Module 8 Option #1: R Portfolio Project

Eric Schmidt

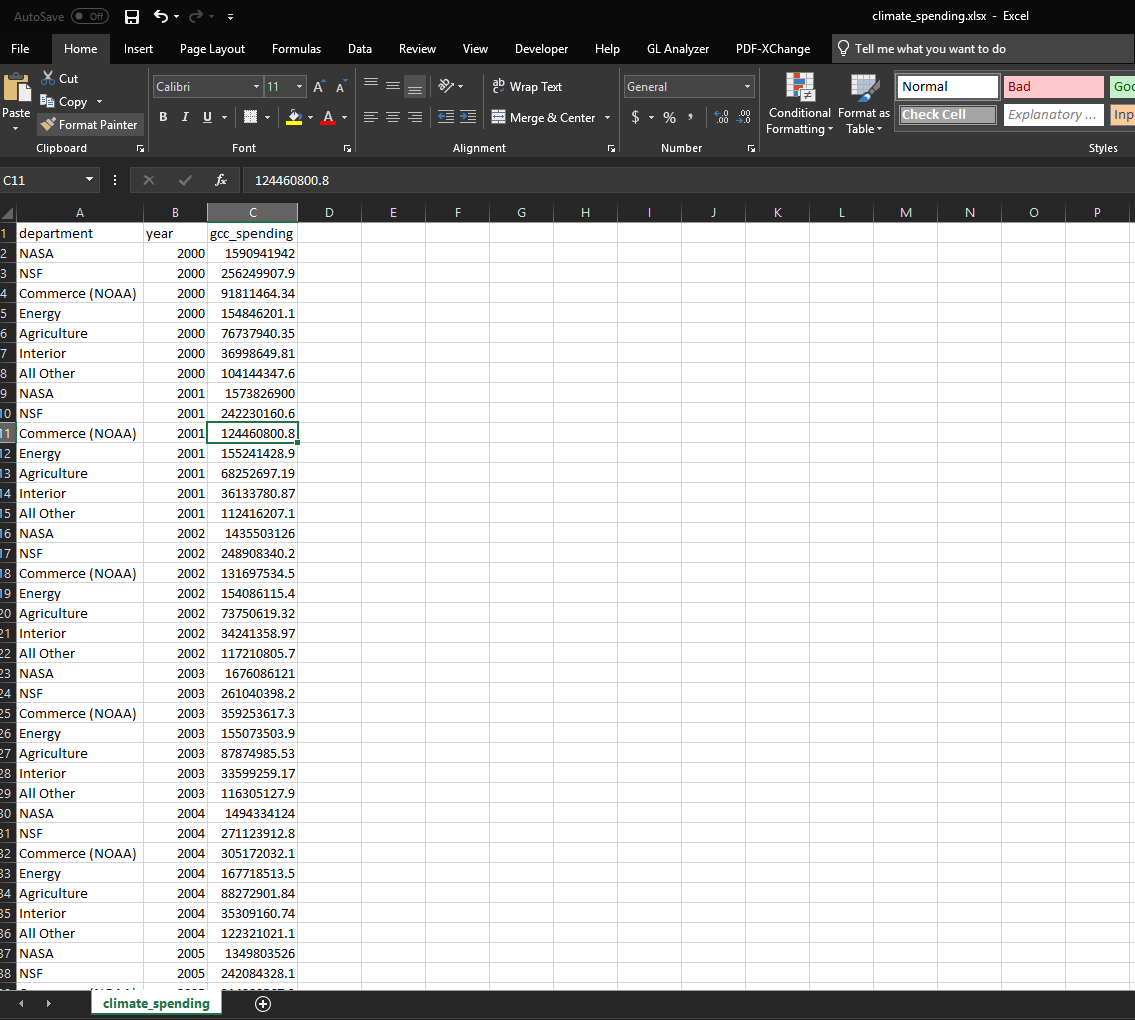
Colorado State University – Global Campus

For this portfolio project, I wanted to expand upon my project milestone and compare data from the US government budgets related to climate change and the potential impact it has on the CO2 emissions we emit as a country. Discussions related to climate change and what we can do as a country are integral to save the Earth before it gets to an irreparable state. Despite this, many news sources and information found online statistics that we are spending less money towards these solutions and therefore not preventing further damage to the Earth. For this project, my hypothesis is that; spending less on climate change and programs that are responsible for managing and lowering our CO2 emission will cause an increase of the greenhouse gas carbon dioxide.

