Survey Research Methods (2017)
Vol. 11, No. 3, pp. 307-327
doi:10.18148/srm/2017.v11i3.6733

ISSN 1864-3361
http://www.surveymethods.org

Integrating Geographic Information into Survey Research: Current Applications, Challenges, and Future Avenues

Matthias Bluemke
GESIS – Leibniz Institute for the Social Sciences
Mannheim, Germany

Bernd Resch
Department of Geoinformatics
University of Salzburg, Austria
and Center for Geographic Analysis
Harvard University, Cambridge, U.S.A.

Clemens Lechner
GESIS – Leibniz Institute for the Social Sciences
Mannheim, Germany

René Westerholt
GIScience Research Group
Institute of Geography
Heidelberg University, Germany

Jan-Philipp Kolb
GESIS – Leibniz Institute for the Social Sciences
Mannheim, Germany

Geographic information science (GIScience) offers survey researchers a plethora of rapidly evolving research strategies and tools for data acquisition and analysis. However, the potential for incorporating geographic information systems (GIS) tools into traditional survey research has not yet been fully appreciated by survey researchers. In this article, we provide a comprehensive overview of recent advances and challenges in leveraging this potential. First, we present state-of-the-art applications of GIS tools in traditional survey research, drawing mainly on examples from psychological survey research (e.g., socioecological psychology). We also discuss innovative GIS tools (e.g., wearables) and GIScience methods (e.g., citizen sensing) that expand the scope of traditional surveys. Second, we highlight a number of challenges and problems (e.g., choice of spatial scale, statistical issues, privacy concerns) and – where possible – suggest remedies. With increasing awareness of the potential that GIS tools hold for survey research, and intensified dialogue between researchers from both sides, more fruitful collaboration appears within reach.

Keywords: GIScience; GIS tools; socioecological psychology; wearables; social media

1 Introduction and Overview

1.1 Relevance of Geospatial Aspects

Recent advances in integrating geographic information science (GIScience) into psychology and survey methodology may be considered evolutionary by some researchers and revolutionary by others. Some observers view these advances as a paradigmatic shift that justifies the term "spatial turn" (e.g., Richardson et al., 2013). At any rate, the sheer number of new research strategies and geographic information system (GIS) tools can be daunting, and it is hard for those outside GIScience to keep pace with recent developments. Perhaps for this reason, psychology has, for the most part, been slow to adopt some of the promising innovations offered by GIScience and related disciplines (see Appendix for a glossary of terms). There is the danger of a widening gap, if not detachment, between scientific communities in terms of concepts, methods, and tools.

To avoid this pitfall, we think it is vital to inform survey researchers in general, and psychological researchers in particular, about the methodological potential that GIS tools (e.g., techniques for acquiring, analyzing, and visualizing geographic data) hold. Although in-depth discussions of single geospatial techniques abound, an up-to-date integrative overview of these innovations is absent from the literature. The present paper is the first to offer such an overview from the perspective of GIScientists and psychological researchers practically involved in the application of GIS tools in survey research. We restrict our overview to those GIScience

Contact information: Matthias Bluemke, GESIS – Leibniz Institute for the Social Sciences, B2/15, 68159 Mannheim, Germany (email: Matthias.Bluemke@gesis.org)

techniques that concern the analysis of survey data, that is, techniques that augment traditional surveys by incorporating geospatial information, or that complement traditional surveys by providing new forms of data and data analysis. We do not consider how GIScience techniques can be used to improve traditional survey methodology (e.g., for monitoring field work and interviewer behavior, or for designing sampling frames and weighting schemes).

The purpose of our paper is twofold. First, it is intended as a starting point for survey researchers interested in applying innovative GIS tools in their work. Despite the rise of other approaches (e.g., neuroscientific methods, computer-based cognitive tests), questionnaires and population surveys are still among the most widely used tools for collecting data on individuals and social groups, notably in psychology. Second, we hope to foster an informed discussion between two inherently methodological disciplines – GIScience and survey research. We encourage researchers from both sides to critically examine and make use of solutions offered by each field, to develop a common theoretical framework, and to adopt each others' insights, tools, and methods. Such an intensified dialogue, we believe, may challenge our traditional understanding of survey data and methodology, thus paving the way for future innovations in survey research (see Arias & Warf, 2009).

Yet why should survey researchers consider spatial aspects at all? Geospatial information might be useful at various stages of the survey design. Surveys can be supplemented with geographic coordinates, such as the location at which a respondent completed the questionnaire (i.e., the point of origin) or where the respondent predominantly lives (i.e., place of residence). Either respondents' exact geographic locations or their approximated locations (e.g., via regional codes) may be available. These geographic coordinates can then be used to incorporate contextual data into subsequent analyses. For instance, socioeconomic data on households in a neighborhood, or regional divorce rates as a proxy for individualization in a society, may be contextual variables and may complement individual-level data analyses (Lechner, Obschonka, & Silbereisen, 2017). Geographic visualization techniques may also provide additional insights into the spatial distribution of survey results. By analyzing and mapping biophysiological data from study participants who wear trackable devices while moving through space (Tröndle, Greenwood, Kirchberg, & Tschacher, 2014), or by plotting Twitter-based information almost in real time to geographic maps (Curini, Iacus, & Canova, 2015), researchers can follow social processes at unprecedented spatiotemporal resolutions.

1.2 Survey Concepts

It is vital to note that the understanding of the term "survey" often differs between survey researchers and GIScientists. For survey researchers, the survey questions – augmented with georeferenced data – are typically meant to reveal information about individuals. Spatial information (e.g., location-related information) is added to better understand humans or social systems. Survey researchers often operate with a survey definition that involves the collection of data from a sample of elements drawn from a well-defined population through the use of a questionnaire (Visser, Krosnick, & Lavrakas, 2000). Hence, "a survey can be seen as a research strategy in which quantitative information is systematically collected from a relatively large sample taken from a population" (de Leeuw, Hox, & Dillman, 2008, p. 3). GIScientists, by contrast, typically focus on extracting information about places and spaces (e.g., street corners, cities, regions, or countries) and the phenomena and developments they undergo over time. For them, survey questions add information from a human perspective to better understand the environment. Yet GIScientists often endorse a rather minimalistic definition of "survey" as a method of gathering information from any sample of individuals (Scheuren, 2004). Their broader term may include gathering data en passant from social media users, rather than collecting answers from survey respondents.

Section 2 illustrates the breadth of potential applications of GIScience methods and GIS tools in survey research. These applications include the augmentation of classic survey data with georeferenced information, on the one hand, and new techniques to obtain spatiotemporally distributed data from GIScience, on the other. In section 3, we outline a number of challenges shared by most of these applications. They include, for example, spatial scale usage, the modifiable areal unit problem, the arbitrary nature of maps and visualization techniques, potential pitfalls in the analysis of contextual data, fallacies when dealing with different levels of analysis (individual and aggregate data); issues surrounding user-generated data, and privacy and data protection issues. Where possible, we make methodological recommendations to deal with these challenges. Finally, also in section 3, we briefly consider how traditional survey data and methodology may be of benefit to GIScience.

2 Recent Applications of GIS Tools to Augment Survey Data

Before we begin our overview of recent applications, let us follow Agnew (2011) and Tuan (1977) and clarify the distinctions between important GIScience terms that readers may or may not be familiar with: space, location, and place. Space is understood as an abstract, non-semantically enriched geographic space spanning planet Earth, in which processes of interest occur. Location demarcates a specific point or area in this space, mostly delimited by crisp boundaries, which can be represented in GIS. In contrast, place is defined as space infused with human meaning. As this meaning is almost never specified with perfect intersubjectivity, its borders are often fuzzy and ambiguous. Several scientific disciplines deal with these concepts. The methodological companion to geography is geographic information science, or GIScience (Goodchild, 1992, 2010). GIScience and GIS tools are closely related, and they provide partly overlapping innovations (see Appendix).

Augmenting classic survey data with georeferenced data represents a first way in which geographic information about individuals and their backgrounds is utilized in the social sciences (Hoffmeyer-Zlotnik, 1994; Okner, 1972; Schnell, 2013b). Survey datasets that are geocoded – that is, datasets that contain one or more variables assigning a geographic location (e.g., an exact location or a more coarse location such as a postal code or an administrative unit) to each response unit – can be merged with geotagged contextual information, thereby greatly enhancing the value of these augmented datasets to investigate new research questions (Meyer & Bruderer Enzler, 2013; Okner, 1972; Schnell, 2013b). To take the Swiss Environmental Survey as an example, regional statistics on environmental factors (e.g., pollution, emissions) were linked to respondents' subjective impressions of environmental stress to gain a better understanding of the relationship between objective contextual variables and participants' subjective responses (Diekmann & Meyer, 2010). By adding contextual information to individual respondent data, cross-level relationships can be explored (Hoffmeyer-Zlotnik, 2013; RatSWD – Rat für Sozial- und Wirtschaftsdaten [German Data Forum], 2012). Rich contextual data are now offered by various public institutions (e.g., register, census, and economic data), private organizations and companies (e.g., operational and customer-tracking data), and accumulated sources (e.g., social media, representative surveys; for a comparison, see Hüttenrauch (2016). Some public-use surveys, such as the European Social Survey (ESS), already include a large number of contextual variables at different geographic levels (e.g., national and regional migration, or unemployment rates) in their data distributions, and make these data readily available to researchers. Furthermore, various kinds of geospatial data have become publicly available, for instance, authoritative topographic data (e.g., OpenStreetMap). The following sub-sections describe possible applications of these contextual data in research.

2.1 Socioecological Psychology

The emerging field of socioecological psychology, also known as geographic(al) psychology (Oishi, 2014; Rentfrow, 2014, for reviews, see), utilizes the new possibilities of integrating contextual information and traditional survey data. Socioecological psychology directs attention to how objective (as opposed to perceived) features of macro-level social ecologies (i.e., physical, interpersonal, economic, or political environments) shape human behavior, cognition, and emotion – and how human behavior, in turn, gives rise to changing social ecologies ("niche construction"). Extant socioecological studies mostly relate individual-level psychological outcomes to socioecological variables that are measured at (not aggregated to) the national or regional level and that assume the role of a predictor of individual-level variability (e.g., Talhelm et al., 2014) or a moderator of individual-level relationships (e.g., Jokela, Bleidorn, Lamb, Gosling, & Rentfrow, 2015; Lechner et al., 2017). For example, Talhelm et al. (2014) were able to show how the agricultural legacy of regions in China shapes the cultural and psychological traits of these regions' inhabitants until the present day; they found that a history of farming rice was linked to more interdependent traits, whereas a history of farming wheat was linked to more independent cultural patterns. Other socioecological studies extend this focus, investigating the spatial distribution of psychological constructs and their contextual-level correlates. The level of analysis in the latter strand of studies is thus a geographic one: Individual survey responses (such as answers to a Big Five personality battery) per geographic unit are aggregated in order to map them to spatial contexts and link them to each other (Rentfrow, Jokela, & Lamb, 2015) or, alternatively, to external data sources such as health statistics (Kitchen, Williams, & Chowhan, 2012) or entrepreneurship rates (Obschonka, Schmitt-Rodermund, Silbereisen, Gosling, & Potter, 2013).

Although socioecological psychology is still in a nascent state, it already exerts a noticeable influence on psychological theorizing. Socioecological studies are contributing to a gradual shift in the traditional focus on the individual toward a more environmentally informed understanding of the discipline's key phenomena. (Arguably, this marks a veritable "spatial turn," especially in the fields of personality psychology and social psychology.) While this development opens up new avenues for collaboration with disciplines at the interface of human behavior and geography (Oishi, 2014; Rentfrow, 2014), it also brings methodological challenges, which will be discussed later.

