



How do people perceive the disclosure risk of maps? Examining the perceived disclosure risk of maps and its implications for geoprivacy protection

Junghwan Kim, Mei-Po Kwan, Margaret C. Levenstein & Douglas B. Richardson

To cite this article: Junghwan Kim, Mei-Po Kwan, Margaret C. Levenstein & Douglas B. Richardson (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, DOI: [10.1080/15230406.2020.1794976](https://doi.org/10.1080/15230406.2020.1794976)

To link to this article: <https://doi.org/10.1080/15230406.2020.1794976>



View supplementary material [↗](#)



Published online: 24 Aug 2020.



Submit your article to this journal [↗](#)



Article views: 1754



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 3 View citing articles [↗](#)



How do people perceive the disclosure risk of maps? Examining the perceived disclosure risk of maps and its implications for geoprivacy protection

Junghwan Kim^a, Mei-Po Kwan^{b,c}, Margaret C. Levenstein^d and Douglas B. Richardson^e

^aDepartment of Geography and Geographic Information Science, University of Illinois at Urbana-Champaign, Urbana, IL, USA; ^bDepartment of Geography and Resource Management, Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Hong Kong, China; ^cDepartment of Human Geography and Spatial Planning, Utrecht University, Utrecht, The Netherlands; ^dInstitute for Social Research and School of Information, University of Michigan, Ann Arbor, MI, USA; ^eCenter for Geographic Analysis, Institute for Quantitative Social Science, Harvard University, Cambridge, MA, USA

ABSTRACT

This research examines how people subjectively perceive the disclosure risk of a map using original data collected in an online survey with 856 participants. The results indicate that perceived disclosure risk increases as the amount of locational information displayed on a map increases. Compared to point-based maps, perceived disclosure risk is significantly lower for kernel density maps, convex hull maps, and standard deviational ellipse maps. The results also revealed that perceived disclosure risk is affected by map scale and the presence of information of other people on a map. For geomasking methods, perceived disclosure risk decreases as aggregation level increases and as relocation distance increases. However, aggregation methods (point to polygon) are more effective in preventing the re-identification of individuals when compared to relocation methods (point to point). Lastly, the perceived disclosure risk of a map that displays socially-vulnerable people is significantly higher than that of a map that displays non-vulnerable groups. Specifically, a map displaying the private locations of elementary school students has the highest perceived disclosure risk. Based on the results, a set of geoprivacy protection guidelines for mapping people's private locations to minimize people's perceived disclosure risk is proposed. Implications for mapping infectious diseases like the COVID-19 are also discussed.

ARTICLE HISTORY

Received 17 February 2020
Accepted 8 July 2020

KEYWORDS

Disclosure risk; geoprivacy; geomasking; perception; COVID-19; survey

Introduction

In recent years, advances in geospatial technologies and GIScience methods have allowed researchers to analyze and visualize geospatial data in great detail (Gutmann et al., 2008; Kwan, 2012; Richardson et al., 2013). For instance, maps have been widely utilized to visualize the complex patterns of diverse social phenomena, such as infectious diseases and human daily mobility (e.g., Kim & Kwan, 2019, 2020; Kim & Lee, 2019; Kwan, 2004; Reich & Haran, 2018). At the same time, however, mapping people's confidential geospatial data (e.g., people's homes, workplaces, or GPS trajectories) may allow an individual's identity (e.g., name and street address) to be identified from a map (A.J. Curtis et al., 2006; Brownstein et al., 2006; A. Curtis et al., 2011; Kounadi & Leitner, 2014; Kounadi & Resch, 2018; VanWey et al., 2005). Therefore, when mapping individual confidential geospatial data, it is crucial to accurately assess the disclosure risk of the map to protect people's privacy from being violated.

Although previous studies have evaluated the disclosure risk of maps (i.e., probability of reidentification),

little is known about how people subjectively perceive disclosure risk when their confidential locations are displayed on maps. In addition to evaluating the risk of identifying a person through spatial reverse engineering (A.J. Curtis et al., 2006; Brownstein et al., 2006; A. Curtis et al., 2011), assessing perceived disclosure risk is also important. This is because it can inform researchers to establish geoprivacy protection guidelines for mapping people's private locations to avoid privacy violations, which cannot be done without obtaining people's perceived disclosure risk since the concept of privacy is socially and culturally constructed (Armstrong & Ruggles, 2005; McLafferty, 2004).

Therefore, to fill this gap, this research examines how people subjectively perceive the disclosure risk of a map using original data collected in a survey ($n = 856$). Specifically, we ask the following three questions: (1) How do different components of a map affect an individual's perceived disclosure risk of a map? (2) How do different geomasking methods affect an individual's perceived disclosure risk of a map? (3) How does an individual perceive

the disclosure risk of a map when the private locations of socially-vulnerable people are visualized?

The results of the study reveal that an individual's perceived disclosure risk is significantly affected by four attributes of a map (i.e., the mapping method, the amount of private locational information displayed on a map, the map scale, and the presence of the information of other people on the map), the application of different geomasking methods, and the presence of socially vulnerable groups on a map. Based on the results, a set of geoprivacy protection guidelines for mapping people's private locations to minimize their perceived disclosure risk is proposed. Moreover, implications for mapping infectious diseases such as COVID-19 are also discussed, with a focus on balancing the need for disease control (by releasing geospatial information about infected persons) and individual geoprivacy protection (by geomasking information).

Background: disclosure risk of a map and privacy violation

Mapping is a useful tool for visualizing detailed geospatial data. But it may violate people's geoprivacy through the process of spatial reverse engineering when their private locations such as homes, workplaces and GPS trajectories are displayed (and thus disclosed) on maps (A.J. Curtis et al., 2006; Brownstein et al., 2006; A. Curtis et al., 2011; Gutmann et al., 2008; Kounadi & Leitner, 2014; Kounadi & Resch, 2018; VanWey et al., 2005). For example, imagine a map that displays the home locations of survey participants to illustrate the point patterns (e.g., whether the locations are clustered or not) (Haley et al., 2016; Kounadi & Leitner, 2014). Due to human subjects protection and related requirements (see The Federal Policy for the Protection of Human Subjects, 45 C.F.R. 46, 1991, also known as "The Common Rule"), responsible researchers would be cautious about publishing maps that explicitly label each point with participants' name and home address. Instead, a researcher publishes a point-based map without labeling. Some may think this point-based map has low privacy disclosure risk because it does not explicitly label people's specific identity (e.g., name and street address). However, this may not be the case because it is possible to accurately estimate the street address of a person from the map and discover the true identity of each person by linking the street address to publicly available data (e.g., white pages or voter lists) (A.J. Curtis et al., 2006; Brownstein et al., 2006; A. Curtis et al., 2011; Gutmann et al., 2008; Kounadi & Leitner, 2014; Kounadi & Resch, 2018; VanWey et al., 2005).

The disclosure of people's private locations through mapping is a serious problem because it may violate personal privacy and data confidentiality requirements. Privacy violation is often considered to be illegal (e.g., the violation of the Health Insurance Portability and Accountability Act [HIPAA] in the United States) (Kar et al., 2013; Kwan et al., 2004). Readers may refer to Kar et al. (2013) for a comprehensive summary of the legal and policy aspects of the disclosure of people's private locations. Moreover, privacy violation is often considered to be unethical as it may cause unintended consequences (e.g., robbery or discrimination) to people whose private locations are disclosed (Kwan et al., 2004). Further, the violation of privacy may discourage people to participate in surveys because people may not be willing to share their private information with researchers when realizing it is possible for their privacy to be violated (Gutmann et al., 2008; Singer, 1978).

Previous studies have assessed the disclosure risk of maps (i.e., probability of reidentification) to help prevent privacy violations. Specifically, studies have shown that, through spatial reverse engineering, an individual's identity may be uncovered even though a point-based map does not explicitly indicate it. For example, Brownstein et al. (2006) showed that a street address of an individual whose home location is displayed as a point in a map may be accurately reverse engineered from the map regardless of its resolution. Moreover, A.J. Curtis et al. (2006) concluded that the accuracy of the spatial reverse engineering of a point-based map depends on the urban form (e.g., the housing pattern on a street).

Furthermore, researchers have evaluated how geomasking methods can reduce the disclosure risk of maps. Geomasking methods refer to methods that deliberately introduce errors to the original points on maps to reduce disclosure risk (Armstrong et al., 1999; Hampton et al., 2010; Kwan et al., 2004; Zandbergen, 2014; Zhang et al., 2017). In general, there are two main geomasking methods: aggregation (point to polygon) and relocation methods (point to point, such as affine transformation and random perturbation). Note that in addition to geomasking methods, there are other methods for reducing the disclosure risk of an individual in a database (e.g., differential privacy) (Abowd & Schmutte, 2019). However, since the primary focus of this paper is on assessing disclosure risk through mapping, these other methods are not discussed in this paper.

First, in aggregation methods (point to polygon), individual locations are aggregated to a polygon that contains a certain number of people rather than displaying the original points. An example is a map showing the total number of cancer patients at the census tract

level instead of displaying their home locations as individual points. Different aggregation levels are used in past studies, but the levels between 1:20 and 1:20,000 are typically used (A. Curtis et al., 2011; VanWey et al., 2005). For instance, the aggregation level of 1:20 means that an individual's private location is displayed in a polygon that also includes 19 other persons, indicating that an individual can be identified with a probability of 1 in 20 (5%). Previous studies have examined how disclosure risk is affected by different aggregation levels. For example, A. Curtis et al. (2011) investigated how the aggregation level affects disclosure risk and concluded that a map at a scale of 0.5 km grid can be sufficient to protect from the disclosure. Moreover, by using the 2001 Canadian census data, Emam et al. (2009) examined aggregation population thresholds (i.e., cutoffs) that contain a large number of people for lowering disclosure risk. They concluded that there is no single threshold for aggregation because the optimal threshold is affected by the characteristics of the variables in question.

Second, relocation methods (point to point) move the original points on a map to nearby new points at a random distance and/or direction. Studies have shown that parameters (e.g., radii) used in relocation methods affect disclosure risk. For example, Kwan et al. (2004) examined three geomasking methods with different perturbation radii (r) using a 1999 dataset of lung cancer deaths in Franklin County, Ohio, and observed a consistent trade-off between analytical accuracy and confidentiality (geoprivacy protection). Moreover, researchers have developed various metrics, such as spatial k -anonymity, to assess the performance of geomasking methods in minimizing disclosure risk (Ghinita et al., 2010; Wang & Kwan, 2020; Zhang et al., 2017).

Although these previous studies have provided invaluable insights into assessing the disclosure risk of maps, they have largely ignored how people perceive the disclosure risk of a map (i.e., perceived disclosure risk). An individual's perceived disclosure risk of a map refers to how the individual subjectively perceives the disclosure risk (e.g., feels uncomfortable) of the map that displays his/her private locations (e.g., home location) (Benisch et al., 2011; Groff et al., 2005; Keßler & McKenzie, 2018; Ketelaar & VanBalen, 2018; Kounadi et al., 2015; Slovic et al., 1980). Even for one map, the perceived disclosure risk can vary among individuals based on one's previous experiences and opinions on locational privacy (Slovic et al., 1980). For example, consider a map where an individual's private location can be identified at a probability of 1 in 100 (1%). Some people may feel comfortable with this level of disclosure risk, while others may not because each individual may

feel differently about disclosure risk based on their subjective assessment.

