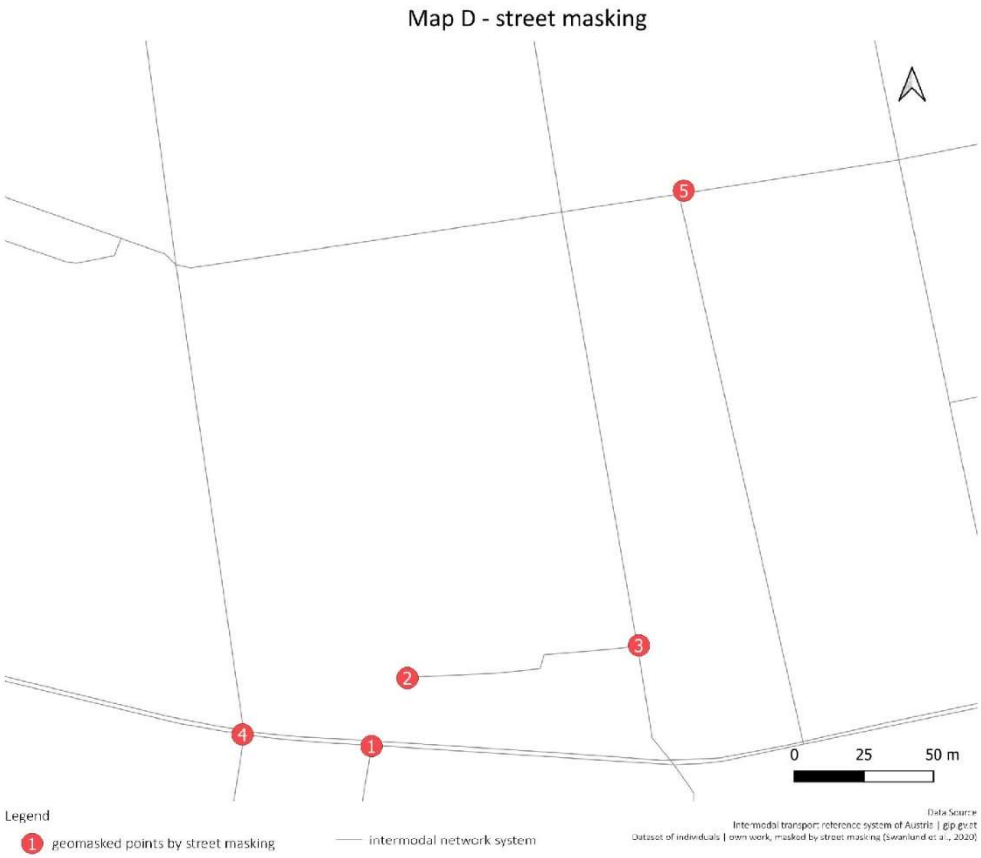


How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

4. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

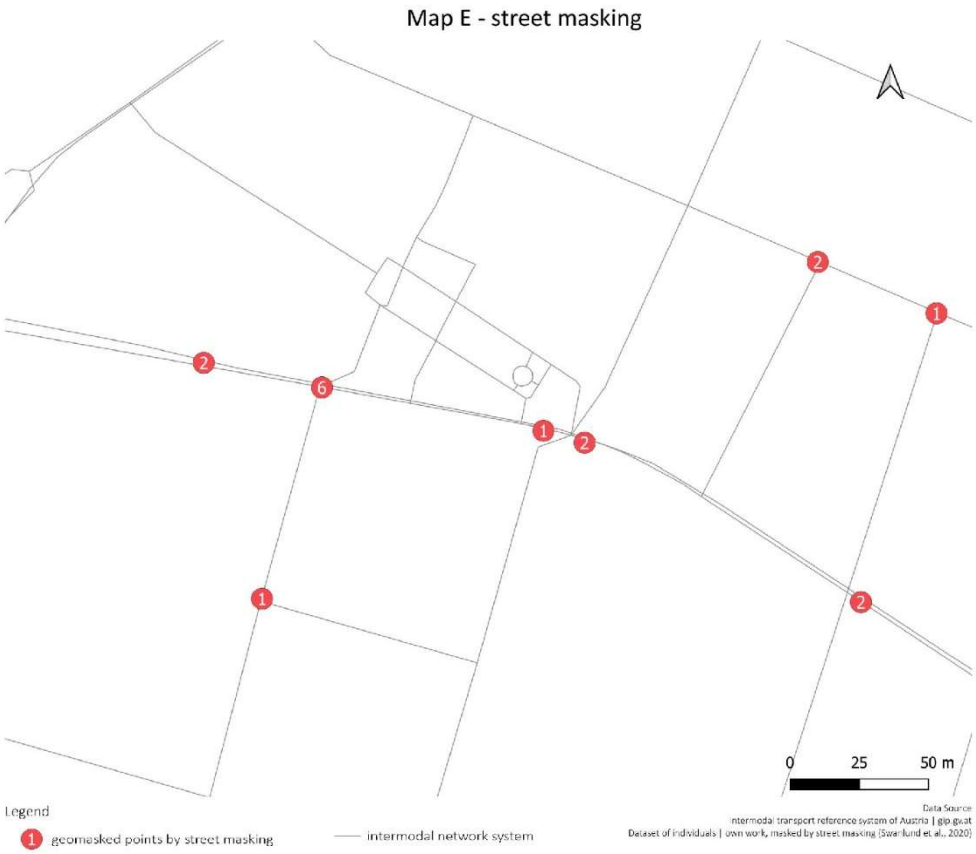


How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

5. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:



How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

Disclaimer of the masking method - Donut masking

The method, which is used to mask the points in the following five maps, is called donut masking (Hampton et al., 2010). Donut masking moves the unmasked point from its original location to a random location within a donut-shaped area around the original position. The outer radius of the donut is the maximum shifting distance and is chosen so that only 14 other people are living inside. The inner radius is the minimum shifting distance and is 20 percent of the outer radius.

Our masked point is randomly placed somewhere in the donut shaped area around the original point, so that it cannot be distinguished from the other individuals inside the area.

Therefore, in a low population density area the point is shifted farther away than in a high population density area.

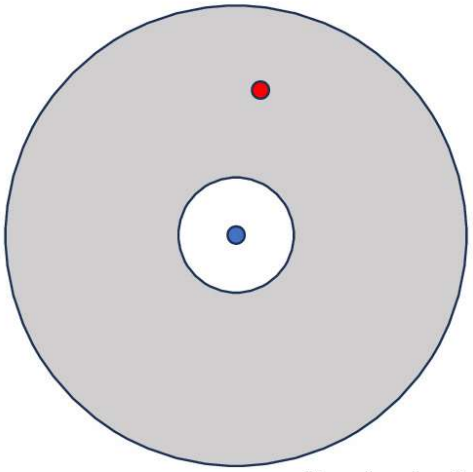


Figure based on Hampton et al., 2010

The blue, unmasked point is in the center of the donut. During the masking procedure it is moved into the shaded area (=red point).

With this method, there is always only one individual in the masked point. This means that there are no numbers on the masked points shown in the following maps.