2.2 Survey Responses as a Function of Georeferenced Indicators

Geographic context can also be used to identify (and correct for) sources of variance in survey responses. Depending on one's focus, such variance may represent either explained variability or nuisance variance in survey responses. For example, participants' life satisfaction scores might be influenced by (a) aspects of the natural environment, such as the greenness of neighborhoods (Leslie, Sugiyama, Ierodiaconou, & Kremer, 2010), (b) the built environment (McGinn, Evenson, Herring, Huston, & Rodriguez, 2007), or (c) circumstances of the survey location, such as indoor versus outdoor interviewing (Iosa, Fusco, Morone, & Paolucci, 2012).

A prime application of this approach is the influence of weather on mood and well-being. With governments increasingly adopting well-being as a policy target, subjective ratings of life satisfaction and happiness are often important indicators that complement panel data on regional and macroeconomic factors (Schyns, 1998; for a recent Eurobarometer analysis, see Brulé & Veenhoven, 2014). For well-being to inform public policy choices, one would like to be sure that any regional differences in average well-being ratings are truly related to economic prosperity and other policy-relevant factors, rather than being driven by "nuisance" factors such as the climate at survey locations (Brulé & Veenhoven, 2015; Rehdanz & Maddison, 2005) or by transient weather conditions during interviews (Schimmack, Diener, & Oishi, 2002).

Laboratory and field evidence has shown that judgments of life satisfaction are influenced by the reported weather conditions (Schwarz & Clore, 1983), and that ambient temperature ratings, in turn, depend on people's current mood (Messner & Wänke, 2011). Although this cross-sectional evidence was challenged by panel data (Lucas & Lawless, 2013; Schmiedeberg & Schröder, 2014), more recent panel data providing detailed information on all relevant weather variables at the precise location and time of an interview have, in fact, revealed variation in life satisfaction scores as a function of weather (Feddersen, Metcalfe, & Wooden, 2016). Beyond well-being, studies have shown that Big Five trait ratings can also be influenced by contextual factors such as weather (Rammstedt, Mutz, & Farmer, 2015). Increasing spatial granularity yields better evidence on climatic and weather influences on survey responses. It may become possible to purge respondents' mood, well-being, or life satisfaction ratings of unwarranted nuisance variance and to obtain unbiased scores that offer a more solid ground for policy decisions.

2.3 Experience Sampling in Dynamic Contexts

Methods of studying individuals in their natural settings – often in real time, on repeated occasions, and free of retrospective biases – offer tremendous potential for survey research. One such method – and one that has recently gained some popularity – is the experience sampling, or event sampling, method (ESM; see H. T. Reis & Gable, 2000), which allows respondents to be surveyed in their natural environments on repeated occasions (Hektner, Schmidt, & Csikszentmihalyi, 2007; Larson & Csikszentmihalyi, 1983). ESM prompts participants (e.g., via mobile devices) to take a survey at fixed time intervals or randomly throughout the day. In this way, the likelihood of events, the base rate of behaviors, or the prevalence of feelings can be surveyed amidst temporal fluctuations of experiences and dynamic transitions between places. The recent emergence of mobile electronic devices allows even large crowds to be observed at nearly any time and place so as to investigate relationships with increased ecological validity (Shiffman, 2007).

There are three typical ESM procedures – signal-contingent (survey after notification via pager or SMS text message), event-contingent (recording data after predefined events have occurred), and interval-contingent (data acquisition after periods of time have passed) – whose respective (dis-)advantages have been described elsewhere (Conner & Barrett, 2012). Here, we would like to stress that adding a spatial layer to this threefold distinction allows context-aware ESM to be used. Augmenting ESM data with location data (e.g., Global Positioning System, GPS, coordinates gathered by users' mobile devices) offers a convenient way of conducting surveys at predetermined locations (which allows further data to be gathered about these locations as socially relevant places). Location data might help to explain individuals' attitudes and behaviors. These data include not only static factors, such as types of buildings or population density, or rather stable influences such as unemployment rates, but also each individual's exposure to noise at specific workplaces, stressful traffic encounters at specific intersections, etc.

So far, traditional population surveys mostly abstract from the dynamic contexts in which respondents generate their responses, or in which they have experiences that they report only later. From this perspective, survey samples must first and foremost mirror the population. However, it is worthwhile reflecting on the fact that any interview represents a mere snapshot of a respondent's state of mind generated within a specific spatiotemporal slice of the environment. Population surveys typically leave such short-term volatility and spatial dynamics of survey responses unmonitored. By linking ESM data to rich contextual information such as location and time, survey research proceeds to the next stage, where human characteristics are explained as a function of idiosyncratic events, personal contexts, and participants' spatial transitions. Research in health-related and occupational fields has started to incorporate these new possibilities (Richardson et al., 2013; Sonnentag, Binnewies, & Ohly, 2013) – for instance, by using ESM to investigate whether environmental factors, such as rare exposure to nature, might influence mental health (Reichert et al., 2016).

2.4 Objective Data Capture by Means of Wearables

While ESM focuses on the subjective experiences that respondents have over any pre-specified time span, these data can be amended with objective data on the same individuals. There has recently been a rapid rise in the use of wearable sensors to measure a number of physiological parameters (e.g., heart rate variability, blood pressure, or skin conductance, Swan, 2012). These sensors, together with the increasing penetration rate of smartphones across age groups, have paved the way toward virtually ubiquitous data acquisition and have opened up new opportunities for obtaining information about the environment (Triantafyllidis et al., 2015).

For example, the so-called quantified-self movement promotes the use of sensor technology for acquiring data about one's own daily life, ranging from concrete physiological parameters to rather abstract parameters such as physical performance and associated affective consequences (e.g., emotional states). This movement is reinforced by the rapid development of wearable sensors that allow for continuous surveillance of everyday activities and daily routines (Swan, 2013). Although people are joining the quantified-self movement mainly to achieve self-awareness through self-monitoring (Ayobi, Marshall, & Cox, 2016), it has also led to rising awareness of physiological sensor devices among the wider public. As a result, citizens' familiarity with the use of sensors has dramatically increased. This is of particular importance for survey research, as most of the sensor-based quantified-self applications are explicitly geolocated, which allows survey data to be complemented with additional data from wearables at high temporal and spatial resolution, thereby yielding information that cannot be obtained by simply asking survey questions. Physiological signals obtained from wearable sensors can then be used to make inferences about individual experiences that are associated with, or can be mapped to, events and places in the environment. For instance, one could compare a survey intended to identify dangerous traffic intersections in a city with a study that captures heart rate and blood pressure data from drivers, cyclists, and pedestrians. This may help to identify city areas or spots of increased stress levels other than those identified by participants' subjective ratings. Sufficiently rich data may allow the emotional experiences of future pedestrians, cyclists, or motorists at the same location to be predicted, thus enabling a more citizen-centric planning of city infrastructure (Resch, Summa, Sagl, Zeile, & Exner, 2015).

2.5 Humans as Proactive Sensors (Rather Than as Respondents)

User-generated data are by no means limited to physiological data from wearable sensors that are collected for a specific purpose and under the researcher's control. A number of new approaches elicit, observe, or analyze information generated by individuals (and groups) that are hardly under the control of a researcher. Instead, participants act more and more as researchers of their own affairs, and thus control over the data-generation process is increasingly left to them. GIScience capitalizes on this trend.

Citizen sensing describes a unique measurement approach in which persons do not merely deliver reports but rather act as non-technical, context-aware sensors with situational intelligence and extensive background knowledge about their present location (Resch, 2013). Specifically, citizens are asked to provide their impressions, perceptions, and observations about a well-defined issue with explicit reference to geographic space. Akin to ESM, people provide their subjective recordings through eDiaries, which are designed to be context-aware. Contextualized reports can be gathered through dedicated smartphone apps (Triantafyllidis et al., 2015).

A recent example is the collection of citizens' subjective feelings and emotions about different places in the city (Resch, Sudmanns, et al., 2015). Participants who move in geographic space are equipped with a smartphone app for reporting sensations and impressions – for instance, about traffic safety or public safety. Each dataset is associated with location and timestamp, which enables spatiotemporal analysis of the data. Apart from an immediate glimpse of a geographic context, this allows for an analysis of changes in ecosystems as continuously monitored through citizen-sensing technologies. Rather than acting as mere respondents to questionnaires, this approach empowers participants to proactively report not only on their spatial transitions but also on changes in ecosystems themselves.

One methodological implication of this survey method is that data are unlikely to be fully reproducible (Sagl & Resch, 2014). From a survey research perspective, data reproducibility may not even be a goal (the data always present slices of information tied to time and context); yet from a GIScience perspective, the aim is to obtain (stable or reliable) information about the environment. Moreover, given users' proactive role in generating responses, the sampling and data generation processes are not necessarily controlled, and the observations are highly idiosyncratic. The reliability of such a measurement procedure is a far cry from that of representative population surveys with regular waves, or calibrated wearable sensors that produce measurements in well-defined physical settings. With high volatility in the sampling process, one problem is how to generalize to a whole population from self-selected samples who themselves determine what snapshots in time to deliver and when. In section 3, we elaborate on challenges common to survey research and GIScience.

2.6 Spatiotemporally Distributed Information in Social Media

Social media represent a useful resource for complementing survey data (Hill, Dean, & Murphy, 2013; Murphy et al., 2014). In contrast to citizen sensing, the analysis of data from social media does not require additional survey infrastructure (eDiary apps, digital surveys, etc.). Rather than surveying individuals about specific locations, this approach analyzes aggregated, anonymized data from collective sources such as Flickr, Twitter, Foursquare, or the mobile phone network (Resch, 2013). In this manner, information can be gained about the situational awareness of human environments and temporal dynamics on the basis of human communication, without attributing data to specific individuals. Social media posts reveal people's thoughts, emotions, or activities in geographic space, time, and linguistic space (Steiger, Resch, & Zipf, 2016).

Although unprompted social media posts cannot be considered to be interviews in the formal sense of the word, people still provide "answers" (to questions that are not asked by an interviewer) by stating their perceptions and opinions. Yet, in contrast to classic surveys, it is more difficult to correctly map the target population. It may be difficult (albeit possible) to gauge opinions among specific subpopulations (Pötzschke & Braun, 2016). However, it is almost impossible to get a representative picture of the entire general population. Therefore, social media data cannot replace targeted and structured surveys. Yet they are a useful extension, and they can yield additional insights (e.g., via content analysis and text mining) that are not bound to pre-formulated questions and researcher-determined response categories. Instead, the focal topic can be determined by the social media user; the information obtained there is not elicited in a "synthetic situation" of a formal interview; and with regard to both the amount of content and its format, the user can express him- or herself freely. To provide an application example, the topic of sentiment analysis is currently gaining momentum. It deals with (the strength of) positive, negative, or neutral sentiments as conveyed by the polarity of words, sentences, or documents chosen by social media users (B. Liu & Zhang, 2012). Newer approaches (Resch, Summa, Zeile, & Strube, 2016) automatically extract from Twitter tweets and posts from other social network sites affective content that corresponds to the fundamental model of basic emotions (anger, disgust, fear, happiness, sadness, and surprise; Ekman & Friesen, 1971) or the refined model of four basic emotions (happiness, sadness, fear, and anger; Jack, Garrod, & Schyns, 2014).

Leveraging user-generated social media data has one major advantage over traditional surveys: the possibility of near real-time analysis. Analyzing user-generated data allows large-scale environmental, social, and geographic developments to be investigated "in the now," rather than after they occur. This kind of continuous cross-sectional monitoring – with unknown changes in the population that produces the data – is far from the quality of surveying a panel repeatedly in waves. However, it partly mitigates some shortcomings of traditional surveys, such as their low temporal resolution. Recent examples demonstrate the suitability of social media data in applications such as earthquake detection (Crooks, Croitoru, Stefanidis, & Radzikowski, 2013; Sakaki, Okazaki, & Matsuo, 2010, 2013) or the analysis of political sentiment (Caldarelli et al., 2014; Vasiliu et al., 2016; Wang, Can, Kazemzadeh, Bar, & Narayanan, 2012).