Perceived disclosure risk provides several important insights into assessing the disclosure risk of a map. First, perceived disclosure risk considers the social and cultural influences on people's risk assessment, which may not be adequately addressed by previous studies that did not consider people's perception. For example, consider a map that displays the home locations of socially vulnerable people, such as people with physical disabilities or mental health disorders, children, pregnant women, and AIDS patients. Concerning their privacy, it is widely known that these people may suffer from more serious harms when their locational privacy is breached than what the non-vulnerable groups may experience (Breslin et al., 2019; Duncan et al., 2016; Fuller et al., 2017; Leitner & Curtis, 2006; Mirzazadeh et al., 2014). However, disclosure risk assessed without considering how people perceive the risk may not adequately tell which socially-vulnerable group has a higher risk. For example, imagine two maps where an individual can be identified at a probability of 1 in 100 (1%), but the first map displays the home locations of elementary school students, while the second map displays the home locations of lung cancer patients. Which map had a higher disclosure risk? The disclosure risk evaluated without considering how people perceive the risk may not be able to fully answer this question because the potential harms associated with spatial reverse engineering are socially and culturally constructed (Armstrong & Ruggles, 2005; McLafferty, 2004). Instead, examining the extent to which people feel uncomfortable with each map (i.e., perceived disclosure risk) can take into account the social and cultural influences on people's subjective risk assessment while attempting to answer this question.

Second, perceived disclosure risk can be used to establish geoprivacy protection guidelines for mapping people's private locations. Recently, studies have called for establishing guidelines that inform researchers when creating maps with higher analytical utility while minimizing disclosure risk (Boulos et al., 2009; Kounadi & Resch, 2018; Leitner & Curtis, 2004; Richardson et al., 2015; VanWey et al., 2005). Although current privacy protection laws (e.g., the HIPAA in the United States) guide researchers to follow certain thresholds of disclosure risk (i.e., probability of reidentification), there is no clear consensus among scientists or society at large about which risk level should be used as the safe threshold (A. Curtis et al., 2011; VanWey et al., 2005). Would 1:20 (5%) or 1:2,000 (0.05%) be acceptable to people? Disclosure risk that is assessed without considering how people

perceive the risk may not fully address this question. In this light, perceived disclosure risk may provide useful supplemental or alternative insight into the extent to which a given level of disclosure risk (i.e., probability of reidentification) is acceptable to people in geoprivacy protection guidelines.

Despite the significance of assessing the perceived disclosure risk of maps, little attention has been paid to it to date. Therefore, to fill this gap, the goal of this research is to investigate how people perceive the disclosure risk of a map by conducting an online survey. Moreover, we propose a set of geoprivacy protection guidelines for mapping private locations that help researchers minimize perceived disclosure risk based on the survey results.

Data and method

Data collection

The data for this study were collected through an online survey, which was implemented in Google Forms. Participants were recruited through distributing a solicitation e-mail message to 12,000 members (6,000 students and 6,000 faculty/staff members) of a university who were randomly selected from the complete list of the university's members in November 2019. Table 1 shows the sociodemographic characteristics of the 856 participants who completed the survey. Note that although the survey participants were recruited from a university and therefore may differ systematically in their beliefs from the population at large, the survey still gives us insight into the perceived disclosure risk of people from diverse backgrounds in terms of their gender, age, and income. The survey protocol and instrument were reviewed and approved by the Institutional Review Board (IRB) of the concerned university.

Survey questionnaire

The questionnaire used in the online survey has 4 parts, each of which addresses one of the research questions mentioned earlier in the paper. In Parts 1 and 2, we ask participants to imagine the hypothetical scenario that their private locations (e.g., home/workplace and GPS trajectories) are displayed on a map published in a journal article that can be accessed by the general public.

Part 1 of the questionnaire aims to examine the first research question: how do different attributes or characteristics of a map affect an individual's perceived disclosure risk of the map? Specifically, we focus on the following four components: (1) the amount of private locational information: home, workplace, and daily GPS

Table 1. Sociodemographic characteristics of the survey participants ($n = 856$).

Sociodemographic variables		n	%
Gender	Female	503	58.8
	Male	302	35.3
	Others	9	1.1
	Prefer not to say	42	4.9
Age	~ 19 years	62	7.2
	20 ~ 29 years	221	25.8
	30 ~ 39 years	176	20.6
	40 ~ 49 years	169	19.7
	50 ~ 59 years	127	14.8
	60 ~ 69 years	69	8.1
	70 years ~	6	0.7
	Prefer not to say	26	3.0
Race/Ethnicity ^a	African American or Black	31	3.6
	Asian	117	13.7
	White or Caucasian	621	72.5
	Hispanic or Latino	50	5.8
	Others	12	1.5
	Prefer not to say	49	5.7
	International	67	7.8
Monthly Income	Yes	789	92.2
	No	188	22.0
	~ \$2,000	205	23.9
	\$2,000 ~ \$4,000	146	17.1
	\$4,000 ~ \$6,000	63	7.4
	\$6,000 ~ \$8,000	31	3.6
	\$8,000 ~ \$10,000	80	9.3
	\$10,000 ~	143	16.7
	Prefer not to say	590	68.9
Affiliation	Staff/Faculty	114	13.3
	Graduate	152	17.8
	Undergraduate	21	2.5
Major	Geography-related majors ^b	835	97.5
	Otherwise		

a Race categories are not mutually exclusive. b) Geography-related majors include Geography, GIScience, Urban and Regional Planning, Spatial Planning, City Planning, Landscape Architecture and Architecture.

trajectories, (2) the type of map: point, flow, kernel density estimation, minimum convex hull, and standard deviational ellipse, (3) the scale of the map: city-level and regional-level, and (4) the number of other people (i.e., survey participants) displayed on the map: single and multiple. The amount of private locational information refers to how much locational information of a person is displayed on a map. For instance, assume we have two maps of the same type and on the same scale. If the first map displays a person's home location while the second map displays a person's daily GPS trajectory points, the second map displays more private locational information than the first map. Table 2 describes the different combinations of these four map attributes of the 10 maps used in the online survey. Figures S1 and S2 and Table S1 in the accompanying Supplementary Material provide the maps and detailed instructions for participants that were used in Part 1 of the survey.

Note that the private locations and street networks used in Maps 1 ~ 10 in the online survey are hypothetical and do not represent any actual private locations of the survey participants. Specifically, while undertaking the survey, a participant reports his/her perceived disclosure risk as if the point in Map 1 (Figure 1(a)) represents his/

Table 2. A detailed description of the 10 maps used in the survey (Part 1).

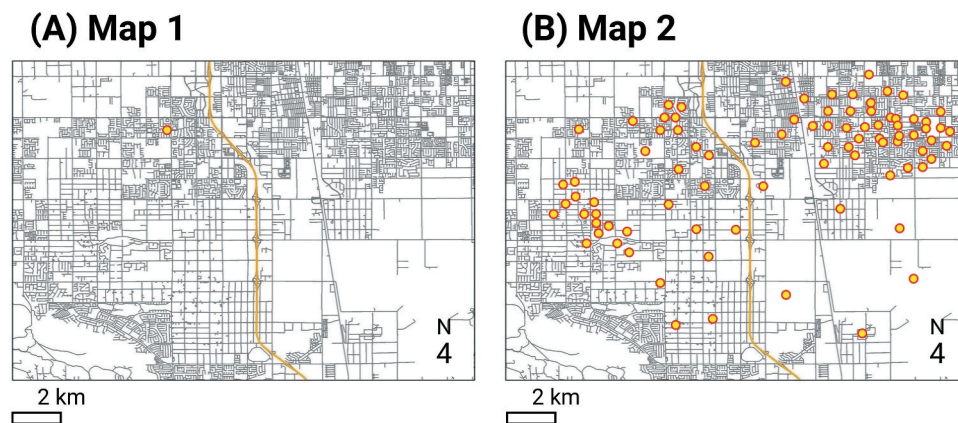
Map	Locational information	Type	Scale	Number of people
Map 1	Home	Point	City	1
Map 2	Home	Point	City	Multiple
Map 3	Home and workplace	Point (flow)	City	1
Map 4	Home and workplace	Point (flow)	City	Multiple
Map 5	Daily GPS trajectory	Point	City	1
Map 6	Daily GPS trajectory	Point	City	Multiple
Map 7	Daily GPS trajectory	Point	Regional	1
Map 8	Daily GPS trajectory	Kernel density estimation	City	1
Map 9	Daily GPS trajectory	Convex hull	City	1
Map 10	Daily GPS trajectory	Standard deviation ellipse	City	1

her actual home location. Similarly, the participant reports his/her perceived disclosure risk as if the points in Map 2 (Figure 1(b)) display his/her home location and those of other survey participants in the given hypothetical scenario. Although it would be ideal if customized maps and participants' actual private locations were used in the survey, we used hypothetical locations because obtaining participants' actual private locations, such as home addresses or daily GPS trajectories, would raise significant privacy concerns among them and increase their cognitive burdens in participating in the survey.

Moreover, note that, for some survey items that might have used technical terms (e.g., kernel density), the survey items are phrased without using the technical terms. This is because most of the survey participants may not be familiar with these technical terms as they may not come from geography-related fields (Table 1). Kernel density estimation is simply represented by "the density of your daily GPS trajectory points." The minimum convex hull is explained as "the smallest polygon that includes all the daily GPS trajectory points." The standard deviational ellipse is represented by "an ellipse that shows the general pattern of your daily GPS trajectory points." The detailed instructions for each map are provided in Table S1 in the accompanying Supplementary Material.

Part 2 of the questionnaire aims to answer the second research question: how do different geomasking methods affect an individual's perceived disclosure risk of a map? Although there are many other existing geomasking methods, such as affine transformation and location swapping (e.g., Armstrong et al., 1999; Zandbergen, 2014), this research purposefully selects two geomasking methods: aggregation (point to polygon) and random perturbation (point to point). This is because each of these methods represents a broad type of geomasking methods. For example, survey items about the areal aggregation method explore the perceived disclosure risk of maps when points are replaced by polygons (i.e., point to polygon). On the other hand, survey items about the relocation method investigate the perceived disclosure risk of maps when points are still displayed as points but at new locations (i.e., point to point). Thus, by focusing on these two geomasking methods, we can study how the type of geomasking methods used may affect the perceived disclosure risk of maps.

Specifically, regarding aggregation methods (point to polygon), we examine six different aggregation levels: 1:20, 1:100, 1:200, 1:2,000, 1:5,000, and 1:20,000. For example, the 1:20 aggregation level indicates that an individual's private location is displayed in a polygon

**Figure 1.** (a) Map 1 and (b) Map 2 of Part 1 of the survey questionnaire.

that also includes 19 other people, and thus the persons' private location can be identified with a probability of 1 in 20 (5%). With respect to relocation methods (point to point), we examine five different relocation distances: 100 ft (30 m), 500 ft (150 m), 1,000 ft (300 m), 2,000 ft (600 m), and 4,000 ft (1,200 m). For example, the relocation distance of 1,000 ft (300 m) indicates that an individual's private location is displayed at a new point that is 1,000 ft (300 m) away from the original point in a random direction. Figures S3 and S4 and Table S2 in the accompanying Supplementary Material provide the maps and detailed instructions for participants that were used in Part 2 of the survey.