6. Map

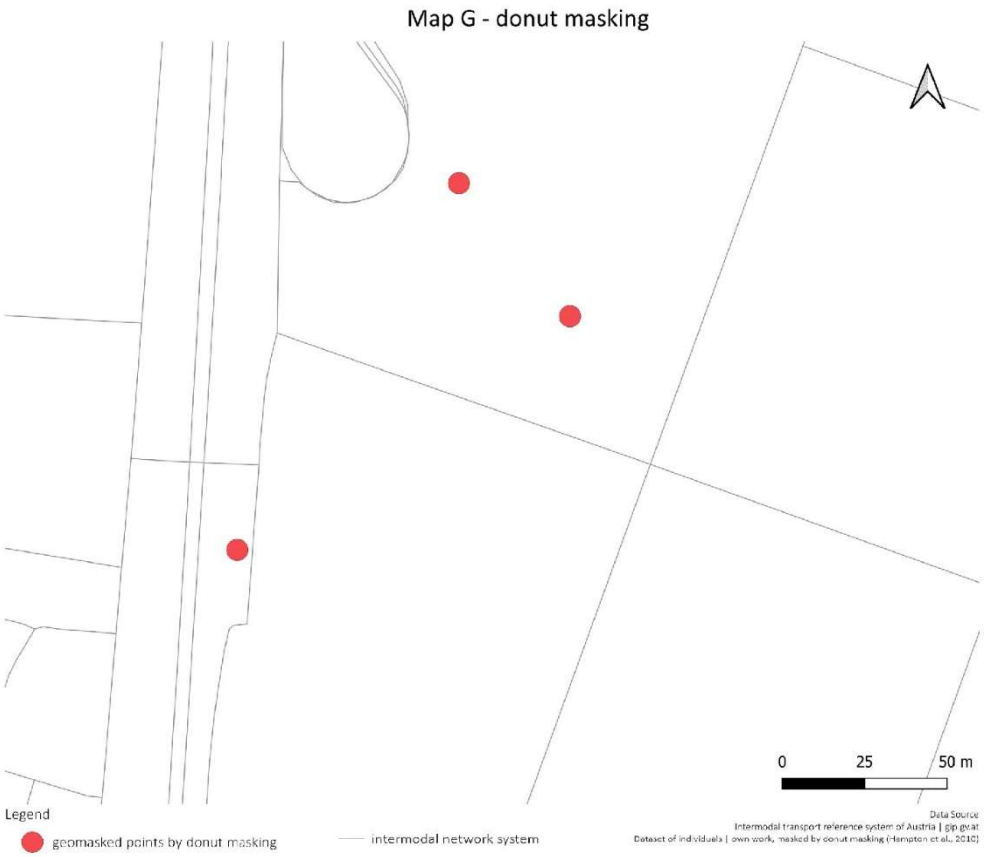
Please figure out the original position of the points you choose on the map below according to the instructions above:

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

7. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:



How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

8. Map

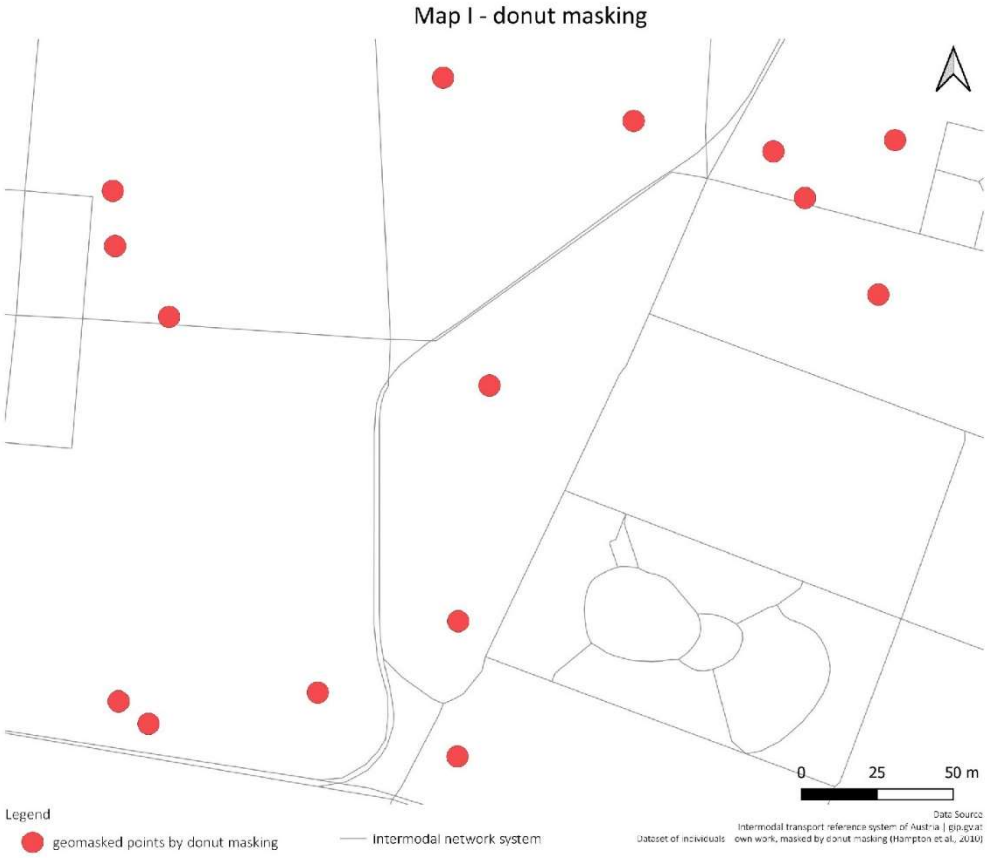
Please figure out the original position of the points you choose on the map below according to the instructions above:

How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

9. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:

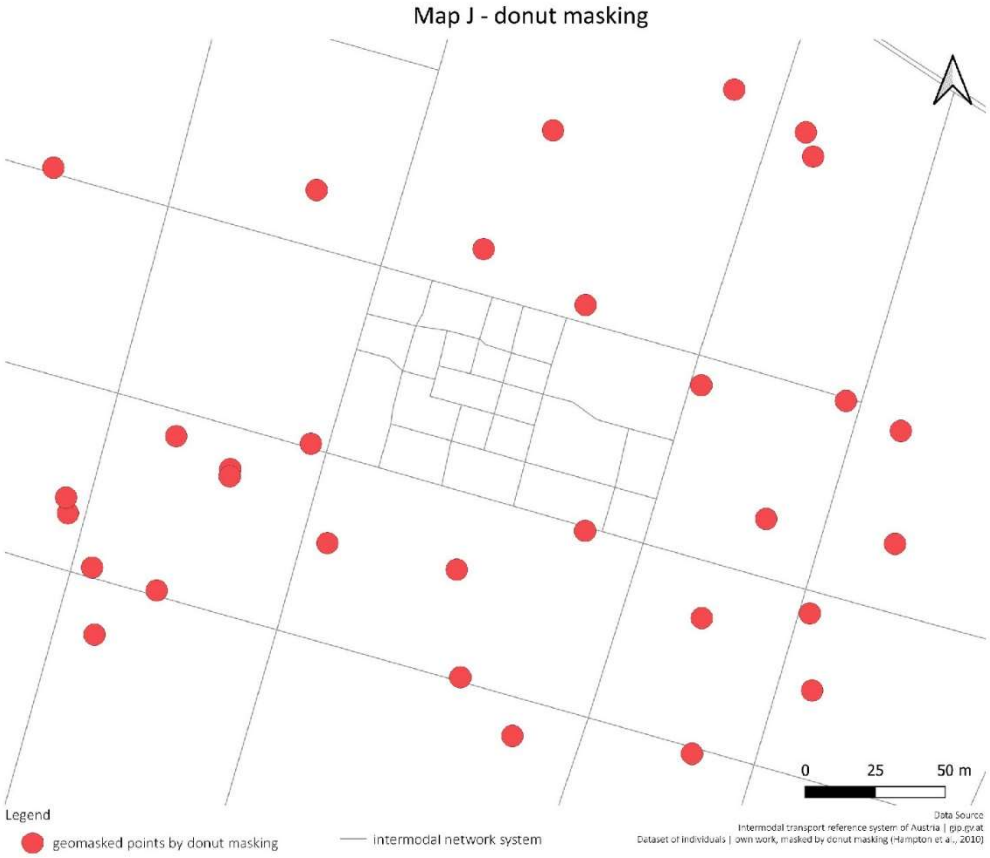


How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

10. Map

Please figure out the original position of the points you choose on the map below according to the instructions above:



How confident did you feel in figuring out the original position?

