Climate Change Analysis Report:

By: Travis Marchok, Adam Whiteside, Eli Aaron, Chylle Gamino

We will analyze the effects of gas emissions on our average temperature, and in turn, the rising sea level. This will be done through 5 parts:

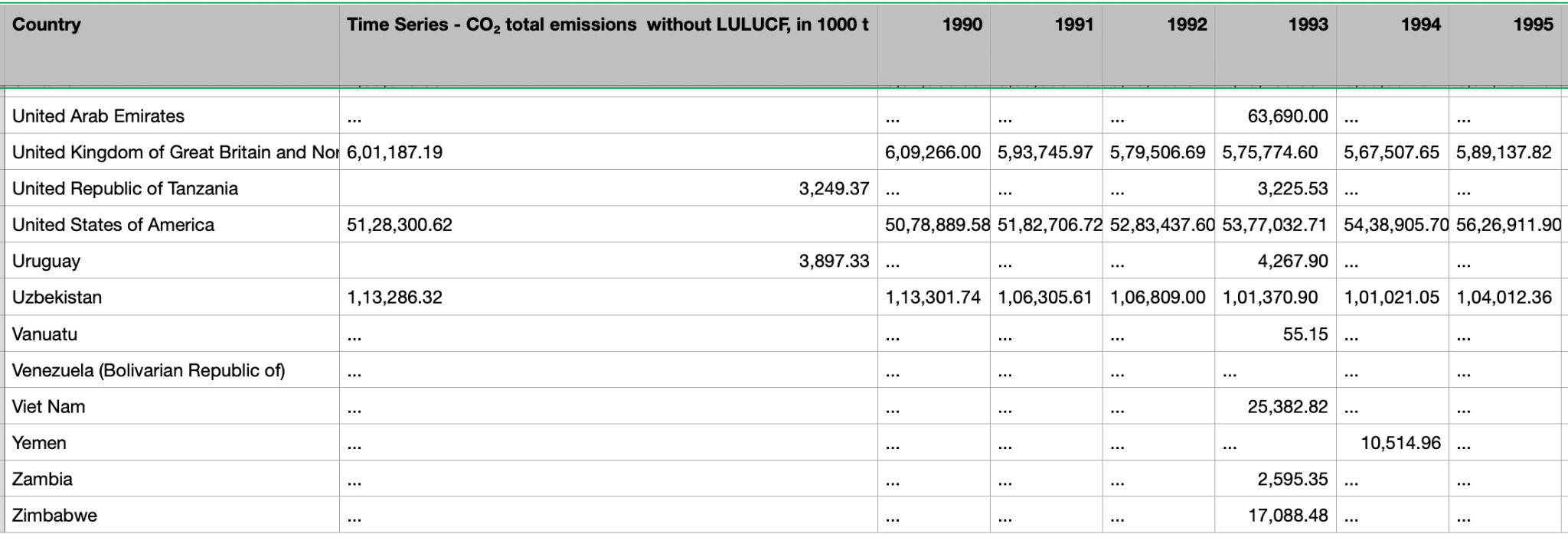
1. Report which gasses have the highest correlation with rising temperature
2. Analyze most important contributing factors to CO2 emissions
3. Determine to what effect emissions have on our average temperature
4. Predict future temperature levels by year using current temperature data
5. Analyze and predict how temperature affects sea level

We will demonstrate that CO2 emissions are causing a dangerous increase in average temperature and sea level, which will have worrisome effects in the future.

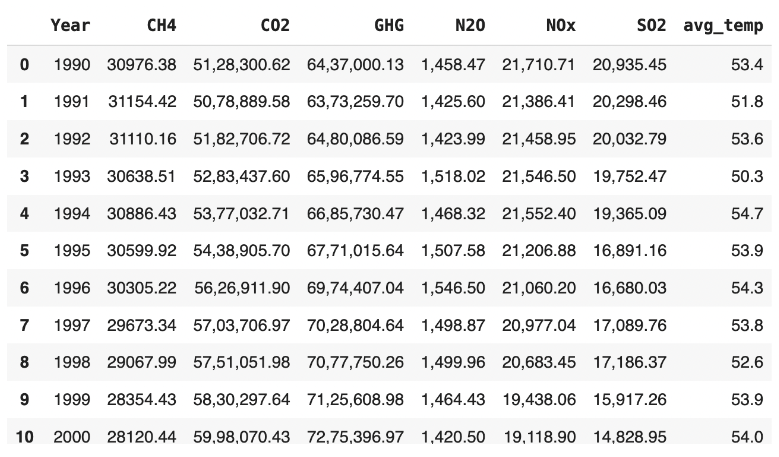
1. Gasses correlation with rising temperature:

