Methods for deriving and calibrating privacy-preserving heat maps from mobile sports tracking application data

Article *in* Journal of Transport Geography · September 2015

DOI: 10.1016/j.jtrangeo.2015.09.001

CITATIONS
8

READS
38

4 authors, including:

Juha Oksanen
Finnish Geospatial Research Institute / Natio…
32 PUBLICATIONS **333** CITATIONS
SEE PROFILE

Cecilia Bergman
Finnish Geospatial Research Institute
3 PUBLICATIONS **10** CITATIONS
SEE PROFILE

Journal of Transport Geography 48 (2015) 135–144

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrg

Methods for deriving and calibrating privacy-preserving heat maps from mobile sports tracking application data

Juha Oksanen [a,*], Cecilia Bergman [a], Jani Sainio [b], Jan Westerholm [b]

[a] Department of Geoinformatics and Cartography, Finnish Geospatial Research Institute/National Land Survey of Finland, P.O. Box 84, FI-00521 Helsinki, Finland
[b] Faculty of Science and Engineering, Åbo Akademi University, Joukahainengatan 3-5, FI-20520 Turku, Finland

ARTICLE INFO

Article history:
Received 19 December 2014
Received in revised form 3 September 2015
Accepted 5 September 2015
Available online 20 September 2015

Keywords:
Cycling
Location-based services (LBSs)
Urban planning
Privacy
Big data
GIS

ABSTRACT

Utilization of movement data from mobile sports tracking applications is affected by its inherent biases and sensitivity, which need to be understood when developing value-added services for, e.g., application users and city planners. We have developed a method for generating a privacy-preserving heat map with user diversity (ppDIV), in which the density of trajectories, as well as the diversity of users, is taken into account, thus preventing the bias effects caused by participation inequality. The method is applied to public cycling workouts and compared with privacy-preserving kernel density estimation (ppKDE) focusing only on the density of the recorded trajectories and privacy-preserving user count calculation (ppUCC), which is similar to the quadrat-count of individual application users. An awareness of privacy was introduced to all methods as a data pre-processing step following the principle of k-Anonymity. Calibration results for our heat maps using bicycle counting data gathered by the city of Helsinki are good ($R^2 > 0.7$) and raise high expectations for utilizing heat maps in a city planning context. This is further supported by the diurnal distribution of the workouts indicating that, in addition to sports-oriented cyclists, many utilitarian cyclists are tracking their commutes. However, sports tracking data can only enrich official in-situ counts with its high spatio-temporal resolution and coverage, not replace them.

1. Introduction

Mobile sports tracking applications have become popular among the public audience, and a large number of smartphone users are willing to collect and compare their workouts privately, as well as to share their data within social networks or even publicly, for all application and Internet users. Key factors in this development have been the maturity of sensor technology, such as an accelerometer, digital compass, gyroscope, and GPS (e.g., Lane et al., 2010), available in nearly all recent mid- and top-range smartphones; and well-documented application programming interfaces for third-party developers to create new applications for mobile platforms.

The aim of our work is to develop methods to enrich workout data from a mobile sports tracking application to create privacy-preserving information about the most popular places to do sports. While our case study focuses on public cycling workouts collected using *Sports Tracker* (http://www.sports-tracker.com/), the developed methods can be used for any other sports recorded using any mobile sports tracking application. The approach chosen for this study relies on visual data mining, which utilizes human perception in data exploration, and combines human flexibility, creativity, and knowledge with a computer's storage capacity, computing power, and visualization capabilities (Keim, 2002). When integrated into a location-based service (LBS), the result of our analysis replies to the end-user's question "*Where have most cyclists continued to from here?*" In addition, we investigate the relation of tracking data and in-situ bicycle counting information in order to compare the quality of the derived heat maps, as well as to calibrate the heat maps based on mobile sports tracking application data, for example, for city planning purposes. We use the term "workout" throughout the paper to denote all recorded trajectories, be they recreational/exercise or utilitarian by purpose.

The idea of generating heat maps from mobile sports app data to communicate the popularity of sports is not new (e.g., Garmin, 2013; Lin, 2012; Strava, 2014), but less attention has been paid to the methods for making the calculation, and concerns about the appropriate understanding of the heat maps have been raised due to new application areas of heat maps, such as city planning (Maus, 2014a) and analysis of eye-tracking data (Bojko, 2009). When creating heat maps, an obvious surrogate for the popularity of sports is the density of workout trajectories, but other surrogates, such as the number of different people doing sports, can also be used. According to the limited information available on existing heat maps, the one provided by Strava uses the number of GPS points as a pixel value (Mach, 2014), whereas in the heat map offered by Nike+, the value at each pixel represents the number of users (Lin, 2012). As we will show in this paper, the two methods can locally result in very different patterns of bike riding.

* Corresponding author. Tel.: +358 40 831 4092.
E-mail address: juha.oksanen@nls.fi (J. Oksanen).

http://dx.doi.org/10.1016/j.jtrangeo.2015.09.001

The use of heat maps as a representation of the intensity of a phenomenon has its roots in spectrometry (e.g., Moon et al., 2009) and the generation of isarithm and dot density maps (Slocum et al., 2009), but in the context of cartography, a textbook definition of a heat map is still missing (e.g., Trame and Keßler, 2011). Heat maps are a common visualization technique in many fields of research where large amounts of data are handled. For example, in human–computer interaction, "attention heat maps" are a popular tool of visual analysis of eye-tracking data (e.g., Blascheck et al., 2014; Bojko, 2009). Related to studies utilizing the increasing volumes of volunteered geographic information (VGI; Goodchild, 2007), heat maps have been used, for example, to reveal attractive/popular places in a region using the density-based spatial clustering of geotagged images (Kisilevich et al., 2010; Kurata, 2012) and videos (Mirković et al., 2010), or to visualize spatio-temporal patterns revealed by the distribution of tweets (e.g., Morstatter et al., 2013; Zeile et al., 2012). The coloring of a heat map is typically selected in such way that the interpretation of the intensity differences becomes intuitive. This is expected to happen when warm colors, in terms of color temperature, are used for high intensities of the represented phenomenon and cool colors for low intensities (e.g., Špakov and Miniotas, 2007).

In addition to the public audience, interest in mobile tracking applications, and enriching, especially, cycling data collected with them has emerged among city planners (Albergotti, 2014; Charlton et al., 2011; Hood et al., 2011). One of the biggest challenges in a city-planning context regarding non-motorized traffic is the lack of documentation on the usage and the demand figures for, for example, cyclists and pedestrians (Lindsey et al., 2014; NBPD, 2014). Traditional approaches for monitoring cycling traffic have been the use of surveys for qualitative results and different types of manual and automatic in-situ counting for quantitative results (Griffin et al., 2014; NBPD, 2014; Rantala and Luukkonen, 2014). Mobile tracking of cyclists has been seen as an attractive, inexpensive, and dynamic alternative to traditional bicycle data collection (Hudson et al., 2012). An early approach to tracking was the development of dedicated platforms, such as CycleTracks, which was developed at San Francisco Municipal Transportation Agency and has since been used at a number of agencies and municipalities in the US (Charlton et al., 2011; Masoner, 2014). In the UK, another crowdsourcing-based application, Cycle Hackney, is expected to provide a cost-effective way to find out where, especially, utilitarian cycling is taking place (CycleStreets, 2014). Recently, in the city of Oulu, Finland, there has been a development project aiming to create a "smoothness navigator" for cyclists, based on 1000 recorded tracks of people participating in the pilot phase (Poikola, 2014). The problem in dedicated tracking platforms appears to be the limited group of people interested in using them voluntarily (SFMTA, 2013). To overcome this problem the potential of utilizing mobile sports tracking data has been recognized, reflecting the idea of utilizing humans as sensors (Goodchild, 2007) and big data analytics (e.g., Russom, 2011). For example, Oregon's Department of Transportation paid $20,000 to use data from the mobile sports tracking application Strava for a year, containing 400,000 individual bicycle trips, totaling 8 million bicycle kilometers traveled (Estes, 2014; Maus, 2014b).