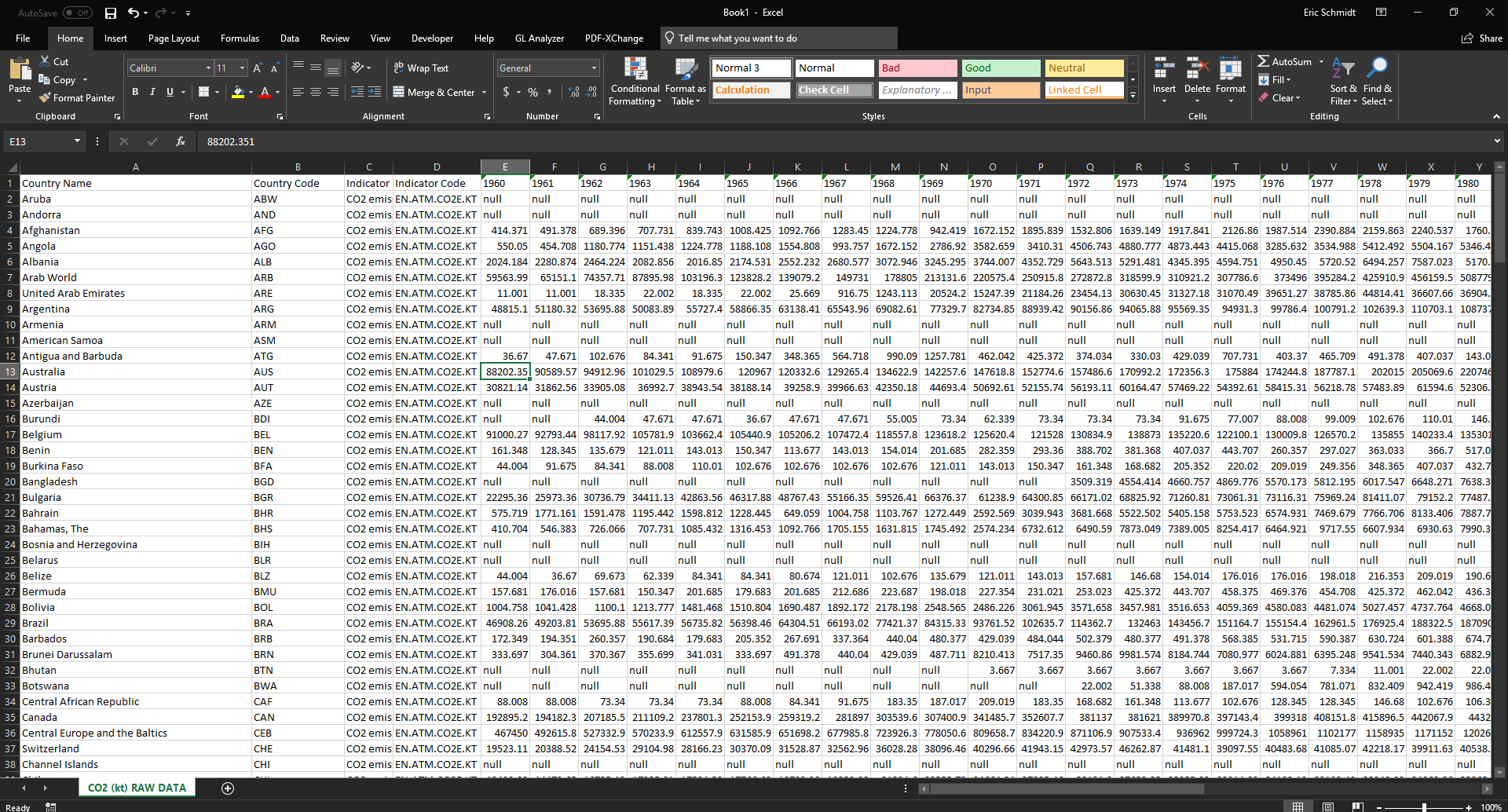
Null Hypothesis: There is little to no relationship between the Department of Energy’s budget and the amount of CO2 gas emitted into the atmosphere.

Alternative Hypothesis: There is a direct relationship between the Department of Energy’s budget and the amount of CO2 gas emitted into the atmosphere.

