Statistics without Numbers: Using Data Visualization to Quantify Trends for Cycling Safety

Jacob Quartuccio[1*] , Simone Erchov[1], Christian Gonzalez[1], David Cades[2]

1. George Mason University
2. Exponent Failure Analysis Associates
*Contact author: jquartuc@gmu.edu

Big data sets can be cumbersome and difficult to understand. User-centered and interactive graphical displays help communicate messages from large and complex data as well as provide a new method to identify data trends outside of tabular or statistical analysis. Not all statistics need to be summarized numerically, sometimes a visualization can convey uncertainty with parameters. Researchers can use data visuals to not only develop questions but also answer them through visual exploration that previously proved difficult. This approach can be especially relevant to the field of surface transportation research where complex plots can incorporate both temporal and geospatial data in an easy-to-digest format. These plots may ease communication of potential issues and solutions between various stakeholders such as policy makers, liability companies, and people using transplantation systems. As a proof of concept, the visualization presented here demonstrates how data showing bike-sharing in 2013 in Chicago and past bicycle collision incidents can meaningfully merge to produce graphical displays that readily identify and communicate potential infrastructure problems for safety. The city of Chicago seemed especially ripe for a safety analysis since the city has actively targeted bicycle safety challenges as the number of cyclists on the road rises [1]. Bicycling routes in the figure were generated via a Bayesian model by estimating the posterior probability that a cyclist traveled between two stations. The model began with an uninformative prior, assuming that every end station had an equal probability of arrival from any other station, and updated with bikeshare trip data via a Bernoulli function in the **LearnBayes** package 2.12. A map of Chicago was acquired from Google to provide a base map for the plots with the **ggmap** v2.3 package. Accident data was overlaid via a heatmap to show areas that have been traditionally prone to cycling accidents; areas in red represent the highest density of collisions, yellow represents a reduction, and cream portions of the image represent lowest levels of collisions. The visualization not only conveys current Bikeshare traffic patterns, but also encourages stakeholders' further exploration of areas that may need further infrastructure development.

References

[1] Chicago Department of Transportation (2013). Chicago Streets for Cycling Plan 2020. Policy Report. Retrieved from http://www.cityofchicago.org/content/dam/city/depts/cdot/bike/general/ChicagoStreetsforCycling2020.pdf.