Analysis of United States Interstate Travel Speeds for Safety Correlation

Vijaysrinivas Rajagopal[1], Jarod Jelinek[2], Abhishek Ravi[3], and Tyler Nguyen[4]

Abstract— **This paper seeks to analyze United States Interstate fatalities to give insight into how speed impacts the overall rate of accidents. Additionally, this paper will discuss the general accident rate trend between states by looking through two main lenses - normalized and cumulative state measurements. Code can be found at: `https://github.com/fdac20/RoadSafety`**

I. INTRODUCTION

Throughout the history of the United States, there has been a struggle with creating a consistent, efficient system to allow citizens to travel across the massive continent. Created in 1956 under the Eisenhower, the Interstate System of roads attempted to decrease the cost of transportation by allowing a continuous path for travelers regardless of the state that they're located in. Although the Interstate System is a federal program, the road construction, maintenance, and laws surrounding the roads are dictated by the states that the roads themselves are situated in. This presents a unique situation where a singular road system has varying attributes, such as speed limits and vehicular traffic. This project uses data collected by the federal government in order to create some insight into how interstate travel differs between different states based on attributes such as speed limits and statistics on vehicle accidents. Additionally, this paper discusses the general accident rates between states by looking at the normalized and cumulative state measurements.

II. PROJECT PROCEDURES

A. Collected Data

In order to analyze accident rates granularly in the United States, this project mainly utilized the Fatality Analysis Reporting System (FARS) dataset, which is created and maintained by the National Highway Traffic Safety Administration (NHTSA). The FARS dataset is actually the raw data that is used in the annual report put out by the NHTSA on road safety and its associated trends. In the case of this project's analysis, the raw CSV version of the FARS datasets of 2010, 2014, and 2018 were used. It is important to note that these datasets contain extensive information on not only the accidents themselves, but many detailed metadata that can categorize and pinpoint each accident. Due to this reason, the analysis was confined to the main document called "ACCIDENTS.csv" (also titled as "ACCIDENTS.CSV"). This CSV file contains useful metadata, such as location data, automobile data, as well as fatality number and type. This project only uses the location data because the importance of the objective is placed on the localization of these accidents.

Apart from the FARS dataset, additional datasets were used. One such dataset was the Wikipedia article called "Speed limits in the United States", which contains a table that details road characteristics such as, Freeway (rural), Freeway (trucks), Freeway (urban). Because some states did not have a rural freeway, this project uses the values under the "Freeway (urban)" column and takes the maximum range of the speed as the general "speed limit" of that given state. This data was manually transcribed to a CSV file for use in the analysis. Finally, some indirect datasets were used to help visualize the accident data from FARS. The GADM provides shape files that allow the project to visualize a heatmap of the United States with a Python package called Cartopy.

B. Algorithms and Packages Used

The algorithms used to formulate the data was a general accumulator and a standard normalization technique. For the accumulator, we simply used our parser API to sum up the total number of accidents in a given state. For normalization, the number of accidents from a certain state was divided by the population of the state to create a general statistic for the automotive travel safety of the state. Several different models were used to display the data extracted from the CSV files. The bar graphs used to display total accidents and normalized accidents were rendered using the data visualization library Seaborn, which is based on Matplotlib. As mentioned before, Cartopy was used to render the heatmap of the United States. NumPy was used to create the 2D mesh for the heatmap, and GeoJSON was used to read the accident data from the FARS dataset.

III. RESULTS

From this analysis, we were able to extract some key results related directly to the FARS report. Mainly, the standard normalized values showed a trend where states with a sparse population exhibited higher rates of accidents compared to more populous regions. This can be seen with Wyoming topping all three selected years and is followed by other similar states such as Montana and South Dakota. Moreover, this can be clearly illustrated in Figure 1.

*This work was not supported by any organization

[1]V. Rajagopal is with the Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN 37916, USA `vrajago2 at vols.utk.edu`

[2]J. Jelinek is with the Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN 37916, USA `jjelinek at vols.utk.edu`

[3]A. Ravi is with the Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN 37916, USA `aravi at vols.utk.edu`

[4]T. Nguyen is with the Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN 37916, USA `tnguye69 at vols.utk.edu`