While mobile sports tracking data may not qualify as 'big data' regarding its volume – except in the sense "bigger than previously" (Goodchild, 2013) – it shares many characteristics with other social media data, often classified as big data. Many of the characteristics follow from the fact that big data is typically not collected with any specific purpose in mind or not used for its original purpose (Kitchin, 2014). In statistics, random sampling is used to guarantee the representativeness of observations, but in big data analytics, the 'sample' is not randomly chosen at all (Goodchild, 2013). Rather, the aim is to use all the data following the principle of exhaustivity in scope (Harford, 2014; Kitchin, 2014). However, considering that only a small and possibly behaviorally biased subset of cyclists use mobile applications to track their routes, the question is how well they represent the whole population of cyclists (e.g., Maus, 2014a; Rantala and Luukkonen, 2014); i.e., as social media data in general, sports tracking data is affected by self-selection bias (Shearmur, 2015). In addition, mobile tracking applications have their differences with respect to appearance and function, such as the available range of sports (multi-sports or single activity type), and may therefore attract different people. As an example, the cycling and running app Strava has the reputation of being used by more competitive or "serious" cyclists (Zahradnik, 2014) and is also targeting people who identify themselves as "athletes" (Strava, 2014). On the other hand, for example, Sports Tracker (ST, 2014) "want[s] to help people train better, connect through sports, and live healthier, happier lives;" HeiaHeia! focuses on the business-to-business sector and work welfare (Kauppalehti, 2013); and Endomondo (2014) is, in its own words, aiming "to motivate people to get and stay active." It has been estimated that 90% of the cyclists who use Strava are male (Usborne, 2013; Vanderbilt, 2013) and in 2012, 75% of all Endomondo users were men (Endoangela, 2012). Furthermore, participation inequality is a known property of VGI and online communities, according to which 90% of community members are followers and do not contribute to the community, whereas 9% contribute from time to time, and 1% account for most contributions (Nielsen, 2006). Although sports tracking applications do not today represent shared projects where people would track and share their workouts to promote the common good, some typical motivations for contribution in VGI, such as social reward and enhanced personal reputation (Coleman et al., 2009), can be identified within their communities as well. These bias issues introduce a major challenge in using mobile sports app data in a city-planning context.

According to Westin's tenet, privacy is an individual's right to have full control over information about themselves, and to decide when, how, and to what extent this information is shared with others (Agrawal et al., 2002). Guaranteeing privacy in LBSs is extremely important, due to the unique characteristics of moving object data (Fung et al., 2010; Montjoye et al., 2013; Verykios et al., 2008). Topics such as anonymization of the original dataset (e.g., Monreale et al., 2010; Pensa et al., 2008), or de-identifying a given LBS-request location (e.g., Bettini et al., 2005; Gedik and Liu, 2004; Gruteser and Liu, 2004), have gained a great deal of attention in trajectory studies but are beyond the scope of this paper. Instead, we approach privacy-preservation from the standpoint of visualization.

The idea behind preserving privacy in visualizations is to generalize or otherwise obfuscate data in such a manner that the disclosed data is still useful in the particular case (Andrienko et al., 2008; Fung et al., 2010). Various methods of geographical masking, first introduced by Armstrong et al. (1999), have been developed with the aim of protecting the confidentiality of individual locations by adding stochastic or deterministic noise to the geographic coordinates of the original data points without substantially affecting analytical results or the visual characteristics of the original pattern (Kwan et al., 2004). Spatial aggregation of individual-level data for administrative areas or other areal units that have a population greater than a chosen cutoff value is a common procedure of preserving confidentiality, for example, in censuses where disclosure control has long been an integral part of the process (Armstrong et al., 1999; Kwan et al., 2004; Leitner and Curtis, 2006; Young et al., 2009). Because aggregation can hide important spatial patterns in the data, various alternative geo-masking techniques, such as random perturbation and affine transformation (translate, rotate, and scale), have been introduced to preserve the disaggregated, discrete nature of the original data (Armstrong et al., 1999; Kwan et al., 2004). Although they have been used mainly with geo-referenced, sensitive health- and crime-related point data (e.g., Leitner and Curtis, 2006; Kounadi and Leitner, 2015), Krumm (2007) and Seidl et al. (2015) have applied them also to GPS trajectory data. In this study, where it was crucial to prevent re-identification of an individual user and trajectory while providing the heat map viewer accurate information about popular cycling paths in their actual locations on the road network, geo-masking techniques as such were, however, not

adequate. Instead, we followed the principle of k-Anonymity in anonymization of the final heat maps. K-Anonymity was originally developed for record data by Samarati and Sweeney (1998), but has since been extended to spatio-temporal movement data by, for example, Abul et al. (2008), and Terrovitis and Mamoulis (2008). Related to visualization techniques, the method has been previously applied to privacy preservation for parallel coordinates (Dasgupta and Kosara, 2011). K-Anonymity refers to the requirement according to which "each release of data must be such that every combination of values of quasi-identifiers can be indistinctly matched to at least k individuals" (Samarati and Sweeney, 1998). By quasi-identifiers, they mean a set of attributes whose publication needs to be somehow restricted. When bringing the k-Anonymity requirement into the heat map context, our goal was that all details interpretable from the final heat map must be such that the result has been borne out from at least a minimum pre-defined number of application users.

This paper presents a novel method for deriving privacy-preserving heat maps from mobile sports application data which takes into account the density of trajectories, as well as the diversity of users, attempting to avoid the bias caused by having few very active sports application users. As a secondary objective, the study presents a method by which heat maps derived from a mobile sports application's cycling data can be calibrated using in-situ field-collected bicycle counting data. With this approach, cities may use the calibrated heat maps as source information for city planning purposes. The remainder of the paper is organized as follows. In Section 2, we describe the data and methods used for deriving privacy-preserving heat maps. In Section 3, we compare the heat maps and discuss the characteristics of different methods, and we also show the relation between our heat maps and real-world in-situ bicycle counting data. Finally, in Section 4, conclusions are drawn and some future directions are pointed out.

2. Materials and methods

The data used for our study was obtained from Sports Tracking Technologies Ltd., the inventor of the first mobile sports tracking application, Sports Tracker, for mobile phones (ST, 2014). While the app turns a smart phone into a sports computer, it is also a complete social platform on which application users share their workouts, photos, and experiences of exercises, with their group of friends or even with everyone. The basic unit of recorded data is a workout, which contains information about the user, sport (28 pre-classified sports, such as cycling, walking, and running), and a 4D (x, y, z, and t) coordinate list recorded at approximately 1-second intervals from the start to the finish of the workout. For this study, the data contained only public workouts from the region of Helsinki, and each user identifier was changed into a pseudo-ID at Sports Tracking Technologies Ltd. before the data delivery.

The total sample data contained 192,597 workouts, of which a subset of 36,757 workouts collected by 2424 users represented cycling inside our approximately 320 km^2 study region. Temporally continuous data was recorded between April 17th 2010 and November 21st 2012, and it consisted of a total of 36,663,190 GPS points. About 82% of the data contained valid time values accepted for further analysis. In a pre-processing phase, the data was systematically thinned to 10-second time intervals, the time data was transformed from POSIX time to Coordinated Universal Time (UTC + 02:00), and coordinates were converted from WGS84 (EPSG:4326) to ETRS-TM35FIN (EPSG:3067) (Anon, 2012). In addition, each point was supplied with information about distance, time, and speed from the previous tracked point. Further filtering of gross errors was done based on thresholds set for distance (<300 m) and speed (<50 m/s) between consecutive GPS observations.

One characteristic of the data was that the number of people tracking and publishing a large number of workouts was small, and most of the users had tracked less than 10 workouts (Fig. 1), which appears to follow the principle of participation inequality (Nielsen, 2006). In the study area and within the time range, 65% of the users had published 5 or fewer workouts, and 87% had published 20 or fewer workouts. While the most active user had published more than 600 workouts, only 3% of the users had published more than 100 workouts. On average, for the sample data, each user had published 15 workouts.

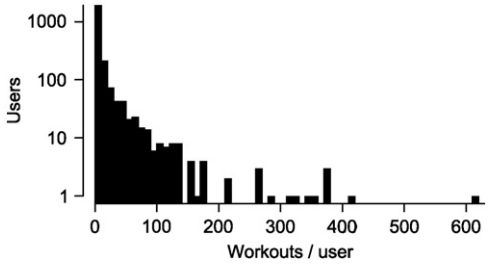

Fig. 1. Number of tracked and published cycling workouts per user.

Further inspection of the data revealed cyclicity at a number of temporal scales. At an annual level, the popularity of cycling increased sharply during the summer season and peaked in August (Fig. 2a and b). In addition, the popularity of tracking cycling increased steadily from 2010 to 2012. When focusing on the weekly pattern, tracking of cycling was at its highest level on Tuesdays and Wednesdays, while the minimum level was reached on Fridays (Fig. 2c). Peaks in the frequencies of tracked points between 7–8 AM and 4–5 PM (Fig. 2d) reveal that many people use Sports Tracker to track their daily commuting trips.

The use of pseudoIDs is not an adequate means for preserving the privacy of the application users since location alone may also reveal sensitive personal data (e.g., Bettini et al., 2005; Samarati and Sweeney, 1998). Privacy filtering of the tracked points was done in a pre-processing phase containing: 1) the generation of trajectories from all points, 2) the generation of a user count raster (number of different users on a 10 m grid within a 15 m search radius) from the trajectories, 3) the extraction of user count values at all points, and 4) the generation of trajectories from the subset of points with a user count value higher than a pre-defined threshold (here 5 users). By this method, we were able to filter privacy-preserving trajectories, which contained only the features that were covered by an adequate number of different users.

