Exercise 03 Report: *Effects of Population and GDP on State CO2 Levels*

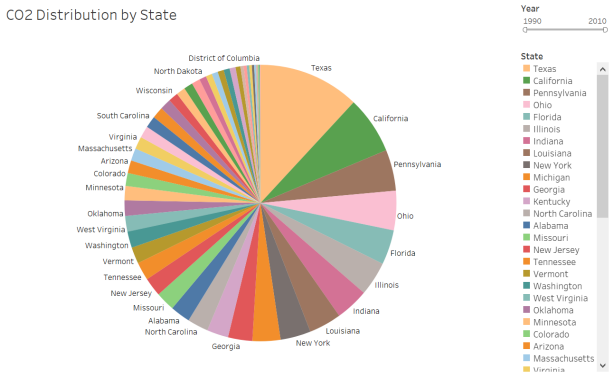
Adelle Driker, Bo Yan, Yuan Hu

Motivation: One of today's most talked-about issues that is affecting the planet is the amount of greenhouse gases, specially carbon dioxide, that are produced around the world. The United States, having a large GDP that is heavily manufacturing-, power plant-, and transportation-based, is considered one of the world's largest producers of carbon dioxide. Further analysis of the population, CO2, and GDP relationships and their trends may be performed at the state and per capita level.

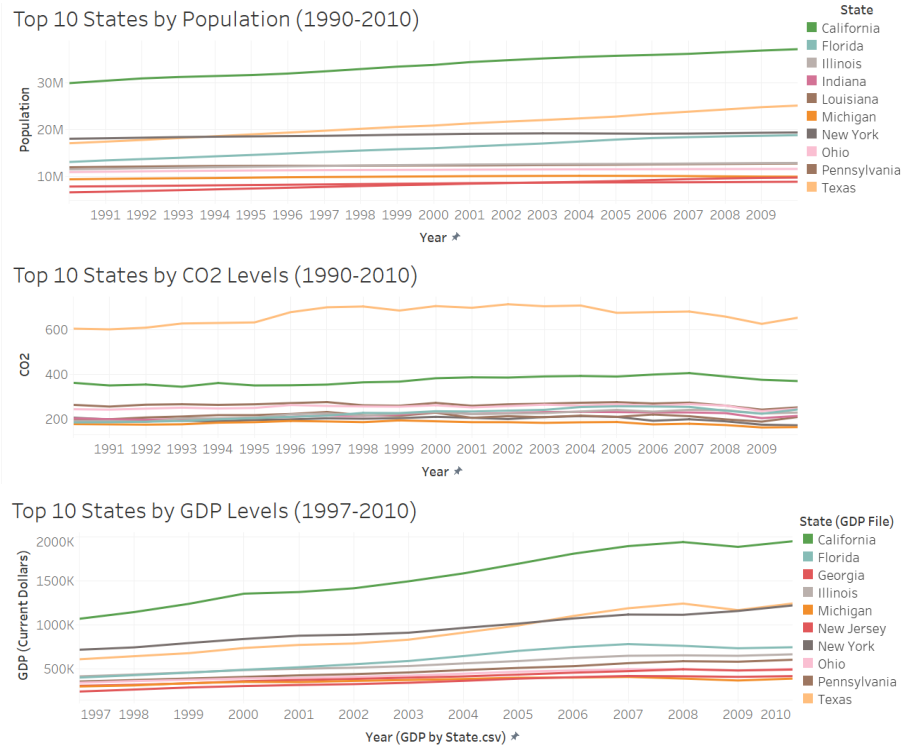
Tasks: Given two datasets, one containing information about a state's population and CO2 levels by year (spanning 1990-2010) and the other consisting of state GDP totals (in today's dollar value) by year (spanning 1997-2010), the following questions arise: which States are the Top 10 producers of CO2? Top 10 by Population totals? By GDP totals? Using that data, can the Per Capita amounts of CO2 give rise to additional insights? What conclusions may be made about the relationships between CO2 vs Population and CO2 vs GDP over time?

Data Preparation/Augmentation: The original dataset (co2-population.csv) contained both CO2 values and Population totals within the same Sector attribute; in order to calculate the per capita totals, the file was adjusted to extract the Population Sector name as its own field and allowed for a new CO2 field containing value totals from non-population sectors. A new calculated field was added as a result, the CO2perCapita, in Millions (= CO2 / (Population/1000000)). Additionally, to add another perspective of causes/effects of CO2 levels, an auxiliary dataset of GDP totals by state has been incorporated to augment the original dataset and another calculated field, CO2/GDP was created. Please note that the GDP dataset includes data beginning in 1997, which is seven years shorter than the original dataset.

