Team Mini Project: Alternative Fueling Locations

Tess Anderson, Caleb Puckett

DS450-01

Data Science Senior Capstone

The decision to use Tableau for this project was primarily based on a desire to deepen our understanding of this tool. One of us had never used Tableau before, and the other had used it briefly. One of us, while never having used Tableau, had experienced an internship with coworkers who frequently used Tableau and who mentioned that it had been a useful tool throughout their careers. One of us has had some experience but wanted to become more capable of using Tableau to represent data efficiently and clearly. The simple and clear functionality of Tableau also appealed to both of us.

Alternate fuel locations in the United States were exclusively represented using maps. A column of regions was added in Python with the Pandas library using a series of “if, then” statements. In this code, the state associated with each data point was used to determine the region. With this, we were able to differentiate each data point into Midwest, Southwest, West, Southeast, and Northeast. Maps were used to represent the distributions of alternate fuel locations within each region, while bar graphs were used to represent the distribution of different fuel types by state. Bar graphs were used here because there were few enough fuel types for the charts to be legible and accurately present a comparison of the number of each type.

Maps were created to show the placement of each electric vehicle network location, colored by different electric vehicle network types. This was done separately for the continental United States, Alaska, Hawaii, the Midwest, Southwest, West, Southeast, and the Northeast. The part to whole graphs used to represent the network distribution amongst each of these seven areas were tree maps. These were chosen because of the numerous different network types. Because of this, bar graphs would not have been effective at representing the data clearly. A bubble could have done the job, but it may have been more confusing to understand than a simple tree map.

The ten most populous states were found in a chart online from the United States Census Bureau (Bureau). This list, in decreasing order, is California, Texas, Florida, New York, Pennsylvania, Illinois, Ohio, Georgia, North Carolina, and Michigan. Maps were used to represent the AFL distribution throughout each state, with a different color representing each fuel type. Bar charts were again used here to show the distribution of fuel types because there were few enough fuel types to make the charts legible and still represent how the number of each type compares.

Although the data were not especially difficult to work with, there were some issues which needed resolution to produce clear and representative visualizations. One issue involved incorrect longitude values. These values were zero, placing them well outside of the United States. One of these values was listed as being in the state of New York and the other was listed as being in the state of Pennsylvania. Knowing that these were most likely incorrect data points, the next step was to remove them from the visualizations for the sake of consistency. Rather than remove the data entirely, a filter was applied to exclude these values in the few spots they were present. Because this issue was not widespread throughout the data, it did not require any further attention.

Another issue that accompanied this project, although caused by user error rather than an error in the data, had to do with efficiency. After manually creating multiple similar dashboards and all individual visualizations for these dashboards, the option to recreate previous work for a new condition was discovered. For instance, this meant that instead of doing the work for each region from scratch, it was possible to complete the work for one region, and then simply copy and apply that work to all other regions. Although it would have been much more convenient to know this from the start, its discovery was still very convenient. This discovery also helped both of us in achieving our goal of understanding more about Tableau and its capabilities.

Another difficulty of this project involved the use of pages for different variables of a graph. Initially, a separate graph was created to represent each different data point category. For instance, the need to represent seven different graphs for a given region, each with only one fuel type, originally involved manually creating seven separate graphs, with each filtered for its desired fuel type. After discovering the option to use pages, however, these were able to be consolidated into one graph with several pages to flip between. Although this was not discovered early in the process, it was another instance of us achieving our goal of increased proficiency in Tableau.

Besides the exclusion of those few flawed data points, there was no need to clean the data. There was no preparation required besides the creation of a new column to identify the regions of each data point based on its state.

*References*

Bureau, US Census. “U.S. Population Trends Return to Pre-Pandemic Norms as More States Gain Population.” *Census.Gov*, 26 Dec. 2023, [www.census.gov/newsroom/press-releases/2023/population-trends-return-to-pre-pandemic-norms.html](http://www.census.gov/newsroom/press-releases/2023/population-trends-return-to-pre-pandemic-norms.html).