At the beginning of our project, we wanted to determine which greenhouse gas has the greatest impact on climate change. We first integrated multiple datasets with information on different types of gas emissions and temperature. We then utilized a machine learning clustering model to determine the gasses relations to temperature and which could have the highest effect on temperature. We were able to produce a visualization of the clustering as well as

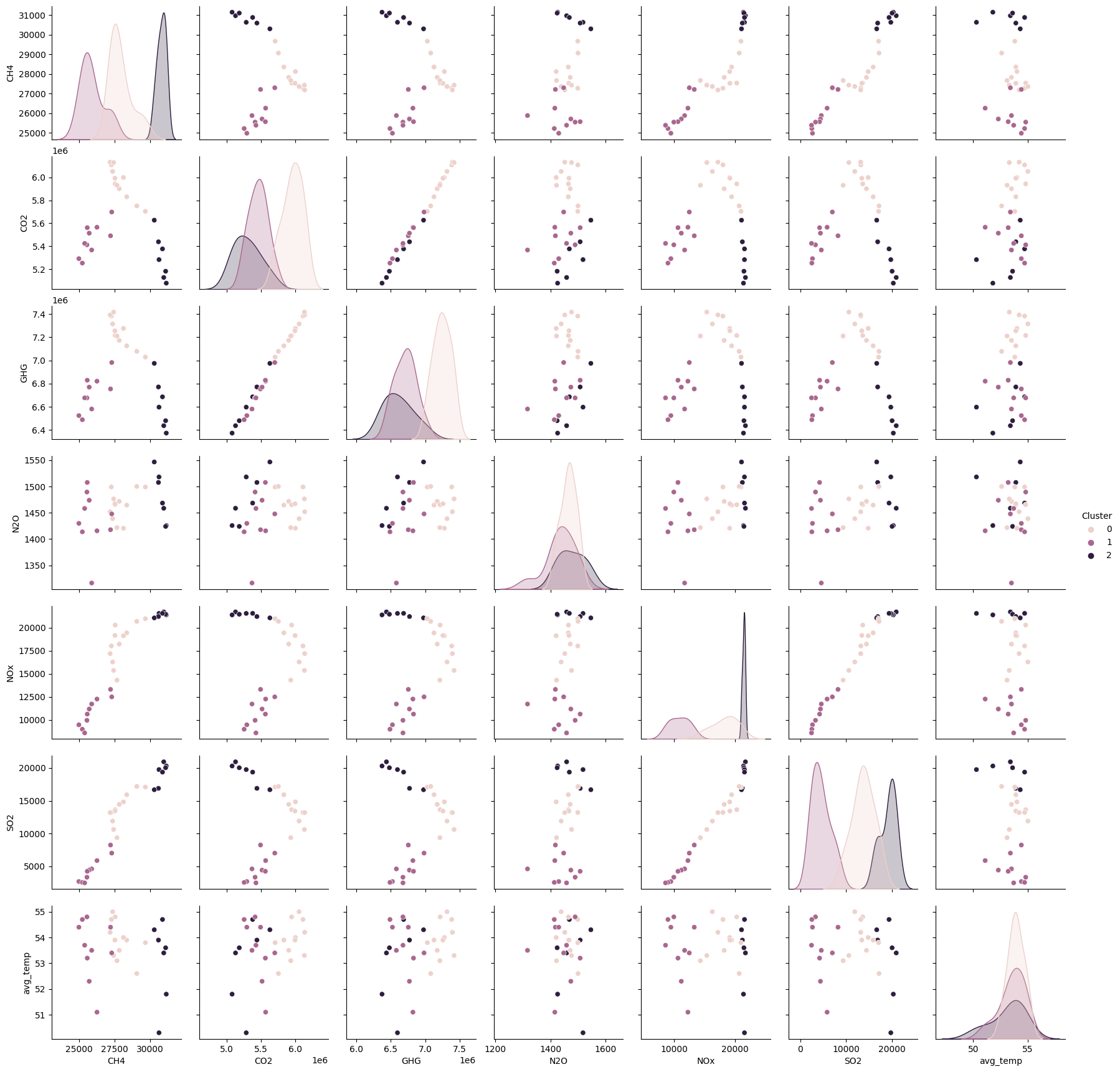
Below is one of the original datasets to be integrated for clustering. This data set had the emission amount for a specific type of gas for each country from 1990-2018. We found six different gasses that seemed significant so we would need to combine the six data sets and filter so only the US data would stay.