Part 3 of the questionnaire aims to investigate the third research question: how does an individual perceive the disclosure risk of a map when the private locations of socially vulnerable people are visualized? Table 3 presents the 10 scenarios used in this part of the study. Note that the survey question presents the hypothetical scenario in which a map displays the fictitious private locations of socially vulnerable people as other people (rather than the survey participants themselves). Note that the study did not recruit socially vulnerable people as survey participants, although it would be ideal to do so. Alternatively, we can ask the survey participants to pretend that they belong to a socially vulnerable group in each scenario (e.g., asking "what if you are a patient . . ."). However, our pilot study that tested the survey questions found that this scenario was not as effective as we expected because people found it difficult or unpleasant as they have rarely imagined such a scenario. Therefore, the research team decided to ask people about their discomfort level when the private locations of "other socially vulnerable people" are displayed on a map.

Lastly, Part 4 of the questionnaire collects basic sociodemographic information from the participants, including gender, age, race/ethnicity, nationality, position (faculty, staff, student), academic major, and income level. Statistical tests are conducted to investigate whether there are significant differences in

perceived disclosure risk between participants with different sociodemographic characteristics.

Measurement of an individual's perceived disclosure risk of a map

The perceived disclosure risk of a map represents the extent to which an individual subjectively feels comfortable with the map that displays his/her private locations (Benisch et al., 2011; Groff et al., 2005; Keßler & McKenzie, 2018; Ketelaar & VanBalen, 2018; Kounadi et al., 2015). Specifically, the feeling can be positive (i.e., comfortable) or negative (i.e., uncomfortable), and the intensity of the feeling can be strong or weak. Following an approach in previous studies (e.g., Benisch et al., 2011; Groff et al., 2005; Kounadi et al., 2015), an individual's perceived disclosure risk in this study is measured on a scale of 1 to 7, where 1 means "the most comfortable" and 7 is "the most uncomfortable." A scale of 1 to 7 is used in the survey questionnaire to fully capture the diverse range of the perceived risk levels of the participants (Bandalos, 2018). Although a scale with fewer response options (e.g., a 5-point scale) may be preferred for simplicity, it may not effectively reflect the wide spectrum of the intensities of the participants' perceived risk. For example, when a scale of 1 to 5 is used, if a participant feels discomfort with a certain map, he/she has only two options (i.e., 4: uncomfortable; 5: the most uncomfortable), which may not effectively capture the various intensities of uncomfortableness. Figure 2 illustrates this scale of perceived disclosure risk used in our online survey.

Moreover, an open-ended question is assigned to each map/scenario in Parts 1 ~ 3 where survey participants briefly explain the reason for their responses. The open-ended questions ask the survey participant why he/she feels comfortable or uncomfortable with a map in each questionnaire item. By categorizing the repeated themes that appeared in the responses, we examine the primary reasons why the survey participants feel comfortable or uncomfortable with a map. Note that, to

Table 3. A detailed description of the 10 scenarios used in the survey (Part 3).

Scenario	Description
S1	Cancer patients
S2	HIV (AIDS) patients
S3	Elementary school students
S4	Pregnant women
S5	Elderly people
S6	People who engage in sex with people of the same sex
S7	People in poverty
S8	High-income earners
S9	People who are in alcohol or substance abuse treatment
S10	Randomly selected people

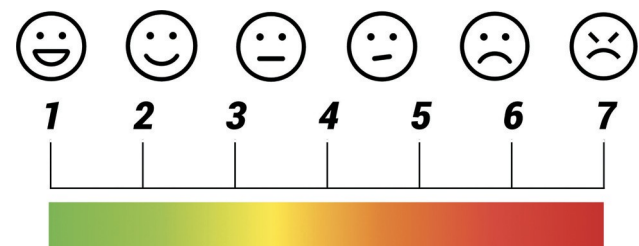


Figure 2. Perceived disclosure risk that is measured on a scale of 1 to 7, using 1 as the most comfortable and 7 as the most uncomfortable.

increase the participation rate, answering the open-ended questions is optional and thus, not every participant responded to the open-ended questions.

Results

The effects of map attributes on the perceived disclosure risk (Part 1)

First, we examine the effects of the amount of private locational information displayed on perceived disclosure risk by focusing on Maps 1, 3, and 5. Recall that these maps are different in the amount of locational information but are the same in map type (i.e., point-based), scale (i.e., city-level), and the number of people displayed (i.e., 1 person) (see Table 2 for a detailed description of each map). Specifically, Map 5 displays the most amount of information, followed by Map 3 and Map 1. As mentioned earlier in the *Survey questionnaire* section, since we did not collect the actual private locations of the survey participants, maps used in the questionnaire represent hypothetical situations. Thus, while undertaking the survey, the participants are asked to imagine that the maps display their actual private locations.

Figure 3 illustrates the bar graphs that show the frequency of the discomfort level for Maps 1, 3, and 5. Overall, the bar graphs show that the percentage of people who feel uncomfortable (i.e., discomfort level 5 ~ 7) is the highest for Map 5 (87%), followed by Map 3 (81%) and Map 1 (74%). Table 4 shows the descriptive statistics. The average discomfort level of the participants is the highest for Map 5 (6.2), followed by Map 3 (5.7) and Map 1 (5.3). Moreover, we investigate whether the pairwise differences between these three maps are significant. The Wilcoxon signed-rank

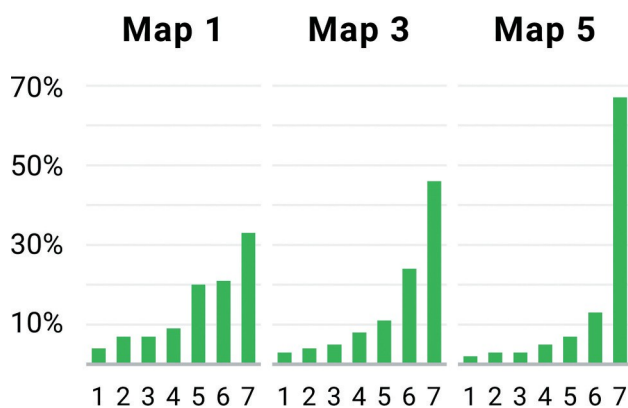


Figure 3. Frequency of the discomfort level in Maps 1, 3, and 5 (Note: 1: The most comfortable, 4: Neutral, 7: The most uncomfortable).

Table 4. Descriptive statistics of the responses to the 10 different maps^{a)}.

Maps ^c	Level of discomfort ^b							M	Mdn	SD
	1	2	3	4	5	6	7			
Map 1	4%	7%	7%	9%	20%	21%	33%	5.3	6	1.8
Map 2	12%	15%	14%	13%	16%	16%	14%	4.1	4	2.0
Map 3	3%	4%	5%	8%	11%	24%	46%	5.7	6	1.6
Map 4	12%	13%	12%	13%	16%	18%	16%	4.3	5	2.0
Map 5	2%	3%	3%	5%	7%	13%	67%	6.2	7	1.5
Map 6	20%	16%	12%	16%	14%	11%	11%	3.6	4	2.0
Map 7	3%	6%	5%	9%	10%	18%	49%	5.7	5	1.7
Map 8	10%	10%	10%	15%	17%	17%	22%	4.6	6	2.0
Map 9	14%	13%	15%	15%	17%	13%	13%	4.0	4	1.9
Map 10	16%	16%	14%	15%	18%	11%	10%	3.8	4	1.9

M denotes Mean, Mdn denotes Median, and SD denotes Standard Deviation.

a) Bold indicates the highest proportion among the responses. b) 1: The most comfortable, 4: Neutral, 7: The most uncomfortable. c) Please refer to Table 2 for the detailed description of each map.

test is used because the responses are on an ordinal scale of measurement and the distribution of participants' responses to each questionnaire item shows strong non-normality as shown in Figure 3–6 (Privitera, 2011). The results indicate that the pairwise differences in perceived disclosure risk between the three maps are statistically significant ($p < 0.001$) even after considering Bonferroni correction.

Therefore, these results suggest that the perceived disclosure risk of a map increases when the map displays more locational information. Moreover, the results support the claim of previous studies that researchers should pay special attention when mapping detailed trajectories obtained from mobile devices or smart sensors (Breslin et al., 2019; Seidl et al., 2016; Fuller et al., 2017; De Montjoye et al., 2013; Wang & Kwan, 2020).

Next, we investigate the effects of the type of map on perceived disclosure risk by focusing on Maps 8, 9, and 10. Recall that these three maps are different map types but are the same in scale (i.e., city-level) and the number

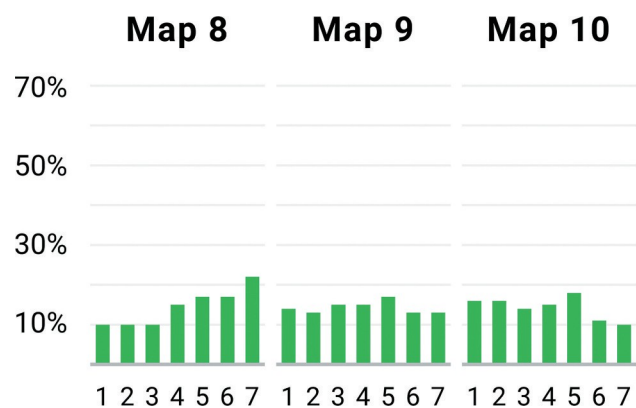


Figure 4. Frequency of the discomfort level in Maps 8, 9, and 10 (Note: 1: The most comfortable, 4: Neutral, 7: The most uncomfortable).

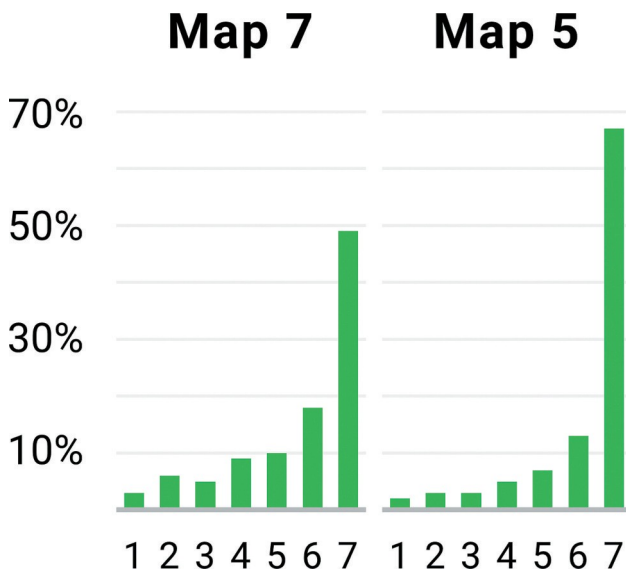


Figure 5. Frequency of the discomfort level in Map 7 (Note: 1: The most comfortable, 4: Neutral, 7: The most uncomfortable).