To start the investigation related to this inquiry, I contacted individuals from USA.gov for any database information they could provide in relation to the governmental budget for the Department of Energy (DOE). The DOE has sectors which are responsible for developing and managing solutions for better climate control. I was able to retrieve data of the governmental budget that was broken out into departments from the years of 2000 to 2017.

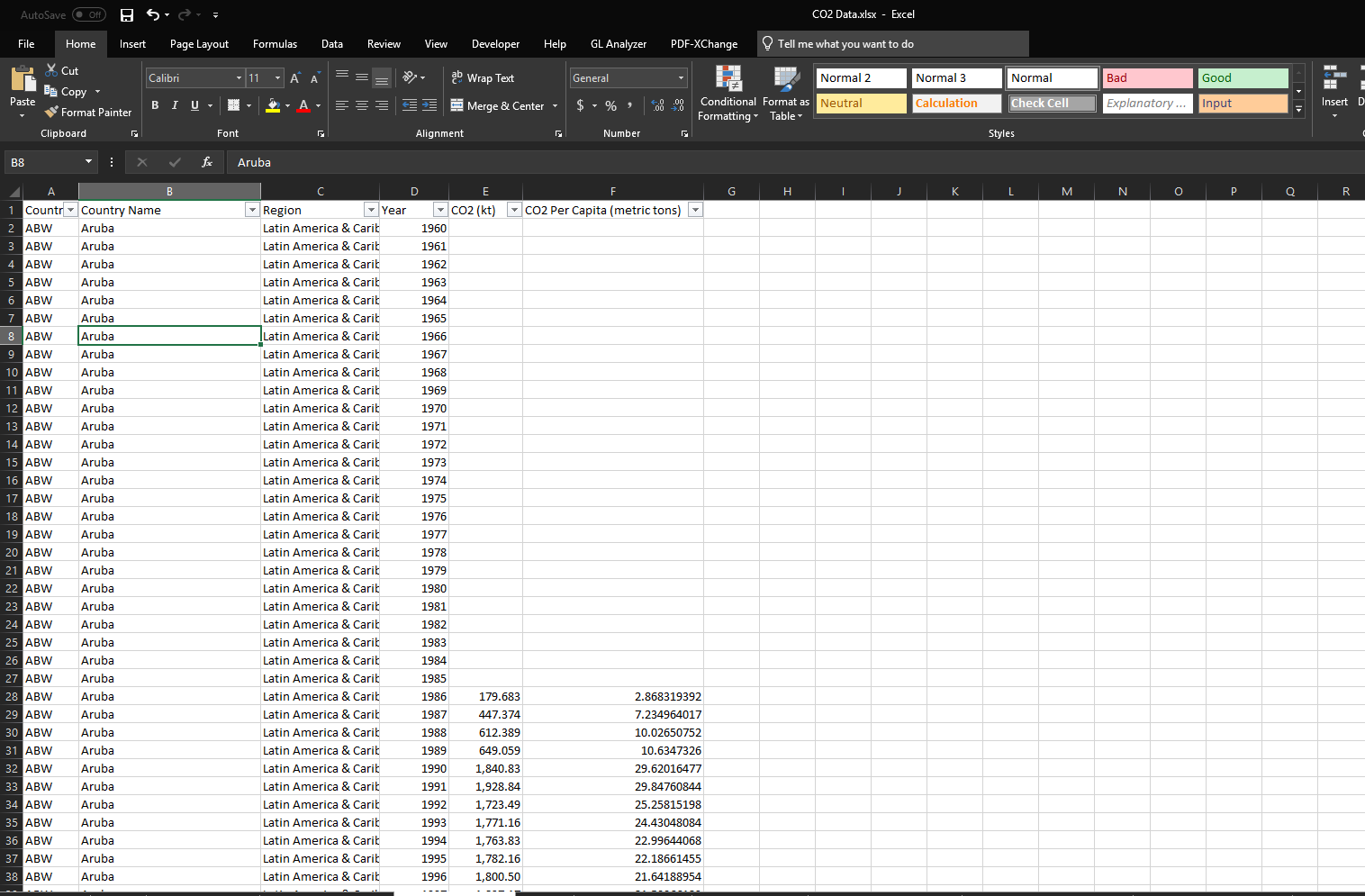
**Raw USA.gov Budget Data** 

To compare this to our country’s impact on the environment, I retrieved a dataset from the TidyTuesday’s public GitHub repository that contained information of CO2 gas emission per country from 1960 – 2011.

