Ronald Balutiu, Jackson Quinn
CMSC 23900
Project One Writeup

Context and Purpose

The purpose of this visualization was to better understand the behavior of drivers in the city of Chicago as a function of day of the week and as a function of the month of the year. Essentially, we sought to answer whether drivers were more likely to engage in risky driving behaviors (specifically running red lights and speeding) on certain weekdays and in certain months. One might hypothesize that drivers are more likely run red lights on weekdays in order to avoid being late to important obligations, or be more likely to speed during the summer due to better visibility and weather conditions. This visualization allows a viewer to answer these questions and more. All of this was accomplished by examining data on the amount of red-light-camera and speed-camera violations as provided by the City of Chicago.

Pre-Processing

The data was downloaded as large CSV files from the City of Chicago data portal. The data included date, location of camera, and number of violations recorded on a given day by a given camera. Using Python and its built in CSV module, each data-row from either of the CSVs was instantiated as a member of a violation class. In each of these violation classes, methods were written that allowed the month or weekday of a violation be returned given the violation's corresponding date-time object. Once the full CSV was instantiated into individual violations, we were able to write queries that would return dictionaries with either weekday or month keys and the correct and corresponding values. The script that accomplished this can be found in /p1/data_extraction/new.py. Also in this folder is main.py, which is an earlier prototype of this script that was harder to write queries for. These Python dictionaries were converted to a JSON object, and then directly imported into the JS file. This data extraction process took nearly 80 megabytes of CSV data and synthesized it down into a 2kb JSON file of just the relevant and already-processed data.

How does it work?

For red light and speed camera violations by weekday, we decided to use a polar bar graph. This offers the ease of comparison of relative values normally found in a bar chart, while also highlighting the inherent cyclicality of this data. We decided against using zero-based bar graphs so that the length of the bars would be different enough that ratio-judgments of relative value could be more easily made. The color choices for each day were arbitrary, but help to visually delineate the graph.

For red light and speed camera violations by month, we decided to use a standard scatter-plot, with lines connecting the data-points in each series. The categorical value months were placed on the x-axis, with the number of violations on the y-axis. This visualization allows for an examination of the number of violations relative to each other, but also relative to other months. We included lines to connect the data-points of each series to help represent the presence of increasing or decreasing violations trends between the months. The color choice for each data-series was arbitrary, but were chosen to be easily distinguishable from one another.

Alternatives and Arbitrariness

This data allows for an incredible range of visualization interpretation. The coupling of a nominal and ratio value for each data point lends itself to be charted in many ways. However, we believed that given the simplicity of this data, to make a visualization too complex or noisy would detract from the simplicity rather than enhance it. For this reason, we chose to stick to very standard visualization choices. The lined scatterplot is a fundamental visualization type. The polar bar graph, while slightly less common, is easily interpreted and analyzed as well. We chose a polar bar graph over an even simpler bar or pie chart because the polar chart allowed us to better represent the implied cyclicality of the data without losing the simplicity of the representation. While there are certainly more complex ways to visualize this data, we believe that in this case the simpler the better.

There is an even balance between the visualization being data-driven and the visualization being comprised of choices. The visualization being focused on the values relative to one another is a product of the data itself. Coloring is the area in which we made the most choices. This visualization, with very few changes, would still be very readable if it were to be printed in black and white. Adding color makes parsing the visualization slightly easier, and makes the visualization itself better-looking, but is not a fundamental necessity for this visualization to work.

Who did what?

Generally, the amount of work done was very even, but there were certain divides on who spent more time doing what. Much of the project was pair-programmed, thus even when one partner was certainly leading for a given problem, the other partner was following along and providing insight. Jackson, being more familiar with Python, spent more time on the data extraction and parsing, and ensuring the parsed data worked correctly when imported as a JSON. Ronald did more of the actual visualization JavaScript and D3 coding, though design choices were always made together. This write-up was written together.