Project 2 report

a) Data description

#1 Visualization:

The final dataset is named “info3300_proj2_dataset_sum.csv”, with “us-smaller.json” to help us with US map projection.

The final dataset we compiled for the visualization consists of several features: carbon emission (in million metric tons) per capita, number of cattle (in 1000 heads) in inventory per capita, and number of automobiles registered per capita for 1980, 1990, 2000, 2010, and 2020 for the 52 states of the US. We got the carbon emission data from US Energy Information Administration’s website (https://www.eia.gov/environment/emissions/state/) which features a complete dataset for per capita energy-related carbon dioxide emissions by state (1970-2020).

Asides from displaying the carbon emission data on the map, we figured the number of cattle and automobiles would be key contributors to carbon emissions since beef produces the most greenhouse gas emissions and road transport accounts for the majority of transport emissions. We extract the cattle inventory data from USDA (https://usda.library.cornell.edu/concern/publications/h702q636h?locale=en#release-items) which release inventory numbers and values of all cattle and calves twice every year. We found the respective dataset for 1980, 1990, 2000, 2010, and 2020. One shortage of the dataset is that the statistics of the District of Columbia are missing. Given the number is likely to be insignificant, we decided to still use this dataset and not use DC for our visualization. Rather than comparing the absolute number for each state, we thought it would be more effective if we look at the data on a per capita basis. To convert the number to a per capita basis, we also looked for the US population by state for the five respective years on the US Census website. Then, I created new columns (per capita basis) for each year. To compile the automobile dataset, we went to the US Department of transportation which features highway statistics (https://www.fhwa.dot.gov/policyinformation/statistics.cfm) including the number of automobiles registered every year. Similarly, I converted the data to a per capita basis for ease of comparison.

#2 Visualization:

The final dataset is named “ev.json”.

To compile the EV (Electric Vehicles) registration dataset, we turned to the US Department of Energy (https://afdc.energy.gov/vehicle-registration) which features the light-duty vehicle registration counts by state and fuel type from 2016 to 2021. I scraped the data from the 5 excels for the visualization.

When selecting data to be used for our visualization, we made sure to use credible sources to ensure the accuracy and consistency of our data.

b) Visual design rationale overview

After examining the datasets, we decided to include visualizations that can present the changes in carbon emission and its relationship with other variables by year and by state.

<u>Visualization #1:</u>
For the first visualization, we are hoping to feature a map to display the amount of carbon emission across states and time, as well as how the number of cattle and automobiles in each state might affect carbon dioxide emission.

To visualize the carbon emission trend by state, we projected the data into a US map. Given that people usually associate pollution with grey, gloomy color, we decided to deploy darker Aegean grey for the states with more emissions and lighter Aegean grey for the state with lesser emissions. After looking a closer look at the dataset, we found that most states emit about 10-28 million metric tons of carbon dioxide per capita. To make sure the variance is visible to the user, we decided to apply the Threshold scale instead of the Ordinal scale. We inspected the data and placed more thresholds in the smaller range of the values instead of having the thresholds evenly distributed. Thus, we magnified the variance of data across states and time. Otherwise, the users will see the same color across a majority of states, which is less user-friendly. However, one disadvantage of this scale is that the colors then do not reflect the variance in data proportionally, which might lead to some confusion. In this map, the visual channels are the different color saturation, which represents the quantity of carbon-dioxide emission. Visual marks are on the outline of each state.

In order to examine any correlations between carbon emission and its contributors (Number of Cattle per capita and Number of Automobiles per capita) by year, we decided to use the scatter plot, which can better capture any statistically meaningful changes and relationships. The visual marks are dark blue and orange circles. The visual channels are the vertical and horizontal aligned positions.

<u>Visualization#2</u>

We noticed that the electric vehicle industry has received high expectations and investment regarding the issue of environmental degradation. Thus, for the second visualization, we are hoping to turn to this prominent future factor, specifically, to see how EV/PHEV/BEV gains popularity over time. We used a line chart to perform such visualization. The visual marks are lines and dots. The visual channels are the vertical and horizontal aligned positions of each dot. In order for the users to differentiate types of vehicles clearly, we used different color hues to represent different types of EVs.

c) Interactivity design rationale

For the first visualization, we added two interactions to the map. One is when mousing over each state, its corresponding information about Carbon Emission per capita, cattle per capita, and automobile per capita will appear in a rectangular text box. The color change itself can't effectively inform users of the specific amount of Carbon Emission in each state. It can give users a direct visualization of different states' performances compared with others. By adding this mouseover interaction, we are able to provide more quantitative information as a complement to the map. We decided to use a black rectangle as the background and white text to show contrast and make the interaction easily discoverable.

The second interaction for the map is a time slider that shows the corresponding map and scatterplots for the specific year. The time slider allows users to change the displayed year by decade, from 1980 to 2020, showing 5 different years of information in total. This visualization is applied to both map and scatter plot visualizations, so the combination of these two can give users a more comprehensive understanding of the Carbon Emission situation both geographically and quantitatively. With the time slider, we are able to provide visualizations that show environmental changes across a large time span, providing more insights than only showing the 2020 data.