**Raw CO2 Data pulled from TidyTuesday’s GitHub** 

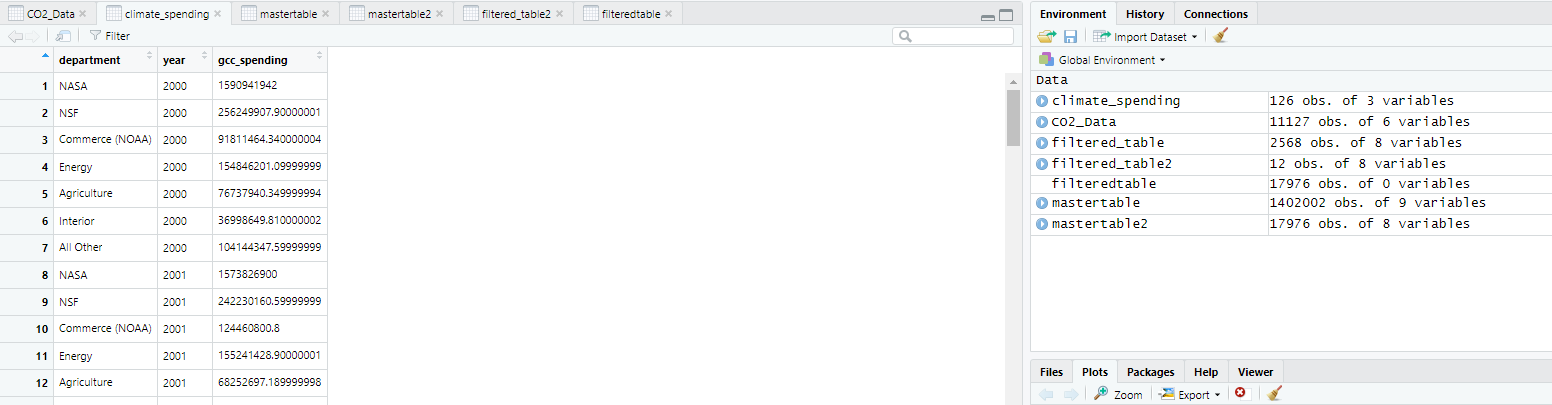
The CO2 data needed to be cleaned up before it was processed (which I performed in Excel directly) so I can better utilize the data for my testing. To cleanse the data, I joined a country with its respective year’s data, and removed null values.

**Cleansed CO2 data**

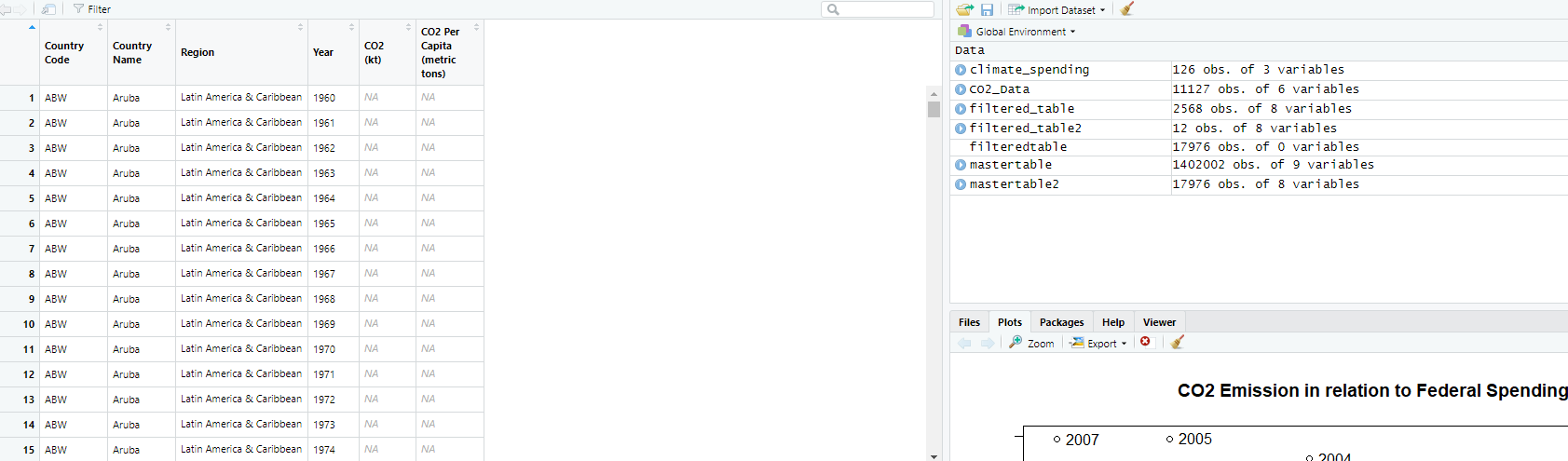


After data cleansing, I entered both files into RStudio and joined the two data files