From the filtered cycling trajectories we generated heat maps by a custom ArcGIS tool using three methods, namely privacy-preserving user count calculation (ppUCC), privacy-preserving kernel density estimation (ppKDE), and privacy-preserving kernel density estimation modified with the user diversity index (ppDIV). ppUCC was the simplest of the applied methods, in which the study region was partitioned into sub-regions of equal area and each area got its value, so-called quadrat counts (Bailey and Gatrell, 1995), from the number of users passing the pixel or a larger calculation window (20 m in our case). Thus, using ppUCC, we created a 2D histogram of individual cyclists within the study area.

Kernel density estimation (KDE) is a family of methods originally developed to obtain smooth estimates of uni- or multivariate densities from observations (Bailey and Gatrell, 1995), but recently KDE has also been widely used to derive heat maps from data representing moving objects (e.g., Krisp and Peters, 2011; Willems, 2011). The commonly used kernel function is described by Silverman (1986):

$$\mathrm{ppKDE}(s) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{s - s_i}{h}\right),$$

where h is the width of the calculation window (bandwidth), n is the number of sample points, and K is the kernel function used for smoothing the estimate. Here, a quartic approximation of a Gaussian kernel (Silverman, 1986) was used. Sample points within the radius h are

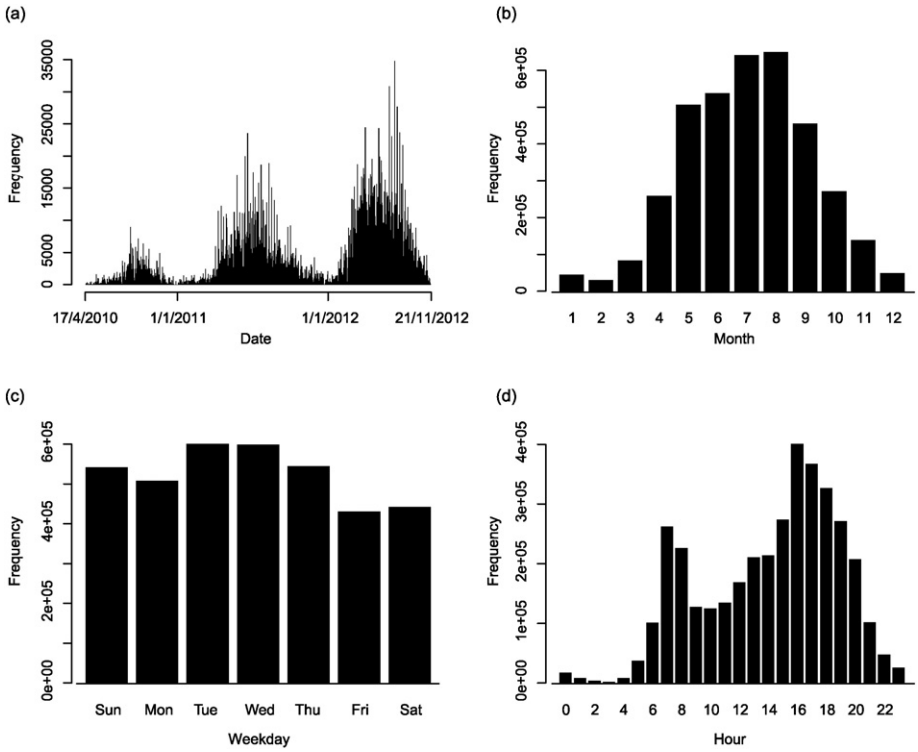

Fig. 2. Frequencies of tracked points a) daily, b) monthly, c) according to weekday, and d) according to the hour of the day.

represented by s_i. When KDE is used for line features, such as trajectories, the result of KDE can be thought of as the result of placing a smooth kernel surface on top of the lines. The bandwidth was chosen as 25 m, to generalize the positioning uncertainty of GPS devices, but simultaneously to preserve the details of the street network along which the cycling has occurred.

In ppDIV the result of ppKDE was scaled with the diversity of users at each quadrat. Diversity is a descriptive statistic of a population with a

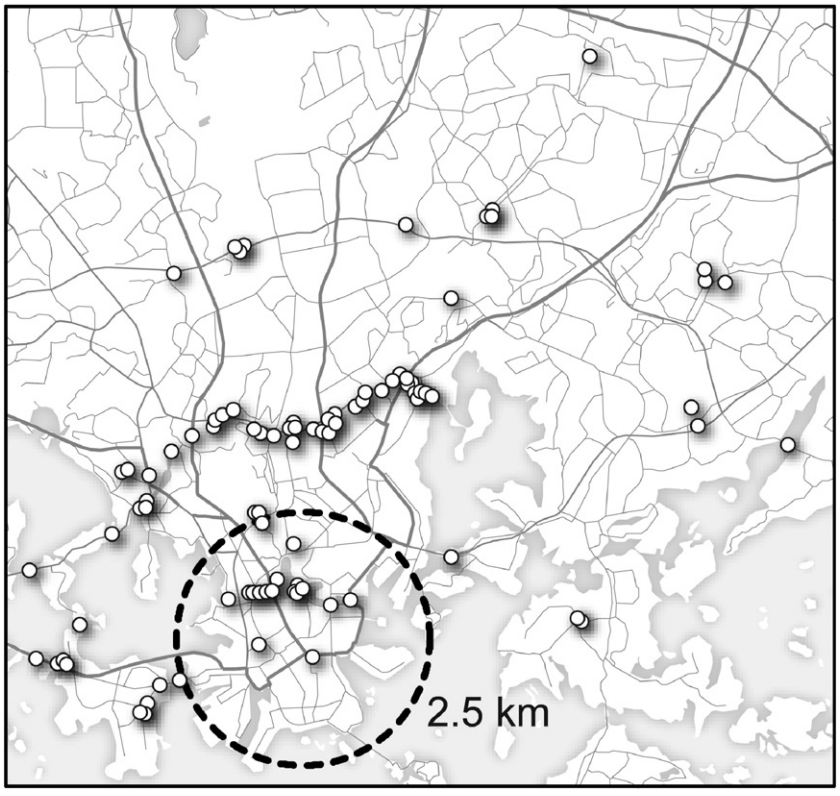

Fig. 3. Locations of the 89 bicycle counting sites in the city of Helsinki in June 2013 (Hellman, 2013) used in the study. The dashed circle represents a distance zone of 2.5 km from the city center. The map contains data from the Topographic database by the National Land Survey of Finland, 7/2012.

class structure (Junge, 1994). Diversity indices are used when we are interested in quantifying how many different classes there are in our data, and simultaneously how evenly the entities are distributed among those classes. The index used in this study is Simpson's Diversity Index D (McDonald and Dimmick, 2003):

$$D(s) = 1 - \sum p(s)_i^2$$

$$ppDIV(s) = ppKDE(s) * D(s)$$

where p_i is the proportion of user i's trajectories among all trajectories in the quadrat s. This resulted in an index that was close to 1 when the user distribution was locally uniform, and close to 0 when skewness of the user distribution was locally high.

Finally, the heat maps were compared with each other, as well as with the official 2013 bicycle counting data gathered by the city of Helsinki (Hellman, 2013). The challenge in a quantitative comparison of densities represented as heat maps based on different calculation methods is that the information varies in terms of the shape and scale of the density zones. Therefore, point sampling was done by extracting values of ppUCC, ppKDE, and ppDIV at the centroids of the Topographic database's (NLS, 2014) road segments from the Helsinki region for

