The Philly Crash Atlas: A Data-driven Visualization of Traffic Collisions in Philadelphia, Pennsylvania

Marissa Defratti, Aaron Dennis, Bill Limpisathian, Maggie Norton, Tyler Regino, Kim Schmid, Derek Shaffer, Pat Stephens

ABSTRACT – The Pennsylvania Department of Transportation records all reported vehicle collisions that occur within the city of Philadelphia. Each collision contains a spatial reference point and many attributes referring to the cause, type, severity, time, etc. of the collision. The Philly Crash Atlas turns these many data points into a useful visualization for those who commute out, to, or within the city of Philadelphia. The application also has potential to be used by infrastructure planners to find dangerous routes, assess the problems of that specific route, and allocate resources accordingly. In order to produce this application, a custom web map was designed to accentuate the data points along specific routes, user enabled filters were coded to easily sort data points, and a data-driven histogram for summarizing the data was made. This application has the potential to be used with collision data from other areas in the country as well as other data types that require similar visualization.

Index Terms – geovisualization, d3, data-driven visualization, histogram, infrastructure, vehicle, PennDOT.

---

1 INTRODUCTION

According to the Federal Highway Administration (FHA), the volume of traffic density in terms of freight movement alone is expected to double from 2006 to 2020 (Tien-Lu, 2006). Traffic has the ability to affect anyone, whether you are an employee on their morning commute, a child on their way to school, or a city planner evaluating the safety and feasibility of road construction in a large city. Typically, datasets containing traffic incident data are extensive, but meaningful patterns go unnoticed. Most states provide incident data online for the public, but often in the form of long and static text reports, with little effort put towards pattern extraction or knowledge creation (Wongsuphasawat, 2009). In order present these patterns, The Philly Crash Map provides an interactive tool that allows users to visualize traffic collisions in Philadelphia. The data used in our application was obtained from the Pennsylvania Department of Transportation (PennDOT) and geocoded by Philadelphia GIS consulting agency Azavea. The data set consists of 53,260 accidents from 2008 to 2013. Each collision point contains attributes referring to specific details of the collision such as time, location, vehicle type, cause of collision, severity of collision, and other relevant collision details.

2 PROBLEM STATEMENT

Our motivation for this project stemmed from the need for a visual aid for traffic incidents. This application will help the many who commute to, from, or through Philadelphia to identify dangerous routes. This application will also be useful for infrastructure planners who need to find problem routes and allocate resources.

3 OUTCOMES

The resulting application provides a map and data visualization for users to explore the data. There are 3 subset tools within the application that can be used to explore the data in greater detail. In a scenario based exercise, the user is expected to enter a starting point and destination. All collisions along that route will then be returned and represented in the map and the histogram. The data can then be filtered by 3 main attributes: time of day, type of vehicle, and cause of collision. After the user has displayed and filtered the data along their chosen route, they can explore the route in greater detail through the histogram, map, and Crash Summary. See Figure 3 for a complete overview of the application interface.

3.1 INTERFACE AND SYMBOLOGY

The goal of the interface and symbology is to accentuate the visualized data while keeping the application intuitive and uncluttered. Accordingly, the sign-vehicles are highly abstract with a low level of iconicity. The visualized data is represented using geometric circles, with significance being represented using logically-corresponding hues. This symbology choice additionally helps to ensure legibility and consistency of the data both on the map and in the associated histogram. Careful considerations were made to ensure that proper contrast and hierarchical-relations between map elements were maintained.

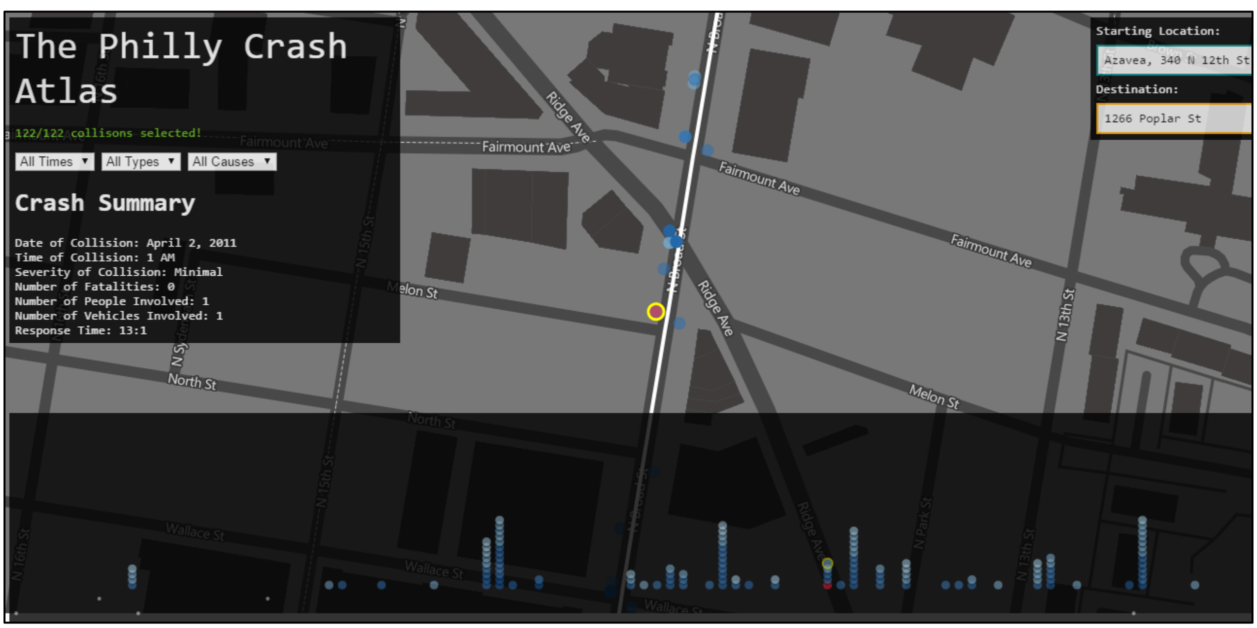

Figure 3: The application interface with a route selected and details on a specific collision point.