Furthermore, in addition to unprecedented temporal granularity, collective sources often provide spatially distributed data. Together with the high temporal resolution, a vast amount of data is available, even though the number of explicitly geolocated messages is limited. For example, depending on the topic under investigation, a maximum of 10% of all Twitter posts are currently georeferenced (Anselin & Williams, 2016). However, the majority of social media posts occur in situ – that is, posted content often refers explicitly to local phenomena even when no geocoded tags accompany the posts (Cuevas, Gonzalez, Cuevas, & Guerrero, 2014).

User-generated data in general – and social media data in particular – are sources of big data. Due to the opt-in mechanism, such data may suffer from strong self-selection and may not represent the population in its entirety; yet they can well be used to monitor urban, social, and environmental processes (Anstead & O'Loughlin, 2015). Consequently, they have been recognized by the World Health Organization as a further means of monitoring health developments at the population level (Conway & O'Connor, 2016).

3 Challenges and Recommendations

Although the applications of geodata and GIS tools discussed thus far open up promising new research avenues, there are a number of challenges and pitfalls that survey researchers interested in applying these applications in their own research must bear in mind. Some are well known among GIScientists, but less so among survey researchers (and vice versa). In this section, we discuss these challenges and pitfalls and, where possible, suggest some remedies.

3.1 Spatial Scale

Spatial scale is a central issue for GIScientists, and thus for spatial data acquisition and analysis. Scale may refer to different components of a geographic analysis, such as the level of geographic detail at which observations are made ("sampling scale"), the spatial range at which processes of interest operate ("phenomenon scale"), or the degree of abstraction of a spatial analysis ("analysis scale"; Dungan et al., 2002; Ruddell & Wentz, 2009). In a more technical sense, scale comprises grain (the smallest distinguishable parts possible) and extent (size of the study area; Turner, Dale, & Gardner, 1989). This technical use has a geometric interpretation of scale. It prevails in physical geography, but is often inappropriate when investigating social processes through surveys. Socially meaningful spatial scales, such as neighborhood, city, region, and nation, are often better suited for surveys (McMaster & Sheppard, 2004).

Spatial scale is not only an objective frame of reference for spatial phenomena but also a property of people's subjective perceptions of space that has a strong bearing on their answers to questions about local geographic phenomena. People perceive their spatial surroundings in unique ways and imbue them with individual meaning (see Dangschat, 2007). Respondents also use idiosyncratic spatial scales that are limited by their spatial perception capabilities (Wender, Haun, Rasch, & Bluemke, 2003). For example, when asked about their "local community," voters in the British Election Study thought of completely different areas, ranging from streets and suburbs, through regions, to whole countries (Fieldhouse, Green, Schmitt, Evans, & van der Eijk, 2014). Different mental systems are involved in perceiving phenomena at different spatial scales(see also Hegarty, Montello, Richardson, Ishikawa, & Lovelace, 2006; Montello & Golledge, 1999; Tversky, Morrison, Franklin, & Bryant, 1999), and a number of intrinsic and extrinsic factors influence the idiosyncratic spatial scale that people use (Witt, Proffitt, & Epstein, 2010), with well-established differences in spatial perception along the lines of gender (after puberty, males tend to perform better at spatial cognition; Weiss, Kemmler, Deisenhammer, Fleischhacker, & Delazer, 2003), age (younger people tend to underestimate distances; Sugovic & Witt, 2013), emotions (impacting on perception; Zadra & Clore, 2011), and properties of the physical environment (visual/acoustic cues; Iosa et al., 2012).

To illustrate, imagine an interviewer asking about an areal region such as an urban green space, a residential neighborhood, or a local community. Respondents will use their subjective representations of the region based on their idiosyncratic conception of space. Using their imaginations, they will mentally construe the region in question. Hence, any information gained when looking at space through the eyes of survey respondents is potentially susceptible to scale differences, because the location, shape, and size of any perceived areas will influence respondents' answers. For example, whether there are enough early childhood education and care centers in a suburb might crucially depend on the correct or incorrect inclusion of an institution into the referenced area of interest, necessitating an accuracy check of respondents' mental representations. One can also try to exploit respondents' expertise. For instance, citizens may include areas in their answers that have not been considered by experts, which may be beneficial in natural hazard analysis when the goal is to identify areas prone to urban floods (Klonner, Marx, Usón, & Höfle, 2016).

The fact that different respondents use different (and highly idiosyncratic) spatial scales when thinking about the physical environment – and that even one and the same respondent may resort to different spatial scales when thinking about his or her surroundings – implies that respondents' answers in any survey on the physical and social environments do not refer to a fixed, objective geographic frame of reference (see Yabiku, Glick, Wentz, Ghimire, & Zhao, 2017, in this issue). Spatial heterogeneity manifests itself as nuisance variance in the data, which increases the total survey error (Groves, 2010). More specifically, heterogeneity in respondents' spatial scales causes instabilities of estimated means of quantitative data (due to spatial trends or discrete spatial regimes) and variances (spatial heteroscedasticity; see Ord & Getis, 2012). Moreover, mixing highly different individual representations of arbitrary regional conceptions may not only render inferences based on such responses unreliable, or even bias-prone, but may even make numerical aggregates of respondents' answers difficult to interpret. This spatial-scale-related heterogeneity contributes to another form of (non-spatial) heterogeneity well-known in survey research, namely variability due to differing respondent and interviewer characteristics, or due to specific interactions between interviewers and respondents (e.g., Gabler & Lahiri, 2009; Schaeffer, Dykema, & Maynard, 2010; West, Kreuter, & Jaenichen, 2013).

There are different ways to address such scale-related issues: First, survey design requires careful construction of questions and questionnaires. All items that refer to a spatial phenomenon (e.g., "your neighborhood") should be as explicit as possible in order to lower the risk of ambiguities. One possible solution is to assist interviewees by providing a map of the area of interest whenever possible (i.e., standardizing the geographic presentation). However, this cannot always be smoothly integrated into the interview process. Moreover, it does not fully rule out the problem of different subjective geographic representations, and the maps provided to respondents restrict their answers to what is displayed on the map. As an alternative solution, mental maps and sketch maps can be used to document respondents' representations of geographic space (Boschmann & Cubbon, 2014). Mental maps are free-form drawings, and sketch maps are accurate maps augmented by the respondents, allowing the researcher to get a clearer picture of the respondent's inherent scale use (e.g., Coulton, Chan, & Mikelbank, 2010). Using mental maps minimizes the risk of accidentally mixing different scales during analysis and interpretation.

3.2 The Modifiable Areal Unit Problem

The modifiable areal unit problem (MAUP Openshaw, 1984) is a well-known issue that occurs when researchers aggregate data to reflect areal units. MAUP describes the fact that the choice of an – often arbitrary – spatial unit for an analysis can influence the outcomes of that analysis. Figure 1 illustrates how the three key characteristics of spatial units – location, shape, and scale – affect the analysis of underlying data points (e.g., from georeferenced surveys). For example, obesity rates can be meaningfully analyzed at the country level or at the state level. Depending on the level, we might see a different statistical pattern, either A or B, and draw the respective conclusions. Yet, even though the shape of geographic units (e.g., state borders) can certainly carry meaning for political and administrative bodies, it is still arbitrary and does not necessarily best reflect the aggregated data and the associated data-generating processes (from a causal or associative point of view). Given that the geographic units are arbitrary, so, too, is their position (encompassing specific locations) and the resulting distribution of data points to be aggregated. Even if the lattice of administrative units were transformed only marginally, the substantive conclusions that a researcher arrives at might change drastically.

MAUP is one of the long-standing and still unresolved issues in GIScience, and its ramifications are vividly discussed throughout different academic fields. Recent examples include investigations of human mobility (Mitra & Buliung, 2012; Xu, Huang, Dong, & Abdel-Aty, 2014), criminology (Gerell, 2017; Vogel, 2016), and forestry (Kozak & Szwagrzyk, 2016; Mas, Pérez Vega, Andablo Reyes, Castillo Santiago, & Flamenco Sandoval, 2015). For instance, Mitra and Buliung (2012) related properties of the built environment to children's mode of active/passive school transportation (i.e., whether they walk or cycle to school rather than taking the bus). Testing six different spatial configurations, they found that the sign as well as the size of the regression coefficients varied across scales and polygon forms used for defining the built-environment variables. Similarly, Vogel (2016) investigated the relationships between environmental factors and violence. Respondents were aggregated to reflect census tracts as well as units at the city block level. While the analyses revealed significant associations of the environmental factors, the effect of the geographic neighborhood did not exist at the block level, but only at the level of census tracts. As these examples show, the effect of MAUP on survey outcomes can be severe. MAUP should be taken into account by testing the replicability across different spatial units.

For survey researchers, MAUP matters (a) for the answers given by respondents on the basis of subjective representations of geography (mental representations of spatial phenomena), and (b) for the objective scale of georeferenced external data. MAUP is thus an important issue when it comes to augmenting classic survey data with external data such as census variables. External data are often in aggregated form, not free from geometric arbitrariness. The choice of the geographic level of analysis in many studies to date appears to have been driven largely by data availability rather than by a priori theoretical considerations of what constitutes meaningful context information. For instance, when analyzing the impact of covariates on respondents' answers, as in the previous example taken from Vogel (2016), it is clear that some information is available only through the census. If surveys release only geoidentifiers that refer to a rather coarse scale (e.g., state or county in the U.S.), researchers cannot look at covariates at a finer scale. In such cases, MAUP is essentially inevitable, but it should be kept in mind when drawing any inferences.

While the available level may be appropriate for many research questions, it would be desirable to devote greater attention to the choice and justification of the geographic level of analysis. Ideally, researchers would consider using the geographic level that appears most appropriate from a theoretical point of view, rather than the level for which data happen to be available. Moreover, they should report whether their substantive findings are robust across different geographic levels of analysis (e.g., Saib et al., 2014).

3.3 Maps, Distortion, Meaning, and Visualization

Another perennial issue in GIScience is the cartographic representation of the results generated through geospatial analysis. Unlike typical charts, maps can be used to bias communication in ways that survey researchers might be less familiar with. As Monmonier (1996) states, cartography has the power to bias the presentation of spatial information by generating "selective truth". Cartographic styles may strongly influence which information is ultimately perceived by the respondent. Ways of biasing maps include, inter alia, the choice of spatial aggregation and scale levels (related to MAUP), but also the selection of suggestive color ramps, the creation of categories and classes according to different criteria (natural statistical breaks, quantiles or units of standard deviations, etc.), the presentation of relative or absolute numbers, the influence of different coordinate reference systems (geographic vs. projected), or the choice of icons that represent geographic features. Figure 2 illustrates these effects by showing the same piece of information (crude U.S. birth rate in 2000) in different ways as originally described by Monmonier (2005). Thus, the information obtained from maps may be subject to arbitrariness, regardless of whether they are used for survey sampling, as visualization aids in a survey, or to draw inferences from an analysis. The complex questions involved have given rise to the scientific endeavor known as "critical cartography" (Crampton, 2010; Crampton & Krygier, 2005).