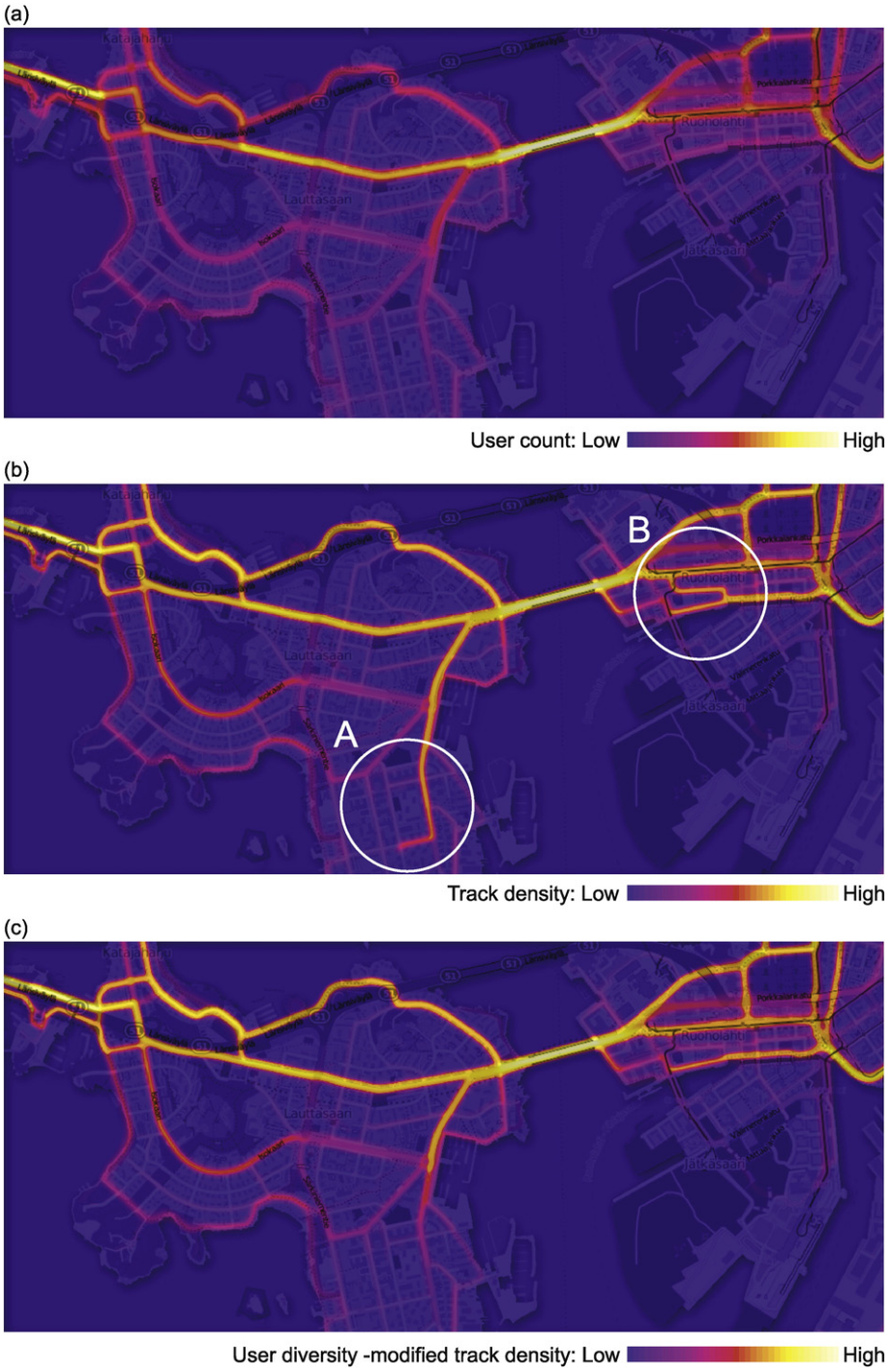

Fig. 4. Heat maps based on (a) privacy-preserving user count calculation (ppUCC), (b) privacy-preserving kernel density estimation (ppKDE), and (c) privacy-preserving kernel density estimation modified with the user diversity index (ppDIV). Major differences between the methods are found in the highlighted regions A and B. The map contains data from the Topographic database by the National Land Survey of Finland, 7/2012; © OpenStreetMap contributors.

regression analysis purposes. The extraction was done by a nearest neighbor assignment for ppUCC, and by a bilinear interpolation for ppKDE and ppDIV. Manual bicycle counting was done at 94 counting points in early June 2013, during one weekday between 7 AM–7 PM, and this was transformed to 24 h counts with a supply from semantically near automatic counting stations (Hellman, 2014). The precise locations of the counting sites were undocumented, but road segments used for the counting were published, and we adjusted the location of the 89 unambiguous points to correspond to the local maxima of the heat maps (Fig. 3).

3. Results and discussion

In general, the three methods for deriving heat maps produced similar looking results, but when focusing on the details, some major differences could be found. In the simplest heat map based on ppUCC (Fig. 4a), patterns of the popularity of cycling in the region became clearly visible, but the problem appeared to be a lack of knowledge about the density of the trajectories. For example, in ppUCC, a road segment with 100 workouts recorded by 10 users got the same heat map score as a segment with 10 workouts recorded by 10 users. In ppKDE (Fig. 4b), the density of the trajectories became visible and the previously mentioned limitation of ppUCC was surpassed, but a single very active user may result in significant bias and overestimation of the route section's popularity (Fig. 4a, highlights A and B). In ppDIV (Fig. 4c), we took into account the density of the workout trajectories and the diversity of users, which resulted in a heat map that closely corresponded to ppKDE but that did not suffer from the bias introduced by very active application users.

To further compare the methods, regression analysis between ppUCC, ppKDE, and ppDIV was performed on the point sample, based on the centroids of all road segments in the study region. In accordance with our expectations the coefficients of determination for regression models were high (Fig. 5), $R^2 = 0.83–0.94$ with the highest coefficient being between ppKDE and ppDIV (Fig. 5c), revealing the similarity of the methods in general. Most often, the low number of users was associated with a small number of individual tracks, and a large number of users also indicated a large number of tracks. However, the most interesting features of the regression models were found from a spatial analysis of the residuals of the models.

By residual analysis we could find answers to questions such as why a large number of tracks were not always related to a large number of users (Fig. 5a), and where were the places where the impact of taking the diversity of users into account in the calculation of densities of trajectories was the biggest (Fig. 5c). Observations below the regression curve in Fig. 5a revealed road segments where the density of trajectories (ppKDE) was lower than might have been expected according to the number of individual users (ppUCC). When plotting those parts of the observations on the map (Fig. 6a, blue dots), we could see that many of these road segments are important through-roads and entry cycling roads to downtown Helsinki. These also appeared to be routes not favored by the most active cyclists, by whom we mean cyclists actively tracking their workouts. In the opposite case (Fig. 6a, red dots), the density of trajectories (ppKDE) was higher than predicted based on the number of individual users (ppUCC). These were found from the routes with very active cyclists, either because of the route's popularity among the enthusiastic sports cyclists or because of a very active user using the Sports Tracking application to track daily commuting trips. While the first group is interesting in terms of finding the most popular routes for cycling, the second group clearly displays bias, and it should be possible to filter it out from the results.

In a similar manner, analyzing residuals from the regression model between ppKDE and ppDIV revealed the road segments where the diversity of the users had the biggest impact on ppKDE. Again, observations below the regression curve in Fig. 5c revealed road segments where a low diversity of users decreased the density value in the heat map (Fig. 6b, blue dots). In other words, they were the road segments where either a single cyclist or very few active cyclists had caused the largest positive bias in ppKDE. In the opposite case (Fig. 6b, red dots), a high diversity of users had resulted in a relative increase in ppKDE. These road segments appeared to be mostly the same important through-roads and entry cycling roads to downtown Helsinki, highlighted in blue in Fig. 6a.

When comparing our heat maps to real-world data, all methods for deriving heat maps performed almost equally well (Fig. 7). Any of the heat maps can be used in predicting 24 h bicycle counting data, keeping in mind the coefficients of determination, $R^2 = 0.49–0.50$ ($p < 0.001$). In practice this means that while, in many places, heat maps and real-world counting data had a clear connection, there are places where predictions based on heat maps fail. This may result at least partly from a temporal mismatch of the datasets and fundamental changes occurring in the cycling infrastructure. From the ten largest absolute residuals in the regression model between ppDIV and 24 h counting data (Fig. 7c, red and blue highlighting), we see that in 80% of the residuals, real-world 24 h counting data was greater compared to the predicted value based on ppDIV, and for only 20% the opposite was true.

When these ten largest residuals were plotted on a map (Fig. 8), we noticed that the maximum (point 26) was located on Baana, a popular cycle path opened on June 12th 2012. Our data covered the time period from April 17th 2010 to November 21st 2012, which means that only the last 6 months of our data contained cyclists using Baana. A similar fundamental change in cycling infrastructure had taken place near measurement point 13 where the new cycle and pedestrian bridge Aurora was inaugurated in late 2012 (HS, 2012). The other large positive residuals could at least partly be explained by the overall increase in cycling (Hellman, 2013), which has been most significant on main entry cycleways to the city (points 7, 16, 21, 28, and 30). Together with the essential cycle-way through downtown Helsinki (point 27), these might also be locations where the difference between everyday cyclists and cyclists using Sports Tracker to record their workouts is largest. At points 7, 16, 30, and 28, manual bicycle counting data has been collected for opposite lanes, and the data was generalized to a single point. Using this method

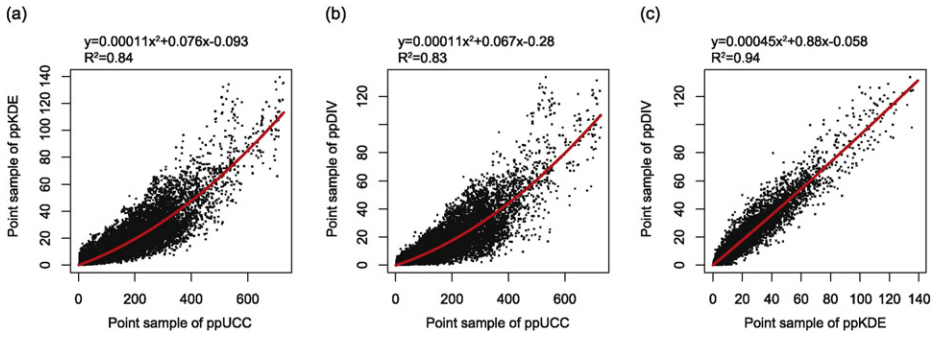

Fig. 5. Regression models between (a) ppUCC and ppKDE, (b) ppUCC and ppDIV, and (c) ppKDE and ppDIV.