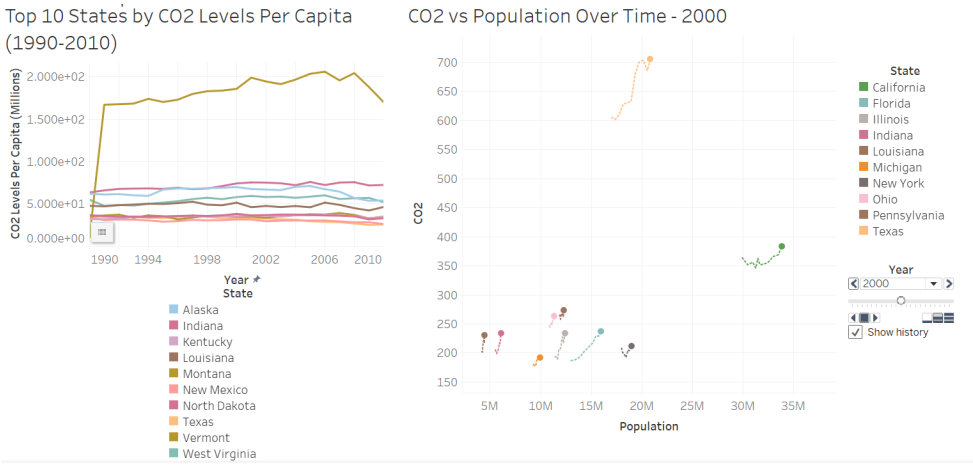
Visualization: To help the user analyze noteworthy trends, the visualization is composed of four dashboards, each with a dedicated purpose. Since the datasets contain a collection of time series data, a Line Plot with Year on the x-axis was chosen to effectively represent the Sequential data. The Categorical data has been modeled using a combination of color schemes to distinctly represent the Top 10 States. The first (shown below) is a Pie Chart, ordered by CO2 levels in a clockwise order, that summarizes the CO2 totals for each state in the union, with the State names appearing as labels and totals displayed upon hovering over a section. The user may use the Year slider in the top right corner to apply a desired Year of interest as well as select a state to inspect using the State legend on the right.



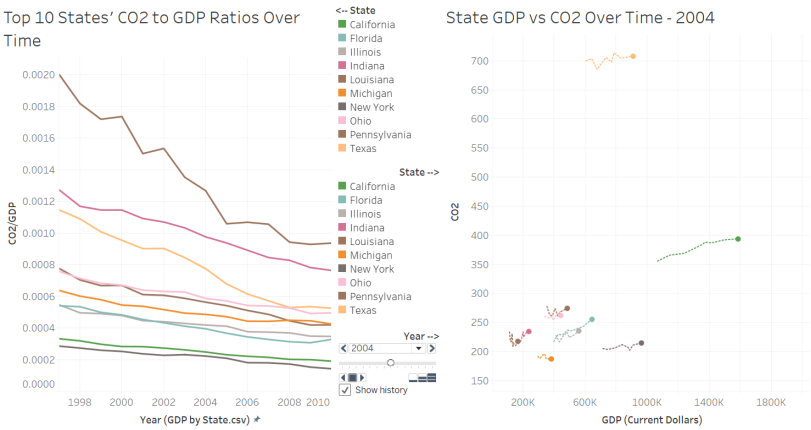
The next dashboard presents the Top 10 States by CO2 levels, Population, and GDP over time and each colored line represents a State's Population/CO2/GDP totals. An interactive legend on the right of the charts allows the user to filter and view individual states, and hovering over a section of the line will highlight the point's values/totals (this is also true for other line plots in the dashboards that follow).



The third dashboard combines the CO2 and Population data to illustrate the Per Capita trends. The left chart in the dashboard shows the Top 10 States by CO2 levels Per Capita in a Line Plot, while the figure on the right portrays how the Population may affect CO2 levels over time with the help of an animation. The user is able to play or pause the animation using the Year legend on the bottom right, and the speed may be adjusted to preference. The figure below shows the animation paused at the year 2000 to show that the current year will be effectively shown as a solid circle, and the previous years' data will be tracked with a dashed line.



The final dashboard focuses on combining the GDP data with the original dataset. The figure on the left illustrates the Top 10 States with the highest ratios of CO2 to GDP over time, which may be used to estimate the impact of GDP on a state's CO2 levels or vice versa. The chart on the right is another animation (similar to the previous dashboard's), which shows the user how the relationship between a state's GDP and CO2 levels changed over time.



Omitted Data and Visualization: Due to the nature of the tasks, the Sector names (besides Population) were extraneous; however, further breakdown by Sector types may be added in future versions of the report. Additionally, the dataset contains totals for the US as a whole, which were removed as well since the visualization was focusing on the state level.

Conclusions: To conclude, the combination of both datasets gave rise to important insights regarding Population totals, CO2 levels, and GDP totals and their relationships with each other across the United States. From the first dashboard, one may observe that the Top 10 States compose nearly 50% of the nation's total CO2 levels, an eye-opening insight. From the second dashboard, the user may deduce that although CA has had a much higher Population and GDP totals throughout 1990-2010, TX had been growing at approximately the same rate but actually had much higher CO2 levels.

As a result of the further Per Capita investigation, Vermont was found to have the highest CO2 per capita levels which was very unexpected, but may possibly be explained by low Population but a relatively more industrial state economy. Some rather interesting conclusions may be made from the first animation (CO2 vs Population) as well - clear outliers like CA and TX show steady trends in the Population and CO2 directions, respectively. Most states begin to show a dip in CO2 levels around 2008-2009, which is most likely due to the Recession that affected states' economies.

Finally, after analysis of the GDP-based dashboard, it is apparent that the CO2/GDP ratios show a negative trend over time for the Top 10 States, with Louisiana being at the very top of the list. This may be due to increasing GDP levels and/or decreasing CO2 levels over time. The animation shows that as Top 10 States' GDPs continued to grow, their CO2 levels stayed relatively constant, with another noticeable dip around 2008-2009 indicating the effects of the Recession. Overall, CO2 levels may be considered a factor affecting a State's GDP levels, but not the main driver.

GDP Data Source: Bureau of Economic Analysis - US Department of Commerce
https://apps.bea.gov/itable/iTable.cfm?ReqID=70&step=1