3.4 Analyzing Context and Using Georeferenced Contextual Data

Several problems exist that relate to the concept of "context". According to Dey (2001), context is defined as implicit or explicit information that is useful to characterize a situation. External, physical contexts are strongly associated with the objective physical environment, typically measured by physical sensors (e.g., room temperature). However, as noted in sections 2.4 and 2.5 above, contexts can also be described through respondents' subjective impressions at an individual level (Hong, Suh, Kim, & Kim, 2009) or by aggregating respondent data from wearables and tracking devices (Bettini et al., 2010; Sagl, Resch, & Blaschke, 2015). The spectrum of available technologies for capturing contextual information allows situational features to be quantified comprehensively and in unprecedented detail. These features include geographic aspects such as current environmental conditions (weather, air quality, etc.), the human perception of urban spaces, and the individual and collective behavioral

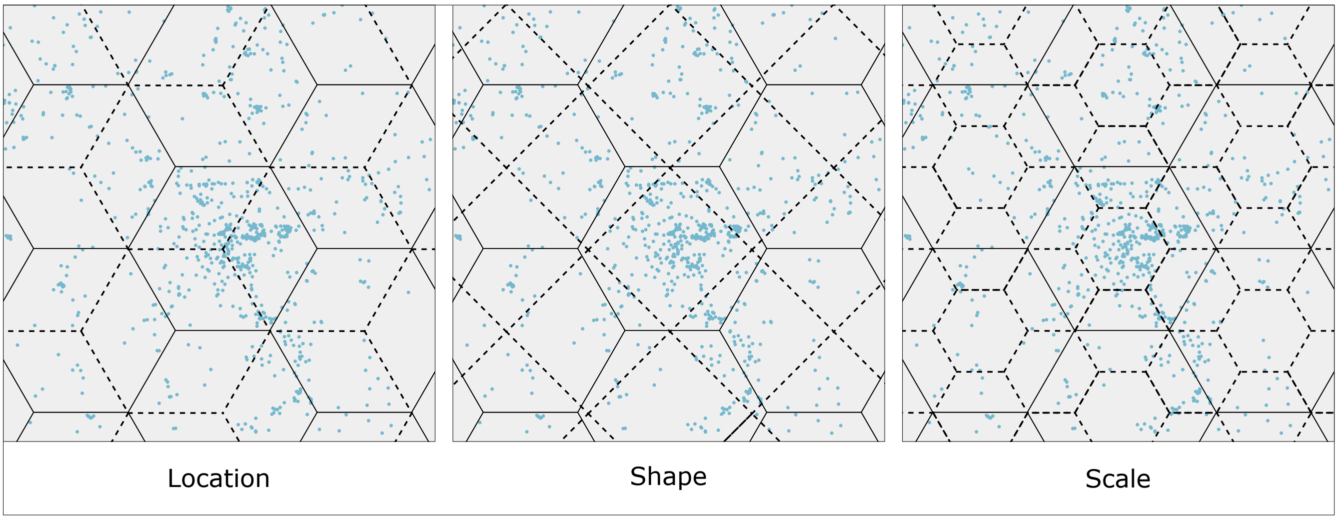

Figure 1. Three researcher-dependent aspects illustrating the influence of MAUP on data analysis: different locations resulting from shifted polygon positions, different distributions as a function of distinct polygon shapes, and different scale resolutions due to diverse polygon sizes (resulting in different aggregation effects).

responses to a range of functional settings including traffic infrastructures, open spaces, neighborhoods, or residential areas. All these settings are of considerable importance for human-environment interactions and citizens' quality of life, yet the number of characteristics with which to describe (and analyze) the impact of these contexts is manifold.

One limitation of most socioecological research to date is that – again due to data availability – it adopts a rather static view of contexts. The contextual information in these studies is often confined to cross-sectional snapshots, with the result that the dynamic nature of contexts goes unnoticed. We would like to challenge survey researchers to aim for a more dynamic conceptualization of contexts. Environments change (as do people). Once chosen for analysis, geographic variables may not represent the same context a few days, months, or years later. For instance, contextual factors, such as weather conditions, traffic density, air pollution, vegetation, etc., are characterized by high spatial and temporal variability. Especially if longitudinal data on individual survey respondents are available, there may be ample opportunity to also treat contextual information as time-varying. Linking changes in ecological variables to variation in individual-level outcomes may stimulate new research questions and also aid in identifying the direction of causal influence.

Another challenge arises when individual survey data are aggregated to a geographic level in order to map them into a spatial context and infer something about the target population or the context. When participating in a survey, individuals may provide answers about their current environment as indicated by a GPS location, and they may appear to be knowledgeable about the reference object. However, their true degree of expertise may be concealed due to the complexities of the question-answering process. For instance, participants might be living in different environments during the week than at the weekend (e.g., commuters). Simply assuming that data reflect information about some location just because a location happens to be available can introduce error of unknown magnitude, especially if respondents' answers are mapped to geographic units to which their answers do not actually belong (e.g., due to imprecise question wording, or participating in a survey on a mobile device at an unintended location that differs from what is reported as the place of residence, etc.).

Finally, any mapping of survey data to geographic units is done on the assumption that geographic units can be validly characterized by individual survey responses. Only then can statistical aggregates across respondents legitimately describe the specified geographic unit – for instance, by means, variances, and other indices of heterogeneity such as fractionalization and polarization (Chakravarty, 2015). For this assumption to be valid, two minimum requirements must be met: (a) For reliable estimates, the number of data points from individual response units (cases) per geographic unit must be large enough in relation to the unit population that the aggregate measure intends to describe; if not, a higher aggregation level (and spatial scale on which conclusions can be drawn) may be required. (b) The survey sample must be sufficiently representative of the target population characteristics in each geographic unit, lest bias arise in characterizing the unit by aggregate measures. Great care must be exerted in testing the extent to which the data can be considered representative of the target population and the contextual variables inferred from them (Doff, 2010), especially if participant self-selection and participants' selective mobility are not controlled for (see Jokela et al., 2015; Rentfrow et al., 2015).

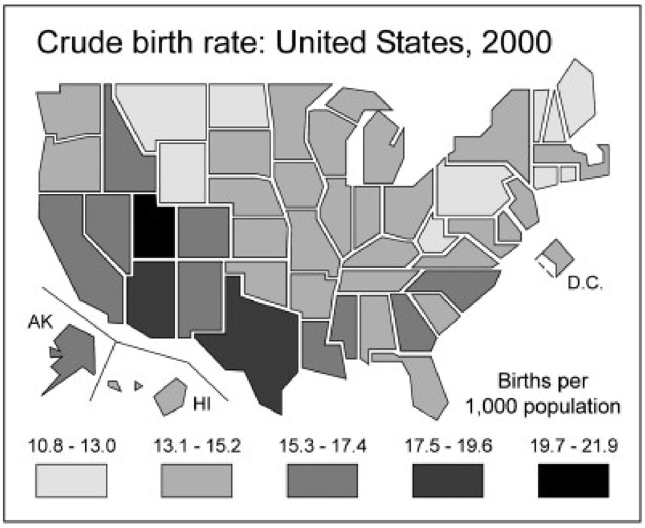

FIG. 2. *Crude birth rates, 2000, by state, based on equal-intervals cut-points and plotted on a visibility base map.*

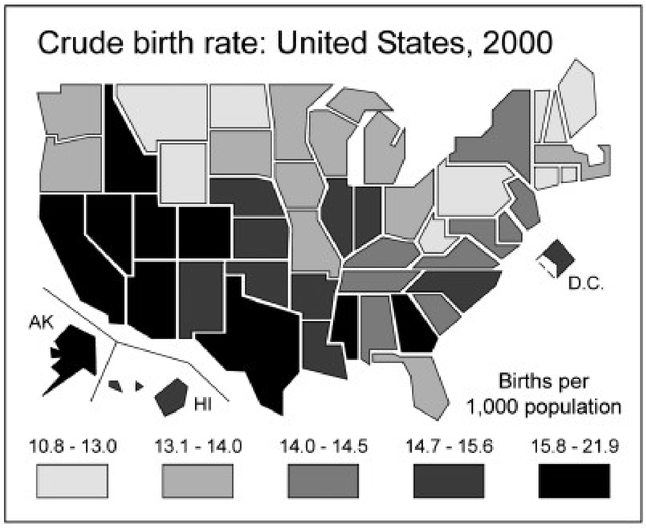

FIG. 3. *Crude birth rates, 2000, by state, based on quantile cut-points and plotted on a visibility base map.*

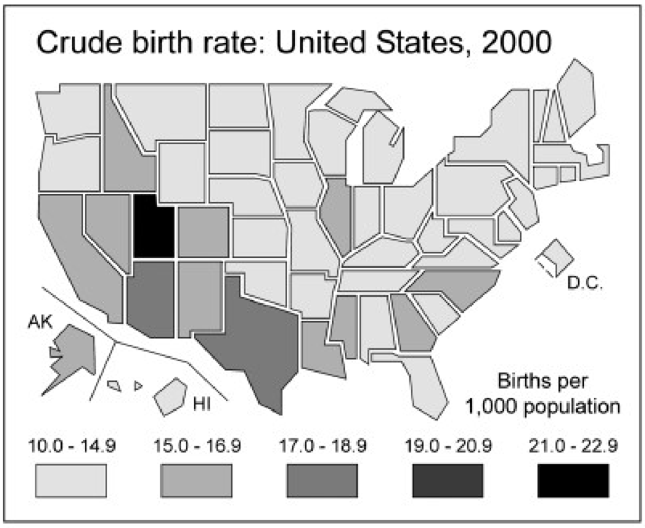

FIG. 5. *Crude birth rates, 2000, by state, categorized to suggest dangerously low rates overall.*

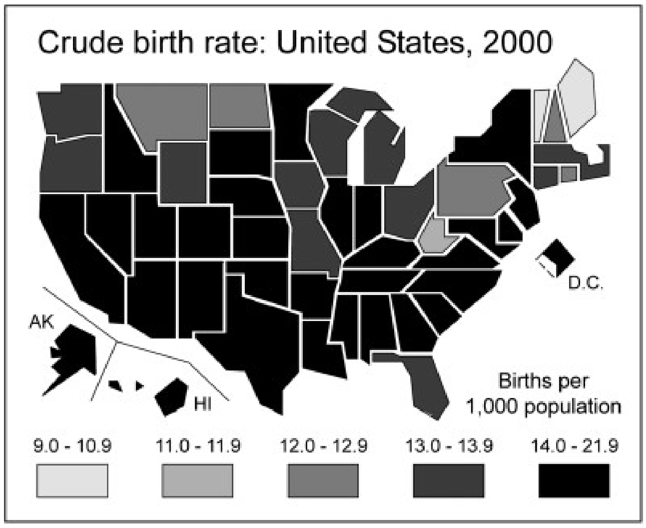

FIG. 6. *Crude birth rates, 2000, by state, categorized to suggest dangerously high rates overall.*

Figure 2. Reprint of Monmonier's (1996, 2005) classic example of "how to lie with maps": Arbitrary choices influence which information is being communicated with, and obtained from, maps.

3.5 Fallacies and Statistical Issues

With the increasing availability of big data and geocoded databases, there is also an increased risk of inferring ecological relationships (based on aggregate data) that may lead to misleading causal inferences if these are extended to the level of individual agents. For instance, some geographic areas may be inhabited by groups of different sizes (majority/minorities), and the same areas may have different likelihoods of showing other characteristics. But that does not mean that the statistical relationship between aggregate properties supplies the right clue to the underlying causal pathways.

For example, it is possible to obtain area data on crime and correlate them with other area-level information (e.g., ethnic composition). The temptation is to natively use the aligned skewed base rates of attributes to infer a relationship between them where none actually exists – a so-called pseudo-contingency (Kutzner & Fiedler, 2017). Higher crime rates may be observable in areas with a higher prevalence of a minority group. And yet the relationship at the aggregate level cannot hold the minority accountable for the crime rates observed in the areas in question. Based on the observed aggregate-level information, pseudo-contingencies provide a legitimate proxy for inferences at the ecological level. However, they are, at best, a heuristic for individual-level inferences. They may reflect genuine contingencies under various conditions, yet pseudo-contingencies are also at risk of inviting the wrong inference level. As the access to databases with high geographic resolution increases, survey researchers and GIScientists have a responsibility to ensure the correct interpretation of their data. We may face an increasing ethical obligation to correct blatant misuse of data (e.g., for political or ideological purposes).