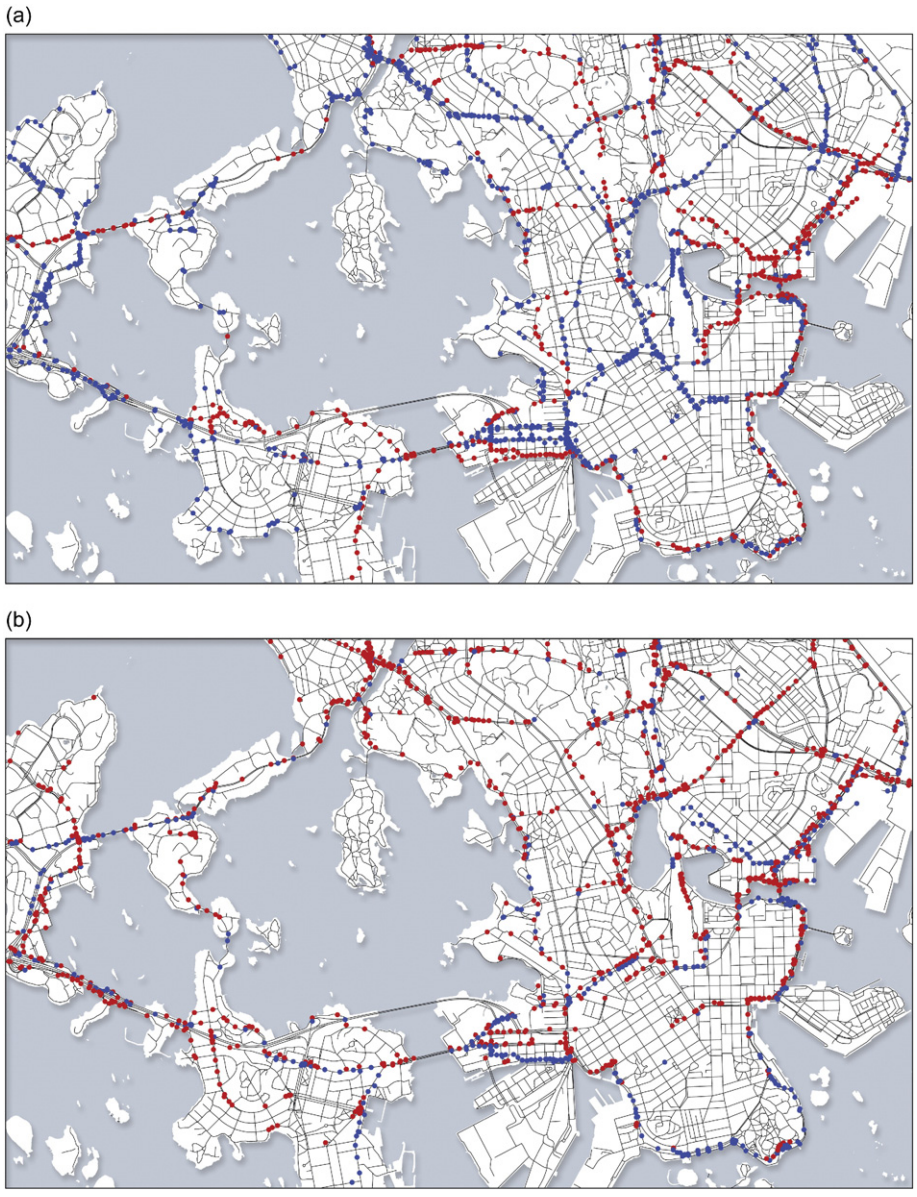

Fig. 6. Spatial distribution of maximum absolute residuals from the regression models between (a) ppUCC and ppKDE, and (b) ppKDE and ppDIV. The lowest 5th percentile of the residuals is represented with blue dots and the highest 5th percentile with red dots. The map contains data from the Topographic database by the National Land Survey of Finland, 7/2012.

to derive heat maps may be insensitive to the traffic on different lanes and underestimates the popularity indicating problems induced by the size of the calculation window. On the other hand, at point 8, the manual counting data has been collected only from one side of the road, and a generalization of the result into a single point resulted in a large positive bias in prediction. The other negative residual (point

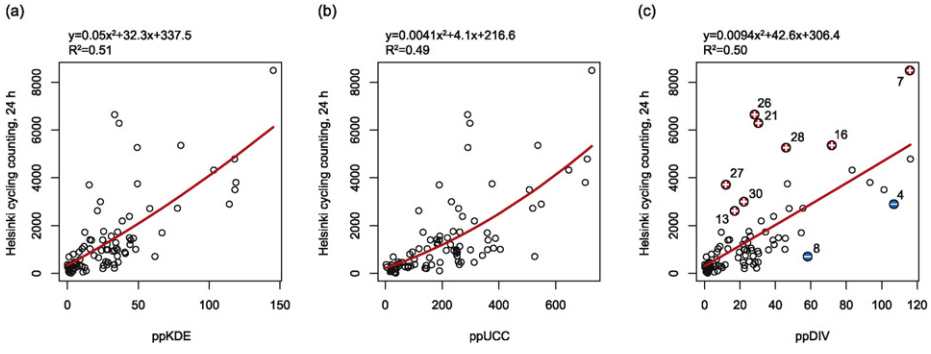

Fig. 7. Regression models between (a) ppKDE, (b) ppUCC, and (c) ppDIV and 24 h bicycle counting data 2013 from the city of Helsinki. In panel (c), the ten largest positive (red, + sign) and negative (blue, – sign) residuals from the regression model are highlighted. The same labels are used in the map in Fig. 8.

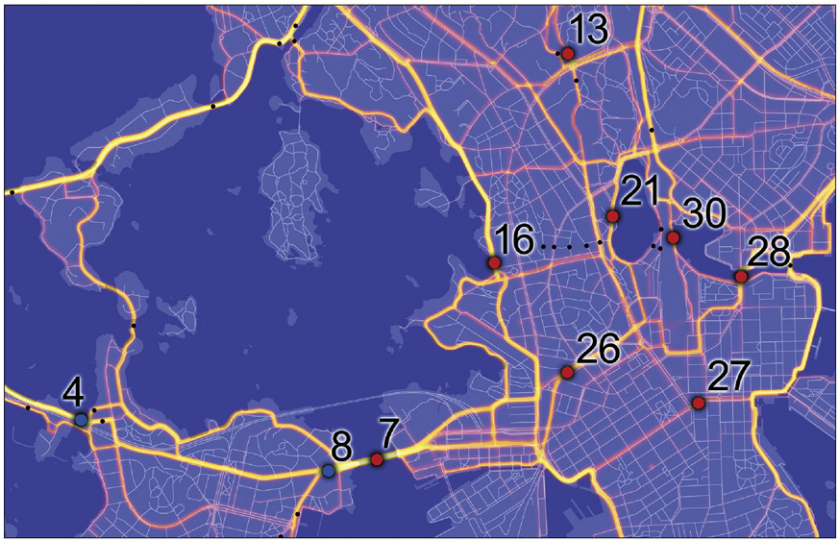

Fig. 8. The ten largest positive (red) and negative (blue) residuals from the regression model between ppDIV and 24 h bicycle counting data (black) in the city of Helsinki (Fig. 7c). The background heat map is based on ppDIV. The map contains data from the Topographic database by the National Land Survey of Finland, 7/2012.

4) might be explained by the construction site of Koivusaari metro station, which affected routes in 2013. When we compared the location of all the above-mentioned positive residuals to the distance zones in Helsinki (Fig. 3), they were all within a 2.5 km radius from the city center.

Finally, when the observations resulting in the ten largest residuals were removed from the data (Fig. 9a), the coefficient of determination for predicting the number of cyclists based on ppDIV rose to $R^2 = 0.76$ ($p < 0.001$). When focusing on the ten largest residuals, and by removing the points where significant changes in cycling infrastructure have occurred (points 13, 26, and 4) and where uncertainty due to a mismatch in in-situ counting and the heat map was the largest (point 8), we also got a high coefficient of determination, $R^2 = 0.72$ ($p = 0.15$) (Fig. 9b). Furthermore, if the points for Töölönranta (point 21) and Kaisaniemenranta (point 30) were removed, the coefficient of determination rose to $R^2 = 0.96$ ($p = 0.20$). This clearly indicates that, in our case study, a calibration of heat maps with the absolute number of cyclists outside a 2.5 km radius can be done with moderate accuracy, and calibration of the city center heat map would be best done as a separate processing step.

