Project Milestone 2

Cannon Lock, Tsz Yau Iris Chow, Madison Dunn, Nicole Athanasiou, Hannah Sun

H_0: CO2 emission does not change during covid in US
H_1: CO2 emission decrease during covid in US

Data Curation - Covid Data

To clean the COVID-19 dataset from the New York Times, we first dropped the unnecessary columns in the dataset, including county, fips, and state by taking a subset of the original dataset, as we are only interested in the country. Then, we grouped the observations by date so that the values for each location were summed together for each day. Completing this step provided us with the total cases and deaths in the United States for each day.

To tidy the dataset to correspond to our ozone data and the analysis that we desired, we had to derive the mortality and infection rates for all of the United States by day. To do this, we derived a new variable, “mortalityRate”, by dividing deaths by cases and using the function “mutate”. In addition, we derived the variable “infection” by taking the difference in cases between days. We will use the derived variables as a comparison to the CO2 rate.

The steps in the curation process of the COVID data were not especially challenging. Most of the steps involved techniques we have learned in class. The most challenging step was computing the difference between days as the step involved an unfamiliar function.

Data Curation - CO2 Emissions Data

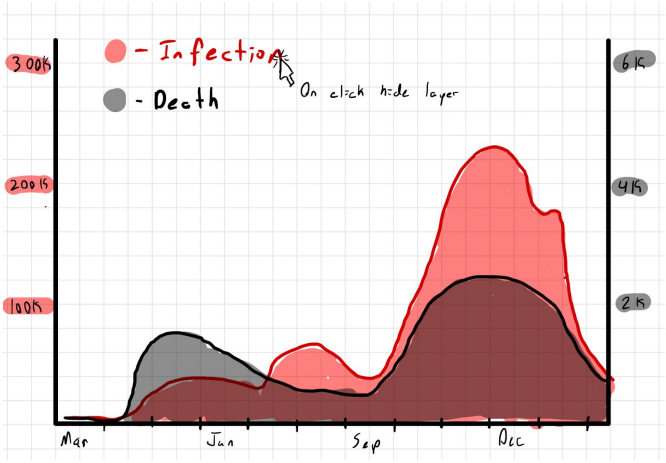
There were a few steps to clean and tidy the CO2 Emissions dataset. First, we removed an unnecessary column from the dataset, the ‘Group of Countries’ column (as the dataset is only looking at the United States). We did this since the column would not be helpful for anything and would just take up space. We next removed rows containing ‘NA’ which could cause cause issues of missing data in our visualizations. The last thing we did to tidy the dataset was modify the dates by using mutate() and group_by(), so that the dates could be used correctly in our

visualizations. The dates need to be formatted in the form "day/month/year" and they need to be recognized as dates to be used correctly in our visualizations.

We made two copies of the data. One focuses on the carbon emissions each day and the other focuses on carbon emissions from each sector.

In general, it was not too challenging to clean and tidy this dataset as the original dataset was mostly well formatted. The most challenging step was utilizing dplyr to format the dates correctly as at first glance to the data, one might not realize it had to be done.

Design Brainstorming - Covid Visualization



The encodings used in this visualization include size and texture, in the form of area under the curve and the corresponding color respectively. The marks used are two lines and areas underneath that show the infection and death rates of Covid-19.

This emphasizes the rate of the virus spreading through the United States which pairs with our analysis of the emissions in the U.S. during the same time period. In this visualization the comparison is not yet made to Covid and Emissions.

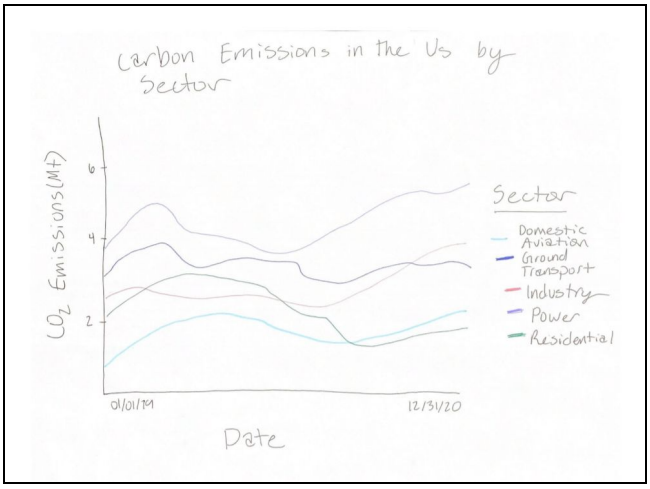
Project Milestone 2

Cannon Lock, Tsz Yau Iris Chow, Madison Dunn, Nicole Athanasiou, Hannah Sun

The strengths of this design are that it makes it easy to see the rate of change and relationship of Covid-19 over 2020. One weakness is that the emissions cannot be easily added to it without making it too busy. To address that visualization we can use a different graph style.

Hover cues that increase opacity of the hovered element is the cue and response.

Design Brainstorming - CO2 Emissions Visualization



The encodings used in this graph are color and position. The different sectors are different colors and the position of the lines correspond to the amount of CO2 emissions for each date from January 1, 2019 to December 31, 2020 in metric tons.

This visualization emphasizes the change in carbon emissions in the United States in comparison to the rate of spread of COVID-19. This visualization also compares the carbon emissions from various sectors, without comparing it to the rate of COVID infections or deaths.

The strength of this design is that it will be clear to see the differences between carbon emissions for the five sectors over time but it may be difficult to compare to COVID infection/death rates unless we do them side by side.

Interactivity could be used to highlight the specific date and carbon rate when hovering over the different sector. When hovering over a point, the specifics will show.