We caution readers that, irrespective of the specific aggregation level chosen, there is always the risk of an aggregation bias, which refers to the difference between results established at the level of the units of analysis (say, states or groups) and results established for lower levels of analysis (say, counties or individuals) when using aggregated data. Making inferences from higher to lower levels runs the risk of committing an ecological fallacy (Piantadosi, Byar, & Green, 1988), whereby an observed association between variables is erroneously taken to operate at a lower aggregation level than the one actually studied (Robinson, 1950, 2009, 2011). Conversely, an individualistic fallacy may result when relationships observed at a micro-level are erroneously extrapolated to a macro-level (Clark & Avery, 1976). For example, the outcome of encounters between social groups cannot be predicted on the basis of how individual group members from different groups interact with each other (Doerr, Plant, Kunstman, & Buck, 2011; Lichter, Parisi, & Taquino, 2012). Drawing conclusions about individuals (the survey units) through aggregated quantities (to match the geographic units) – and vice versa – requires drawing inferences carefully and properly (Grotenhuis, Eisinger, & Subramanian, 2011), which usually necessitates multilevel data analysis (Nezlek, 2008).

Furthermore, when using georeferenced data, many statistical methods are no longer suitable. Spatial autocorrelation – that is, the degree to which one object is similar to other spatially nearby objects (Goodchild, 2009) – jeopardizes the independence requirement of many statistical techniques. The phenomenon refers to the common finding that observations with a higher proximity in geographic space tend to be more similar to each other than those at a greater distance; this often results in patterns such as gradients or clusters. Such patterns may also be found among survey data. Using spatially distributed data that are either externally linked to, or gathered from, surveys requires methods of data analysis that detect, describe (i.e., quantify), and, if necessary, adjust for the presence of spatial autocorrelation (Assuncao & E.-A. Reis, 1999; Banerjee, Carlin, & Gelfand, 2015; Getis, 2010; Oden, 1995; Waldhör, 1996). We refer interested readers to an online introduction[1] and to recent accessible treatises of applied spatial analysis (Fischer & Getis, 2010; Ward & Gleditsch, 2008).

3.6 Analyzing User-Generated and Spatiotemporally Distributed Data

Another set of challenges arises when integrating new methods of data collection such as citizen sensing (i.e., acquiring people's feedback through dedicated technologies such as smartphone apps) or linking collective data sources (e.g., mobile phone or social networks) with traditional survey approaches. Traditional ways of analyzing geospatial data mostly presume a well-defined data acquisition process and follow Tobler's 1970 first law of geography, according to which processes happening close to each other have a stronger influence than distant ones. However, Tobler's law may not hold for most user-generated data. For example, social media posts about a large-scale sports event or a national election may be related in time (when they are posted) and semantics (the content of the posts), but not in geographic space (as they may be sent from users in different places throughout the world). "Virtual neighborhoods" (e.g., people someone follows on Twitter) may have more influence on a person's attitudes than spatial neighbors. The reason is that the data-generating process for social media posts (as opposed to "traditional" spatial data like demographic data or transportation infrastructure data) is not standardized, nor is it under the control of a researcher. Instead, the mechanisms generating user-driven data are unpredictable and technically arbitrary; user motivations are often hidden, but they are likely to be context-bound. This non-standardized, uncontrolled data-generating process also implies that representativeness for the whole population may be impossible to achieve with data from wearables, citizen sensing, or social media (although targeting more specific populations may be realistic; Pötzschke & Braun, 2016). This issue has been largely neglected in previous research, and it constitutes a potentially high-impact research gap, even though first attempts at overcoming it within spatial analyses are being made (for the case of social media data, see Westerholt, Resch, & Zipf, 2015, 2016).

Another still largely unresolved question is how participants' responses are influenced by repeatedly interacting with technical devices (smartphones), especially if they frequently encounter dedicated survey questions. From a psychological viewpoint, besides typical memory errors, this may induce several kinds of biases. First, conditioning effects may occur such that people become conditioned to specific locations and provide pre-determined answers that they have learned to automatically associate with the loca-

[1] https://docs.aurin.org.au/portal-help/analysing-your-data/spatial-statistics-tools/introduction-to-spatial-autocorrelation

tion when prompted for responses. Given the frequency of recurring situations, they may not be motivated to engage with the question as expected. Due to the cooperative principle that governs effective communication (Grice, 1975), some respondents may alter their statements when answering questions repeatedly – although their opinions have not really changed – because they think that new information must be provided; other respondents may stick with what they answered earlier in order to appear consistent and not contradict themselves. Second, it may not always be possible to gain reports immediately at the location of interest. However, delayed responding may introduce retrospective bias (e.g., inaccurate recall, recency effects, false memories) into respondents' cognitive representations (e.g., Steffens & Mecklenbräuker, 2007). Survey researchers can offer specific advice on how to minimize the impact of such biases on survey quality.

Still, as the availability of data from wearables, social media, and other new data sources increases, greater research efforts will be necessary to resolve questions of survey designs, data quality, representativeness, and potential biases, and to link these new data to traditional surveys. Here, GI-Scientists can benefit from the expertise of psychologists and other social scientists with regard to traditional surveys, and we call on these disciplines to jointly tackle the aforementioned issues. Compared to traditional surveys, surveys in the domain of GIScience often encompass user-generated data, or they comprise a strong technological component (e.g., GPS receivers, physical assessments, advanced spatial analysis techniques).

Whatever survey concept applies, full documentation of the survey design and its quality is required because only this approach permits estimating, and potentially correcting for, sources of sampling related error (Dever, Rafferty, & Valliant, 2008; Gabler & Quatember, 2013). The documentation should include intended and actual populations under study (to determine over- and under-coverage) as well as the sampling design, the obtained sample size, and the reasons for any missing data (e.g., non-response, drop-out; Little & Rubin, 2002). With new forms of data, such as data from social media, this information may be unavailable, so that the quality of the data collection cannot be assessed (Brickman Bhutta, 2012). However – depending on the study goal – this information may be an indispensable requirement (Pötzschke & Braun, 2016; Rothman, Gallacher, & Hatch, 2013). Moreover, most current approaches in geospatial analysis rely on well-defined data structures with known degrees of uncertainty and small error margins, although these requirements are not met by vast portions of user-generated data (Steiger, Westerholt, Resch, & Zipf, 2015). Guidelines may help researchers to minimize total survey error (Groves, 2010) and improve total survey quality under budgetary constraints (Biemer, 2010), for instance, those published by the German Data Forum (RatSWD – Rat für Sozial- und Wirtschaftsdaten [German Data Forum], 2015), AAPOR[2], and ESOMAR[3].

3.7 Privacy Concerns and Data Protection Issues

The last challenge we highlight is the use of any personal – including geocoded – data and researchers' ethical obligation to protect users' privacy (Goebel, Wagner, & Wurm, 2010; Goebel, Wurm, & Wagner, 2010). Typical privacy risks are presence leakage (an attacker might identify individuals present in, or absent from, the database) and association leakage (an attacker might unambiguously associate individuals with sensitive information). The risk of deductive disclosure – identifying a person by a combination of personal characteristics – is a challenge for GIS research (due to the inclusion of geocodes, tracking of individuals, and data linkage). This issue calls for technologies and legal frameworks to protect data against deductive disclosure of participants' identities, unintended transfer, or other misuse by third parties (Barcena, Wueest, & Lau, 2014).

Legislation that ensures a degree of data safety (keeping data available in the future) and data security (limiting access to data) varies from country to country. Consequently, researchers sharing sensitive data in international collaborations may have to deal with diverging legal requirements and policies for raw and derived data across various countries. Moreover, respondents' willingness to voluntarily share highly personal data with scientists differs across individuals and settings. However, support for research is usually closely linked to trust in the security of data and their protection against misuse. Ironically, many users willingly share private information in other places and do not actively try to conceal or protect it, even if they claim that they are concerned about their privacy (Acquisti, John, & Loewenstein, 2013). A striking example is the vast amount of sensitive data (including rich location data) that people share on social media platforms such as Facebook, where they typically have little influence on data collection and processing policies. Likewise, estimates show that one-third of the free smartphone apps collect location information, yielding numerous possibilities of analyzing geographic data and extracting information from them (Kersten & Klett, 2012). Apparently, operators and service providers – whose business models often rest on collecting and selling customer/user data (e.g., Google) – effectively insinuate that less privacy is the new social norm, and that it means better services for the user (Johnson, 2010).

On the researcher's side, several means exist to protect participants and their rights, including privacy (**kounadi15**).

[2] http://www.aapor.org/Standards-Ethics/Best-Practices.aspx

[3] http://www.esomar.org/knowledge-and-standards/codes-and-guidelines/guideline-on-opinion-polls-and-published-surveys.php

First, all participants should participate voluntarily in data-rich scientific studies through an opt-in agreement, after a thorough briefing (informed consent) – something that is rather self-evident from the perspective of survey research. Principal investigators and researchers must enter into a data-sharing agreement about which data will be collected, analyzed and stored, where and for how long, and who will have access to them.

When data are to be stored and made available to other researchers, a possible way of allaying concerns about privacy is to restrict access to sensitive data. For instance, it might be feasible to use different levels of access privileges to sensitive datasets in a data archive ("data enclave"; Lane, Stodden, Bender, & Nissenbaum, 2014). However, such archives usually involve increased levels of burden. Sometimes, only aggregate query results can be obtained, or access might be limited to eligible researchers in a controlled, secure environment with high-security data storage facilities (e.g., GESIS' Secure Data Center with an on-site safe room:[4]. Sharing multi-site research data safely (via the cloud) during collection requires technical solutions that are still in their infancy, and new standards will have to be developed and enforced (Palanisamy & L. Liu, 2015; Veena & Devidas, 2014).

For applications that require the collection and storage of personal information, the Electronic Frontier Foundation and others recommend anonymizing data and using strong cryptographic protocols at various stages of data transmission and handling. However, trajectories of people moving through space (at specified times) can still undermine anonymity. In the case of spatial information, more specifically, previously anonymous users can be re-identified relatively easily by their spatial profiles because personal geodata are highly unique to an individual. Indeed, de Montjoye, Hidalgo, Verleysen, and Blondel (2013) showed that only four random positions from a person's GPS track can be enough to identify an individual. In this context, the concept of location privacy describes the ability of an individual to move in public space without their geographic location being collected or stored. The most restrictive way to achieve location privacy and to prevent misuse of personal data is to opt out of research altogether and to prohibit the collection of any data (Blumberg & Eckersley, 2009) – which is not usually a scientifically viable option. If possible, trajectory data should be analyzed and shared at an aggregated rather than an individual level. Furthermore, privacy should also be protected by splitting trajectories into sub-paths so that they cannot be reconstructed. Although this involves a certain amount of information loss, the restoration of identities is prevented (see Sattar, Li, Ding, Liu, & Vincent, 2013; Wang, 2010).

It is often necessary to georeference the survey data so that they can be mapped to relevant spatial units in order, for example, to link individual respondents' data to contextual data. Several methods of georeferencing exist. Direct georeferencing requires that exact locations be collected via spatial coordinates (e.g., 2D or 3D, GPS). Indirect georeferencing assumes that relevant spatial units are inferred from postal codes, administrative units, etc. However, the use of online geocoding services for converting terms such as ZIP codes to locations such as latitude, longitude, and elevation by means of direct georeferencing (e.g., via Google services) should not, in our view, be the first choice, and – depending on laws or policies, typing in sampled addresses into Google maps may be even probited. Geocoding involves the risk that non-aggregated scientific use files might become de-anonymized, as geocodes potentially undermine a user's location privacy. Even though single locations are not revealing in themselves, complete time-stamped location patterns may be used to identify an individual (especially if a company knows more about that individual than the information contained in the scientific data). Moreover, reverse geocoding – back coding of latitude and longitude to a comprehensible address – involves the risk that an individual's identity will be leaked – even from mere dots representing individuals on a published map. Identity leakage from maps can be prevented by aggregating data points prior to drawing the map or by skewing the presentation of individual data points (Brownstein, Cassa, & Mandl, 2006).

Instead of using geocoding services from companies with commercial interests, we suggest using public geocoding services such as that provided by the German Federal Agency of Cartography and Geodesy (BKG). This service allows users to tag any geographically identifiable object (e.g., on the basis of available address information) with precise geographic coordinates (reverse geocoding is possible, too). It is usually available only to federal authorities. However, under an agreement between GESIS and the BKG, it may also be used by other institutions (Schweers, Kinder-Kurlanda, Müller, & Siegers, 2016).[5] Some specialized centers provide software, services, and support for linking databases while observing privacy-preserving record linkage (e.g., German Data Linkage Center Schnell, 2013a)[6]. The U.S. Census Bureau maintains an application programming interface (API), the Census Geodcoder, for real-time and batch geocoding of residential addresses – a service that is free, fast, and accurate[7].

