Applied Geography 88 (2017) 144–150

Contents lists available at ScienceDirect

Applied Geography

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

Geographic age and gender representation in volunteered cycling safety data: A case study of BikeMaps.org

Colin Jay Ferster[a,*], Trisalyn Nelson[b], Meghan Winters[c], Karen Laberee[a]

[a] Department of Geography, University of Victoria, David Turpin Building A247, PO BOX 3060, STN CSC, Victoria, BC V8W 3R4, Canada
[b] School of Geographical Sciences & Urban Planning, Arizona State University, Mailcode: 5302, 975 S Myrtle Ave, Tempe, AZ, 85281, USA
[c] Faculty of Health Sciences, Simon Fraser University, BLU 11522, 8888 University Drive, Burnaby, BC V5A 1S6, Canada

ARTICLE INFO

Keywords:
Citizen science
Participation
Data informatics
Safety
Active transportation

ABSTRACT

There has been growing interest in using volunteered geographic information (VGI) for transportation planning, such as route data from fitness tracking applications and route mapping smartphone applications, as a compliment to traditional data collection approaches. In particular, cycling safety data from traditional sources are limited since bike crashes are under-reported and there are no central mechanisms for recording near misses. BikeMaps.org is a globally available website for cycling safety VGI, with a focus on spatial analyses of previously unrecorded near misses and collisions. The goal of this paper is to understand how age and gender are related to the use of BikeMaps.org compared to broader ridership and the geographic distribution of incidents for the Capital Regional District (CRD), British Columbia, Canada. Males aged 24–35 reported more incidents than other cohorts, which had similar relative proportions to the regional cycling population in origin-destination survey data. In general, there were higher levels of interaction with the website by younger people (i.e. submitting incidents vs. viewing incidents). Females and people under 35 years of age reported more incidents in central urban areas. People over 35 years of age reported incidents that were more spatially dispersed and covered a broader extent. These findings are indicative of both cycling behaviour and the use of technology. A target group for growth in cycling, females may benefit from improved cycling facilities in city centers. Older people may ride in cities more often if better facilities are provided. Understanding gaps in representation can help target more deliberate campaigns to complete data and inform the effective use of complementary data sources by planners.

1. Introduction

There has been growing interest in volunteered geographic information (VGI) as a new source of cycling data. For example, smartphones applications have been used to record cycling routes (Hood, Sall, & Charlton, 2011; Hudson, Duthie, Rathod, Larsen, & Meyer, 2012), cycling routes have been obtained from fitness applications (Griffin & Jiao, 2015; Jestico, Nelson, & Winters, 2016), and recently, information about crashes and near-misses has been collected over the Internet (Aldred, 2016; Nelson, Denouden, Jestico, Laberee, & Winters, 2015). These data can be collected rapidly, cover broad spatial extents, and provide information that would be expensive and difficult to collect using traditional approaches (Romanillos, Zaltz Austwick, Ettema, & De Kruijf, 2015). The Internet is a medium capable of engaging large audiences using VGI (Goodchild, 2007). VGI typically does not require official credentials to participate and can be generated by people who are geographically close to a phenomenon and are directly impacted and invested in management outcomes (Feick & Roche, 2013). Further, VGI can be more rapidly responsive and incorporate greater local knowledge than government monitoring (Johnson & Sieber, 2013). When transportation VGI datasets are used in a planning context, they give visibility to where people are riding, where crashes happen, and provide authority to arguments, which in turn may be used to allocate public resources to build bike facilities and reduce exposure to risk (Le Dantec, Asad, Misra, & Watkins, 2015).

Well-designed transportation systems can minimize real and perceived exposure to risk and make cities that are attractive and safe for cyclists (Buehler & Pucher, 2012). However, cycling crashes are under-reported in hospital, police, and automobile insurance records, in particular, when they do not involve automobiles or serious injuries (Cryer et al., 2001; Winters & Branion-Calles, 2017). Further, there are no centralized mechanisms for reporting, collecting, and analyzing near miss incidents. Recently, tools have been developed to collect VGI cycling safety data, including previously unreported crashes that do not involve motor vehicles nor insurance claims, minor crashes that did not involve injury, or near misses (Nelson et al., 2015). Near misses provide

* Corresponding author.
E-mail addresses: cferster@uvic.ca (C.J. Ferster), Trisalyn.Nelson@asu.edu (T. Nelson), mwinters@sfu.ca (M. Winters), klaberee@uvic.ca (K. Laberee).

http://dx.doi.org/10.1016/j.apgeog.2017.09.007
Received 23 January 2017; Received in revised form 7 July 2017; Accepted 20 September 2017

an important precautionary signal and are related to public perceptions that cycling is dangerous, thus deterring uptake (Aldred, 2016). Proponents of active transportation strategies emphasize the importance of providing services to diverse demographics of people (Aldred, Woodcock, & Goodman, 2015), and in particular, the cohorts who will benefit the most from the health and access gains (Lee, Sener, & Jones-Meyer, 2016). Since VGI is collected voluntarily and the people who chose to participate may not represent broader populations (Romanillos et al., 2015). Equitable distribution of public resources may require consideration of interests that are not represented within the data (Le Dantec et al., 2015). Research is needed to understand the impacts of demographic representation on volunteered cycling data so they can be used effectively to make equitable planning decisions.

BikeMaps.org is a global tool to collect anonymous information from the public about collisions, near misses, hazards, and thefts, providing an opportunity to evaluate representation within cycling safety VGI (Nelson et al., 2015). At the time of writing, BikeMaps.org documented more than 4000 reports of collisions (20%), near misses (38%), hazards (32%), and thefts (10%). Users are asked to provide demographic information for collisions, near-misses, and hazards (globally, 61% of respondents provided complete information for near-misses and collisions).