☐ completely confident ☐ fairly confident ☐ somewhat confident ☐ slightly confident ☐ not at all confident

Demographics

Age:

☐ 0-14 ☐ 15-19 ☐ 20-34 ☐ 35-49 ☐ 50-64 ☐ 65 and older

Gender:

☐ male ☐ female ☐ other

School leaving qualification:

☐ elementary/ primary school ☐ secondary school without university entrance qualification ☐ secondary school with university entrance qualification ☐ University or/and Fachhochschule

Do you work in the field of Geography, in a related field or with spatial data (cartography, geoinformatics, spatial/landscape/city planning, forestry, astronomy, geology, geo science)?

☐ Yes. ☐ No.

Did you recognize any of the areas shown in the map section?

☐ Yes. ☐ No.

Did you find the instructions of the masking methods helpful?

☐ Yes. ☐ No.

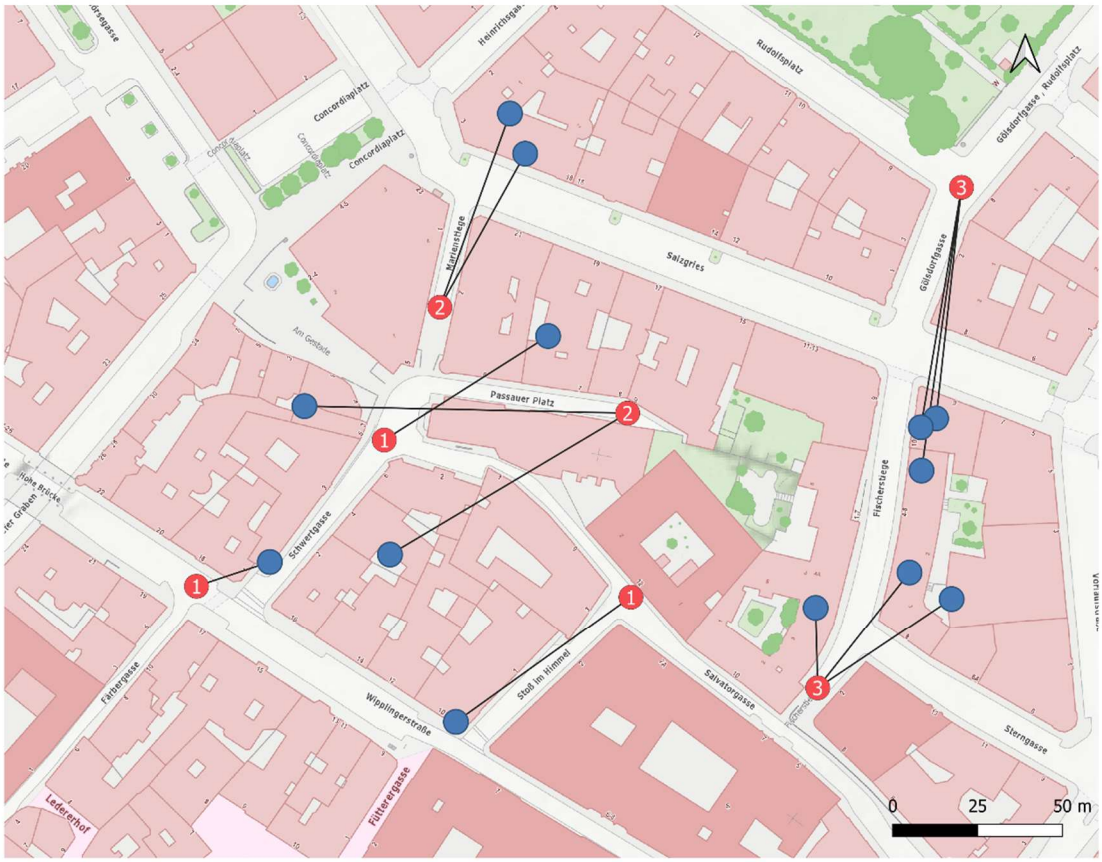
15. Appendix I

This appendix includes all maps with the original and masked points for the survey.

