

Michael Coombs

6/3/24

Data Visualization Mini Project 1 Report

This project examined the EPA's Fuel Economy data from the National Vehicle and Fuel Emissions Laboratory. Specifically, this project focuses on the later 2000's, mostly from 2012 to 2017.

I wanted to demonstrate the improvement of vehicle efficiency over time, while not inundating the reader with data or overestimating and boiling down the data too much. I decided to examine the top 3 most fuel-efficient cars available each year from 2012 to 2017. I split this analysis by how the car is powered, either gas or electric, such that one reading the data could see how each technology improved over this time span. In tandem with this, I plotted the efficiency of the most-efficient car in the dataset for each year between 2000 and 2017, to give some insight into the availability and adoption of high-efficiency technology over this time span.

This project demonstrates several key ideas.

The core "story" these visualizations communicate is that consumer cars have significantly improved in efficiency over the course of the 2000's, especially with the advent of electrification.

It shows also that both gas-powered and electric cars improved in efficiency between 2012 and 2017, and shows that this time frame was where the most efficient cars became available to consumers, with cars above 100 mpg being available to consumers every year from 2011 to 2017. This also demonstrates that the Electric powertrain is key to this level of efficiency.

It also shows that, during the early 2000's, multiple carmakers tried their hand at making a highly economical vehicle above 60 mpg, but they did not stick around for more than 1 or 2 years, and availability of cars at this level of efficiency was spotty at best. This is shown by the fact that for only 4 years prior to 2010, a car was tested that broke 60mpg. Before 2010, the longest running vehicle over 60 mpg was the Toyota Rav4 EV, which was represented in the graph for 2 years, before the Honda Insight became the leader at under 60 mpg. This indicates that the Rav4EV was no longer sold, or no longer tested.

I applied multiple principles of data visualizations and design for this assignment, and they were implemented in R with a few key methods.

Firstly, the multiplots were expanded in size to allow for legibility and avoid crowding. This was done by setting the `fig.width` in each `r` chunk to 12. Next, labels were added that clearly described the value each bar was meant to represent in the multiplots. I also used colors and fills to create aesthetically-pleasing and legible plots. For the line graph, repelling labels were made and all lines were drawn to their respective points to ensure that each datapoint is clearly labeled with the leading car of that year.

To examine this data, I first loaded in my data and libraries. I then broke this data down by its power-type (elec/gas) and year, and assembled the top 3 cars for each year from 2012 to 2017 for each type of car. I chose this time span as this can show a good idea of the improvement over time of each type of car without oversaturating the reader with information. Lastly, I regrouped each dataset by its year, and graphed it in a multiplot. To examine the efficiency over time, I simply took the highest combined mpg car from each year and plotted the efficiencies. I cut the data off at 2000 for this analysis as most cars prior were below 40 mpg, with little improvement until the late 90's.