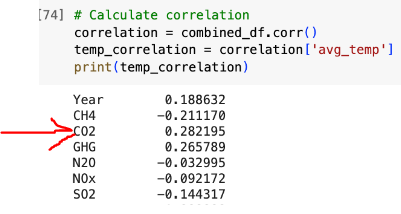
Below is the integrated dataset for US level of emissions for each gas type over time after conducting all the data preprocessing and cleaning.



Below is the pairplot of the clustering model. There is not a lot of insight to take away from this for a couple of reasons. First off there frankly isn’t enough data to draw a lot of conclusions from these clustering results. However one thing that be shown is that the United States has made strides in starting to reduce CO2 emissions. We can tell because relative to the other gas emissions, CO2 has an inverse parabola, showing how throughout the 90s CO2 was increasing, then peaked around mid 2000s, and are now making strides to lower emissions as seen by the trajectory heading downwards.



Below is correlation of emissions to temperature with the highest correlation being CO2:

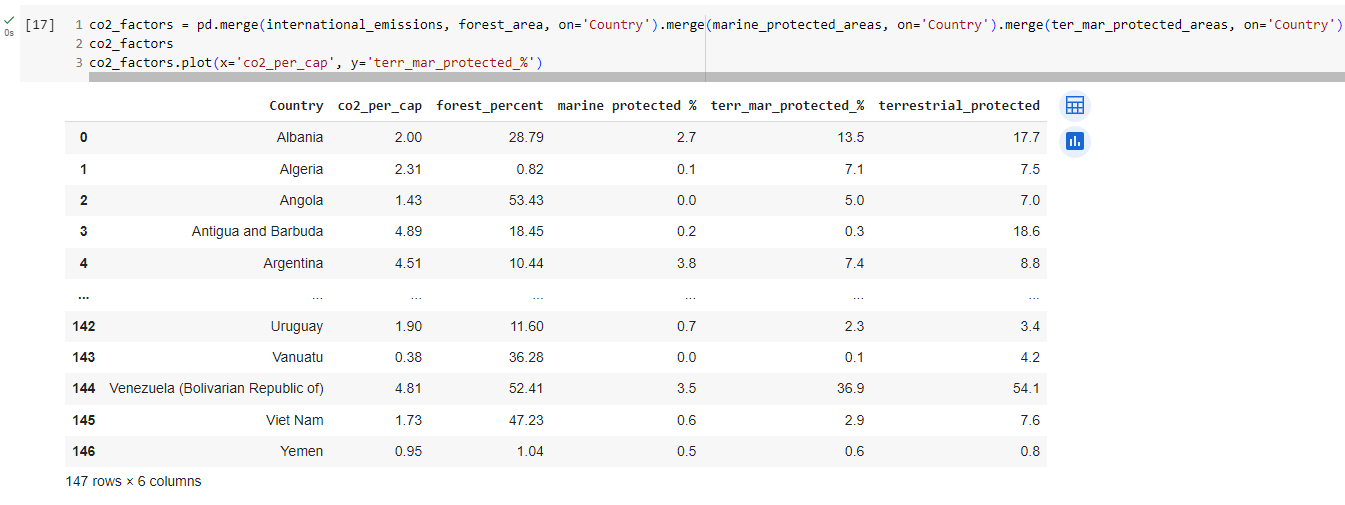


Using this correlation model, we are able to determine that CO2 has the highest correlation to temperature. We will now analyze what factors and entities cause the most CO2 emissions and analyze the correlation between CO2 emissions, average temperature, and sea level.

