Lauren Riker, Alexa Prague, Michael Fischgrund
INFO 3300 - Project 2 Final Report

Description of Data

The first dataset we used can be found here. We obtained this dataset from Data.gov, where it was available in various formats—RDF, CSV, JSON, and XML. We chose the CSV format because it seemed the most straightforward for visualization. The dataset includes six variables: year, state, Auto, Bus, Truck, and Motorcycle, covering all 50 states in the United states from 1900 to 2020. The Auto, Bus, Truck, and Motorcycle columns represent the number of each type of vehicle registered for a given year and state. The dataset also included information about other places like the District of Columbia but we only included those data points in the line plot, not the bar chart.We encountered an issue with this dataset where some state variables included numbers in parentheses. For example, a 2019 data point for New Jersey appeared as "New Jersey (2)." We discovered this after completing the visualizations and noticing that some data wasn't appearing. Since this was an uncommon occurrence and only affected a few data points, we manually removed the parentheses and numbers to resolve the issue. When the audience loads the page, the default year and default state for the bar chart are 1960 and Alabama. We choose Alabama since it is the first alphabetical state and we choose 1960 since it is the middle year between 1900 and 2020.

The second dataset we used can be found here. We obtained this dataset from the 3300 github page. In this dataset there were multiple variables including type (indicating a topology format), objects (geometrical objects present in the topology), states (geometric data for each state), arcs (references to arc indexes), and id (each states numeric code). We wanted to display the state name in our visualization, not their numerical code, so we altered the id variable by going into the dataset and changing the numerical code to the state's actual name. We did this manually, but if we were dealing with a number much greater than 50 states we probably should have used some sort of a loop. We connected our two datasets through the state name/id.

Visual Design Rationale

Throughout creating our project 2, our visual design was prioritized. At the top of our page we included a line plot that shows the number of total registered vehicles in the US, with a unique line for each type of vehicle: Auto, Bus, Truck and Motorcycle. A line plot was the best graph to show for this visualization because it shows the evolution of vehicles throughout the years and a clear representation of the difference between the types of vehicles each year.

We then included a map of the United States and a slider for the years. Depending on what state the user clicks on the map and year the user chooses on the slider, the bar chart shows that data. For example, if the user clicks on California and the year 2000, the bar chart will show the data for the number of registered vehicles in California in 2000 with a unique bar for each type of vehicle: Auto, Bus, Truck and Motorcycle. A bar chart was the best graph to show for

this visualization because it shows a clear comparison between the number of registered vehicles for each of the different types of vehicles for that specific state and year.

Also, when the user clicks on the state, that state on the map will turn black to provide feedback back to the user which state was clicked. We choose to have all of the states be gray, with the selected one black, because gray is known for being a null state/default state color while black is known for being a selected/completed color. The Red bar color in the Bar chart for Auto matches the Red line color in the line plot for Auto and the Purple bar color in the Bar chart for Motorcycle matches the Purple line color in the line plot for Motorcycle etc. All four of the vehicles in the line plot are color coordinated to match the colors on the bar chart.

In order to help make sure that our project was intuitive and easy for the user to understand, we added a small summary in the top right corner that explains what each graph is. We also included detailed labels for our three graphs. Also, having the position of the slider in between the map and the bar chart, helps show the user that those two graphs are connected and involve the slider.

Overview of Interactive Elements

The user has complete interactivity with the map of the United States and the bar graph. While the user is hovering over each of the US states, a textbox/tooltip will appear with that state's name. When the user clicks on a state, that state will then turn black to provide feedback back to the user which state was clicked. The user is also able to interact with the slider and use it to choose a year between 1900 and 2020. Depending on what state the user clicks on the map and year the user chooses on the slider, the bar chart shows that data. For example, if the user clicks on California and the year 2000, the bar chart will show the data for the number of registered vehicles in California in 2000 with a unique bar for each type of vehicle: Auto, Bus, Truck and Motorcycle. If the user moves the slider quickly, they will be able to visually see the vehicle bars for that state constantly and drastically changing while sliding. If the user clicks randomly on many states, they will be able to see how the number of vehicles throughout those states is drastically different for a specific year. The user is in control over what state and year is selected and they can view what they choose as a text box div box above the map, as well as on the bar chart.

We made the interactivity discoverable by having a little summary on the top right corner of the project to show a small explanation of how the graphs work and connect to one another. Also, by having the hover feature with the state name, it makes it intuitive and provides a small sneak peak that there could also be a connected clicking interactivity attached. By having the slider have a blue color with a circle, making it intuitive that moving the circle would change the amount filled by the blue color. Also, having the position of the slider in between the map and the bar chart, helps show the user that those two graphs are connected and involve the slider.

The story

Our visualization helps tell the audience how the number of registered vehicles (Bus, Auto, Truck, and Motorcycle) across the USA, and specifically within each state, have changed throughout the years of 1900-2020. When looking at the line plot, it is interesting that up until around 2012, the most popular vehicle in the USA was Auto, but since then, now the most popular vehicle is a Truck. When looking at the bar chart, we were surprised to notice that the number of buses is relatively low across all states for all of the years, as we assumed that they would be higher in the more recent years with all of the school buses.

Overall, the dramatic increase in vehicles, especially automobiles and trucks, represents the advancements in vehicle and highway development, as well as the rapid suburban development that took place throughout the 20th century. Once highway and road networks became more widespread, and there were safe, personal, and efficient vehicles that citizens could use for transportation, the population unsurprisingly shifted toward vehicles such as buses and cars, and away from railroads, trolleys, and even walking.

It's fascinating to see how specific historical time periods impacted these graphs. For example, after World War II at around 1950, many states, like Alabama and New York for instance, saw a boom in the number of registered automobiles in the bar chart. This is because after the war, the USA's economy flourished which led to more citizens being able to afford homes in the suburbs, more cars being produced, and more advancements in highway infrastructure.

It's also interesting to see the drastic differences in vehicle registrations between different states. States with larger populations, better economies, and dense cities, like New York and California, had significantly higher vehicle registration throughout the years than more rural states. This is interesting because sparsely populated states like Wyoming and Montana have more of a need for cars and motorcycles, due to the high amount of spacious rural area and lack of public transportation. From these graphs, it's clear that population and the strength of an economy have much more of an effect than ruralness on the total number of registered vehicles, even though rural states may have more vehicle registrations per-capita.

Team Member Contributions

Overall, Alexa, Michael, and Lauren contributed equally to the project, with each member playing an integral role in various aspects. After successfully collaborating on Project 1, we decided to team up again together for this project. We used github to collaborate, allowing us to share and refine code efficiently. We would have zoom meetings to plan out our ideas and were very active in our text group message to talk through ideas and collaborate.

- Michael: attended the zoom presentation meeting, contributed to all visualizations and the report.
- Alexa: contributed to all visualizations and the report, primarily focusing on creating the line plot and bar graph and connecting them to the map of the US and slider.

- Lauren: contributed to all visualizations and the report, primarily focusing on the map of the US, and creating the descriptions associated with the visualizations.

The visualizations took the most time to develop due to their complexity, with many moving parts that required iterative troubleshooting. Whenever any team member encountered a challenge, we collaborated through GitHub and reached out to each other to solve the issues.