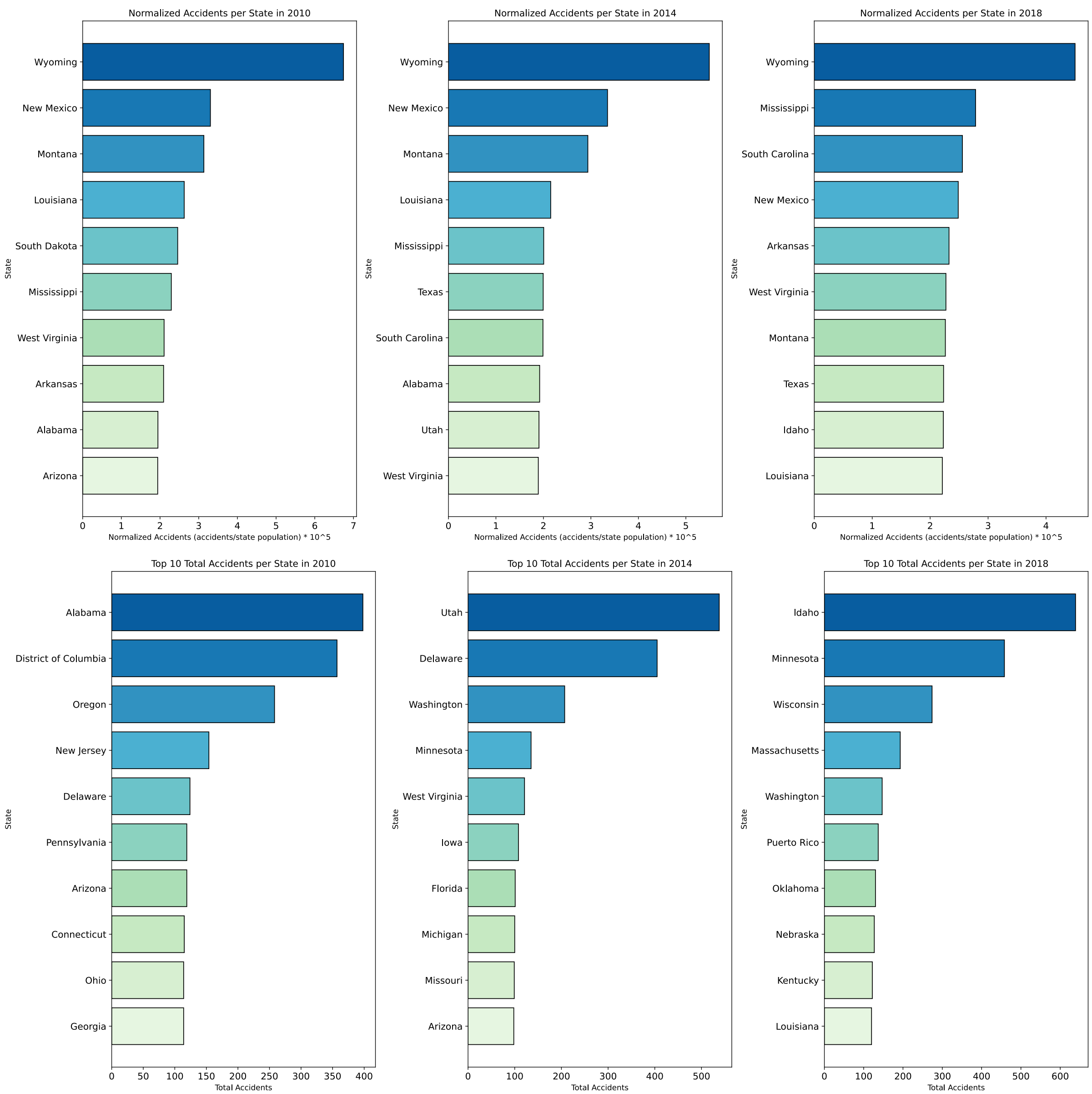

Fig. 1. The top row of bar graphs show the top 10 states with the highest accident rates (normalized total accident occurrences with state population from that year). The bottom row of bar graphs show the top 10 states with the highest cumulative accidents. The top 10 were chosen for the sake of clarity in this report. Please refer to the full Python notebook for graphs containing all states, Puerto Rico, and District of Columbia.

Apart from the analysis of FARS accident trends, there is also some results related to the relationship between speed limits and accident rates. The results of this analysis were mixed. Occasionally, the data showed some correlation between the states with higher speed limits and higher accidents, but the correlation was weak. The correlation was weak due the fact that most of the states with highest speed limits did not consistently show up in the normalized and cumulative graphs. This can be seen more clearly in Figure 2. The existence of this inconsistency indicates that predicting any sort of automotive safety cannot use speed limits alone.

IV. FUTURE WORK

For the future of this project, more of the information provided in the FARS datasets could be used to improve the accuracy of the extracted crash data and provide details on the intensity of the crash. Examples of the data that could be used to help determine a crash's intensity are the type of accident, type of vehicles involved, number of vehicles involved, and number of resulting casualties. The data could then be compared to the general road condition of the state, the time of year, and the weather conditions. The opportunities for future work are nearly endless given the countless number of factors that can be attributed to road safety.