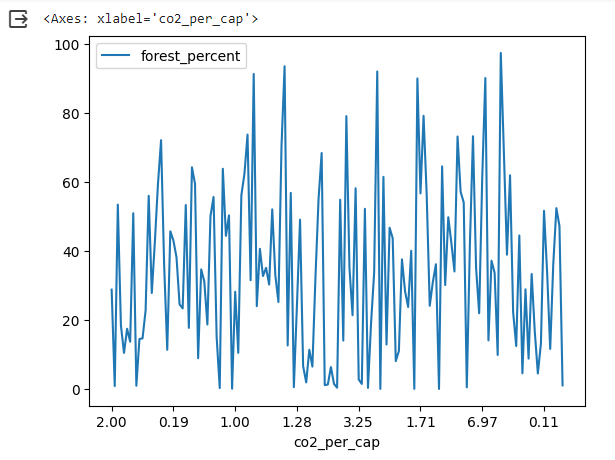
1. Analyzing contributing factors to CO2 emissions:

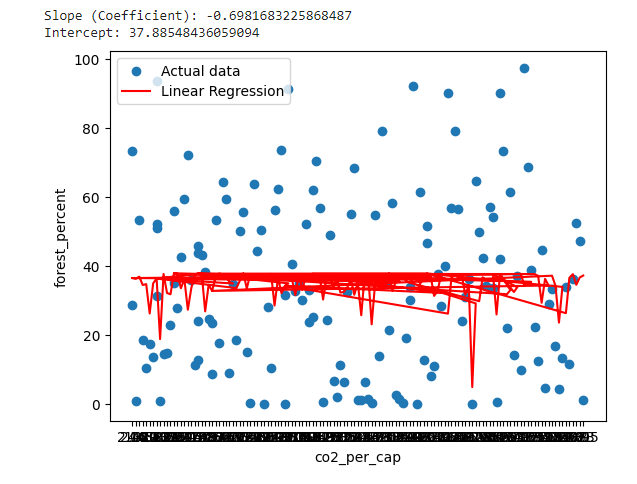
Before analyzing trends between CO2 emissions, temperature, and sea level, we wanted to identify the factors that determine how much CO2 a country emits. To do this, we found datasets that contain data on each country’s CO2 emissions, combined with other relevant data. The first dataset we integrated with CO2 was forest\_area. This dataset contains each country’s percentage of total area that is covered by forests. Our hypothesis is that the more forest area a country has, the less CO2 it will emit. The second dataset we integrated with CO2 emissions is marine\_protected\_areas. This dataset contains data on the amount of marine protected areas a country has. We predicted that the more marine protected areas a country has, the less CO2 that country will emit. This is because we believe that countries with more protected areas will likely place a heavier focus on environmentalism and clean energy, and thus emit less CO2. The last two datasets we integrated with CO2 emissions are terrestrial\_protected\_areas and terrestrial\_marine\_protected\_areas. These two datasets contain more information on each country’s protected areas. We hypothesized again that countries with higher percentages of protected land will have a lower total level of CO2 emissions.

We integrated the datasets and combined each country’s data. In our integrated dataset, each case represents one country. Each case contains 5 features: *co2\_per\_cap, forest\_percent, marine protected %, terr\_mar\_protected\_%*, and *terrestrial\_protected*. We attempted to predict co2\_per\_cap using the other 4 features. Our final integrated dataset looks like this:



We then attempted to find a correlation between any of the four variables and train a model to predict a country’s CO2 emissions based off of the other four features. Despite our ambitious goals, we ran into some challenges when trying to fit a model based on these four features. None of the features had a strong correlation with CO2 emissions, as seen in the following graphs:

