**SARC 5400 – Data Visualization Final Project**

Tyler Hutcherson

[tch6zf@virginia.edu](mailto:tch6zf@virginia.edu)

*How, When, and Where Are Car Crashes Occurring in Virginia?*

This project draws its inspiration from my 2017 DSI Capstone project, working alongside the DMV to better predict fatal car crashes by using varying statistical techniques. The data contains a record of every reported car crash to the Virginia DMV from 2011-2015. With hundreds of variables, a wide array of messy data, and other issues, we spent months as a team combing through the data and performing cleaning and manipulation. Fortunately, for this project, most of the data was aggregated in the way that I needed: at the crash level.

The question I decided to attack was a three pronged question that overlaps, and is intertwined with the others respectively. The three main questions when identifying risky regions in terms of fatal crashes are: How, When, and Where? My visualization allows the User to explore each of these questions through a few interactive ways.

First, I plotted a grid that crosses two categorical variables in the data set that capture cause and potentially location. These can be switched upon User click of the drop down boxes. The size of the circles represents the total number of crashes that occurred in the intersection of the two respective categories. Upon user click, the Virginia map to the right is filtered, displaying only the crashes falling in the selected “bin”. The size of the points on the map tool relate to the total number of cars reported with that crash. This took a lot of preprocessing to put the data into the correct format in little csv files that I could then dynamically access. Because the data is so large, having d3 reach into the main file a bunch of times proved to be a very slow endeavor. Additionally, the nested structure in d3 library was difficult to use and caused much headache (and countless hours). This work around allowed me to use jQuery to read in the files straight, and access the variable names/values. I added some aesthetic on click/hover effects to add to the user experience. Additionally, a sliding scale allows the User to grab data in a certain year. The default value was 2011 because it was the first year we had data for.

In my static visualization, I also included some annotations and a small area plot. I would have like to incorporate this area plot in the dynamic visualization, however, I did not find a way to manipulate path lengths, slopes, and directions, in a time efficient manner.

In the future, I would have liked to include an elegant().enter().remove().update().exit() routine that is faster and more appealing to the eye than having to clear the svg and reload the appropriate data completely. This approach counts the length of the data that circles will be appended to, and then adds/subtracts the correct number from the current plot, then it moves them to their nearest location, and it looks awesome. I also would have like to include more features that allow the user to play with the clicks and perhaps do subsequent clicks that append the map rather than clear it. All of these tasks proved to be too much for the .html file and given the volume of data, and bandwidth of my ability to dig into heavy jQuery/javascript, seemed out of scope, and a part of the “next” layer. I would have also liked to include a pop up histogram of the crash hour over the course of the average day for each of the “bins” selected.

I was surprised by how easily I was able to confirm several assumptions that we make about crashes. For example, upon exploration of crashes caused by a rear end, much to our expectation, these crashes involved more cars than the average case. This was seen by the size of the points on the map tool. However, the slice of data I chose to highlight in the static visualization were Fixed Object-Off Road crashes. These tend to happen all over without the typical density clustering that correlates with state population. This was intriguing to me, and it turns out many of these were caused by alcohol and speed. This enabled me to explore those crashes further by digging into the location by year, and see the influence on fatal crashes over time with the area plot.

I would like to thank the DMV, my capstone team, the DSI, Eric Field, and Naveen Iyer for their help in producing a paper for the SIEDs 2017 conference, pushing me to learn d3 and native javascript in unforeseen ways, guiding me, and most importantly helping me learn. There is much left to do, but I think this is a good first stab at “visualizing complexity”.