Normalized Accidents in 2010

	State
9	Wyoming
8	New Mexico
7	Montana
6	Louisiana
5	South Dakota
4	Mississippi
3	West Virginia
2	Arkansas
1	Alabama
0	Arizona

Cumulative Accidents in 2010

	State
9	Idaho
8	Minnesota
7	Wisconsin
6	Massachusetts
5	Puerto Rico
4	Vermont
3	Louisiana
2	Kansas
1	Washington
0	Kentucky

Speed Limits

	State
9	Wyoming
8	Texas
7	South Dakota
6	North Dakota
5	New Mexico
4	Nebraska
3	Kansas
2	Colorado
1	Wisconsin
0	Virginia

Normalized Accidents in 2014

	State
9	Wyoming
8	New Mexico
7	Montana
6	Louisiana
5	Mississippi
4	Texas
3	South Carolina
2	Alabama
1	Utah
0	West Virginia

Cumulative Accidents in 2014

	State
9	Idaho
8	Minnesota
7	Wisconsin
6	Massachusetts
5	Louisiana
4	Washington
3	Nebraska
2	Vermont
1	Kansas
0	Kentucky

Speed Limits

	State
9	Wyoming
8	Texas
7	South Dakota
6	North Dakota
5	New Mexico
4	Nebraska
3	Kansas
2	Colorado
1	Wisconsin
0	Virginia

Normalized Accidents in 2018

	State
9	Wyoming
8	Mississippi
7	South Carolina
6	New Mexico
5	Arkansas
4	West Virginia
3	Montana
2	Texas
1	Idaho
0	Louisiana

Cumulative Accidents in 2018

	State
9	Idaho
8	Minnesota
7	Wisconsin
6	Massachusetts
5	Washington
4	Puerto Rico
3	Oklahoma
2	Nebraska
1	Kentucky
0	Louisiana

Speed Limits

	State
9	Wyoming
8	Texas
7	South Dakota
6	North Dakota
5	New Mexico
4	Nebraska
3	Kansas
2	Colorado
1	Wisconsin
0	Virginia

Fig. 2. These tables show the top 10 states with the highest urban speed-limits as well as the corresponding top-10 states in both normalized and cumulative accident rates. The yellow highlights indicate a match between a given normalized or cumulative graph with the state speed limit table. From this table, it can be seen that there is not a strong enough correlation between speed limits and accident rates to indicate a solid relationship.

V. IMPORTANT ISSUES

During the collection, parsing, and display of the automotive crash data, some issues were encountered. The first main issue revolves around Cartopy. Using Cartopy was rationalized based on the fact that much of the project was planned and developed on an interactive Python notebook (IPYNB) and many of the map visualizing libraries and techniques applied closer to web development and Javascript. Originally, we were planning on using a Google Maps API that was usable as a Jupyter widget, but the deal-breaking downside was that the widget did not render at all when the Python kernel was active or had ran it previously.

This made situations where viewing online and without a kernel connected (such as viewing on GitHub) impossible to do. Cartopy has an extensive range of capabilities, but its code and additional data needed to render a terrain-view of the Earth was verbose and expensive. Additionally, working with NumPy's mesh API on top of Cartopy resulted in us having to take into account things such as normalizing the coordinates, determining maximum coordinate extents, and accumulation rates for each mesh index.

The primary issue encountered when using Cartopy was that the mesh did not extend across the entire map, leaving a portion unavailable for viewing. Additionally, the data extracted from the FARS from 2010 showed some inconsistencies, which resulted in the heatmap for 2010 rendering incorrectly. Finally, a small issue encountered while parsing the FARS files was the misalignment of the data and the state_id.

VI. ORG CHART

The following list the estimated timeline for major project milestones (weeks start by the second following week of the submission of this proposal):

- **Week 1** - Establish list of reliable, parsable datasets related to road systems in United States and on automotive accidents.
- **Week 2** - Start development of code that can parse and store chosen datasets (FARS) from **Week 1**.
- **Week 3** - Implement some sort of map API for heatmap of accidents that renders in Jupyter notebook.
- **Week 4** - Start development of website design and wrap up the majority of the dataset parsing
- **Week 5** - Continue development of website and have a rough MVP done by the end of this week. **This week is also when we hopefully see some solid correlations between the data we have scrapped and processed**.
- **Week 6** - Start compiling conclusions into usable code for website.
- **Week 7** - Finalize website design and functionality for showcase.

The following is a list of project members and their responsibilities:

- **Vijaysrinivas Rajagopal** - Developed functions to parse FARS data and analyze multi-year accident reports.
- **Jarod Jelinek** - Worked on implementing Cartopy functions and rendering heatmap
- **Abhishek Ravi** - Worked with Vijaysrinivas to develop Matplotlib graphs of FARS data analysis
- **Tyler Nguyen** - Developed parsing code and helped with analysis of speed limits with accident rates.