To summarize the methods (Table 1), it appears that the advantages of ppUCC are the simplicity of the calculation and the intuitive quantity of the result, the number of different users per quadrant. In addition, ppUCC automatically reduces the impact of very active cyclists and therefore, for example, diminishes the bias caused by active commuters. The disadvantage of ppUCC is that it ignores the density of trajectories and therefore brings, for example, mass sports events with a large number of individual athletes into the heat maps. In ppKDE, the density of trajectories is calculated using the well-established kernel density estimation, but individual active cyclists may introduce significant bias into the heat map. In addition, the unit of the heat map is not intuitive, even though in a visual interpretation of the results, the unit of the quantity is noncritical. A novel method introduced in this paper is ppDIV, which combines the best properties of the previously mentioned methods. It takes into account the density of the workout trajectories, as well as the diversity of the users who have tracked the workouts. Therefore, an individual athlete has no significant impact on the resulting heat map, and the computational biases of ppUCC and ppKDE are eliminated. A disadvantage of all the methods is that the results depend on the subjective definition of the size of the calculation window. This decision should be based, on one hand, on the positioning uncertainty of the workout trajectories, and on the other hand, on the desired level of detail of the final results.

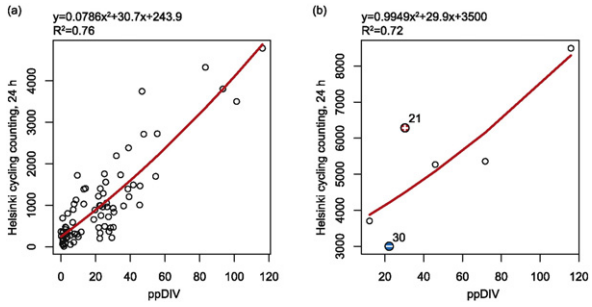

Fig. 9. Regression models between ppDIV and 24 h bicycle counting data 2013 from the city of Helsinki, when (a) observations in the city center (shown in Fig. 8) have been removed and (b) when focusing only on the observations in the city center. In panel (b), the two largest positive (red, + sign) and negative (blue, – sign) residuals from the regression model are highlighted.

4. Conclusions

In this paper, we have introduced a privacy-preserving diversity method (ppDIV) for deriving heat maps from mobile sports tracking application data, which takes into account the density of the trajectories and the diversity of the users. The method was applied to Sports Tracker's public cycling workouts and compared with privacy-preserving kernel density estimation (ppKDE) and privacy-preserving user count calculation (ppUCC). In addition, we demonstrated a method for calibrating the cycling heat map with in-situ bicycle counting data.

Table 1
Performance summary of the properties of the heat map generation methods (+ = poor, ++ = moderate, +++ = good).

	ppUCC	ppKDE	ppDIV
Simplicity of calculation	+++	++	+
Sensitivity to density of trajectories	+	+++	+++
Sensitivity to the number of individual users	+++	+	++
Sensitivity to diversity of users	+	+	+++
Intuitiveness of the measurement unit of the result	+++	+	+
Sensitivity to the size of the calculation window	+	+	+
Ability to filter out very active users	+++	+	+++
Ability to filter out mass sports events	+	+	+

The results show that ppUCC, ppKDE, and ppDIV reveal different aspects of the popularity of cycling, and they all are sensitive in different ways to bias issues in the mobile sports tracking data. The order of superiority between them depends on the requirements set for the final result, but for general purposes, ppDIV offers the most neutral view on the popularity of cycling, and hides the bias issues related to, for example, very active application users. The final added-value application that would inspire and support visual reasoning with the aim of choosing attractive routes for cycling could be further improved by allowing the end-user to filter the heat map based on subjective preferences on, for example, speed and time of day or year.

A balance between location privacy and data accuracy is an important but thus far largely uninvestigated topic in the context of heat maps. In our study area, k-Anonymity-based privacy preservation of the heat maps could be seen as a valid technique to guarantee the non-disclosure of individual users without diminishing the value of the service to the end-users. Still, an open question remains whether the value k should be a function of the density of the population or the underlying road network to protect the confidentiality of sensitive locations even in areas where there are small variations in the routes. A similar issue has been discussed in many geographical masking studies that have suggested weighting the displacement distance by population density because in less densely populated areas the risk of re-identification is higher (e.g., Armstrong and Ruggles, 2005; Kwan et al., 2004; Murad et al., 2014; Seidl et al., 2015).

When considering the usability of our heat maps from a city-planning perspective, the chosen approach for big data analytics appears to be promising. Despite the recognized bias issues, due to a selected group of cyclists tracking their workouts, as well as participation inequality, calibration results for our heat maps are surprisingly good. When doing regression modeling between in-situ bicycle counting data and ppDIV heat map scores, by splitting the data based on a 2.5 km distance from the city center, the coefficients of determination rise to $R^2 > 0.7$. Most likely, the coefficients of determination would be even higher when bicycle counting data and the heat map are derived from the same period of time. This clearly shows the potential of utilizing heat maps in a city-planning context by using the in-situ bicycle counting data to get the absolute scale and the heat map for getting a high level of detail and a large spatial coverage of cycling activities. The use of big data from mobile sports apps also offers a chance for significant savings in the investments made in light traffic counting. In the long term this could significantly help to improve cycling infrastructure development and planning. Still, it is worth noting that big data analytics does not replace the need for in-situ bicycle counting since uncalibrated heat maps suffer from the bias issues caused by a behaviorally biased subset of cyclists using mobile applications to track their routes, as well as participation inequality.

Acknowledgments

We thank Sports Tracking Technologies Ltd. for the possibility to use their public workout data in our research, and Mrs. Susanne Suvanto for assistance in data processing. The study is a part of the research project SUPRA (Revolution of Location-Based Services: Embedded data refinement in Service Processes from Massive Geospatial Datasets) funded by Tekes, the Finnish Funding Agency for Innovation (grants 40261/12 and 40262/12). In addition, Oksanen's funding from the Academy of Finland (grant 251987) is gratefully acknowledged.

References

Abul, O., Bonchi, F., Nanni, M., 2008. Never walk alone: uncertainty for anonymity in moving objects databases. Proceedings of the 24th IEEE International Conference on Data Engineering (ICDE'08), Cancun 7–12 April 2008, pp. 376–385.

Agrawal, R., Kiernan, J., Srikant, R., Xu, Y., 2002. Hippocratic databases. Proceedings of the 28th International Conference on Very Large Databases (VLDB'02), pp. 143–154.

Albergotti, R., 2014. Strava, popular with cyclists and runners, wants to sell its data to urban planners. Digits - Tech News and Analysis from the Wall Street J. (http://blogs.wsj.com/digits/2014/05/07/strava-popular-with-cyclists-and-runners-wants-to-sell-its-data-to-urban-planners/, accessed 29th July 2014).

Andrienko, G., Andrienko, N., Kopanakis, I., Ligtenberg, A., Wrobel, S., 2008. Visual analytics methods for movement data. In: Giannotti, F., Pedreschi, D. (Eds.), Mobility, Data Mining and Privacy — Geographic Knowledge Discovery. Springer, Berlin, pp. 375–410.

Anon, 2012. ETRS89 -järjestelmään liittyvät karttaprojektiot, tasokoordinaatistot ja karttalehtijako (map projections, projected coordinate system, and map sheet division related to ETSR89, in Finnish). Julkishallinnon suosituksia (JHS) 154 (http://www.jhs-suositukset.fi/suomi/jhs154 (accessed 15th December 2014)).

Armstrong, M.P., Ruggles, A.J., 2005. Geographic information technologies and personal privacy. Cartographica 40 (4), 63–73. http://dx.doi.org/10.3138/RU65-81R3-0W75-8V21.

Armstrong, M.P., Rushton, G., Zimmerman, D.L., 1999. Geographically masking health data to preserve confidentiality. Stat. Med. 18 (5), 497–525. http://dx.doi.org/10.1002/(SICI)1097-0258(19990315)18:5<497::AID-SIM45>3.0.CO;2-#.

Bailey, T.C., Gatrell, A.C., 1995. Interactive Spatial Data Analysis. Pearson Education, Harlow, UK.

Bettini, C., Wang, X.S., Jajodia, S., 2005. Protecting privacy against location-based personal identification. In: Jonker, W., Petkovíc, M. (Eds.), Proceedings of the 2nd VDLB international conference on Secure Data Management (SDM'05)Lecture Notes in Computer Science 3674. Springer, Berlin, pp. 185–199. http://dx.doi.org/10.1007/11552338_13.

Blascheck, T., Kurzhals, K., Raschke, M., Burch, M., Weiskopf, D., Ertl, T., 2014. State-of-the-art of visualization for eye tracking data. In: Borgo, R., Maciejewski, R., Viola, I. (Eds.), State of the Art Report, Eurographics Conference on Visualization (EuroVis'14) (http://www.visus.uni-stuttgart.de/fileadmin/vis/pdf_s_fuer_Publikationen/State-of-the-Art_of_Visualization_for_Eye_Tracking_Data.pdf, accessed 15th December 2014).