The application uses an extensively-customized dark grey base map that provides sufficient contrast with the range of hues and saturations being used to represent the visualized data. Dynamic labeling symbolization was implemented with zoom level-dependent city, neighborhood, and street labeling. Hue and saturation choices for the incident data were made with similar considerations, with each saturation level having adequate distinguishability from other levels, while also collectively maintaining the essential contrast with the background elements on the map. A 10-class blue color-ramp was developed to represent the severity level of the accident with an additional red hue to represent fatalities. Furthermore, the use of a white route line helps to provide visual contrast and breakage between the foreground symbols and background elements. All incident symbols were assigned a 15 percent transparency value to maintain the visibility of stacked icons. Additional distinct but complementary hues were used to represent the start and end points and the implemented data brushing.

3.2 FILTERS

The attributes that were selected for filtering contained the most thoroughly reported information with the least amount of missing data. Through data exploration and research, it was found that representing accident severity was something that should inherently symbolized, rather than exist as an attribute that could be filtered. After establishing that accident severity should be the point symbol on the map and histogram, the other selected attributes fell into three categories: time (day of month, day of week, and time of day), means of transportation involved, and cause of collision. These attributes were compiled into their respective groups, and drop-down lists serve as the medium with which the map-user filters data.

Figure 3.2: Number of collisions on selected route shown after filtering and details on demand of a single selected collision.

3.3 HISTOGRAM

Histograms can assist the user in constructing hypotheses and noticing patterns in large datasets (Musdholifah, A. and Hadhim S., 2013). Creating a histogram required translating the geojson file returned from the route into information that could be read in a histogram. The information can then be linked to the map by hovering over one of the dots within the histogram. Linking is an important visualization tool and helps to highlight the data in multiple components (Adrienko & Adrienko, 2009). Each dot represents a collision and is colored based on the severity of that collision. See Figure 3.3 for a highlighted image of the histogram.

Figure 3.3: The histogram showing the number of collisions along a certain route and segment of that route.

3.4 DETAILS ON DEMAND

After a user has filtered their data and found a collision of interest, the user can click on the dot to print summary information about that individual collision. Summary information includes attributes that cannot easily be displayed through icons or colors such as: date of collision, exact time of collision, response time, numbers of vehicles involved, etc. The Crash Summary provides details that cannot be easily be filtered, displayed on the map, or displayed in the histogram. See Figure 3.2 to see how the details are printed.

4 CONCLUSIONS

Although the application does provide insight into the spatial component and patterns within the data, there is still preprocessing that could enhance the data. For example, Azavea, the company who originally geocoded the PennDOT data, found that collision counts are high over roads travelled most frequently. This will make roads that are more travelled appear to be more dangerous. A more robust method would be to assess the collisions per vehicle on a given route segment, rather than the total amount of collisions on a given route segment.

This application was designed to provide a useful visualization of vehicle collision data acquired from PennDOT over the course of five years in the city of Philadelphia. The application was designed for two types of end users, commuters within the city of Philadelphia and planners responsible for the safety and planning of city infrastructure. The application provides commuters and planners with a visualization of the amount of collisions on different routes. It then allows them to filter the data to most closely match commuting patterns and access details on demand for specific collisions. The application visualizes patterns and displays information that would not otherwise be available or recognizable to data viewers.

5 ACKNOWLEDGEMENTS

We would like to thank our professor Dr. Alan MacEachren and our teaching assistant Alexander (Sasha) Savelyev for their assistance in the design and implementation of this application.

6 AUTHORS' CONTRIBUTIONS

MD compiled and wrote the poster, aided in the presentation, and aided in the d3 coding. AD coded the majority of the application interface, geojson retrieval, and found the data. TR did background research on the data attributes and coded the filters. DS did background research on filtering. BL designed the base map and color scheme for the histogram, map, and interface. MN designed the presentation, compiled the video, and aided in the d3 coding. KS conducted background research, managed project information, and aided in writing this paper. PS coded the majority of the d3 histogram and aided in writing and editing this paper.

7 REFERENCES

Musdholifah, A. and Hadhim S., 2013: Mining Spatio-Temporal Data of Fatal Accidents. International Journal of Computer Applications, Vol 63, No. 8.

Wongsuphasawat, K. et al., 2009. Visual Analytics for Transportation Incident Data Sets. IEEE International Conference on Information Reuse & Integration.

Andrienko, G. L. & N. V. Andrienko. 1999. 'Interactive maps for visual data exploration', International Journal of Geographic Information Science, 13(4), pp. 355-374.

Chen, J., MacEachren, A. M., & Guo, D. (2008). Supporting the process of exploring and interpreting space–time multivariate patterns: The visual inquiry toolkit. Cartography and geographic information science, 35(1), 33-50.

Rosson, M.B. and Carroll, J.M. 2002: Scenario-based design. In Jacko, J. and Sears, A., editors, Chapter 53 in The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications: Lawrence Erlbaum Associates, 1032