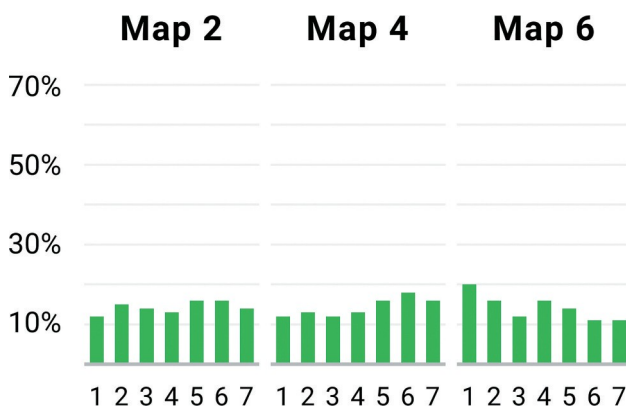


Figure 6. Frequency of the discomfort level in Maps 2, 4, and 6 (Note: 1: The most comfortable, 4: Neutral, 7: The most uncomfortable).

of people displayed (i.e., 1 person). Specifically, compared to Map 5 that displays the hypothetical GPS trajectories of one participant, each map type is different: Map 8 (kernel density estimation), Map 9 (minimum convex hull), and Map 10 (standard deviation ellipse).

Figure 4 shows the frequency of the discomfort level for Maps 8, 9, and 10. Compared to Map 5, the perceived disclosure risks are lower in these three maps. The percentage of people who feel uncomfortable (i.e., discomfort level 5 ~ 7) is the lowest for Map 10 (39%), followed by Map 9 (43%) and Map 8 (56%). Moreover, as Table 4 illustrated, the average discomfort level is the lowest for Map 10 (3.8), followed by Map 9 (4.0), Map 8 (4.6), and Map 5 (6.2). The pairwise Wilcoxon signed-rank test

indicates that the pairwise differences between the four maps are statistically significant ($p < 0.001$) even after considering Bonferroni correction. Therefore, these results suggest that map type is associated with perceived disclosure risk.

These results imply that, given the current research practice that visualizes people's private locations using points (Haley et al., 2016; Kounadi & Leitner, 2014), selecting the proper map type can reduce perceived disclosure risk. For example, converting a point-based map to a kernel density map can be one of the suitable options because it provides useful analytical information as well as reduces the disclosure risk (Boulos et al., 2009; Kounadi & Resch, 2018). However, one caveat is that the kernel density surface may not always be the best answer because the original points can still be accurately recovered when the parameters (bandwidth and cell size) are not carefully selected (Boulos et al., 2009; Lee et al., 2019; Wang et al., 2019).

In what follows, we examine the effects of map scale on perceived disclosure risk by comparing Map 5 and Map 7. These two maps are different in scale (i.e., Map 5: city level, Map 7: regional level) but provide the same amount of locational information (i.e., daily GPS trajectory), use the same map type (i.e., point), and display the same number of people (i.e., 1 person). Figure 5 shows the frequency of the discomfort level in Map 7. Compared to Map 5 (87%), it shows that many participants (77%) still feel uncomfortable but the percentage of Map 7 is lower than that of Map 5. Table 4 showed that the average score of Map 7 (5.7) is lower than that of Map 5 (6.2). The pairwise Wilcoxon signed-rank test illustrates that the pairwise differences between Map 5 and Map 7 are statistically significant ($p < 0.001$). Therefore, the results indicate that the perceived disclosure risk of a regional-scale map is smaller than that of a city-scale map.

Lastly, we investigate the effects of the presence of the other people in a map on perceived disclosure risk by examining three map pairs: Maps 1 and 2, Maps 3 and 4, and Maps 5 and 6. The maps in each pair of these maps have the same amount of locational information and are of the same map type and scale. Figure 6 shows the frequency of the discomfort level in Maps 2, 4, and 6. By comparing this with the results revealed in Figure 3, we observe that fewer people feel uncomfortable with a map when the hypothetical locations of the other survey participants are also displayed on a map. For example, 74% of the survey participants feel uncomfortable with Map 1, while 46% of them feel uncomfortable with Map 2. Moreover, 81% of the survey participants feel uncomfortable with Map 3, while 50% of them feel uncomfortable with Map 4. Lastly, 87% of the survey participants

feel uncomfortable with Map 5, while 34% of them feel uncomfortable with Map 6. The pairwise Wilcoxon signed-rank test shows that the pairwise differences for each of these three map pairs are statistically significant ($p < 0.001$). Thus, the results indicate that perceived disclosure risk is lower when people's private locations are "diluted" by the locations of the others.

Further, we explore why people feel uncomfortable with a map that discloses their home location by analyzing their responses to an open-ended question. Although the open-ended question is assigned to each map (Map 1 ~ 10), we present the results on Map 1 as it has the largest number of responses. Since we designated this open-ended question as an optional question, about half of the survey participants provided answers. We read through all responses and categorized them into several groups that appear repeatedly in the responses. Figure 7 is a pie chart that illustrates the categorized primary reason why people feel

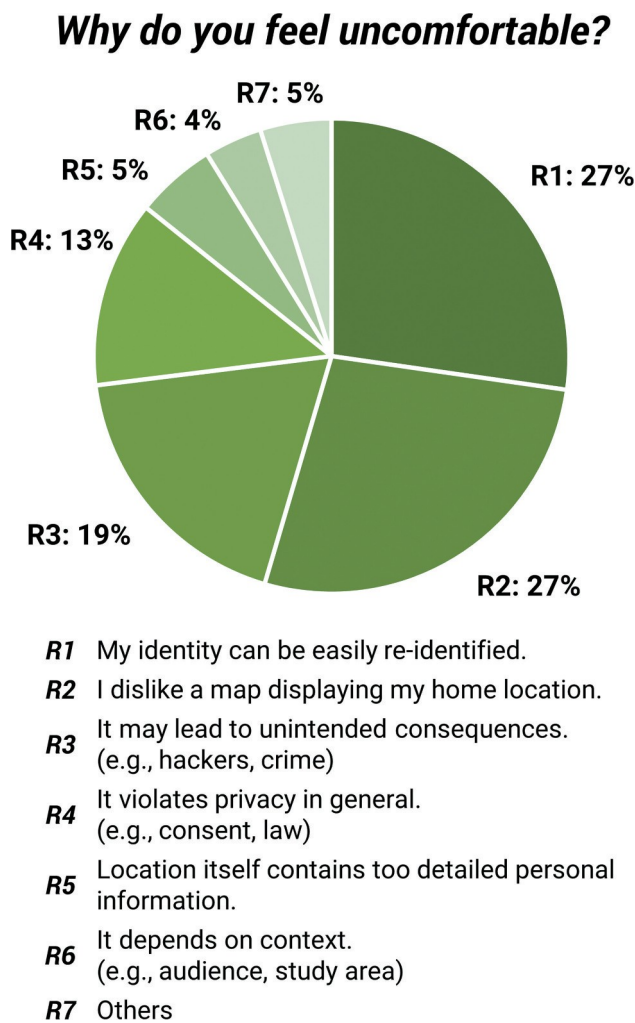


Figure 7. The primary reason why the survey participants feel uncomfortable with Map 1 ($n = 330$).

uncomfortable (i.e., discomfort level: 5 ~ 7) with Map 1 that displays his/her hypothetical home location ($n = 330$). A couple of important observations can be made here.

First, about 30% of the survey participants feel uncomfortable because they think their identity can be easily re-identified. Some participants mentioned that by using web search engines or online map services, anyone can easily identify the specific location and street address. Some of these participants also stated that since their home location represents their identity, they consider that a map that indicates their home location discloses their identity even if the map does not specify their name or other detailed demographic information. Second, about 20% of these participants reported that they worry about possible unintended consequences. For example, some mentioned that strangers may be able to visit their private locations, which may lead to safety issues (e.g., crime). They said this situation makes them feel uncomfortable. Lastly, others mentioned that their perceived disclosure risk depends on the contexts of a map, such as audiences and study areas. Specifically, some mentioned that they would feel more uncomfortable if the map is presented in mass media, which is more accessible by the general public. Others mentioned that the perceived disclosure risk depends on the study area where they live. People who said they live in a small rural community or a suburban single-family neighborhood mentioned they particularly feel uncomfortable with a map because they can easily be identified from their small neighborhood. On the contrary, people who said they live in a high-density urban area or temporary housing (e.g., a dormitory for students) reported that they feel less uncomfortable. They mentioned that since many other people live in the same building (e.g., apartment complex) with them or their home location is not permanent as they frequently move to new places, they are less concerned.

The effects of geomasking methods on the perceived disclosure risk (Part 2)

We first examine how perceived disclosure risk is affected by the aggregation level using the aggregation geomasking method. Figure 8 illustrates the frequency of the discomfort level regarding the six different aggregation levels. Table 5 shows the descriptive statistics. As the aggregation level becomes higher (i.e., from 1:20 to 1:20,000), the perceived disclosure risk decreases. Recall that the 1:20 aggregation level means that an individual's private location is displayed in a polygon that also includes 19 other people, and thus the private location can be identified with a probability of 1 in 20 (5%).