Bojko, A., 2009. Informative or misleading? Heatmaps deconstructed. In: Jacko, J.A. (Ed.), Human–Computer Interaction International 2009, Part I. Lecture Notes in Computer Science 5610, pp. 30–39.

Charlton, B., Hood, J., Sall, E., Schwartz, M., 2011. Bicycle route choice data collection using GPS-enabled smartphones. Transportation Research Board 90th Annual Meeting, Washington DC, 23–27 Jan 2011 (10 pp.).

Coleman, D.J., Geogiadou, Y., Labonte, J., 2009. Volunteered geographic information: the nature and motivation of produsers. Int. J. Spat. Data Infrastruct. Res 4 (http://ijsdir.jrc.ec.europa.eu/index.php/ijsdir/article/view/140/198, accessed 25th June 2015).

CycleStreets, 2014. Cycle Hackney app created by CycleStreets. http://www.cyclestreets.net/blog/2014/07/06/cycle-hackney-app/ (accessed 26th November 2014).

Dasgupta, A., Kosara, R., 2011. Adaptive privacy-preserving visualization using parallel coordinates. IEEE Trans. Vis. Comput. Graph. 17 (12), 2241–2248.

Endoangela, 2012. Popular Endomondo sports tracker mobile app hits 10 million user milestone and 320 million miles logged. http://blog.endomondo.com/2012/06/26/popular-endomondo-sports-tracker-mobile-app-hits-10-million-user-milestone-and-320-million-miles-logged/ (accessed 7th August 2014).

Endomondo, 2014. What we do. https://www.endomondo.com/about (accessed 7th August 2014).

Estes, A.C., 2014. Why a fitness-tracking app is selling its data to city planners. GIZMODO. http://gizmodo.com/why-a-fitness-tracking-app-is-selling-its-data-to-city-1572964149 (accessed 2nd September 2014).

Fung, B.C.M., Wang, K., Chen, R., Yu, P.S., 2010. Privacy-preserving data publishing: a survey of recent developments. ACM Comput. Surv. 42 (4). http://dx.doi.org/10.1145/1749603.1749605 (Article 14).

Garmin, 2013. Garmin connect is heating up with heat maps. http://garmin.blogs.com/my_weblog/2013/03/garmin-connect-is-heating-up.html#.U6FbGbFABa4 (accessed 16th December 2014).

Gedik, B., Liu, L., 2004. A customizable k-anonymity model for protecting location privacy. Technical Report GIT-CERCS-04-15. Georgia Institute of Technology (12 pp., https://smartech.gatech.edu/xmlui/bitstream/handle/1853/100/git-cercs-04-15.pdf, accessed 14th October 2014).

Goodchild, M.F., 2007. Citizens as sensors: the world of volunteered geography. GeoJournal 69, 211–221. http://dx.doi.org/10.1007/s10708-007-9111-y.

Goodchild, M.F., 2013. The quality of big (geo)data. Dialogues Hum. Geogr. 3 (3), 280–284. http://dx.doi.org/10.1177/2043820613513392.

Griffin, G., Nordback, K., Götschi, T., Stolz, E., Kothuri, S., 2014. Monitoring bicyclist and pedestrian travel and behavior — current research and practice. Transportation Research Circular E-C183 (32 pp., http://onlinepubs.trb.org/onlinepubs/circulars/ec183.pdf, accessed 26th November 2014).

Gruteser, M., Liu, X., 2004. Protecting privacy in continuous location-tracking applications. IEEE Secur. Priv. 2 (2), 28–34.

Harford, T., 2014. Big data: are we making a big mistake? The Financial Times. http://www.ft.com/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html#axzz3GqbLWgPw (accessed 22nd October 2014).

Hellman, T., 2013. Polkupyöräliskennat Helsingissä 2013 (bicycle counting in Helsinki 2013, in Finnish). Memorandum 28th October 2013. City of Helsinki. City Planning Department, Transportation and Traffic Planning Division (http://www.hel.fi/hel2/ksv/Aineistot/Liikennesuunnittelu/Liikennetutkimus/pyoralaskennat_2013.pdf, accessed 2nd September 2014).

Hellman, T., 2014. Personal Communication, June 13th 2014.

Hood, J., Sall, E., Charlton, B., 2011. A GPS-based bicycle route choice model for San Francisco, California. Transp. Lett. 3, 63–75.

HS, 2012. Neljän miljoonan euron Auroransilta avautui vihdoin ulkoilijoille (four million euro Aurora's bridge finally open for citizens, in Finnish). Helsingin Sanomat, 24th November 2012. http://www.hs.fi/kaupunki/a1305621575898 (accessed 12th November 2014).