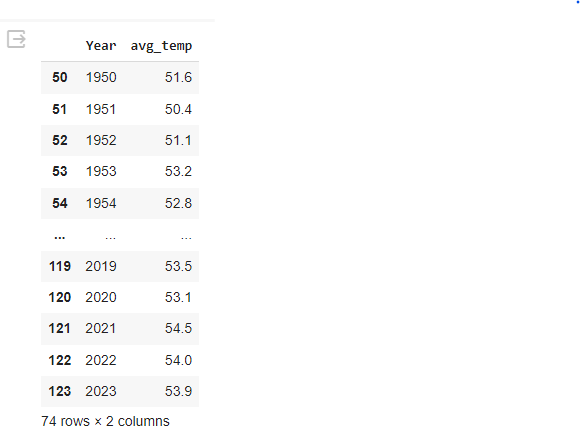




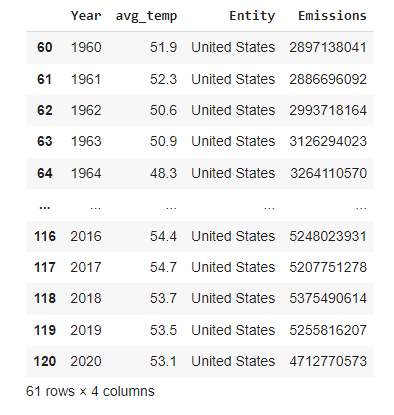
These graphs show the lack of a trend between forest space and CO2 emissions for each country. The results were similar for each of the four predictor features. Despite not finding a correlation between any of the variables and CO2 emissions for each country, we gained valuable experience integrating these 5 datasets together and learning how to determine which data is necessary for us to keep, and which data can be disregarded for the final integrated dataset. We also learned how to use features to attempt to build a linear regression model.

1. Determine how CO2 emissions affect average U.S. temperature

We then analyzed what effect CO2 emissions have on average U.S. temperature. We analyzed a CO2 emissions dataset and an average U.S. temperature dataset. The CO2 emissions dataset contained data on irrelevant countries and years, so we had to perform data cleaning to transform the data into a usable form. We then integrated the emissions dataset with the temperature dataset and produced the following results:

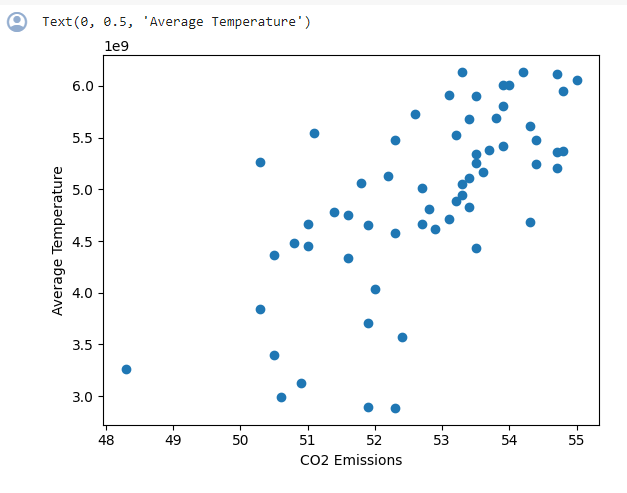
Original Dataset 1 (CO2 emissions by country) Original Dataset 2 (average U.S. temperature)



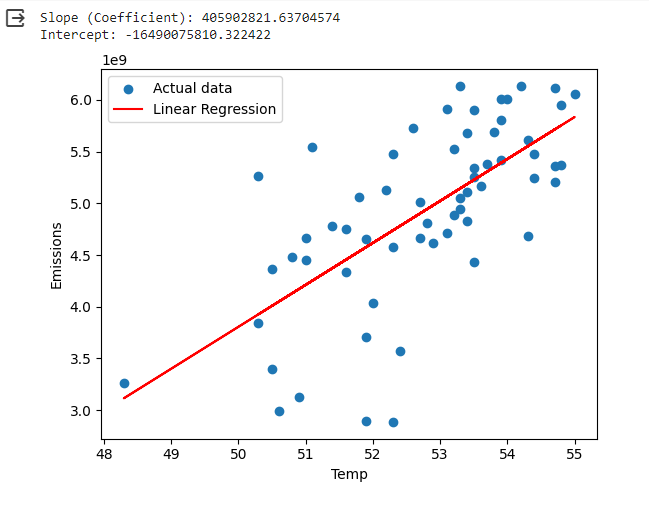
Integrated temperature/CO2 emissions dataset

After integrating dataset 1 and 2 together by year, we looked at the correlation between CO2 emissions and average temperature. Using linear regression, we found that carbon emissions and average temperature have a positive correlation.