In this paper, we evaluated who is using BikeMaps.org and who is missing. First, we compared by age and gender the rates of ridership, web visits to BikeMaps.org to view data, and incidents reported to BikeMaps.org to understand the demographic representation of the data. Second, we evaluated whether age and gender were related to the geographic distribution of VGI on BikeMaps.org. This paper provides insight into both the processes of participation in VGI and how VGI may be applied for transportation planning.

2. Methods

2.1. VGI data

The Capital Regional District (CRD), which includes Victoria where BikeMaps.org was first launched, was chosen as the study extent to provide a defined boundary to compare against demographic information. All BikeMaps.org reports of collisions and near misses were extracted for the CRD (reporting started in October 2014 and the incidents were downloaded 17 July 2016), totaling 610 records: 29% collisions, 71% near misses); 62% had complete demographics and were selected for further analysis (117 collisions, 260 near misses for a total of 377 observations). Web traffic for BikeMaps.org was obtained using Google Analytics (Google Inc., 2016). Google Analytics makes classifications of gender (male or female) and age in six classes based on user-supplied information in social media profiles and patterns in web browsing habits. The counts of unique visitors were queried for BikeMaps.org from 1 Oct 2014 to 17 July 2016 including the age category, gender, and city. Caution is warranted in interpreting Google Analytics results as validation data are not available. No less, this tool was still considered useful to provide a general indication of patterns and another corroborating line of evidence that has increasingly been used in scientific study (Burgess et al., 2016; Plaza, 2009).

2.2. Ridership data

Information about ridership in the region was obtained from an origin-destination (O-D) survey conducted by Malatest Program Evaluation and Market Research for the CRD in 2011 (R.A. Malatest & Associates, 2012). This was a 24-h recall survey for weekdays (Monday through Friday) excluding holidays collected over the period of 5 October 2011 to 11 December 2011. Respondents were asked to provide household demographic information and details for all trips in the previous 24 h for all members of the household. Households were sampled based on random telephone dialing, which would capture all households in the region in the sample frame, aside from those without land-line telephones and institutionalized housing. Responses were collected from people 11 years of age and older, only trips for people over 18 were used in this study. There was a 36% response rate. Expansion factors were calculated by the survey company based on household counts, dwelling types, household size, population, and age and gender relative to 2011 census data. After data cleaning by the survey company, the survey data included 6172 households, 13,986 persons, and 40,973 trips, representing 3% of all households and a target margin of error of 5%. Cycling trip data were released by the CRD aggregated by gender and age category with, origins and destinations within 19 sub-municipal regions, which approximately correspond with Canadian Census Subdivisions. To add context about population changes in the CRD between the collection of the O-D data and the more recent BikeMaps.org data, we also discussed demographic changes between the 2011 National Household survey and the recently released 2016 Canada Census (Statistics Canada, 2017).

2.3. Statistical analysis

First, to describe the demographics of BikeMaps.org, incidents were summarized by age class (18–24; 25 to 34; 35 to 44; 45 to 54; 55 to 64; and 65+) and gender (male or female). Percentages were calculated for each age class and gender for the total count for trips in the O-D data, unique web visitors to BikeMaps.org, and incidents submitted to BikeMaps.org. To test for differences between age and gender categories between the CRD O-D cycling trips, web visits, and BikeMaps.org incidents by age and gender categories, we used a three Chi-squared tests of multiple proportions: two ordered Chi-squared tests comparing the distribution of age classes across the data sources for males and females with scores assigned as the mid-point of the age categories; and one standard Chi-squared test comparing gender distributions for all age classes combined across the data sources. The ordered Chi-squared tests were completed using the R package "coin" version 1.2 (Hothorn, Hornik, van de Wiel, & Zeileis, 2008). We used 35, the approximate median age of incidents submitted to BikeMaps.org to define two age groups. This corresponded with a natural break in the distribution of reported ages for BikeMaps.org incidents and maintained proportional balance. To describe the geographic distribution of incidents, the mean nearest neighbor distance (MNN) was calculated as the arithmetic mean distance to the nearest point in the same age group and gender. A higher MNN indicates that incidents are more dispersed and a lower MNN means the incidents are closer together. To describe the location of incidents, the proportion of incidents was calculated within the boundaries of the City of Victoria for each age group and gender. Maps were created of BikeMaps.org incidents using the quartic kernel density with a 1 km search distance. The MNN and city center metrics were calculated using R version 3.3.0 (R Core Team., 2015) and the spatstat package (Baddeley & Turner, 2005). The density maps were created using Quantum GIS (QGIS) version 2.14.3, and the QGIS Heatmap Plugin (Quantum GIS Development Team, 2016).

3. Results

There were statistically significant differences in age for males (Chi-squared $p = 0.005$ for males and $p = 0.751$ for females) but not for gender (Chi-squared $p = 0.569$) in the distributions of bike trips from O-D data, BikeMaps.org web views, and BikeMaps.org data submissions (Fig. 1). For ridership in the O-D survey, the majority of trips were made by males (68%), and people over 35 years of age (70%). Overall, the most frequent age category for O-D trips was 35–44 years of age. Females had higher ridership in younger age classes, while males had higher ridership in the older age classes. The majority of web visits to BikeMaps.org were by males (62%) and people over 35 years of age (66%). For web visits, the most common age category was 25–34 years of age overall, 45–54 years of age for males, and 25–34 years of age for