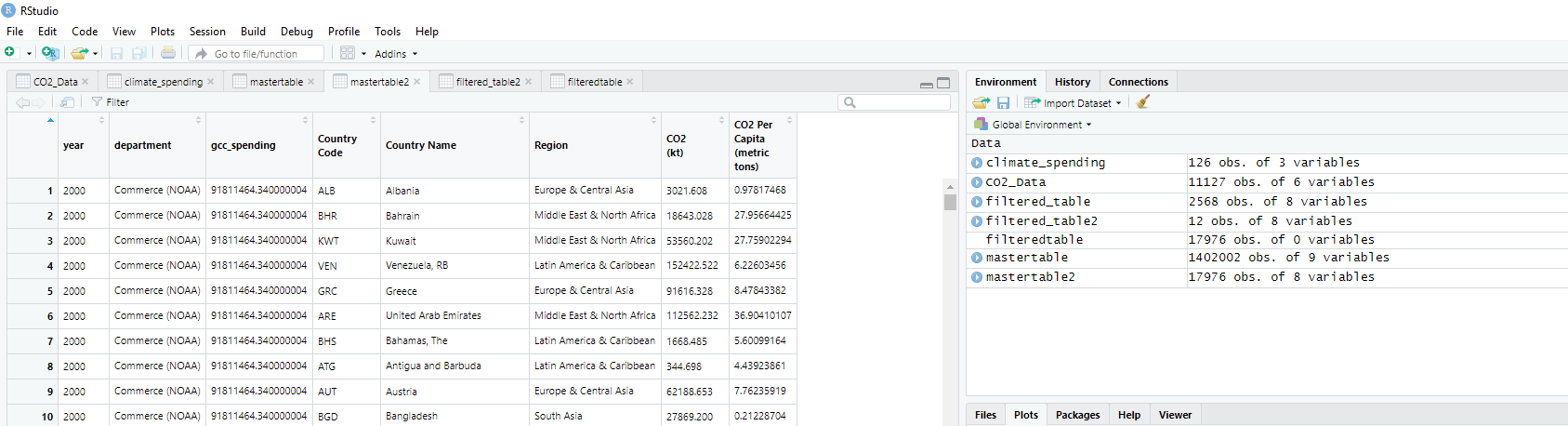
**Climate Spending Data Table in RStudio**



**CO2 Emission Data Table in RStudio**



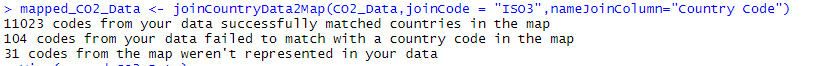
**Full Joined (on the Year column) Dataset in RStudio**