Map A - street masking

Map B - street masking

Map C - street masking



Legend

1 geomasked points by street masking

unmasked points

connection

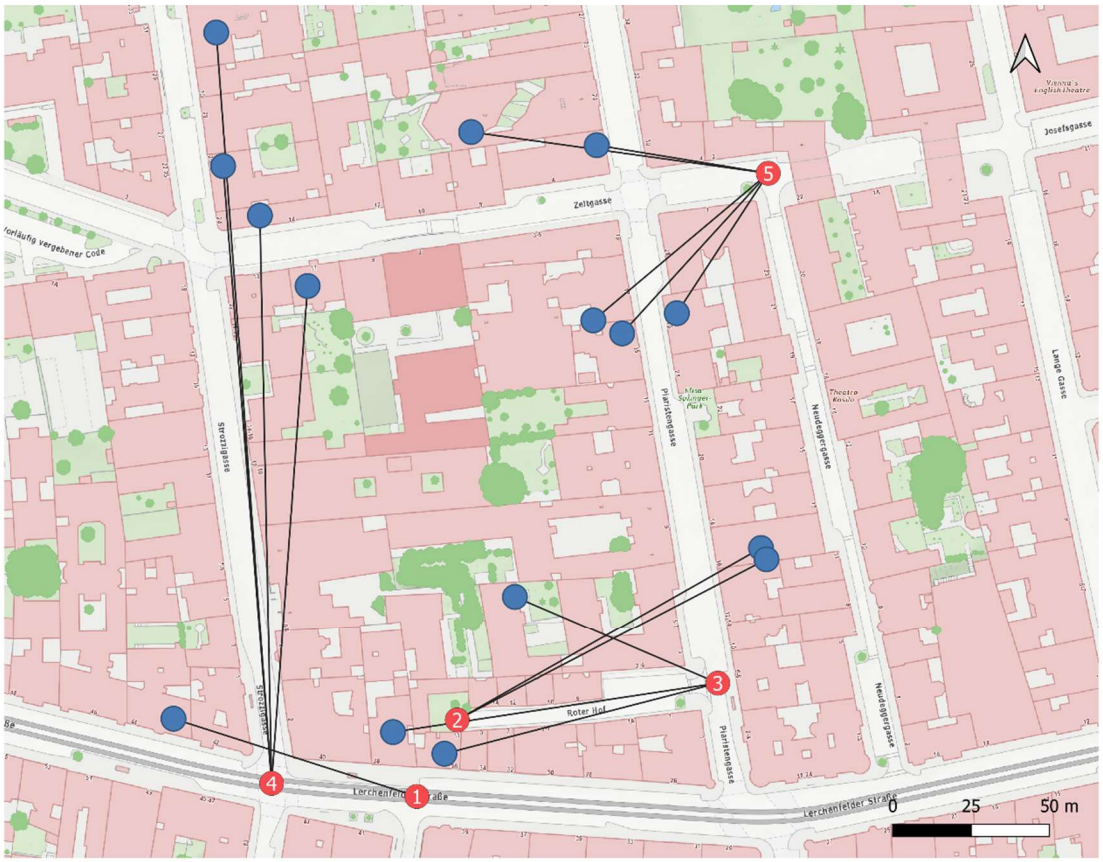
Geoland basemap color

Data Source

Raster data of the multi-purpose map of Vienna | Stadt Wien- https://data.wien.gv.at

Dataset of individuals | own work, masked by street masking (Swanlund et al., 2020)