Note that bias can be introduced later on when linking the datasets (Sakshaug, Couper, Ofstedal, & Weir, 2012). Respondents who are used to being interviewed about sensitive issues (e.g., political attitudes) may not be willing to consent to their data being linked to additional databases, and those who are willing may not be representative of the popu-

[4] http://www.gesis.org/en/services/data-analysis/data-archive-service/secure-data-center-sdc

[5] http://www.gesis.org/forschung/drittmittelprojekte/projektuebersicht-drittmittel/georefum

[6] http://www.record-linkage.de

[7] https://coding.geo.census.gov/geocoder

lation. GIScience participants may agree to be tracked (e.g., resulting in trajectories), yet this may be due to previous self-selection (which would be accompanied by overall lower response rates). Initial self-selection and subsequent selective dropout may introduce bias into a combined dataset.

We conclude this discussion by encouraging researchers to reflect on the ethical implications and the long-term societal impact of fine-grained spatial analyses. For example, terms such as "air quality" or "pollutant dispersion" are only surrogates for more direct and far-reaching influences on individuals, such as life expectancy, respiratory diseases, or quality of life (Resch, Britter, & Ratti, 2012). Knowledge about these phenomena at high geospatial resolutions may affect relevant aspects in people's lives, such as health insurance rates or real estate prices. Researchers' ethical responsibility to find the appropriate spatial granularity level when providing information and communicating research findings has never been more acute. The scientific drive to provide ever more accurate, possibly finer-grained, and complete information competes with other ethical principles surrounding privacy concerns and prevention of misleading conclusions.

4 Conclusion

In this article, we have discussed the promising new opportunities of integrating GIScience tools into survey research in general, and psychological survey research in particular, and the challenges associated with these opportunities. In so doing, we have focused mainly on how survey research can profit from incorporating recent advances in GIScience. We highlight, however, that GIScientists can also profit greatly from the accumulated wisdom in survey research methodology, for example, when thinking about measurement and assessment, data quality, or representativeness. We are certain that intensified interdisciplinary dialogue holds great potential for future research. In our view, both survey researchers and GIScientists would benefit from incorporating each others' traditions into their own theorizing and methodologies. In this process, survey researchers can act as consultants to GIScientists just as much as GIScientists can inspire survey researchers with new advancements.

A stronger integration of the research traditions will also enable highly inspirational interdisciplinary research. Future research at the intersection between survey research methodology and GIScience may even blur the very boundaries of the survey concept and bring us closer to studying the person – context transactions that are deemed crucial in shaping individual behavior and development (Bronfenbrenner, 1979; Lerner, 1991). Thanks to the progress that has been made in GIScience, the study of the current environment that Lewin (1936) once envisioned can now include precise temporal and spatial aspects. Context information can increasingly be incorporated in real time, and it may be based on subjective as well as objective contextual characteristics of individual situations.

Obviously, the fruitfulness of future research enterprises depends on the engagement of researchers from both sides, their growing awareness of the tools, methods, and concepts they offer each other, and of the goals and challenges associated with each of them. We hope that this overview will be instrumental in fostering dialogue between survey research and GIScience.

Acknowledgements

We would like to express our gratitude to the Austrian Science Fund (FWF - Der Wissenschaftsfonds) for supporting the project "Urban Emotions", reference number I-3022. Furthermore, we would like to thank the Doctoral College GIScience (DK W 1237-N23) at the Department of Geoinformatics – Z_GIS, University of Salzburg, Austria, funded by the Austrian Science Fund (FWF) for their support.

References

Acquisti, A., John, L. K., & Loewenstein, G. (2013). What is privacy worth? *The Journal of Legal Studies*, *42*, 249–274.

Agnew, J. A. (2011). Space and place. In J. Agnew & D. N. Livingstone (Eds.), *The SAGE handbook of geographical knowledge* (pp. 316–330). London, UK: Sage.

Anselin, L. & Williams, S. (2016). Digital neighborhoods. *Journal of Urbanism*, *9*, 305–328.

Anstead, N. & O'Loughlin, B. (2015). Social media analysis and public opinion: the 2010 UK general election. *Journal of Computer-Mediated Communication*, *20*, 204–220.

Arias, S. & Warf, B. (2009). *The spatial turn: interdisciplinary perspectives*. Routledge Studies in Human Geography, Vol. 26. London, UK: Routledge.

Assuncao, R.-M. & Reis, E.-A. (1999). A new proposal to adjust Moran's I for population density. *Statistics in Medicine*, *18*, 2147–2162.

Ayobi, A., Marshall, P., & Cox, A. L. (2016). Reflections on 5 years of personal informatics: rising concerns and emerging directions. In *Conference on human factors in computing systems proceedings* (pp. 2774–2781). San Jose, CA: Association for Computing Machinery.

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2015). *Hierarchical modeling and analysis for spatial data* (2nd ed.). Boca Raton, FL: Chapman and Hall/CRC Press.

Barcena, M. B., Wueest, C., & Lau, H. (2014). *How safe is your quantified self?* Mountain View, CA: Symantech.

Bettini, C., Brdiczka, O., Henricksen, K., Indulska, J., Nicklas, D., Ranganathan, A., & Riboni, D. (2010). A survey of context modelling and reasoning techniques. *Pervasive and Mobile Computing*, *6*, 161–180.

Biemer, P. P. (2010). Total survey error: design, implementation, and evaluation. *Public Opinion Quarterly*, *74*, 817–848.

Blumberg, A. J. & Eckersley, P. (2009). *On locational privacy, and how to avoid losing it forever.* San Francisco, CA: Electronic Frontier Foundation.

Boschmann, E. E. & Cubbon, E. (2014). Sketch maps and qualitative GIS: using cartographies of individual spatial narratives in geographic research. *The Professional Geographer*, *66*(2), 236–248.

Brickman Bhutta, C. (2012). Not by the book: facebook as a sampling frame. *Sociological Methods and Research*, *41*, 57–88.

Bronfenbrenner, U. (1979). *The ecology of human development: experiments by nature and design.* Cambridge, MA: Harvard University Press.

Brownstein, J. S., Cassa, C. A., & Mandl, K. D. (2006). No place to hide. Reverse identification of patients from published maps. *New England Journal of Medicine*, *355*, 1741–1742.

Brulé, G. & Veenhoven, R. (2014). Average happiness and dominant family type in regions in Western Europe around 2000. *Advances in Applied Sociology*, *4*, 271–288.

Brulé, G. & Veenhoven, R. (2015). Geography of happiness: configurations of affective and cognitive appraisal of life across nations. *International Journal of Happiness and Development*, *2*, 101–117.

Caldarelli, G., Chessa, A., Pammolli, F., Pompa, G., Puliga, M., & … & Riotta, G. (2014). A multi-level geographical study of italian political elections from Twitter data. *PLOS ONE*, *9*, e95809.

Clark, W. A. & Avery, K. L. (1976). The effects of data aggregation in statistical analysis. *Geographical Analysis*, *8*, 428–438.

Clifford, N., Holloway, S., Rice, S. P., & Valentine, G. (Eds.). (2009). *Key concepts in geography* (2nd ed.). Thousand Oaks, CA: Sage.

Conner, T. S. & Barrett, L. F. (2012). Trends in ambulatory self-report: the role of momentary experience in psychosomatic medicine. *Psychosomatic Medicine*, *74*, 327–337.

Conway, M. & O'Connor, D. (2016). Social media, big data, and mental health: current advances and ethical implications. *Current Opinion in Psychology*, *9*, 77–82.

Coulton, C. J., Chan, T., & Mikelbank, K. (2010). *Finding place in making connections communities: applying GIS to residents' perceptions of their neighborhoods.* Washington, DC: The Urban Institute.

Crampton, J. W. (2010). *Mapping: a critical introduction to cartography and GIS.* New York, NY: Wiley-Blackwell Publishers.

Crampton, J. W. & Krygier, J. (2005). An introduction to critical cartography. *ACME*, *4*(1), 11–33.

Crooks, A., Croitoru, A., Stefanidis, A., & Radzikowski, J. (2013). #earthquake: Twitter as a distributed sensor system. *Transactions in GIS*, *17*, 124–147.

Cuevas, R., Gonzalez, R., Cuevas, A., & Guerrero, C. (2014). Understanding the locality effect in twitter: measurement and analysis. *Personal and Ubiquitous Computing*, *18*, 397–411.

Curini, L., Iacus, S., & Canova, L. (2015). Measuring idiosyncratic happiness through the analysis of Twitter: an application to the Italian case. *Social Indicators Research*, *121*, 525–542.

Dangschat, J. S. (2007). Raumkonzept zwischen struktureller Produktion und individueller Konstruktion. *Ethnoscripts*, *9*, 24–44.

de Lange, N. (2013). *Geoinformatik in Theorie und Praxis (Vol. 3).* Heidelberg: Springer.

de Leeuw, E. D., Hox, J. J., & Dillman, D. A. (2008). *International handbook of survey methodology*. EAM book series. New York, NY: Erlbaum.

Dever, J. A., Rafferty, A., & Valliant, R. (2008). Internet surveys: can statistical adjustments eliminate coverage bias? *Survey Research Methods*, *2*, 47–62.

Dey, A. K. (2001). Understanding and using context. *Personal and Ubiquitous Computing*, *5*, 4–7.

Diekmann, A. & Meyer, R. (2010). Demokratischer Smog? Eine empirische Untersuchung zum Zusammenhang zwischen Sozialschicht und Umweltbelastungen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *62*, 437–457.

Doerr, C., Plant, E. A., Kunstman, J. W., & Buck, D. (2011). Interactions in black and white: racial differences and similarities in response to interracial interactions. *Group Processes and Intergroup Relations*, *14*, 31–43.

Doff, W. (2010). *Puzzling neighbourhood effects: spatial selection, ethnic concentration and neighbourhood impacts.* Amsterdam, Netherlands: IOS Press.

Dungan, J.-L., Perry, J.-N., Dale, M.-R.-T., Legendre, P., Citron-Pousty, S., Fortin, M.-J., & Rosenberg, M. (2002). A balanced view of scale in spatial statistical analysis. *Ecography*, *25*, 626–640.

Ekman, P. & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*, 124–129.

Feddersen, J., Metcalfe, R., & Wooden, M. (2016). Subjective wellbeing: why weather matters. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *179*, 203–228.

Fieldhouse, E., Green, J., Schmitt, H., Evans, G., & van der Eijk, C. (2014). *The 2015 British Election Study: vot-*

ers in context. Presentation at the EPOP 2014 Conference, September 2014, University of Edinburgh, UK.

Fischer, M. M. & Getis, A. (2010). *Handbook of applied spatial analysis.* Heidelberg, Germany: Springer.

Gabler, S. & Lahiri, P. (2009). On the definition and interpretation of interviewer variability for a complex sampling design. *Survey Methodology*, *35*, 85–99.

Gabler, S. & Quatember, A. (2013). Repräsentativität von Subgruppen bei geschichteten Zufallsstichproben. *AStA Wirtschafts-und Sozialstatistisches Archiv*, *7*, 105–119.

Gerell, M. (2017). Smallest is better? The spatial distribution of arson and the modifiable areal unit problem. *Journal of Quantitative Criminology*, *33*(2), 293–318.

Getis, A. (2010). Spatial autocorrelation. In M.-M. Fischer & A. Getis (Eds.), *Handbook of applied spatial analysis* (pp. 255–278). Heidelberg, Germany: Springer.