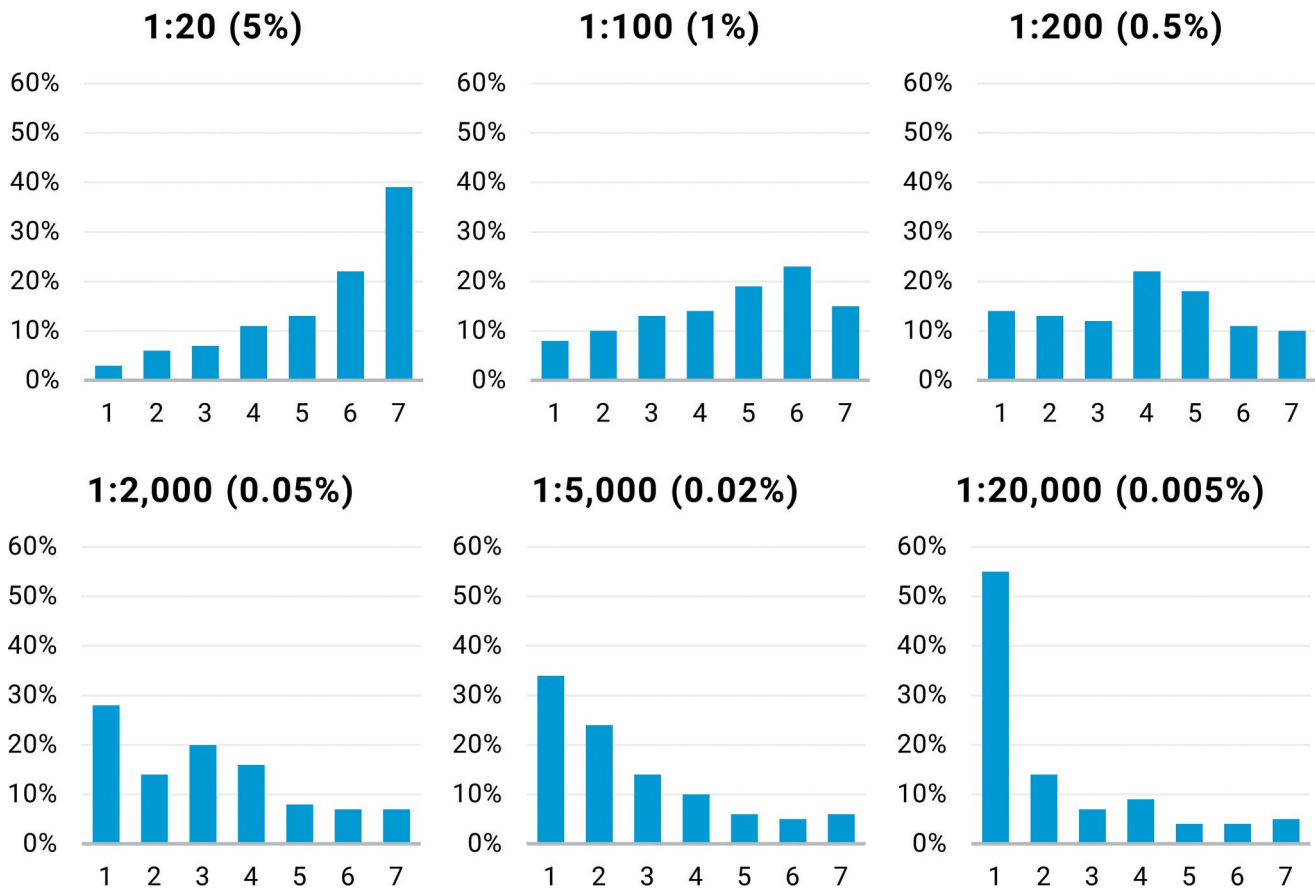


Figure 8. Frequency of the discomfort level regarding 6 different aggregation levels (Note: 1: The most comfortable, 4: Neutral, 7: The most uncomfortable).

Table 5. Descriptive statistics of the responses regarding different aggregation levels^a.

Aggregation Levels	Level of discomfort ^b							M	Mdn	SD
	1	2	3	4	5	6	7			
1:20	3%	6%	7%	11%	13%	22%	39%	5.4	6	1.7
1:100	8%	10%	13%	14%	19%	23%	15%	4.5	5	1.8
1:200	14%	13%	12%	22%	18%	11%	10%	3.9	4	1.8
1:2,000	28%	14%	20%	16%	8%	7%	7%	3.1	3	1.9
1:5,000	34%	24%	14%	10%	6%	5%	6%	2.7	2	1.8
1:20,000	55%	14%	7%	9%	4%	4%	5%	2.3	1	1.8

M denotes Mean, Mdn denotes Median, and SD denotes Standard Deviation. a) Bold indicates the highest proportion among the responses. b) 1: The most comfortable, 4: Neutral, 7: The most uncomfortable.

The results reveal that, at the 1:20 (5%) aggregation level, 74% of the survey participants feel uncomfortable (i.e., discomfort level: 5 ~ 7), but only 13% of them feel uncomfortable with the 1:20,000 (0.005%) aggregation level. Similarly, the average perceived disclosure risk of the 1:20,000 aggregation level is 2.3, which is smaller than that of the 1:20 level, which is 5.4.

Second, we investigate how perceived disclosure risk is affected by relocation distance using the relocation geomasking method. Figure 9 illustrates the frequency of the discomfort level regarding five different relocation distances. Table 6 shows the descriptive statistics. The results indicate that perceived disclosure risk

decreases as relocation distance increases. For instance, 77% of the participants who responded to this part of the survey feel uncomfortable with a 100 ft (30 m) relocation distance, while 32% of them feel uncomfortable with a 4,000 ft (1,200 m) relocation distance. Similarly, the average perceived disclosure risk of the 4,000 ft (1,200 m) distance is 3.4, which is smaller than that of the 100 ft (30 m) distance (5.7).

Based on these results, we suggest tentative aggregation levels or distances that may seem to be acceptable to some people. For each aggregation level (or distance), we calculate the ratio of the number of the survey participants who feel uncomfortable (i.e., discomfort level 5 ~ 7) to that of



Figure 9. Frequency of the discomfort level regarding 5 different geomasking relocation distances.

Table 6. Descriptive statistics of the responses regarding different geomasking relocation distances^a.

Geomasking Distances	Level of discomfort ^b							<i>M</i>	<i>Mdn</i>	<i>SD</i>
	1	2	3	4	5	6	7			
100 ft (30 m)	4%	4%	5%	9%	14%	15%	48%	5.7	6	1.7
500 ft (150 m)	5%	5%	7%	14%	14%	23%	32%	5.2	6	1.8
1,000 ft (300 m)	9%	8%	12%	16%	18%	15%	22%	4.6	5	1.9
2,000 ft (600 m)	15%	15%	13%	16%	14%	11%	16%	3.9	4	2.0
4,000 ft (1,200 m)	30%	13%	12%	14%	11%	8%	13%	3.4	3	2.1

M denotes Mean, *Mdn* denotes Median, and *SD* denotes Standard Deviation. a) Bold indicates the highest proportion among the responses. b) 1: The most comfortable, 4: Neutral, 7: The most uncomfortable.

the survey participants who feel comfortable (i.e., discomfort level 1 ~ 3). If the ratio is smaller than 1.0, it indicates that the number of people who feel comfortable is higher than that of people who feel uncomfortable.

For illustration purposes, we assume that a ratio of smaller than 1.0 indicates a tentative acceptable aggregation level or relocation distance. It is worth mentioning that we do not intend to provide a universal principle for aggregation levels or relocation distances. Specifically, acceptable aggregation levels or relocation distances based on this assumption (i.e., 1.0 as a threshold) may be used as a reference. If the ratio is 1.0, it indicates that about half of the people are uncomfortable at the given aggregation level or relocation distance. Therefore, thresholds that

are smaller than 1.0 would be preferred for studies that specifically need more rigorous privacy protection.

Figure 10 shows the ratio with respect to aggregation level and relocation distance. In Figure 10(a), the ratio is smaller than 1.0 when the aggregation level is higher than 1:200. Thus, it suggests that the range of acceptable levels can be higher than 1:200. This result thus implies that census block group (containing 2,000 people on average) and census tract (containing 5,000 people on average) may be suitable options for aggregation methods. Moreover, in Figure 10(b), the ratio is smaller than 1.0 when the relocation distance is greater than 2,000 ft (600 m). Therefore, it suggests that acceptable relocation distances are those that are greater than 2,000 ft (600 m). However, one caveat is

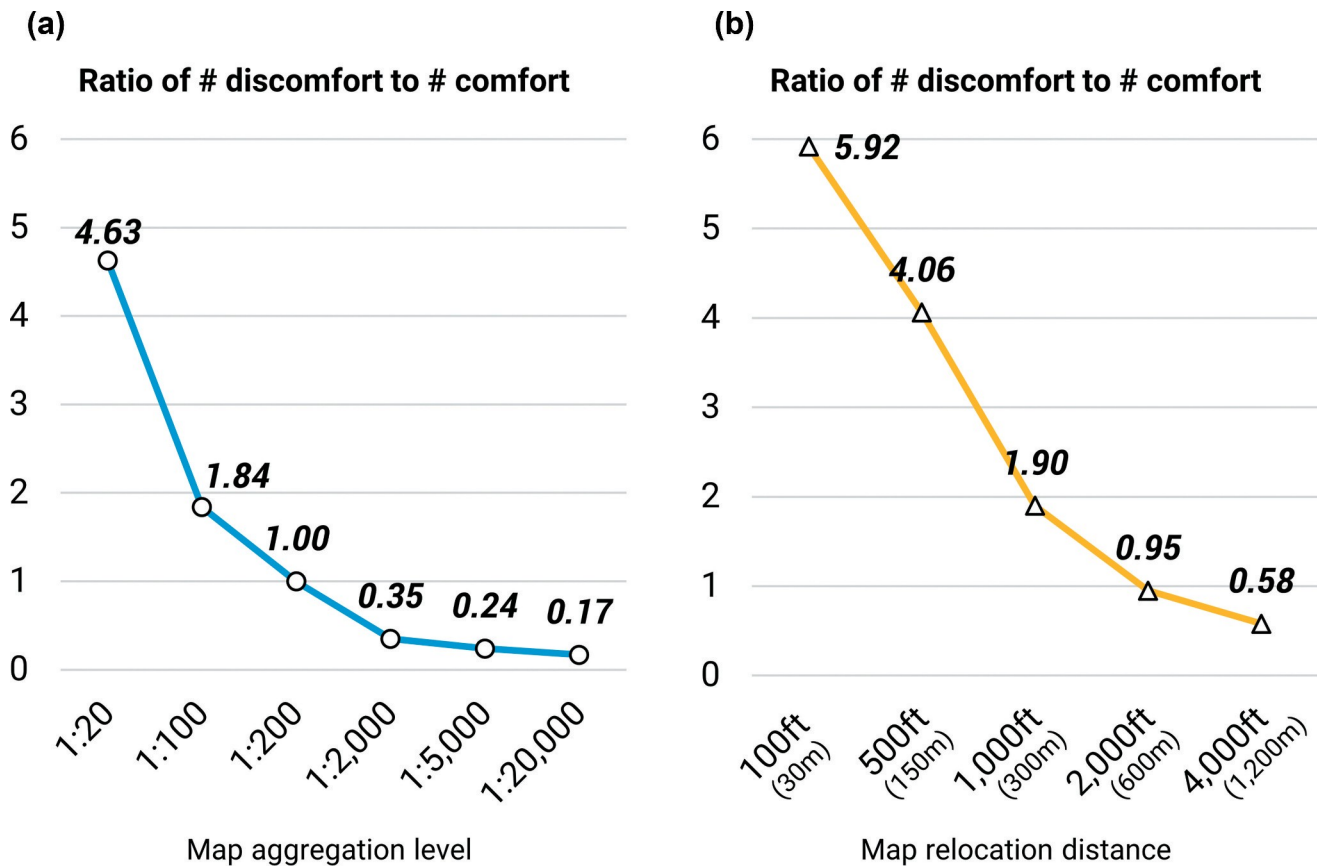


Figure 10. The ratio of the number of the survey participants who feel uncomfortable (i.e., discomfort level 5 ~ 7) to that of the survey participants who feel comfortable (i.e., discomfort level 1 ~ 3) regarding (a) the aggregation level and (b) the relocation distance.

that this is not a golden rule that should always be followed (Boulos et al., 2009). Recall that, for instance, there are still 13% of the participants who feel uncomfortable with a 1:20,000 level of aggregation, which is the highest aggregation level used in the questionnaire. In this light, one possible data collection practice can be that researchers ask survey participants (whose locational information is being collected) about which aggregation level or relocation distance is acceptable to them. Although different people may have different levels, by obtaining information about the acceptable levels from all study participants, researchers can systematically determine an aggregation level or relocation distance rather than determining them arbitrarily.