For the second visualization, we focus more on the amount and components of different types of Electric Vehicles (EVs). The lines can reflect the general increasing trend for different vehicle types. In order to illustrate specific amounts of vehicles, we include the mouseover functions for all the circles every year. In this way, the users can hover their mouse over the points to know detailed information about a specific type of vehicle for the corresponding year. The design can make the visualization more informative and clearer.

d) The story

As COP 27 takes place in Egypt, it again directed our attention to climate change and our commitment to the Paris Agreement. Most countries are struggling to meet the goal as outlined by the Paris agreement, the US is no exception. In our visualization, we are hoping to look at the carbon-related emission over the past few decades, as well as key contributors to greenhouse gas.

The first COP meeting was held in 1995, we were hoping to look at the carbon emission data from 1980 to 2000 to see whether measures are in effect to counter climate change.

The map explored the carbon dioxide emission trend over different geographical regions in the US from 1980 to 2020. We see there is an increase in carbon emissions overall from 1980-2000, as represented by the darker colors in many states. But after 2000, many states exhibit lighter colors, which means their carbon emissions decreased. We thought the change could be aligned with the fact that more carbon-neutralization measures have been deployed by state governments after 2000.

Looking at specific states, we found that in general California has very low carbon emissions per capita, which might be due to strict fuel standards and its cap-and-trade program. In contrast, Alaska has very high carbon emissions per capita. The reason for this is extensive oil extraction in the Arctic and transportation, as many communities can only be accessed by river or air. Additionally, Wyoming and North Dakota have had very high carbon emissions per capita throughout 1980 -200 because they generate a large portion of coal in the country.

Knowing that transportation and our diet contribute to a large portion of carbon dioxide emission, we wanted to take a closer look at the relationship between the number of cattle, the automobile registered and the amount of carbon dioxide emission in each state. We hypothesized that carbon emissions would increase with the number of cattle, and the number of automobiles registered. Looking at the scatterplots, we can see that there is a weak positive relationship between the number of cattle and carbon dioxide emission, meaning the carbon dioxide emission generally increases with the number of cattle in the respective state. Looking at the pattern of the dots over time, I found that the number of cattle in inventory has been decreasing over time. Similarly, we also found a weak positive relationship between the number of automobiles per capita and carbon emissions across different years, meaning carbon emission generally increases with the number of vehicles. This result is expected as automobiles are one of the most important factors contributing to carbon emissions. Additionally, I found it to be interesting that the number of automobiles per capita has been decreasing recently (throughout 2010 – 2020). For future investigation, we can also look at transportation in general (including automobiles, buses, motorcycles, etc.), electricity production, food consumption, etc.

Knowing that electric vehicles are one of the most significant technological transformations to help mitigate carbon emissions, we decided to explore the different types of EVs in the second visualization. There are three different types of electric vehicles: EV, PHEV, and HEV. EVs are pure electric vehicles, HEVs are hybrid electric vehicles, and PHEVs are plug-in electric vehicles, which use batteries to power an electric motor and gasoline to power an internal combustion engine. The number of HEVs has always been significantly higher than the number of EVs or PHEVs from 2016 to 2021, and it also shows the largest increase from 2020 to 2021 among the different types of EVs. For future research, we can explore the amount of CO2 produced by different types of EVs in comparison to traditional petrol cars. With gas prices hitting a record high as a result of the Russian-Ukraine war, many more people are considering electric vehicles as an alternative, hoping to save fuel costs in the long run. One reason for the popularity of HEVs is that they achieve a good balance between saving fuel costs and having a reasonable mileage range. They also require less expensive battery packs, so switching to HEVs could often be an easier transition than switching to EVs for many drivers. Based on the information reflected by the charts, we can conclude that more users prefer Hybrid Electric Vehicles (HEV) and it has a greater increase (steeper slope) compared to other types of vehicles in the market. However, starting in 2020, electric vehicles (EV) also present a steep increase in the registered number of cars. The statistical conclusion is also aligned with the fact that Hybrid Electric Vehicles (HEV) have better fuel economy without sacrificing performance in a way to utilize both energy sources effectively. So, in the future, we can predict a continuous increase in the HEV type, which is likely to be the mainstream electric vehicle people will choose. PHEV is likely to become obsolete as charging infrastructure expanded.

Team member contributions:

- Isabella Zhi: Visualization 1, Final report
- Jessica Zhou: Data cleaning, Visualization 1, Visualization 2, Final report
- Kayla Yang: Visualization 1, Final report
- Jasmine Wang: Visualization 2, Final report

Time Commitment:

- Brainstorming: 3hr
- Data Gathering & Cleaning: 4hr
- Visualization#1: 15hr
- Visualization#2: 5hr
- Final Report: 6hr