Hudson, J.G., Duthie, J.C., Rathod, Y.K., Larsen, K.A., Meyer, J.L., 2012. Using smartphones to collect bicycle travel data in Texas. Final Report of the UTCM Project #11-35-69 (http://utcm.tamu.edu/publications/final_reports/Hudson_11-35-69.pdf, accessed 26th November 2014).
Junge, K., 1994. Diversity of ideas about diversity measurement. Scand. J. Psychol. 35 (1), 16–26. http://dx.doi.org/10.1111/j.1467-9450.1994.tb00929.x.
Kauppalehti, 2013. Suomalainen liikuntasovellus: 100 yritystä 12 maassa (Finnish sports application: 100 companies in 12 countries, in Finnish). Kauppalehti 17th March 2013. http://www.kauppalehti.fi/omayritys/suomalainen+liikuntasovellus+100+yritysta+12+maassa/201303381093 (accessed 16th December 2014).
Keim, D.A., 2002. Information visualization and visual data mining. IEEE Trans. Vis. Comput. Graph. 7 (1), 100–107.
Kisilevich, S., Mansmann, F., Bak, P., Keim, D., 2010. Where would you go on your next vacation? A framework for visual exploration of attractive places. Second International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOPROCESSING), 2010, pp. 21–26. http://dx.doi.org/10.1109/GEOProcessing.2010.11.
Kitchin, R., 2014. Big data, new epistemologies and paradigm shifts. Big Data Soc. 1 (1). http://dx.doi.org/10.1177/2053951714528481.
Kounadi, O., Leitner, M., 2015. Defining a threshold value for maximum spatial information loss of masked geo-data. ISPRS Int. J. Geo-Inf. 4 (2), 572–590. http://dx.doi.org/10.3390/ijgi4020572.
Krisp, J.M., Peters, S., 2011. Directed kernel density estimation (DKDE) for time series visualization. Ann. GIS 17 (3), 155–162. http://dx.doi.org/10.1080/19475683.2011.602218.
Krumm, J., 2007. Inference attacks on location tracks. In: LaMarca, A., Langheinrich, M., Truong, K.N. (Eds.), Proceedings of the 5th International Conference on Pervasive Computing (Pervasive'07)Lecture Notes in Computer Science 4480. Springer, Berlin, pp. 127–143. http://dx.doi.org/10.1007/978-3-540-72037-9_8.
Kurata, Y., 2012. Potential-of-interest maps for mobile tourist information services. In: Fuchs, M., Ricci, F., Cantoni, L. (Eds.), Information and Communication Technologies in Tourism 2012. Springer, Vienna, pp. 239–248. http://dx.doi.org/10.1007/978-3-7091-1142-0_21.
Kwan, M.P., Casas, I., Schmitz, B., 2004. Protection of geoprivacy and accuracy of spatial information: how effective are geographical masks? Cartographica 39 (2), 15–28. http://dx.doi.org/10.3138/X204-4223-57MK-8273.
Lane, N.D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., Campbell, A.T., 2010. A Survey of mobile phone sensing. IEEE Commun. Mag. 140–150 September.
Leitner, M., Curtis, A., 2006. A first step towards a framework for presenting the location of confidential point data on maps—results of an empirical perceptual study. Int. J. Geogr. Inf. Sci. 20 (7), 813–822. http://dx.doi.org/10.1080/13658810600711261.
Lin, D., 2012. How is Nike+ Heat Map calculated? http://howtonike.blogspot.fi/2012/06/how-is-nike-heat-map-calculated.html (accessed 16th December 2014)
Lindsey, G., Nordback, K., Figliozzi, M.A., 2014. Institutionalizing bicycle and pedestrian monitoring programs in three states: progress and challenges. 93rd Annual Meeting of the Transportation Research Board, Washington, DC, January 12–16.
Mach, P., 2014. What do 220,000,000,000 GPS data points look like? http://engineering.strava.com/global-heatmap/ (accessed 20th August 2014)
Masoner, R., 2014. Santa Cruz County: log your bike rides for transportation planning. http://www.cyclelicio.us/2014/santa-cruz-county-log-your-bike-rides-for-transportation-planning/ (accessed 16th December 2014).
Maus, J., 2014a. A closer look at Strava's 'heat map' for the Portland region. BikePortland.org (Blog Post) (http://bikeportland.org/2014/04/30/a-closer-look-at-stravas-heat-map-for-the-portland-region-105280, accessed 29th July 2014).
Maus, J., 2014b. ODOT embarks on "big data" project with purchase of Strava dataset. BikePortland.org (Blog Post) (http://bikeportland.org/2014/05/01/odot-embarks-on-big-data-project-with-purchase-of-strava-dataset-105375, accessed 29th July 2014).
McDonald, D.G., Dimmick, J., 2003. The conceptualization and measurement of diversity. Commun. Res. 30 (1), 60–79. http://dx.doi.org/10.1177/0093650202239026.
Mirković, M., Aulibrk, D., Milisavljević, S., Crnojević, V., 2010. Detecting attractive locations using publicly available user-generated video content — central Serbia case study. Proceedings of 18th Telecommunications Forum (TELFOR'10), Serbia, Belgrade, November 23–25, 2010, pp. 1089–1092.
Monreale, A., Andrienko, G., Andrienko, N., Giannotti, F., Pedreschi, D., Rinzivillo, S., Wrobel, S., 2010. Movement data anonymity through generalization. Trans. Data Privacy 3, 91–121 (http://www.tdp.cat/issues/tdp.a045a10.pdf, accessed 15th December 2014).
Montjoye, Y.-A., Hidalgo, C.A., Verleysen, M., Blondel, V.D., 2013. Unique in the crowd: the privacy bounds of human mobility. Sci. Rep. 3. http://dx.doi.org/10.1038/srep01376.
Moon, J.-Y., Jung, H.-J., Hee Moon, M., Chul Chung, B., Ho Choi, M., 2009. Heat-map visualization of gas chromatography-mass spectrometry based quantitative signatures on steroid metabolism. J. Am. Soc. Mass Spectrom. 20 (9), 1626–1637. http://dx.doi.org/10.1016/j.jasms.2009.04.020.
Morstatter, F., Kumar, S., Liu, H., Maciejewski, R., 2013. Understanding Twitter data with TweetXplorer. Proceedings of the 2013 ACM SIG KDD International Conference on Knowledge Discovery and Data Mining (KDD'13), ACM, August 11–14, 2013, Chicago, IL, USA, pp. 1482–1485 http://dx.doi.org/10.1145/2487575.2487703.
Murad, A., Hilton, B., Horan, T., Tangenberg, J., 2014. Protecting patient geo-privacy via a triangular displacement geo-masking method. In: Kessler, C., McKenzie, G.D., Kulik, L. (Eds.), Proceedings of the 1st ACM SIGSPATIAL International Workshop on Privacy in Geographic Information Collection and Analysis (GeoPrivacy'14), ACM, November 4–7, 2014, Dallas/Fort Worth, TX, USA http://dx.doi.org/10.1145/2675682.2676399.
NBPD, 2014. National bicycle and pedestrian documentation project. Alta Planning & Design. Institute of Transportation Engineers (ITE) Pedestrian and Bicycle Council (http://bikepeddocumentation.org/, accessed 12th November 2014).
Nielsen, J., 2006. The 90-9-1 Rule for Participation Inequality in Social Media and Online Communities. Nielsen Normal Group. (http://www.nngroup.com/articles/participation-inequality/, accessed 10th October 2014).
NLS, 2014. The Topographic database. National Land Survey of Finland. http://www.maanmittauslaitos.fi/en/digituotteet/topographic-database (accessed 14th October 2014).
Pensa, R.G., Monreale, A., Monreale, A., Pinelli, F., Pedreschi, D., 2008. Pattern-preserving k-anonymization of sequences and its application to mobility data mining. In: Bettini, C., Jajodia, S., Samarati, P., Wang, X.S. (Eds.), Proceedings of the 1st International Workshop on Privacy in Location-Based Applications (PiLBA '08), Malaga, Spain, October 9, 2008 (http://ceur-ws.org/Vol-397/paper4.pdf, accessed 16th December 2014).
Poikola, A., 2014. Sujuvuusnavigaattorin pilotointi (piloting of the smoothness navigator). Open knowledge Finland. http://bit.ly/sujuvuusnavi_kalvot (accessed 12th November 2014).
Rantala, T., Luukkonen, T., 2014. Bicycle and Pedestrian Traffic Monitoring — Guide to Creating an Indicator Toolbox. Finnish Transport Agency, Planning Department, Helsinki (37 pages, 2 appendices. http://www2.liikennevirasto.fi/julkaisut/pdf8/lts_2014-15_kavelyn_pyorailyn_web.pdf, accessed 30th September 2014).
Russom, P., 2011. Big data analytics. TDWI Best Practices Report, Fourth Quarter 2011 (35 pp., http://tdwi.org/research/2011/12/sas_best-practices-report-q4-big-data-analytics.aspx?tc=page0, accessed 14th October 2014).
Samarati, P., Sweeney, L., 1998. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Proceedings of the IEEE Symposium on Research in Security and Privacy (S&P). May 1998, Oakland, CA, pp. 384–393 (http://dataprivacylab.org/dataprivacy/projects/kanonymity/paper3.pdf, accessed 16th December 2014).
Seidl, D., Jankowski, P., Tsou, M.-H., 2015. Spatial obfuscation of GPS travel data. CSU Geospatial Review vol. 13 (http://csugis.sfsu.edu/CSU_Geospatial_Review/2015.pdf (accessed 25th June 2015)).
SFMTA, 2013. CycleTracks for iPhone and Android. San Francisco County Transport Authority. http://www.sfcta.org/modeling-and-travel-forecasting/cycletracks-iphone-and-android (accessed 2nd September 2014).
Shearmur, R., 2015. Dazzled by data: big data, the census and urban geography. Urban Geogr. http://dx.doi.org/10.1080/02723638.2015.1050922.
Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman & Hall, London.
Slocum, T.A., McMaster, R.B., Kessler, F.C., Howard, H.H., 2009. Thematic Cartography and Geovisualisation. Prentice Hall, New Jersey, NJ.
Špakov, O., Miniotas, D., 2007. Visualization of eye gaze data using heat maps. Electron. Electr. Eng. 2, 55–58.
ST, 2014. Introducing Sports Tracker. Sports Tracking Technologies Ltd. (http://www.sports-tracker.com/blog/about/, accessed 20th August 2014).
Strava, 2014. About us. Strava Inc. (http://www.strava.com/about, accessed 7th August 2014).
Terrovitis, M., Mamoulis, N., 2008. Privacy preservation in the publication of trajectories. Proceedings of the 9th International Conference on Mobile Data Management (MDM '08), 27–30 April 2008, pp. 65–72. http://dx.doi.org/10.1109/MDM.2008.29.
Trame, J., Keßler, C., 2011. Exploring the lineage of volunteered geographic information with heat maps. Abstracts of GeoViz 2011, Hamburg, Germany, March 10–11, 2011. http://www.geomatik-hamburg.de/geoviz11/abstracts/28_TrameKessler_Abstract_GeoViz2011.pdf (accessed 12th November 2014).
Usborne, S., 2013. Can cycling app Strava change the way we ride? The Independent, 4th July 2013. http://www.independent.co.uk/life-style/gadgets-and-tech/features/can-cycling-app-strava-change-the-way-we-ride-8685996.html (accessed 7th August 2014).
Vanderbilt, T., 2013. How Strava is changing the way we ride. Outside Magazine, January 2013. http://www.outsideonline.com/fitness/biking/How-Strava-Is-Changing-the-Way-We-Ride.html (accessed 7th August 2014).
Verykios, V.S., Damiani, M.L., Gkoulalas-Divanis, A., 2008. Privacy and security in spatio-temporal data and trajectories. In: Giannotti, F., Pedreschi, D. (Eds.), Mobility, Data Mining and Privacy, pp. 213–240.
Willems, C.M.E., 2011. Visualization of vessel traffic (PhD Thesis) Eindhoven University of Technology.
Young, C., Martin, D., Skinner, C., 2009. Geographically intelligent disclosure control for flexible aggregation of census data. Int. J. Geogr. Inf. Sci. 23 (4), 457–482 (http://dx.doi.org/10.1080/13658810801949835).
Zahradnik, F., 2014. Strava Cycling app with unique social and record-keeping features. About.com. http://gps.about.com/od/sportsandfitness/fr/Strava-Cycling-App-Review.htm (accessed 7th August 2014).
Zeile, P., Memmel, M., Exner, J.-P., 2012. A new urban sensing and monitoring approach: tagging the city with the RADAR SENSING app. Proceedings REAL CORP 2012 Tagungsband 14–16 May 2012, Schwechat (http://programm.corp.at/cdrom2012/papers2012/CORP2012_104.pdf (26th September 2014)).