**Queries**

* *Which years saw the greatest increase in emissions? Can we tie this increase to a particular event?* We examined this question by subsetting the data by Year-Over-Year percent change values of over 100%; that is, years where a country’s emissions at least doubled. This resulted in 354 observations, a bit too much to examine all at once, so we used the pd.Series.value\_counts() method to get the number of times each Year appeared in the resulting subset. We found that 1950 appeared 26 times, meaning 26 countries’ emissions doubled from 1949 to 1950. Many of these countries were in regions of the world that were rapidly industrializing at the time like Latin America, Southeast Asia, and Eastern Europe, which could explain why 1950 was such a common year to see these sorts of increases.
* *When did countries decrease their emissions? Do these moves correlate with increased awareness of the effects of emissions on the climate?* We examined this section by subsetting the data by Year-Over-Year percent change values that were negative. This resulted in nearly 5000 observations, and we similarly used pd.Series.value\_counts() to see when the decreases occurred. 106 countries saw decreases in 2009, and in examining the possible causes we considered a link between the late-2000s financial crisis and a decrease in general economic activities like travel and manufacturing which could have precipitated the decreases.
* *Are there any instances of countries with carbon-zero years in the modern era?* We examined this question by subsetting the dataset pertaining to observations after 1963 for Emissions values of 0. While some observations met the criteria, they were several sequential years from Moldova and Kyrgyzstan, indicating data may have simply still been unavailable for the earliest years in the range we considered. Therefore it appears there have been no genuinely carbon-zero years for any country in the modern industrial area.
* *What proportion of global emissions did each country make up in 2017?* As 2017 is the most recent year in the dataset, we decided to use it as the year in which to examine the breakdown of global emissions by country. We accomplished this by subsetting the data and sorting emissions values in descending order. We then produced a pie chart; with all countries shown, the chart was impossible to read, so we created a function which takes a value from the user to choose how many of the top countries to show individually, then aggregates remaining countries into an “other” category before producing the chart. We found that the largest industrialized economies (China, United States, India) accounted for over half of global emissions in 2017.

**Conclusions**

The knowledge that human activity can, and does, shape the ecosystem of our planet is perhaps the most vital information to our future as a species to come from scientific research. No activity on Earth can escape the influence of the environment, and its sensitivity to activities like emission of greenhouse gasses must be met with equal sensitivity on the part of our societies. Our group’s investigation into historic data pertaining to CO2 emissions has, if anything, emphasized to us that the present problem cannot be solved by individuals acting alone, or even individual nations, but is inherently a function of the actions and efforts of the world’s population and industries as a whole. We have seen how large-scale human events like wars, economic fluctuations, and growth and decline of industries can affect emissions on the scale of a nation, as well as the timeframe over which the current state of things has developed, and can thereby gain an understanding of the scale of effort required to alleviate these effects through prevention and mitigation.

Our program has allowed us to more easily discover and visualize trends in this vital data that has not only improved our understanding of the subject but has led us to postulate about causes and effects of changes in particular periods. We have been able to determine periods of significant change in emissions and discuss their causes, as well as gain an overview of the current state of global emissions. Were we to extend the scope of this project, we might try to incorporate other datasets that measure other climate metrics or provide important context like population and demographic information. This sort of investigation would allow us to understand key issues related to emissions and its effects more deeply and perhaps apply our findings more concretely. We could then expand this functionality into the user-facing portion of the project by supporting more data and types of visualization in the visualization tool. We feel this investigation has improved our understanding of working with and querying large datasets and has provided us a starting point for investigating complex problems of this scale including important aspects like planning, exploring large datasets, and validating our work.