Map D - street masking



Legend

1 geomasked points by street masking

unmasked points

connection

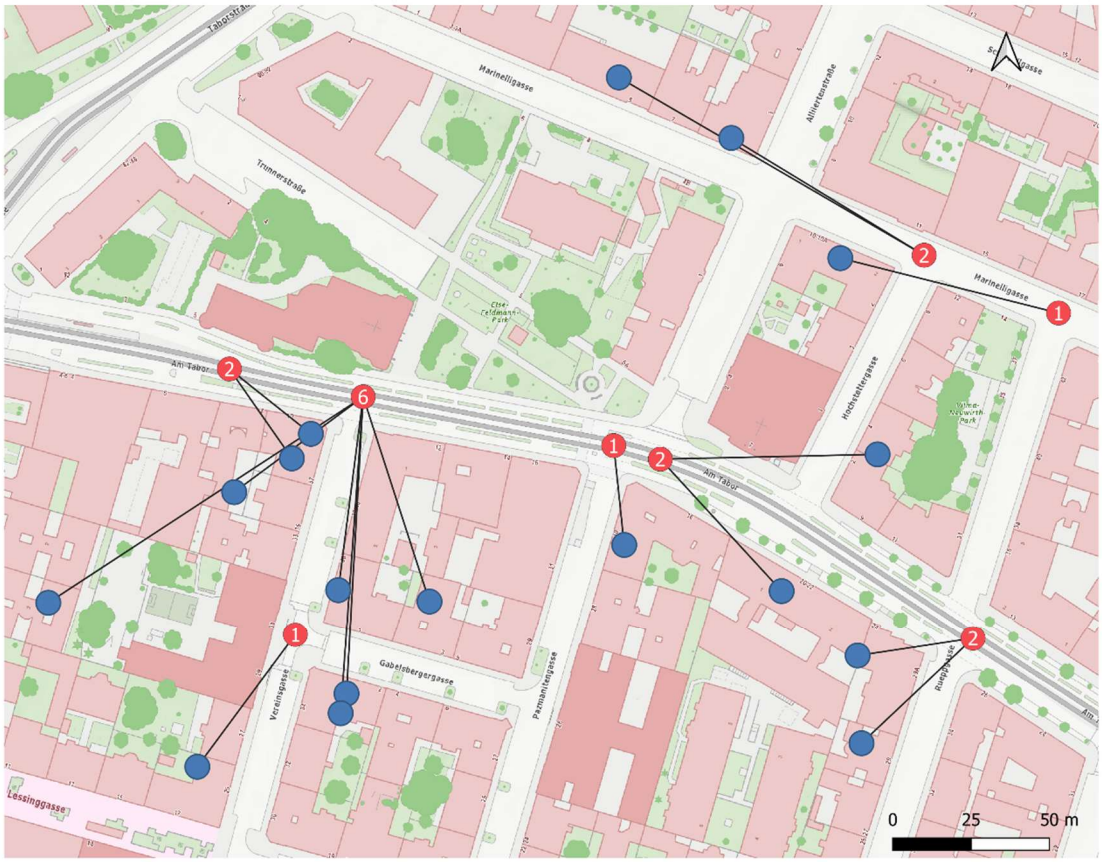
Geoland basemap color

Data Source

Raster data of the multi-purpose map of Vienna | Stadt Wien- https://data.wien.gv.at

Dataset of individuals | own work, masked by street masking (Swanlund et al., 2020)

Map E - street masking



Legend

1 geomasked points by street masking

unmasked points

connectionE

Geoland basemap color

Data Source

Raster data of the multi-purpose map of Vienna | Stadt Wien- https://data.wien.gv.at

Dataset of individuals | own work, masked by street masking (Swanlund et al., 2020)

Map F - donut masking

Map G - donut masking

Legend

geomasked points by donut masking

unmasked points

connection

Geoland basemap color

Data Source

Raster data of the multi-purpose map of Vienna | Stadt Wien- https://data.wien.gv.at

Dataset of individuals | own work, masked by donut masking (Hampton et al., 2010)

Map H - donut masking

Legend

geomasked points by donut masking

unmasked points

connection

Geoland basemap color

Data Source

Raster data of the multi-purpose map of Vienna | Stadt Wien- https://data.wien.gv.at

Dataset of individuals | own work, masked by donut masking (Hampton et al., 2010)

Map I - donut masking

Legend

- geomasked points by donut masking
- unmasked points
- connection

Geoland basemap color

Data Source

Raster data of the multi-purpose map of Vienna | Stadt Wien- https://data.wien.gv.at
Dataset of individuals | own work, masked by donut masking (Hampton et al., 2010)

Map J - donut masking

Legend

- geomasked points by donut masking
- unmasked points
- connection

Geoland basemap color

Data Source

Raster data of the multi-purpose map of Vienna | Stadt Wien- https://data.wien.gv.at
Dataset of individuals | own work, masked by donut masking (Hampton et al., 2010)