Original scatter plot for the carbon emissions:



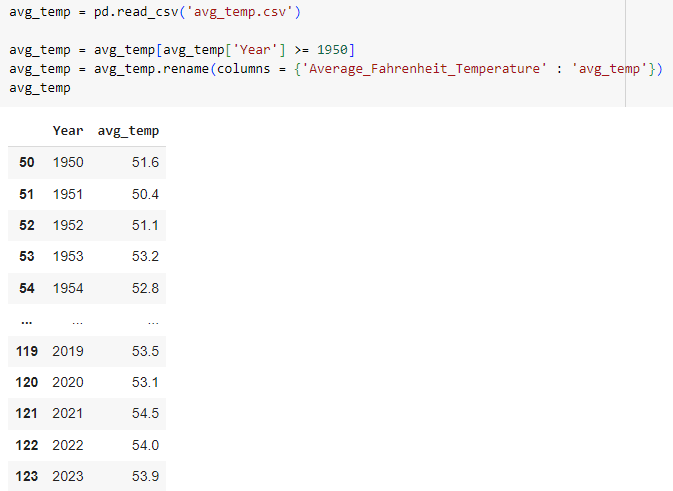
Scatter plot with regression line:



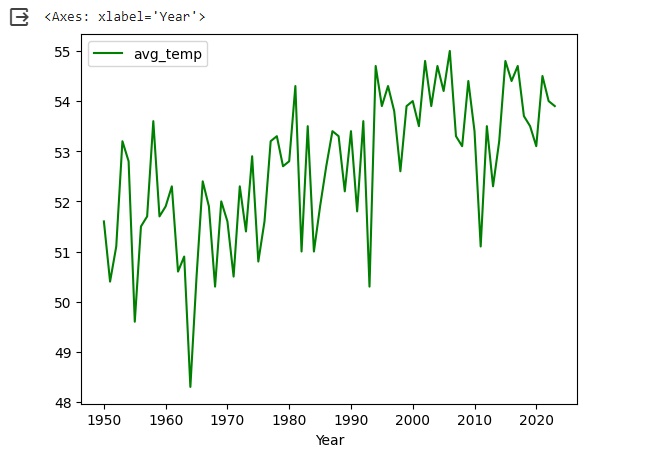
Result: We found through data integration that an increase in carbon emissions lead to an increase in average temperature as can be seen with the regression line which indicates a positive correlation.

1. Predict future temperature levels by year using current temperature data

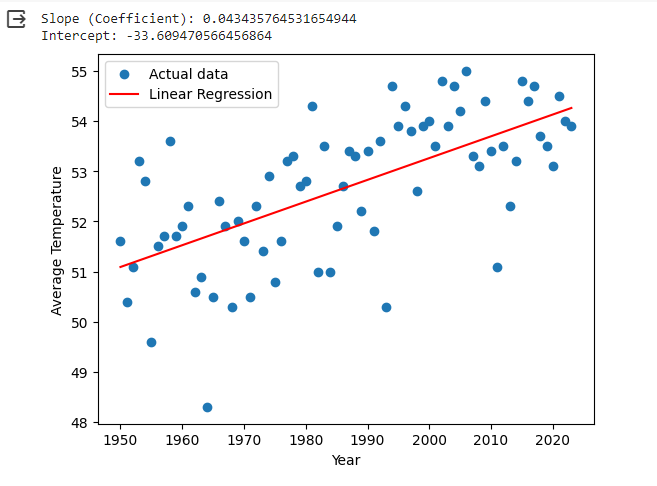
For this dataset, no integration was needed to be done since the dataset already contains all the necessary information. Each case in this dataset represents the average temperature for a given year.