Lastly, we explore why people feel uncomfortable with a certain aggregation level and relocation distance by analyzing their responses to the two optional open-ended questions in Part 2 of the survey. 161 participants responded to the open-ended question about the aggregation method. About 70% of them mentioned that higher aggregation levels are better because it is less likely their identity can be estimated. More people (174 participants) responded to the second open-ended question about the relocation distance, which may imply their stronger concerns about this method.

For example, about 40% of them stated that the relocation method itself may be inadequate because of two major reasons. One reason is that the distances listed in the survey (100 ft [30 m] ~ 4,000 ft [1,200 m]) are not large enough, and relocated points can still be located in the same neighborhood where the original points are located, which makes them feel uncomfortable. The second reason, which is often ignored by researchers, is that a randomly-located point may harm or endanger the innocent people who are falsely indicated by the relocated point, which causes additional concerns to the participants. These observations also corroborate the findings of previous studies that false identification is one of the critical issues in geo-masking methods and should be avoided (Seidl et al., 2018; McLafferty, 2004).

The perceived disclosure risk of a map when socially vulnerable people are displayed (Part 3)

In what follows, we examine how an individual's perceived disclosure risk is affected when a map displays the private locations of socially vulnerable people. Recall that the survey presented 10 different hypothetical scenarios related to socially vulnerable populations (see Table 3).

Table 7. Descriptive statistics of the responses regarding 10 different scenarios^a.

Scenarios ^c	Level of discomfort ^b							M	Mdn	SD
	1	2	3	4	5	6	7			
S1	5%	5%	4%	18%	14%	17%	37%	5.3	6	1.8
S2	3%	4%	3%	13%	11%	17%	48%	5.7	6	1.7
S3	4%	4%	2%	10%	9%	14%	56%	5.8	6	1.7
S4	4%	4%	4%	15%	12%	16%	45%	5.5	6	1.7
S5	4%	6%	5%	15%	12%	16%	42%	5.4	7	1.8
S6	5%	4%	2%	15%	8%	16%	50%	5.7	7	1.7
S7	4%	4%	6%	17%	14%	20%	35%	5.3	5	1.7
S8	7%	7%	6%	20%	12%	17%	32%	5.0	6	1.9
S9	2%	4%	3%	13%	13%	20%	45%	5.7	6	1.6
S10	13%	7%	5%	26%	11%	13%	24%	4.5	4	2.0

M denotes Mean, Mdn denotes Median, and SD denotes Standard Deviation.

a) Bold indicates the highest proportion among the responses. b) 1: The most comfortable, 4: Neutral, 7: The most uncomfortable. c) Please refer to Table 3 for the detailed description of each scenario.

Table 7 shows the descriptive statistics, and Figure 11 illustrates the frequency of the discomfort level. In terms of the average discomfort level, a scenario for which people are most uncomfortable is a map about elementary school students (S3; 5.8), followed by HIV/AIDS patients (S2; 5.7), people who engage in sex with people of the same sex (S6; 5.7), people who are under alcohol or substance abuse treatment (S9; 5.7), pregnant women (S4; 5.5), elderly people (S5; 5.4), cancer patients (S1; 5.3), people in poverty (S7; 5.3) and high-income earners (S8; 5.0). In addition, the average perceived disclosure risk of these 9 scenarios is higher than that of the randomly selected people (S10; 4.5). Moreover, the pairwise Wilcoxon signed-rank test results illustrate that the pairwise differences between each scenario (S1 ~ S9) and S10 are statistically significant ($p < 0.001$) even after Bonferroni correction. Therefore, the results indicate that the perceived disclosure risk of a map displaying the private locations of socially vulnerable people is higher than the maps that display non-socially vulnerable populations (e.g., randomly selected people).

We further explore why people reported higher perceived disclosure risk regarding socially vulnerable people by analyzing the optional open-ended question in Part 3 of the questionnaire. About half of the 236 participants who responded to this question expressed their worry about unintended negative outcomes, such as hate crime, discrimination, and stigmatization. Some stated that the socially-vulnerable groups can be easily targeted because they cannot defend themselves, which calls for special attention when mapping their locations.

Survey participants particularly reported that they feel the most discomfort (i.e., highest perceived disclosure risk) with a map that displays the private locations of younger children among all the scenarios investigated in the survey. They reported that this is because younger children are more vulnerable to unintended negative outcomes (e.g., a crime like kidnapping or molesting)

than the other socially-vulnerable populations investigated in the survey. However, others stated that some maps can be useful for public policy purposes. For example, a map of cancer patients may provide useful insights into public health, and a map of the poor or the elderly may be informative for knowing the people in need of certain social services.

Therefore, the results imply that researchers should exercise special care when mapping the private locations of the socially-vulnerable populations (Breslin et al., 2019; Duncan et al., 2016; Fuller et al., 2017; Leitner & Curtis, 2006; Mirzazadeh et al., 2014). Specifically, by comparing the 10 different scenarios, a map displaying the private locations of elementary school students reported the highest perceived disclosure risk. Considering the recently growing attention in children's behavior studies (e.g., McCarthy et al., 2017; Stark et al., 2018), the results imply that researchers should be more careful when visualizing the home or activity locations of children.

Perceived disclosure risk with respect to sociodemographic characteristics (Part 4)

In this subsection, we examine the association between perceived disclosure risk and sociodemographic characteristics. To begin with, we calculate each participant's composite perceived disclosure risk score by adding each response item in Map 1 through Map 10 in Part 1. We did not include items from Part 2 and Part 3 because the main objectives of those parts are specific (e.g., the effects of geomasking and the socially-vulnerable people) rather than investigating the perceived disclosure risk of maps in general. The possible maximum composite score is 70 because each item is measured on a scale of 1 to 7 (i.e., 7×10), while the possible minimum composite score is 10 (i.e., 1×10). Cronbach's alpha of the 10 items is 0.924, indicating that a set of items are closely related as a group (i.e., high internal consistency). We calculate the perceived disclosure risk score of the 856 participants. The average score is 47.19, the median score is 49.00, and the standard deviation is 14.3. The minimum score is 10, while the maximum score is 70.

Next, we investigate the perceived disclosure risk score in terms of sociodemographic characteristics of the survey participants. Since the results of Shapiro-Wilk's test suggest that the perceived disclosure risk scores are not normally distributed ($p < 0.001$), we employ the Mann-Whitney U test to investigate the effects of gender, race, position (faculty/staff or student), residence status (U.S.-born people or foreigners), and major (Privitera, 2011). For investigating the effects of

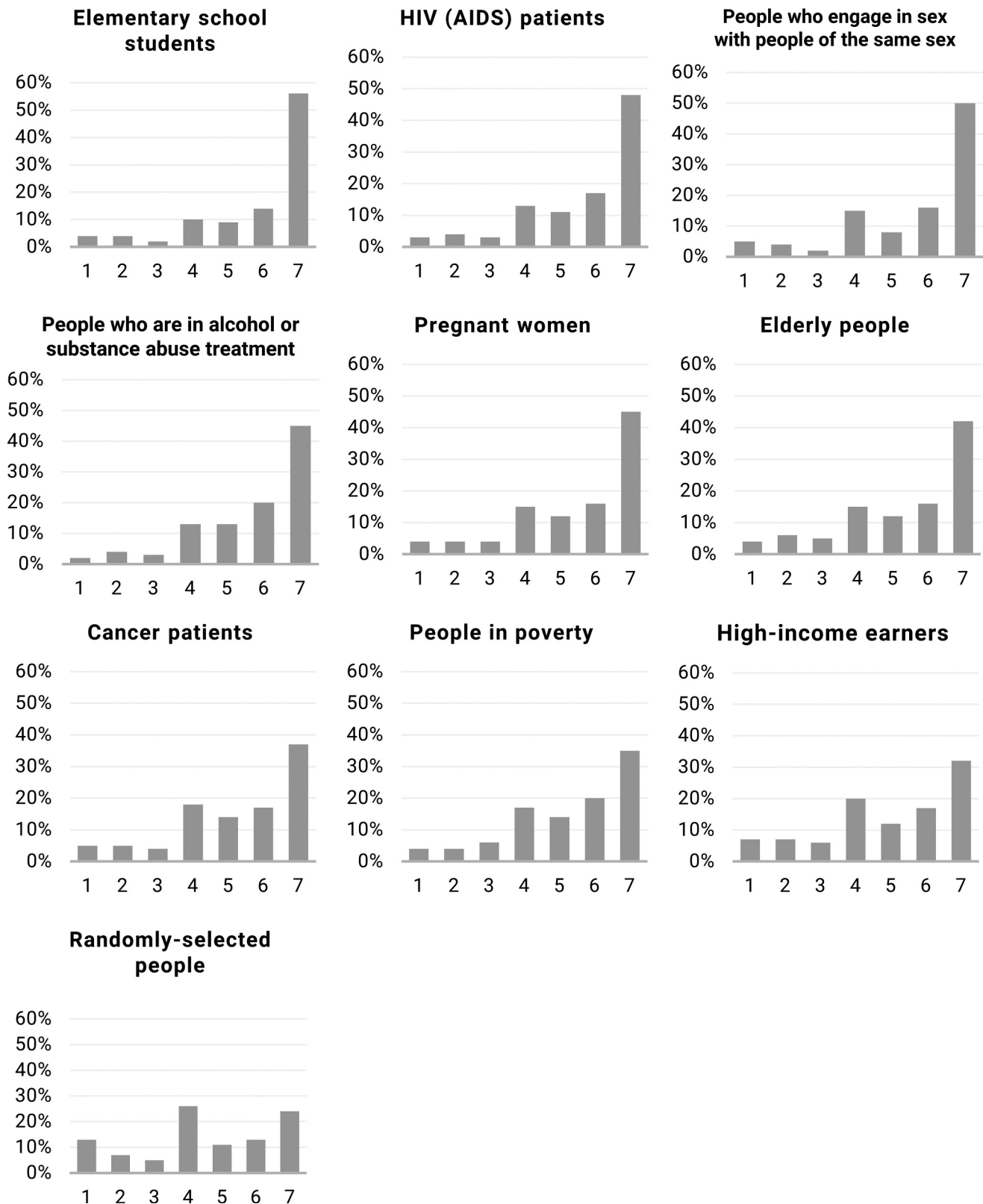


Figure 11. Frequency of the discomfort level regarding 10 different scenarios.

age and income level, we calculate the Spearman correlation coefficient (Privitera, 2011). Table 8 shows the results.

The results of the Mann-Whitney U test in Table 8 show that the perceived disclosure risk score is significantly greater among faculty/staff members compared

Table 8. Descriptive statistics of the perceived risk score in terms of participants' sociodemographic characteristics and results of the Mann-Whitney U tests.