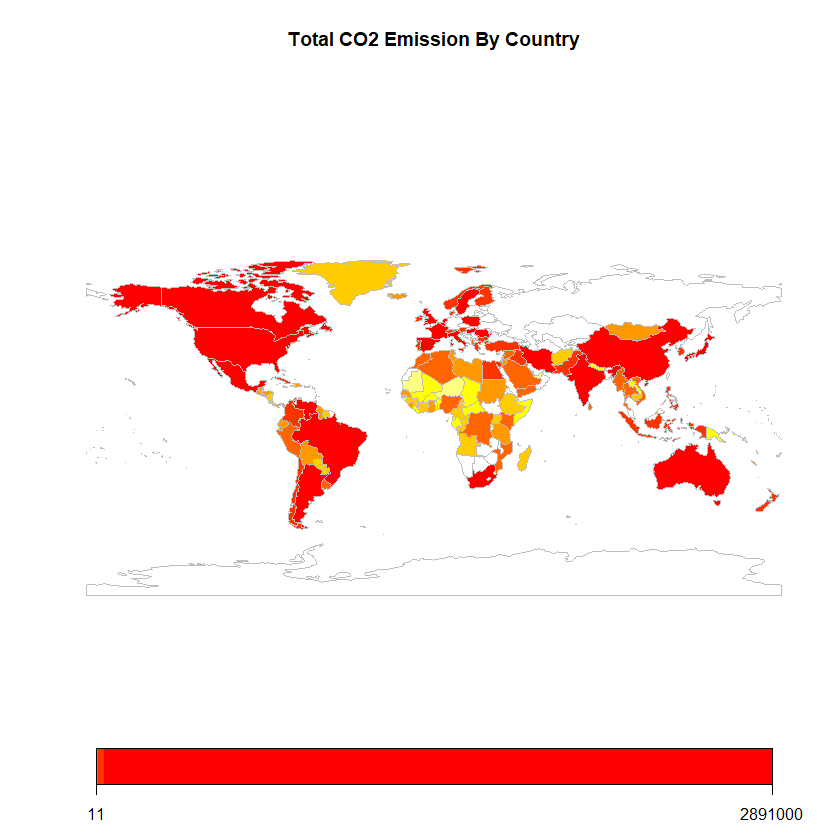
I first wanted to how the US is in comparison to other countries around the world in terms of CO2 emission. To do this, I installed the rworldmap package, which allows me to map data according to county codes.



After this was installed, I put the CO2 emission data in a mappable table, which was joined based on the ISO3 country codification that was part of my original table.

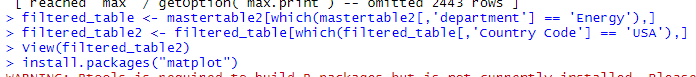


Once I had a mapped CO2 emission table, I used the mapCountryData function from the rworldmap library to map the data for all the country data I had available.



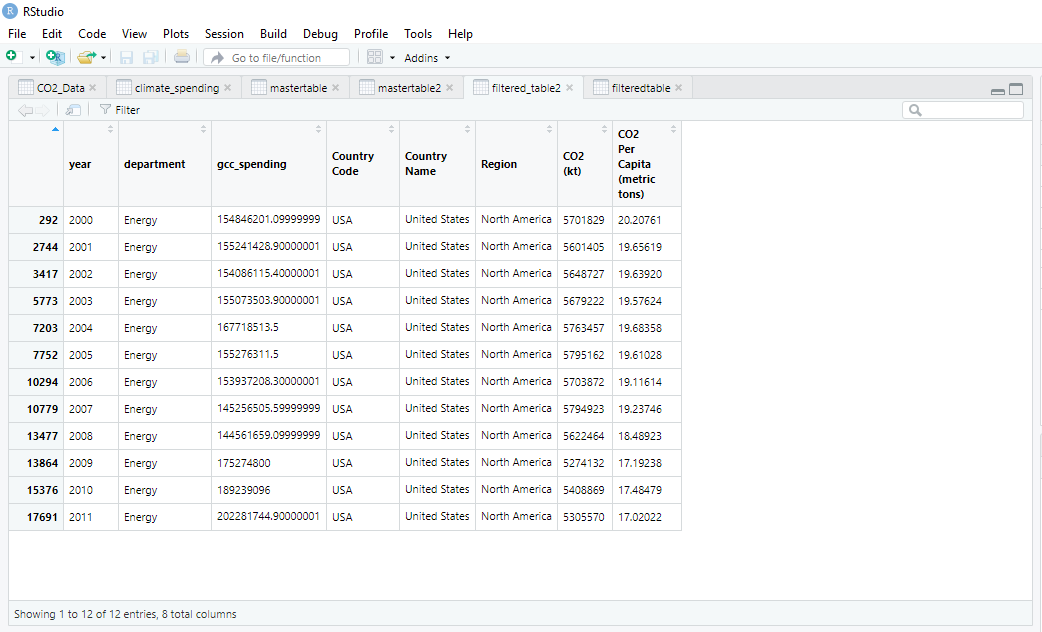
When looking at the plotted graph, I noticed the legend was difficult to understand, but realized this was due to it compounding all the data for all years for each country. I could not figure out a way to visualize the data over a time period, but the graphic did support my original hypothesis that the US is a major pollutor when it comes to CO2 emission.

After mapping the CO2 emission data, I joined the two datasets. I then filtered the data for the entries I wanted to analyze. These were budget data for the Department of Energy and the CO2 emission data for the USA.



