

Final Project Report

2a. Visualization:

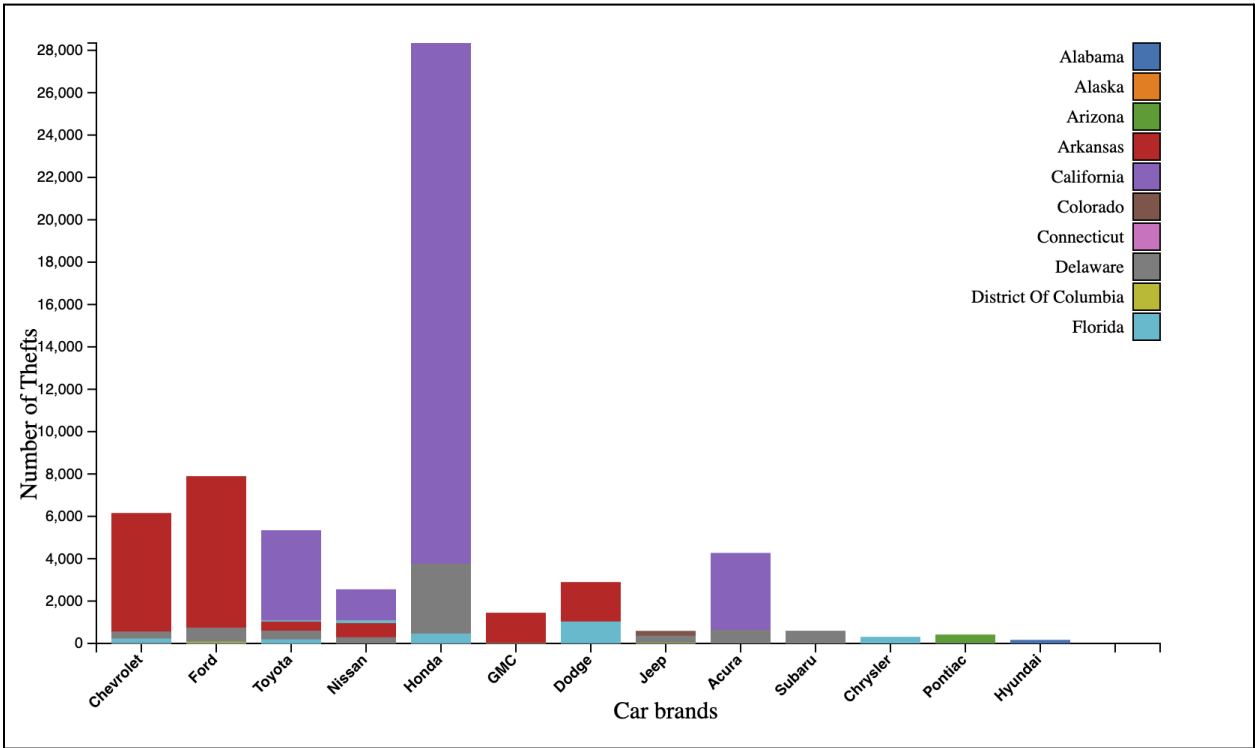


Figure 1: The image displays a visualization of the Top 10 States for Car Thefts in 2015

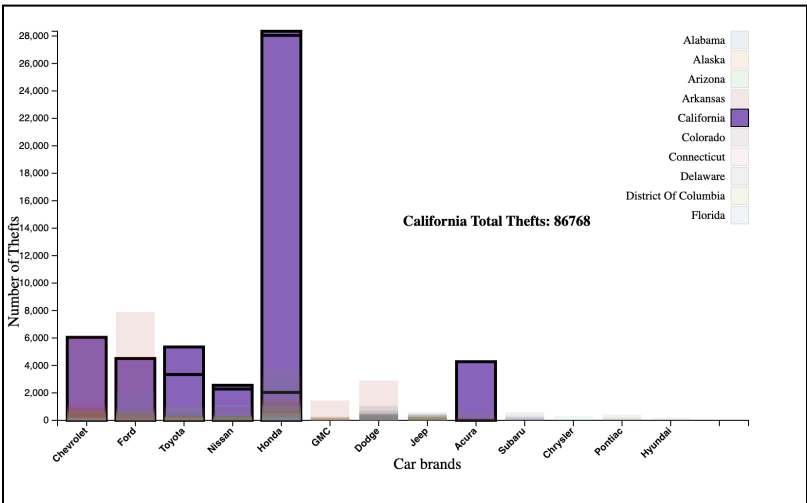


Figure 2: The image displays a visualization of the totaling feature in our Graph

2b. A description of data: The two datasets that we have collected are from Kaggle. In the beginning, we used a dataset that had less information about car thefts. It only covers car models from the year 2015, and thefts from the year 2015—in addition to the number of thefts—and the sample size is 26. The second dataset that we are currently using has the state in which the car was stolen, and the ranking of the states, and it covers a wider number of car model years—up to 2015—as well as theft years long before 2015. The sample size of this dataset is 509 covering a larger range of car theft cases and overall being more robust. We did not use the ranking of the states or the car model year because we were mainly looking to understand the most prominently stolen car and for each state that the number of thefts corresponded to.

A clearer representation of our dataset's variables:

- State: The US state in which the vehicle theft occurred. (String)
- Rank: The rank order of reported thefts from highest to lowest for each make/model within a state. (Integer)
- Make/Model: The make and model of the vehicle that was stolen. (String)
- Model Year: The year the vehicle model was stolen. (Integer)
- Thefts: The number of thefts recorded for that car model in each state. (Integer)

2c. An overview of your design rationale: For **position**, the bars along the horizontal axis represent different car brands. The height of each bar along the vertical axis represents the number of thefts for that brand. For **shape**, the chart uses stacked bar charts, where each bar is divided into segments of different colors. This shows the cumulative contribution of various states to the total number of thefts per car brand. For **color**, different colors represent states, as indicated by the legend on the right side of the chart. When you hover over one of the states on their color-coded legend, it shows the total number of thefts across all car models. For example, as seen in **Figure 2**, California has 86,768 car thefts across all car models.

2d. The story: The data expresses a startling problem in the US as many cars are stolen. This is especially true for Hondas as they have the highest counts of thefts. This could be used for insurance companies to understand that a Honda car in California has an increasingly higher chance of being stolen. It could also inform security managers at Honda that they have more work to do. Our visualization spells a picture that tells consumers what cars are the safest and which need more work.

Team Contribution

Stephen Syl-Akinwale: Worked on cleaning data, and creating and recreating the visualization.

Joshua Jaquez: Worked on the description of data, and the design rationale.

Anthony Lewis: Worked on reviewing the final submission.

Lukman Moyosore: Worked on Graph interactivity.