	<i>M</i>	<i>Mdn</i>	<i>SD</i>	<i>M</i>	<i>Mdn</i>	<i>SD</i>	<i>p</i> ^a
Gender	47.0	Female (n = 503) 48.0	13.1	46.3	Male (n = 302) 48.0	15.8	0.563
Race	47.2	White (n = 599) 49.0	13.9	44.9	Nonwhite (n = 206) 46.0	14.9	0.025
Affiliation	48.0	Faculty/Staff (n = 590) 50.0	14.5	45.3	Student (n = 266) 46.5	13.6	0.001
International	47.5	U.S.-born (n = 789) 49.0	14.0	43.8	Foreigner (n = 67) 45.0	16.5	0.042
Major ^b	47.2	Non-geography (n = 835) 49.0	14.3	46.8	Geography-related majors (n = 21) 50.0	13.8	0.788

M denotes Mean, *Mdn* denotes Median, *SD* denotes Standard Deviation, and *p* denotes p-value. a) Mann-Whitney U test results. b) Geography-related majors include Geography, GIScience, Urban and Regional Planning, Spatial Planning, City Planning, Landscape Architecture and Architecture.

to students ($p < 0.01$). One possible explanation is that faculty/staff members may be more concerned and aware of (geo)privacy issues because they may have participated in ethics training or are involved in the IRB process more regularly and actively than students. Moreover, the results indicate that the perceived disclosure risk score of U.S.-born participants is significantly greater than that of non-U.S.-born participants (foreigners) ($p < 0.05$), and the perceived disclosure risk score of white people are also significantly greater than that of nonwhite people ($p < 0.05$). One possible explanation is that faculty/staff members largely consist of white and U.S.-born people. For example, among the 266 students in our sample, 12% ($n = 31$) of them are foreigners, and 42% ($n = 112$) of them are nonwhite. On the contrary, among the 590 faculty/staff members in the sample, only 6% ($n = 35$) of them are foreigners, and only 16% ($n = 94$) of them are nonwhite. In other words, the proportions of students who are foreigners and nonwhite people are higher than those of faculty/staff members.

Furthermore, there is no significant difference in the perceived risk score in relation to gender and major. One might expect the opposite result in terms of major because people with a geography-related major may have fundamental mapping and spatial analysis backgrounds, which may lead to their high perceived disclosure risk. One possible explanation is that the number of people with a geography-related major is too small in our sample ($n = 21$, 2.5% of all participants), which may not allow us to capture the trend of the geography-related participants.

Lastly, the Spearman correlation coefficients show that the correlation between the perceived disclosure risk score and age is weak but significant ($r_s = 0.076$, $p < 0.05$), but the correlation between the perceived disclosure risk score and income level is insignificant. One possible explanation is that faculty/staff members, who reported high perceived disclosure risk, are older than students.

However, one caveat is that this analysis is based on a limited number of sociodemographic characteristics of

the participants. Since the sample is recruited from the university population, the percentage of high-income and educated people in the sample may be higher than that of the general public. Also, some groups include only a smaller number of participants (e.g., people with geography-related majors), which may lead to limited representativeness of these groups in our study. Moreover, the linear regression model results (not presented here) illustrate that there are no significant associations between the perceived disclosure risk score and each sociodemographic variable when all other variables are controlled. Therefore, future studies should address this issue by using participants with more diverse backgrounds to investigate the association between perceived disclosure risk and sociodemographic characteristics.

Conclusion

This research examines how people subjectively perceive the disclosure risk of maps by conducting an online survey. The results indicate that perceived disclosure risk increases as the amount of locational information displayed on a map increases. Compared to point-based maps, perceived disclosure risk is significantly lower for kernel density maps, convex hull maps, and standard deviational ellipse maps. The results also reveal that perceived disclosure risk is affected by map scale and the presence of the locational information about other people on a map. With regard to the effects of geomasking methods, perceived disclosure risk decreases as aggregation level increases and as relocation distance increases. Lastly, the perceived disclosure risk of a map that displays socially-vulnerable people is significantly higher than that of a map that displays non-vulnerable groups. Specifically, a map displaying the private locations of elementary school students has the highest perceived disclosure risk.

Based on these results, we propose some tentative guidelines for geoprivacy protection that consider people's perceived disclosure risk. Researchers also should be

encouraged to fully understand the unique contexts of their studies in which people's private locations are mapped. (1) To visualize the point patterns of the location of people (e.g., survey participants), a kernel density map may be recommended rather than a point-based map that directly displays the points. However, we urge researchers to pay special attention to carefully selecting parameters (e.g., bandwidth and cell size) and trying different combinations of parameters to reduce the possibility of successful spatial reverse engineering which may occur if the parameters are improperly selected. (2) To visualize the flow of people (e.g., commuting pattern), using aggregated cells or polygons that each includes enough number of people for the origin/destination points of a flow may be recommended. Based on the results of this study, including at least 200 people may be recommended, but researchers should carefully consider the unique context of their study. (3) To display the daily GPS trajectories of one person as an illustrative example, a standard deviational ellipse map may be the most recommended, followed by a minimum convex hull map and a kernel density map. However, when using a kernel density map, researchers should pay attention to the caveat mentioned in Guideline (1) above.

We also offer the following suggestions related to geomasking methods: (1) The aggregation method may be preferred when compared to the relocation method (e.g., affine transformation and random perturbation) for avoiding the false identification of individuals. (2) For the aggregation method, aggregation levels higher than 1:200 may be acceptable. For example, census block groups (containing 2,000 people on average) and census tracts (containing 5,000 people on average) may be one of the acceptable options. However, researchers should carefully consider the unique context of their study. (3) For the relocation method, relocation distances greater than 2,000 ft (600 m) may be one of the acceptable options. However, researchers should carefully consider the unique context of their study. (4) The best practice may be to ask survey participants (whose locational information is collected) in advance to know which level or distance is acceptable to them.

Lastly, we provide some guidelines for mapping socially-vulnerable people: (1) The aggregation method (point to polygon) with a proper aggregation level may be recommended rather than the relocation method (point to point). (2) The socially-vulnerable groups that need the most attention are elementary school students, followed by HIV/AIDS patients, people who engage in sex with people of the same sex, people who are in alcohol or substance abuse treatment, pregnant women, elderly people, cancer patients, people in poverty, and high-income earners. However, researchers

should carefully consider the unique context of their study.

The results of our research also provide some important insights into mapping infectious diseases such as the COVID-19 (the novel coronavirus disease of 2019) pandemic in 2020. During the pandemic, mapping the infectious disease is one of the effective ways for public health authorities to communicate with the general public (Blendon et al., 2008; Boulos & Geraghty, 2020). For example, some countries like South Korea publicly released maps with great detail (e.g., building-level resolution) that display a patient's trajectories of several days before he/she is diagnosed with COVID-19 (Sonn & Lee, 2020). The rationale behind releasing such maps is that, by referring to the maps, people who might be in close contact with the infected person would know that they might be infected too and thus should be tested, which may play a critical role in controlling the pandemic (Wu & McGoogan, 2020). On the contrary, despite the potential social benefits of mapping patients' trajectories and making such maps publicly available, some countries decided not to release the trajectories because such disclosure may seriously violate patients' geoprivacy. For instance, as suggested by the results of Part 3 in this research, the geoprivacy of disease patients is especially important and should be carefully protected to avoid unintended negative consequences, such as stigmatization and discrimination.

All things considered, to effectively control a pandemic like COVID-19, it may be critical to balance between promoting public health (by releasing information) and protecting geoprivacy (by geomasking information) rather than exclusively choosing one over the other (Halpern, 2020; Harari, 2020). In this light, conducting surveys that are similar to ours may provide important insights into how public health agencies can wisely balance between disease control and geoprivacy protection. For example, as suggested by the results of Part 2 in our survey, public health agencies may consider applying geomasking methods to maps in order to protect the geoprivacy of patients while still providing useful information to the general public. However, we do not intend to argue that our results can be directly applied in public health contexts. Instead, we suggest that public health agencies may consider conducting surveys that are similar to our research before the next pandemic so that they can be fully prepared before an outbreak.

However, this research has several limitations that future studies should address. First, this research examined the perceived disclosure risk of a map displaying the private locations of the socially-vulnerable populations in a hypothetical setting. However, the responses may be biased because they do not reflect the perceived disclosure

risk of actual socially-vulnerable people. Second, this research did not evaluate the impact of other geomasking methods, such as affine transformation and location swapping, on participants' perceived disclosure risk. People may perceive disclosure risk differently when a different geomasking method is applied to a map. For example, the location swapping relocation method selects new relocated points that are already inhabited (Zhang et al., 2017). Thus, considering the survey results that people particularly feel discomfort with this kind of maps because of the possibility of falsely identifying innocent people, people may have higher perceived disclosure risks with the location swapping method than the random perturbation method. Therefore, future studies should examine the effect of various geomasking methods on the perceived disclosure risk of maps. Moreover, the perceived disclosure risk measured in the survey may be inflated because people's opinions on locational privacy may be reinforced while conducting the survey (Seidl et al., 2020). Future studies might explore experiments to test alternative sets of wordings of the survey items to optimize results and to test for any bias.

Acknowledgments

Junghwan Kim would like to thank the University of Illinois at Urbana-Champaign (UIUC) Space-Time Analysis and Research (STAR) Lab members (Lirong Kou, Dong Liu, Wataru Morioka, and Shuangshuang Qiu), Yonsei University Gateway to City (GTC) Society members (Younghun Bahk, Hoyeon Hwang, Minseok Kim, and Youngjoon Kim), Soomin Kim, Rebecca Martin, Swati Rastogi, and Minsoo Sung for their helpful feedback during a pilot study. The authors thank the anonymous reviewers for their thoughtful comments, which helped improve the paper considerably. The authors are particularly grateful for the editor's and the reviewers' efforts during the COVID-19 pandemic.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by a grant from the U.S. National Science Foundation (BCS-2025783).

Data availability statement

Due to the nature of this research, participants of this study had not agreed that their data can be shared publicly. As a result, the data used in this study cannot be shared with or made available to others.