The filtering listed above resulted in the following table of data that can be used for my testing:

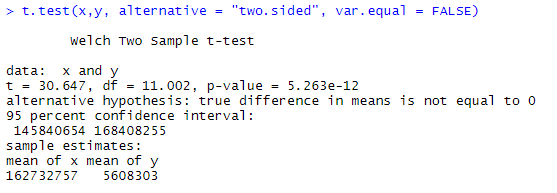
**Filtered Table for US CO2 Emission and Budget Information**



For my hypothesis testing, I first set my variables x and y equal to gcc\_spending (governmental budgeted value) and CO2 emissions measured in kilotons respectfully.

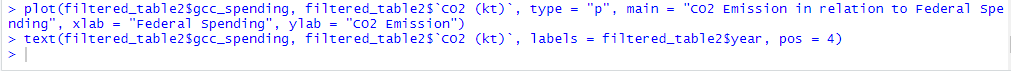


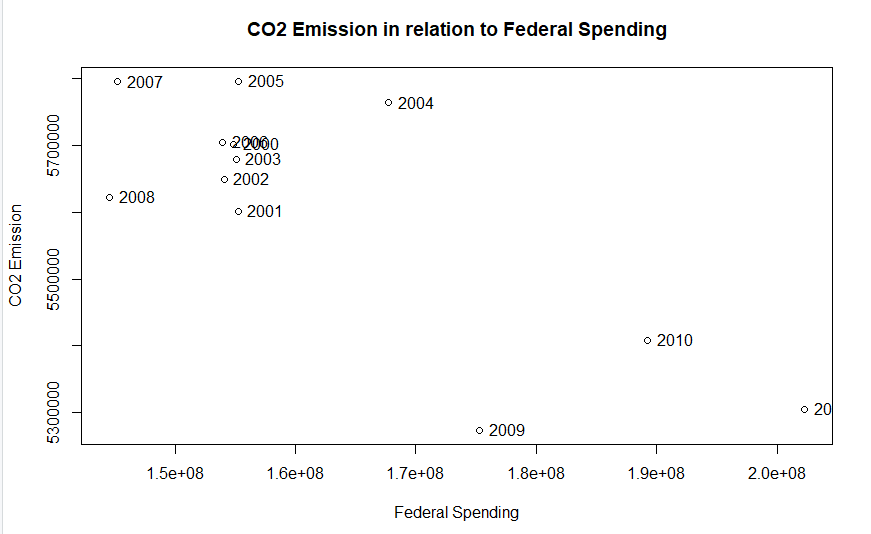
After setting my variables in RStudio, I ran a two sided T-Test to see the correlation.



Because of this test, I received a p-value very close to 0 (5.263 x 10-12). As a result, I can reject the null hypothesis, which states there was no relationship between the Department of Energy’s budget and CO2 emissions. Once I have rejected the null hypothesis, I wanted to see what type of relationship exists between these variables.

From this master data table, I plotted the data with the X-axis being Federal Spending and the Y-Axis being CO2 emission. I also added labels to each dataplot to see when each of these plots occurred.





Some of the data points either overlapped or went offscreen due to the scaling of the axes. Reflecting on the data shown, approaching 2011 (the rightmost datapoint that was slightly cut off) you see as federal spending within the Department of Energy increases CO2 emissions (measured in Kilotons) declined. Therefore, supporting my initial hypothesis that an increase of federal spending has an inverse relationship with CO2 emissions.

When looking through the data shown above, even while it supported my original hypothesis, I was surprised at how significantly the data would support my hypothesis. The change from 2008 to 2009 was a significant change, as federal spending increased by roughly $30M decreased CO2 emissions nearly 300,000 Kilotons.

To further study the relationship between the amount spent towards the Department of Energy and the CO2 emissions, I would have liked to see an increased amount of data related to budgets to get a wider range of information to test from. I inquired the number listed within the Department of Energy but was met with an automated response that was unable to assist me.

References:

<https://github.com/rfordatascience/tidytuesday/tree/master/data/2019> For CO2 emissions data:



<https://www.usa.gov/chat> For Budget data:



Chakravarty, S., Chikkatur, A., De Coninck, H., Pacala, S., Socolow, R., & Tavoni, M. (2009). Sharing global CO2 emission reductions among one billion high emitters. *Proceedings of the National Academy of Sciences*, *106*(29), 11884-11888.