Goebel, J., Wagner, G. G., & Wurm, M. (2010). *Exemplarische Integration raumrelevanter Indikatoren auf Basis von Fernerkundungsdaten in das Sozio-Oekonomische Panel (SOEP) (Technical report)*. Berlin, Germany: DIW Berlin/The German Socio-Economic Panel.

Goebel, J., Wurm, M., & Wagner, G. (2010). *Exploring the linkage of spatial indicators from remote sensing data with survey data: the case of the socio-economic panel (SOEP) and 3d city models*. SOEPpapers on Multidisciplinary Panel Data Research 283. Berlin, Germany: DIW Berlin/The German Socio-Economic Panel (SOEP).

Goodchild, M.-F. (1992). Geographic information science. *International Journal of Geographical Information Systems*, *6*(1), 31–45.

Goodchild, M.-F. (2009). What problem? Spatial autocorrelation and geographic information science. *Geographical Analysis*, *41*, 411–417.

Goodchild, M.-F. (2010). Twenty years of progress: GIScience in 2010. *Journal of Spatial Information Science*, *2010*(1), 3–20.

Goodchild, M.-F. & Kemp, K.-K. (1992). NCGIA education activities: the core curriculum and beyond. *International Journal of Geographical Information Systems*, *6*, 309–320.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics (vol. 3, speech acts)* (pp. 41–58). New York, NY: Academic Press.

Grotenhuis, M., Eisinger, R., & Subramanian, S. V. (2011). Robinson's ecological correlations and the behavior of individuals: methodological corrections. *International Journal of Epidemiology*, *40*, 1123–1125.

Groves, L., R. M. & Lyberg. (2010). Total survey error: past, present and future. *Public Opinion Quarterly*, *74*, 849–879.

Hegarty, M., Montello, D.-R., Richardson, A.-E., Ishikawa, T., & Lovelace, K. (2006). Spatial abilities at different scales: individual differences in aptitude-test performance and spatial-layout learning. *Intelligence*, *34*, 151–176.

Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (Eds.). (2007). *Experience sampling method: measuring the quality of everyday life.* Thousand Oaks, CA: Sage.

Hill, C. A., Dean, E. F., & Murphy, J. J. (Eds.). (2013). *Social media, sociality, and survey research.* Hoboken, NJ: John Wiley & Sons.

Hoffmeyer-Zlotnik, J. H. (1994). Regionalisierung von Umfragen. *ZUMA Nachrichten*, *34*(18), 35–57.

Hoffmeyer-Zlotnik, J. H. (2013). *Regionalisierung sozialwissenschaftlicher Umfragedaten: Siedlungsstruktur und Wohnquartier.* Wiesbaden, Germany: Springer VS.

Hong, J., Suh, E.-H., Kim, J., & Kim, S. (2009). Context-aware system for proactive personalized service based on context history. *Expert Systems with Applications*, *36*, 7448–7457.

Hüttenrauch, B. (2016). *Targeting using augmented data in database marketing: decision factors for evaluating external sources.* Wiesbaden, Germany: Springer Fachmedien.

Iosa, M., Fusco, A., Morone, G., & Paolucci, S. (2012). Walking there: environmental influence on walking-distance estimation. *Behavioural Brain Research*, *226*, 124–132.

Jack, R. E., Garrod, O. G. B., & Schyns, P. G. (2014). Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Current Biology*, *24*, 187–192.

Johnson, B. (2010). Privacy no longer a social norm, says facebook founder. The Guardian, January 11. Retrieved from https://www.theguardian.com/technology/2010/jan/11/facebook-privacy

Jokela, M., Bleidorn, W., Lamb, M. E., Gosling, S. D., & Rentfrow, P. J. (2015). Geographically varying associations between personality and life satisfaction in the London metropolitan area. *Proceedings of the National Academy of Sciences*, *112*, 725–730.

Kersten, H. & Klett, G. (2012). *Mobile device management.* Heidelberg, Germany: Hüthig Jehle Rehm.

Kitchen, P., Williams, A., & Chowhan, J. (2012). Sense of community belonging and health in Canada: a regional analysis. *Social Indicators Research*, *107*, 103–126.

Klonner, C., Marx, S., Usón, T., & Höfle, B. (2016). Risk awareness maps of urban flooding via OSM field papers – case study Santiago de Chile. In A. H. Tapia, P. Antunes, V. A. Bañuls, K. Moore, & J. Porto de Albuquerque (Eds.), *Proceedings of the International Conference on Information Systems for Crisis Response* *and Management (ISCRAM) 2016*. Rio de Janeiro: Federal University of Rio de Janeiro.

Kozak, J. & Szwagrzyk, M. (2016). Have there been forest transitions? Forest transition theory revisited in the context of the modifiable areal unit problem. *Area*, *48*(4), 397–532.

Kutzner, F. & Fiedler, K. (2017). Stereotypes as pseudo-contingencies. *European Review of Social Psychology*, *28*(1), 1–49.

Larson, R. & Csikszentmihalyi, M. (1983). The experience sampling method. *New Directions for Methodology of Social and Behavioral Science*, *15*, 41–56.

Lechner, C. M., Obschonka, M., & Silbereisen, R. K. (2017). Who reaps the benefits of social change? Exploration and its socioecological boundaries. *Journal of Personality*, *85*(2), 257–269.

Lerner, R. M. (1991). Changing organism-context relations as the basic process of development: a developmental contextual perspective. *Developmental Psychology*, *27*, 27–32.

Leslie, E., Sugiyama, T., Ierodiaconou, D., & Kremer, P. (2010). Perceived and objectively measured greenness of neighbourhoods: are they measuring the same thing? *Landscape and Urban Planning*, *95*, 28–33.

Lewin, K. (1936). *Principles of topological psychology*. New York, NY: McGraw-Hill.

Lichter, D. T., Parisi, D., & Taquino, M. C. (2012). The geography of exclusion: race, segregation, and concentrated poverty. *Social Problems*, *59*, 364–388.

Little, R. & Rubin, D. (2002). *Statistical analysis with missing data* (2nd ed.). New York, NY: Wiley.

Liu, B. & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In C. C. Aggarwal & C. X. Zhai (Eds.), *Mining text data* (pp. 415–463). Berlin, Germany: Springer.

Lucas, R. E. & Lawless, N. (2013). Does life seem better on a sunny day? Examining the association between daily weather conditions and life satisfaction judgments. *Journal of Personality and Social Psychology*, *104*, 872–884.

Mas, J.-F., Pérez Vega, A., Andablo Reyes, A., Castillo Santiago, M.-A., & Flamenco Sandoval, A. (2015). Assessing modifiable areal unit problem in the analysis of deforestation drivers using remote sensing and census data. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XL-3*, 77–80.

McGinn, A. P., Evenson, K. R., Herring, A. H., Huston, S. L., & Rodriguez, D. A. (2007). Exploring associations between physical activity and perceived and objective measures of the built environment. *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, *84*, 162–184.

McMaster, R.-B. & Sheppard, E. (2004). Introduction: scale and geographic inquiry. In E. Sheppard & R.-B. McMaster (Eds.), *Scale and geographic inquiry: nature, society, and method.* Oxford, UK: Blackwell.

Messner, C. & Wänke, M. (2011). Good weather for Schwarz and Clore. *Emotion*, *11*, 436–437.

Meyer, R. & Bruderer Enzler, H. (2013). Geographische Informationssysteme (GIS) und ihre Anwendung in den Sozialwissenschaften am Beispiel des Schweizer Umweltsurveys. *Methoden, Daten, Analysen*, *7*, 317–347.

Mitra, R. & Buliung, R.-N. (2012). Built environment correlates of active school transportation: neighborhood and the modifiable areal unit problem. *Journal of Transport Geography*, *20*, 51–61.

Monmonier, M. (1996). *How to lie with maps* (2nd ed.). Chicago, IL: University of Chicago Press.

Monmonier, M. (2005). Lying with maps. *Statistical Science*, *20*, 215–222.

Montello, D.-R. & Golledge, R.-G. (1999). Scale and detail in the cognition of geographic information. In *Report of the Specialist Meeting of Project Varenius, May 14–16, 1998*. Santa Barbara, CA. Retrieved from http://escholarship.org/uc/item/1hf6d3fx

Murphy, J. J., Link, M. W., Childs, J. H., Tesfaye, C. L., Dean, E., Stern, M., … Harwood, P. (2014). *Social media in public opinion research: report of the AAPOR task force on emerging technologies in public opinion research.* Oakbrook Terrace, IL: American Association for Public Opinion Research.

Nezlek, J. (2008). An introduction to multilevel modeling for social and personality psychology. *Social and Personality Psychology Compass*, *2*, 842–860.

Obschonka, M., Schmitt-Rodermund, E., Silbereisen, R. K., Gosling, S. D., & Potter, J. (2013). The regional distribution and correlates of an entrepreneurship-prone personality profile in the United States, Germany, and the United Kingdom: a socioecological perspective. *Journal of Personality and Social Psychology*, *105*, 104–122.

Oden, N. (1995). Adjusting Moran's I for population density. statistics in medicine. *14*, 17–26.

Oishi, S. (2014). Socioecological psychology. *Annual Review of Psychology*, *65*, 1–29.

Okner, B. A. (1972). Constructing a new data base from existing microdata sets – the 1966 merge file. *Annals of Economic and Social Measurement*, *1*, 325–342.

Openshaw, S. (1984). *The modifiable areal unit problem.* Norwich, UK: Geobooks.

Ord, J.-K. & Getis, A. (2012). Local spatial heteroscedasticity (LOSH). *The Annals of Regional Science*, *48*, 529–539.

Palanisamy, B. & Liu, L. (2015). Privacy-preserving data publishing in the cloud: a multi-level utility controlled approach. In C. Pu & A. Mohindra (Eds.), *CLOUD* (pp. 130–137). IEEE. Retrieved from http://dx.doi.org/10.1109/CLOUD.2015.27

Piantadosi, S., Byar, D.-P., & Green, S.-B. (1988). The ecological fallacy. *American Journal of Epidemiology*, *127*, 893–904.

Pötzschke, S. & Braun, M. (2016). Migrant sampling using Facebook advertisements. a case study of Polish migrants in four European countries. *Social Science Computer*, *Online first*. doi:10.1177/0894439316666262

Rammstedt, B., Mutz, M., & Farmer, R. (2015). The answer is blowing in the wind. weather effects on personality ratings. *European Journal of Psychological Assessment*, *31*, 287–293.

RatSWD – Rat für Sozial- und Wirtschaftsdaten [German Data Forum]. (2012). *Georeferenzierung von Daten. Situation und Zukunft der Geodatenlandschaft in Deutschland.* Berlin, Germany: Scivero.

RatSWD – Rat für Sozial- und Wirtschaftsdaten [German Data Forum]. (2015). *Quality standards for the development, application, and evaluation of measurement instruments in social science survey research.* RatSWD Working Paper No. 245, prepared and written by the Quality Standards Working Group).

Rehdanz, K. & Maddison, D. (2005). Climate and happiness. *Ecological Economics*, *52*, 111–125.

Reichert, M., Törnros, T., Hoell, A., Dorn, H., Tost, H., Salize, H.-J., . . . Ebner-Priemer, U. W. (2016). Using ambulatory assessment for experience sampling and the mapping of environmental risk factors in everyday life. *Die Psychiatrie*, *13*, 94–102.

Reis, H. T. & Gable, S. L. (2000). Event-sampling and other methods for studying everyday experience. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 190–222). Cambridge, UK: Cambridge University Press.

Rentfrow, P. J. (2014). *Geographical psychology: exploring the interaction of environment and behavior.* Washington, D.C.: American Psychological Association.

Rentfrow, P. J., Jokela, M., & Lamb, M. E. (2015). Regional personality differences in great britain. *PLOS ONE*, *10*, e0122245.

Resch, B. (2013). People as sensors and collective sensing – contextual observations complementing geo-sensor network measurements. In J. M. Krisp (Ed.), *Lecture notes in geoinformation and cartography: progress in location-based services* (pp. 391–406). Berlin, Germany: Springer.