ORCID

Junghwan Kim  <http://orcid.org/0000-0002-7275-769X>
 Mei-Po Kwan  <http://orcid.org/0000-0001-8602-9258>
 Margaret C. Levenstein  <http://orcid.org/0000-0002-9641-2725>
 Douglas B. Richardson  <http://orcid.org/0000-0003-1111-3249>

References

- Abowd, J. M., & Schmutte, I. M. (2019). An economic analysis of privacy protection and statistical accuracy as social choices. *American Economic Review*, 109(1), 171–202. <https://doi.org/10.1257/aer.20170627>
- Armstrong, M. P., & Ruggles, A. J. (2005). Geographic information technologies and personal privacy. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 40(4), 63–73. <https://doi.org/10.3138/RU65-81R3-0W75-8V21>
- 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-%23](https://doi.org/10.1002/(SICI)1097-0258(19990315)18:5<497::AID-SIM45>3.0.CO;2-%23)
- Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. Guilford Publications.
- Benisch, M., Kelley, P. G., Sadeh, N., & Cranor, L. F. (2011). Capturing location-privacy preferences: Quantifying accuracy and user-burden tradeoffs. *Personal and Ubiquitous Computing*, 15(7), 679–694. <https://doi.org/10.1007/s00779-010-0346-0>
- Blendon, R. J., Koonin, L. M., Benson, J. M., Cetron, M. S., Pollard, W. E., Mitchell, E. W., Weldon, K., & Herrmann, M. J. (2008). Public response to community mitigation measures for pandemic influenza. *Emerging Infectious Diseases*, 14(5), 778–786. <https://doi.org/10.3201/eid1405.071437>
- Boulos, M. N. K., Curtis, A. J., & AbdelMalik, P. (2009). Musings on privacy issues in health research involving disaggregate geographic data about individuals. *International Journal of Health Geographics*, 8(46), 1–8. <https://doi.org/10.1186/1476-072X-8-46>
- Boulos, M. N. K., & Geraghty, E. M. (2020). Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: How 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. *International Journal of Health Geographics*, 19(8), 1–12. <https://doi.org/10.1186/s12942-020-00202-8>
- Breslin, S., Shareck, M., & Fuller, D. (2019). Research ethics for mobile sensing device use by vulnerable populations. *Social Science & Medicine*, 232, 50–57. <https://doi.org/10.1016/j.socscimed.2019.04.035>
- Brownstein, J. S., Cassa, C. A., Kohane, I. S., & Mandl, K. D. (2006). An unsupervised classification method for inferring original case locations from low-resolution disease maps. *International Journal of Health Geographics*, 5(56), 1–7. <https://doi.org/10.1186/1476-072X-5-56>

- Curtis, A., Mills, J. W., Agustin, L., & Cockburn, M. (2011). Confidentiality risks in fine scale aggregations of health data. *Computers, Environment and Urban Systems*, 35(1), 57–64. <https://doi.org/10.1016/j.compenvurbsys.2010.08.002>
- 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(44), 1–12. <https://doi.org/10.1186/1476-072X-5-44>
- De Montjoye, Y. A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific Reports*, 3(1376), 1–5. <https://doi.org/10.1038/srep01376>
- Duncan, D. T., Kapadia, F., Regan, S. D., Goedel, W. C., Levy, M. D., Barton, S. C., . . . Halkitis, P. N. (2016). Feasibility and acceptability of global positioning system (GPS) methods to study the spatial contexts of substance use and sexual risk behaviors among young men who have sex with men in New York City: A p18 cohort sub-study. *PloS One*, 11(2), 1–19. <https://doi.org/10.1371/journal.pone.0147520>
- Emam, K. E., Brown, A., & AbdelMalik, P. (2009). Evaluating predictors of geographic area population size cut-offs to manage re-identification risk. *Journal of the American Medical Informatics Association*, 16(2), 256–266. <https://doi.org/10.1197/jamia.M2902>
- The Federal Policy for the Protection of Human Subjects (The Common Rule), 45 C.F.R. 46 (1991). <https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/index.html>
- Fuller, D., Shareck, M., & Stanley, K. (2017). Ethical implications of location and accelerometer measurement in health research studies with mobile sensing devices. *Social Science & Medicine*, 191, 84–88. <https://doi.org/10.1016/j.socscimed.2017.08.043>
- Ghinita, G., Zhao, K., Papadias, D., & Kalnis, P. (2010). A reciprocal framework for spatial k-anonymity. *Information Systems*, 35(3), 299–314. <https://doi.org/10.1016/j.is.2009.10.001>
- Groff, E. R., Kearley, B., Fogg, H., Beatty, P., Couture, H., & Wartell, J. (2005). A randomized experimental study of sharing crime data with citizens: Do maps produce more fear? *Journal of Experimental Criminology*, 1(1), 87–115. <https://doi.org/10.1007/s11292-004-6465-8>
- Gutmann, M. P., Witkowski, K., Colyer, C., O'Rourke, J. M., & McNally, J. (2008). Providing spatial data for secondary analysis: Issues and current practices relating to confidentiality. *Population Research and Policy Review*, 27(6), 639–665. <https://doi.org/10.1007/s11113-008-9095-4>
- Haley, D. F., Matthews, S. A., Cooper, H. L., 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>
- Halpern, S. (2020, April 27). Can we track COVID-19 and protect privacy at the same time? *The New Yorker*. <https://www.newyorker.com/tech/annals-of-technology/can-we-track-covid-19-and-protect-privacy-at-the-same-time>
- 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>
- Harari, Y. N. (2020, March 20). The world after coronavirus. *Financial Times*. <https://www.ft.com/content/19d90308-6858-11ea-a3c9-1fe6fedcca75>
- Kar, B., Crowsey, R. C., & Zale, J. J. (2013). The myth of location privacy in the United States: Surveyed attitude versus current practices. *The Professional Geographer*, 65(1), 47–64. <https://doi.org/10.1080/00330124.2012.658725>
- Keßler, C., & McKenzie, G. (2018). A geoprivacy manifesto. *Transactions in GIS*, 22(1), 3–19. <https://doi.org/10.1111/tgis.12305>
- Ketelaar, P. E., & VanBalen, M. (2018). The smartphone as your follower: The role of smartphone literacy in the relation between privacy concerns, attitude and behaviour towards phone-embedded tracking. *Computers in Human Behavior*, 78, 174–182. <https://doi.org/10.1016/j.chb.2017.09.034>
- Kim, J., & Kwan, M.-P. (2019). Beyond commuting: Ignoring individuals' activity-travel patterns may lead to inaccurate assessments of their exposure to traffic congestion. *International Journal of Environmental Research and Public Health*, 16(1). Advance online publication. <https://doi.org/10.3390/ijerph16010089>
- Kim, J., & Kwan, M.-P. (2020). How neighborhood effect averaging may affect assessment of individual exposures to air pollution: A study of ozone exposures in Los Angeles. *Annals of the American Association of Geographers*, 1–20. <https://doi.org/10.1080/24694452.2020.1756208>
- Kim, J., & Lee, B. (2019). More than travel time: New accessibility index capturing the connectivity of transit services. *Journal of Transport Geography*, 78, 8–18. <https://doi.org/10.1016/j.jtrangeo.2019.05.008>
- 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., & 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., & 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>
- Kwan, M.-P. (2004). GIS methods in time-geographic research: Geocomputation and geovisualization of human activity patterns. *Geografiska Annaler. Series B, Human Geography*, 86(4), 267–280. <https://doi.org/10.1111/j.0435-3684.2004.00167.x>
- Kwan, M.-P. (2012). The uncertain geographic context problem. *Annals of the Association of American Geographers*, 102(5), 958–968. <https://doi.org/10.1080/00045608.2012.687349>
- 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>
- Lee, M., Chun, Y., & Griffith, D. A. (2019). An evaluation of kernel smoothing to protect the confidentiality of individual locations. *International Journal of Urban Sciences*, 23(3), 335–351. <https://doi.org/10.1080/12265934.2018.1482778>
- Leitner, M., & Curtis, A. (2004). Cartographic guidelines for geographically masking the locations of confidential point data. *Cartographic Perspectives*, 49(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>
- McCarthy, L., Delbosc, A., Currie, G., & Molloy, A. (2017). Factors influencing travel mode choice among families with young children (aged 0–4): A review of the literature. *Transport Reviews*, 37(6), 767–781. <https://doi.org/10.1080/01441647.2017.1354942>
- McLafferty, S. (2004). The socialization of GIS. *Cartographica*, 39(2), 51–53. <https://doi.org/10.3138/F333-6V74-815U-4631>
- Mirzazadeh, A., Grasso, M., Johnson, K., Briceno, A., Navadeh, S., McFarland, W., & Page, K. (2014). Acceptability of Global Positioning System technology to survey injecting drug users' movements and social interactions: A pilot study from San Francisco, USA. *Technology and Health Care*, 22(5), 689–700. <https://doi.org/10.3233/THC-140838>
- Privitera, G. J. (2011). *Statistics for the behavioral sciences*. Sage.
- Reich, B. J., & Haran, M. (2018). Precision maps for public health. *Nature*, 555(7694), 32–33. <https://doi.org/10.1038/d41586-018-02096-w>
- Richardson, D. B., Kwan, M.-P., Alter, G., & McKendry, J. E. (2015). Replication of scientific research: Addressing geoprivacy, confidentiality, and data sharing challenges in geospatial research. *Annals of GIS*, 21(2), 101–110. <https://doi.org/10.1080/19475683.2015.1027792>
- Richardson, D. B., Volkow, N. D., Kwan, M.-P., Kaplan, R. M., Goodchild, M. F., & Croyle, R. T. (2013). Spatial turn in health research. *Science*, 339(6126), 1390–1392. <https://doi.org/10.1126/science.1232257>
- 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., Clarke, K. C., & Nara, A. (2020). Please enter your home location: Geoprivacy attitudes and personal location masking strategies of internet users. *Annals of the American Association of Geographers*, 110(3), 1–20. <https://doi.org/10.1080/24694452.2019.1654843>
- Seidl, D. E., Jankowski, P., & Tsou, M. H. (2016). Privacy and spatial pattern preservation in masked GPS trajectory data. *International Journal of Geographical Information Science*, 30(4), 785–800. <https://doi.org/10.1080/13658816.2015.1101767>
- Singer, E. (1978). Informed consent: Consequences for response rate and response quality in social surveys. *American Sociological Review*, 43(2), 144–162. <https://doi.org/10.2307/2094696>
- Slovic, P., Fischhoff, B., & Lichtenstein, S. (1980). Facts and fears: Understanding perceived risk. In R. Schwing & W. Albers (Eds.), *Societal risk assessment* (pp. 181–216). Springer. https://doi.org/10.1007/978-1-4899-0445-4_9
- Sonn, J., & Lee, J. K. (2020). The smart city as time-space cartographer in COVID-19 control: The South Korean strategy and democratic control of surveillance technology. *Eurasian Geography and Economics*. Advance online publication. <https://doi.org/10.1080/15387216.2020.1768423>
- Stark, J., Bartana, I. B., Fritz, A., Unbehauen, W., & Hössinger, R. (2018). The influence of external factors on children's travel mode: A comparison of school trips and non-school trips. *Journal of Transport Geography*, 68, 55–66. <https://doi.org/10.1016/j.jtrangeo.2018.02.012>
- VanWey, L. K., Rindfuss, R. R., Gutmann, M. P., Entwisle, B., & Balk, D. L. (2005). Confidentiality and spatially explicit data: Concerns and challenges. *Proceedings of the National Academy of Sciences*, 102(43), 15337–15342. <https://doi.org/10.1073/pnas.0507804102>
- Wang, J., & Kwan, M.-P. (2020). Daily activity locations k-anonymity for the evaluation of disclosure risk of individual GPS datasets. *International Journal of Health Geographics*, 19(7), 1–14. <https://doi.org/10.1186/s12942-020-00201-9>
- Wang, Z., Liu, L., Zhou, H., & Lan, M. (2019). How is the confidentiality of crime locations affected by parameters in kernel density estimation? *ISPRS International Journal of Geo-Information*, 8(12), 1–12. <https://doi.org/10.3390/ijgi8120544>
- Wu, Z., & McGoogan, J. M. (2020). Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. *JAMA*, 323(13), 1239–1242. <https://doi.org/10.1001/jama.2020.2648>
- Zandbergen, P. A. (2014). Ensuring confidentiality of geocoded health data: Assessing geographic masking strategies for individual-level data. *Advances in Medicine*, 2014(567049), 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>