To begin our exploratory data analysis on the temperature dataset, we plotted average U.S. temperature since 1950 with respect to time. The graph shows a positive correlation between year and temperature.



We plotted a regression line to further demonstrate the positive correlation:



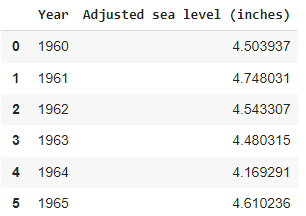
Y-intercept = -33.609470566456864

Slope = 0.043435764531654944

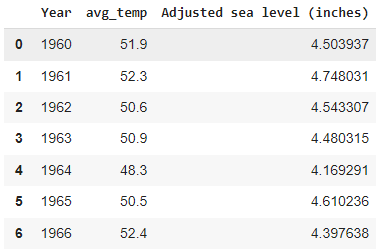
Result: We find that the regression line has a moderate positive correlation which indicates that there seems to be a trend in temperature rising throughout the years. We were then able to predict that in 2100, the temperature will be 57.6°F. We will use this predicted value in our next section.

1. Analyze and predict how temperature affects sea level

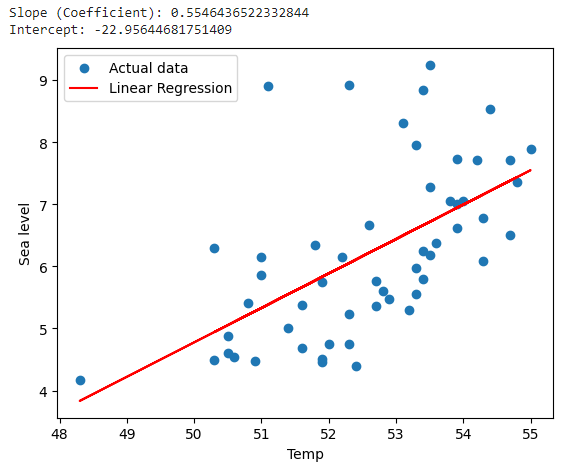
Sea level dataset:



Integrated sea level and temperature dataset:



We then created a regression model of how temperature affects sea level:



We recall that our linear regression model predicted the U.S. temperature to be 57.6°F in 2100. Using our temperature/sea level model, we can predict sea level when given a temperature. If temperatures do rise to 57.6°F as we have predicted, global sea levels will rise to 9 meters, which would start to flood and destroy coastal cities. Coastal ecosystems would be destroyed, causng animals to be displaced from their natural habitats. The effects of climate change will be wide-ranging and devastating to our planet and civilization, which is why we must begin to plan for the future now by reducing CO2 emissions.

Contributions:

Travis Marchok: Integrated datasets in part 2, performed exploratory data analysis to determine trends between each of the variables. Integrated and analyzed datasets in part 3; plotted the CO2 emissions with average temperature data, then fit a linear regression model to predict future temperature if CO2 emissions continue to increase.

Adam Whiteside: Integrated Sea Level and Temperature datasets; plotted sea level based on temperature; created linear regression model for sea level based on temperature; made calculations for future sea level predictions; documented screenshots and findings in report and presentation

Eli Aaron: Integrated and transformed 6 different gas emission data sets utilizing filters, melting, outer joins, and left joins in order to cluster and analyze which emissions had the most impact on US temperatures.

Chylle Gamino: Helped analyze data that was integrated by most of the team and wrote reports for part 3 and 4. Uploaded project to the Github repository.

Git-Hub link:

<https://github.com/Chylle-gamino/DS320Project>

Sources:

<https://www.kaggle.com/datasets/programmerrdai/co-and-greenhouse-gas-emissions/>

<https://www.kaggle.com/datasets/somesh24/sea-level-change>

<https://www.epa.gov/climate-indicators/climate-change-indicators-sea-level>

<https://www.kaggle.com/datasets/giabchnguyn/average-temperature-from-1900-to-2023>

<https://deepchecks.com/free-climate-environment-datasets/> (# 9 Global Environmental Indicators)