Resch, B., Britter, R., & Ratti, C. (2012). Live urbanism: towards senseable cities and beyond. In P. Pardalos & S. Rassia (Eds.), *Sustainable environmental design in architecture: impacts on health* (pp. 175–184). New York, NY: Springer.

Resch, B., Sudmanns, M., Sagl, G., Summa, A., Zeile, P., & Exner, J.-P. (2015). Crowdsourcing physiological conditions and subjective emotions by coupling technical and human mobile sensors. *GI-Forum – Journal for Geographic Information Science*, *1*, 514–524.

Resch, B., Summa, A., Sagl, G., Zeile, P., & Exner, J.-P. (2015). Urban emotions – geo-semantic emotion extraction from technical sensors, human sensors and crowdsourced data. In G. Gartner & H. Huang (Eds.), *Lecture notes in geoinformation and cartography: progress in location-based services 2014* (pp. 199–212). Berlin, Germany: Springer International Publishing.

Resch, B., Summa, A., Zeile, P., & Strube, M. (2016). Citizen-centric urban planning through extracting emotion information from Twitter in an interdisciplinary space-time-linguistics algorithm. *Urban Planning*, *1*, 114–127.

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.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *15*, 351–357.

Robinson, W. S. (2009). Ecological correlations and the behavior of individuals. *International Journal of Epidemiology*, *38*, 337–341.

Robinson, W. S. (2011). Erratum. *International Journal of Epidemiology*, *40*, 1134.

Rothman, K. J., Gallacher, J. E., & Hatch, E. E. (2013). Why representativeness should be avoided. *International Journal of Epidemiology*, *42*, 1012–1014.

Ruddell, D. & Wentz, E.-A. (2009). Multi-tasking: scale in geography. *Geography Compass*, *3*, 681–697.

Sagl, G. & Resch, B. (2014). Mobile phones as ubiquitous social and environmental geo-sensors. In Z. Yan (Ed.), *Encyclopedia of mobile phone behavior* (pp. 1194–1213). Hershey, PA: IGI Global.

Sagl, G., Resch, B., & Blaschke, T. (2015). Contextual sensing: integrating contextual information with human and technical geo-sensor information for smart cities. *Sensors*, *15*, 17013–17035.

Saib, M.-S., Caudeville, J., Carre, F., Ganry, O., Trugeon, A., & Cicolella, A. (2014). Spatial relationship quantification between environmental, socioeconomic and health data at different geographic levels. *International Journal of Environmental Research and Public Health*, *11*, 3765–3786.

Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes Twitter users: real-time event detection by social sensors. In M. Rappa, P. Jones, J. Freire, & S. Chakrabarti (Eds.), *Proceedings of the 19th International Conference on World Wide Web* (pp. 851–860). New York City, NY: ACM.

Sakaki, T., Okazaki, M., & Matsuo, Y. (2013). Tweet analysis for real-time event detection and earthquake reporting system development. *IEEE Transactions on Knowledge and Data Engineering*, *25*, 919–931.

Sakshaug, J. W., Couper, M. P., Ofstedal, M. B., & Weir, D. R. (2012). Linking survey and administrative records: mechanisms of consent. *Sociological Methods and Research*, *41*, 535–569.

Sattar, A. S., Li, J., Ding, X., Liu, J., & Vincent, M. (2013). A general framework for privacy preserving data publishing. *Knowledge-Based Systems*, *54*, 276–287.

Schaeffer, N. C., Dykema, J., & Maynard, D. W. (2010). Interviewers and interviewing. In J. D. Wright & P. V. Marsden (Eds.), *Handbook of survey research (2nd ed.)* Bingley, U.K.: Emerald Group Publishing Limited.

Scheuren, F. (2004). *What is a survey?* (2nd ed.). Washington, D.C.: American Statistical Association.

Schimmack, U., Diener, E., & Oishi, S. (2002). Life-satisfaction is a momentary judgment and a stable personality characteristic: the use of chronically accessible and stable sources. *Journal of Personality*, *70*, 345–384.

Schmiedeberg, C. & Schröder, J. (2014). Does weather really influence the measurement of life satisfaction? *Social Indicators Research*, *117*, 387–399.

Schnell, R. (2013a). *Getting big data but avoiding big brother.* Working Paper No. WP-GRLC-2013-02 (German Record Linkage Center Working Paper Series). German Record Linkage Center.

Schnell, R. (2013b). *Linking surveys and administrative data.* Working Paper No. WP-GRLC-2013-03 (German Record Linkage Center Working Paper Series). German Record Linkage Center.

Schwarz, N. & Clore, G.-L. (1983). Mood, misattribution, and judgments of well-being: informative and directive functions of affective states. *Journal of Personality and Social Psychology*, *45*, 513–523.

Schweers, S., Kinder-Kurlanda, K., Müller, S., & Siegers, P. (2016). Conceptualizing a spatial data infrastructure for the social sciences: an example from Germany. *Journal of Map and Geography Libraries*, *12*, 100–126.

Schyns, P. (1998). Cross-national differences in happiness: economic and cultural factors explored. *Social Indicators Research*, *43*, 3–26.

Shiffman, S. (2007). Designing protocols for ecological momentary assessment. In A. A. Stone, S. Shiffman, A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture: self-reports in health research* (pp. 27–53). New York, NY: Oxford University Press.

Sonnentag, S., Binnewies, C., & Ohly, S. (2013). Event-sampling methods in occupational health psychology. In R. R. Sinclair, M. Wand, & L. E. Tetrick (Eds.), *Research methods in occupational health psychology* (pp. 208–228). New York, NY: Routledge.

Steffens, M. & Mecklenbräuker, S. (2007). False memories: phenomena, theories, and implications. *Journal of Psychology*, *215*, 12–24.

Steiger, E., Resch, B., & Zipf, A. (2016). Exploration of spatiotemporal and semantic clusters of Twitter data using unsupervised neural networks. *International Journal of Geographical Information Science*, *30*, 1694–1716.

Steiger, E., Westerholt, R., Resch, B., & Zipf, A. (2015). Twitter as an indicator for whereabouts of people? Correlating Twitter with UK census data. *Computers, Environment and Urban Systems*, *54*, 255–265.

Sugovic, M. & Witt, J.-K. (2013). An older view on distance perception: older adults perceive walkable extents as farther. *Experimental Brain Research*, *226*, 383–391.

Swan, M. (2012). Sensor mania! The internet of things, wearable computing, objective metrics, and the quantified self 2.0. *1*, 217–253.

Swan, M. (2013). The quantified self: fundamental disruption in big data science and biological discovery. *Big Data*, *1*, 85–99.

Talhelm, T., Zhang, X., Oishi, S., Shimin, C., Duan, D., Lan, X., & Kitayama, S. (2014). Large-scale psychological differences within China explained by rice versus wheat agriculture. *Science*, *344*(6184), 603–608.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, *46*, 234–240.

Triantafyllidis, A., Velardo, C., Salvi, D., Shah, S. A., Koutkias, V., & Tarassenko, L. (2015). A survey of mobile phone sensing, self-reporting and social sharing for pervasive healthcare. *IEEE Journal of Biomedical and Health Informatics. 21*(1), 218–227.

Tröndle, M., Greenwood, S., Kirchberg, V., & Tschacher, W. (2014). An integrative and comprehensive methodology for studying aesthetic experience in the field: merging movement tracking, physiology, and psychological data. *Environment and Behavior*, *46*, 102–135.

Tuan, Y.-F. (1977). *Space and place: the perspective of experience.* Minneapolis, MN: University of Minnesota Press.

Tversky, B., Morrison, J.-B., Franklin, N., & Bryant, D.-J. (1999). Three spaces of spatial cognition. *The Professional Geographer*, *51*, 516–524.

Vasiliu, L., Freitas, A., Caroli, F., Handschuh, S., Ross McDermott, M., & … & Cavallini, A. (2016). *In or out? Real-time monitoring of BREXIT sentiment on Twit-*

ter. Paper presented at the SEMANTICS 2016 Conference., Leipzig, Germany.

Veena, N. & Devidas, B. (2014). Data anonymization approaches for data sets using map reduce on cloud: a survey. *International Journal of Science and Research*, *3*, 308–311.

Visser, P. S., Krosnick, J. A., & Lavrakas, P. (2000). Survey research. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social psychology*. New York, NY: Cambridge University Press.

Vogel, M. (2016). The modifiable areal unit problem in person-context research. *Journal of Research in Crime and Delinquency*, *53*, 112–135.

Waldhör, T. (1996). The spatial autocorrelation coefficient Moran's I under heteroscedasticity. *Statistics in Medicine*, *15*, 887–892.

Wang, H. (2010). Privacy-preserving data sharing in cloud computing. *Journal of Computer Science and Technology*, *25*, 401–414.

Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012). A system for real-time Twitter sentiment analysis of 2012 U.S. presidential election cycle. In M. Zhang (Ed.), *Proceedings of the ACL 2012* (pp. 115–120). Stroudsburg, PA: Association for Computational Linguistics.

Ward, M. & Gleditsch, K. S. (2008). *Spatial regression models.* Thousand Oaks, CA: Sage.

Weiss, E.-M., Kemmler, G., Deisenhammer, E.-A., Fleischhacker, W.-W., & Delazer, M. (2003). Sex differences in cognitive functions. *Personality and Individual Differences*, *35*, 863–875.

Wender, K. F., Haun, D., Rasch, B., & Bluemke, M. (2003). Context effects in memory for routes. In C. Freksa, W. Brauer, C. Habel, & K. F. Wender (Eds.), *Spatial cognition III. Routes and navigation, human memory and learning, spatial representation and spatial learning* (pp. 209–231). Lecture Notes in Computer Science, Vol. 2685. Heidelberg, Germany: Springer.

West, B. T., Kreuter, F., & Jaenichen, U. (2013). ''Interviewer" effects in face-to-face surveys: a function of sampling, measurement error, or nonresponse? *Journal of Official Statistics*, *29*, 277–297.

Westerholt, R., Resch, B., & Zipf, A. (2015). A local scale-sensitive indicator of spatial autocorrelation for assessing high-and low-value clusters in multiscale datasets. *International Journal of Geographical Information Science*, *29*, 868–887.

Westerholt, R., Steiger, E., Resch, B., & Zipf, A. (2016). Abundant topological outliers in social media data and their effect on spatial analysis. *PLOS ONE*, *11*(9), e0162360.

Witt, J.-K., Proffitt, D.-R., & Epstein, W. (2010). When and how are spatial perceptions scaled? *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 1153–1160.

Xu, P., Huang, H., Dong, N., & Abdel-Aty, M. (2014). Sensitivity analysis in the context of regional safety modeling: identifying and assessing the modifiable areal unit problem. *Accident Analysis and Prevention*, *70*, 110–120.

Yabiku, S., Glick, J., Wentz, E., Ghimire, D., & Zhao, Q. (2017). Comparing paper and tablet modes of retrospective activity space data collection. *Survey Research Methods*, *11*(3), XX.

Zadra, J.-R. & Clore, G.-L. (2011). Emotion and perception: the role of affective information. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*, 676–685.

Appendix
Tables

Table A1
Glossary of Terms

Label	Explanation	References
Geography	The discipline dealing with the interactions between humans (or natural systems) and space. The endeavor is limited to the humanly comprehensible scale.	Clifford, Holloway, Rice, and Valentine (2009)
Geographic Information Science (GIScience)	The discipline that investigates theoretical issues regarding the nature, acquisition, storage, analysis, and presentation of geospatial information and data, while abstracting these from specific geographic questions.	Goodchild (1992, 2010)
Geoinformatics	Largely overlaps with GIScience; preferred among German-speaking scholars; stronger technological focus, as it accentuates the development and application of methods and technology.	de Lange (2013)
Geographic Information System (GIS)	A GIS is a system of hardware, software, and procedures to support the capture, management, manipulation, analysis, modeling, and display of spatially-referenced data for solving complex planning and management problems.	Goodchild and Kemp (1992)