**RDB Summary Notes**

* I really would have liked to have a data table presented with my geo map. I have the dataframe and can even format the dataframe. The only way seems to ‘cheat’ and use a screenshot of the dataframe and paste into PowerPoint, or to go to an external application like Word or Excel. Real world uses tables all the time so I can’t imagine this is impossible in Python, or Word/Excel would be used.
* I could not find a clean way to output a numeric table, like a plot. There is a way through matplotlib to create a table can be saved as a png file. Unfortunately editing that table is unwieldy. Table looks terrible.
* Some datasets had the same data (eg. ‘time vs. population’) that would have equivalent values. Whereas some datasets had quite different values. Emitted carbon dioxide would have very different values depending where one looked, and what that dataset was being used to convey.
* Turns out emitted carbon dioxide (CO2) can be defined as manufactured, excluding land use only, or even total greenhouse gas equivalents (in CO2). Cement has it’s own CO2 emissions just because of how much CO2 (equivalents) is generated. Even transportation datasets have claimed total CO2 emissions.
* Choosing the project and parameters was difficult. There were a number of issues encountered as the project targets kept moving.
  + Getting agreement on the project title and scope
  + Finding relatable datasets to a common theme
  + Finding datasets that could fit our dataframe requirements, or be transformed within reason, without a great many transforms (pivot, unpivot, transpose, remove and merge)
  + Meeting the group requirements as well. Were the group interests addressed (financial, environmental, health data to work with)
  + Getting any common datasets into matching column titles/headers (renaming)
* Group communication was varied. Only communication tool that was NOT used was the phone. Used Slack, email, Zoom, Git.
* I used an Excel generated, CSV exported geo coordinates file for the geo maps. I started with the Citipi library, and tried the Google API for geo location calls. Although all of those still require a starting list of locations.
* The Google API appears challenging to use. I didn’t get far. There is no generic pulling data API which seems odd.
* Several columns of data were added after mathematical operations. Per capita columns.
* I could have created the geo locations in the Jupyter notebook but wanted to try a file merge to add to my original data CSV file. Seemed to work fine.
* Asked Jason for assistance with the geo map. Initial plots worked fine, but the size parameter seemed to make the plot look like the ocean. All blue when using the population data. Jason correctly identified the values were way to high, and scaling needed to be applied. Which worked.
* I created a variety of geo maps, but THEY ARE NOT RELATED TO EACH OTHER. The maps only directly relate to each other on each plot. All the maps have different scales (GDP 0.00001 and Cement CO2 per Capita is 40).
* The saving plot output as an image (plt.savefig) for the geo maps, using matplotlib, result in blank white images. There is nothing. However, with Jason’s approval, the geo maps do have a save icon that creates bokeh\_plot.png in the Downloads directory.
* The entire Jupyter script seems to run OK in either PythonData or Python3 kernal. It seems the script works without importing the CSV dependencies.
* Some of the datasets encountered had little to no information. Some of the datasets would reference an external file to describe the datasets. Might have been done to simplify importing CSV datasets?
* Not only did some column names not match (country, Country, Country name), but the units of measure were different.
  + Tons, metric tons, pounds, kilograms
  + Dollars, euros
  + Russia, Russian Federation; China, Mainland China, Republic of China – Mainland and Hong Kong
* Because there was so much freedom with the project, a great deal of time was spent brainstorming, starting and stopping, reversing and changing course. The timing would be different if a group of workers is given a defined (or parameters to define) project to work on. Then the members can split up the duties up front and continue on for the project.

Watching all the presentations, I was surprised at the wide variety of topics, and how the topics were presented. It appeared some groups assigned various tasks to group members (data reduction, plotting, analyzing) while some, like ours, took a sub-topic and did those tasks as well. We may have been overly redundant.

Good planning and task comprehension would probably speed things up greatly.

There is a lot of data to find, but it may not fit the project without some additional work. A round peg in a smaller square hole. How much effort (data manipulation) should be spent getting the peg into the hole?

I thought the team did well together. Most of my time was spent with Rachel, working on the environmental aspect of the project – which makes sense. Interests representing each group member were found and utilized. A well rounded project.

Robert Bentz

**Outline:**

**Global Corporate Environmental Impact??**

**Team Members:** Rachel Le Grand, Arunkumar Sridharan, Steffi Yang, Robert Bentz, Jack Lowry

**Project Description/Outline:**

Environmental, Health and Fiscal Impacts by specific industries.(sic codes)

**Research Questions To Be Answered:**

1. Environmental - Industries with highest/lowest environmental impact?
2. **Variance by Industry code. Correlation by certain industries?**
3. Countries with Highest/lowest environmental impact?
4. Fiscal - GDP By Country equal to higher/lower environmental impact?
5. **Company Level – Operating Income**
6. Population – Higher population equal higher environmental impact?
7. Lower Population equal lower impact?
8. Health – Avg Life Expectancy for countries with high/low environmental impact?
9. **Disease, rare outliers by age in countries.**
10. Timeline – TBD once data is plotted. Possible range of 2010-2018.

\*\*5-10 Countries, compare by different continents/geographic location for environmental effect.

\*\*

**Datasets To Be Used:**

**-** [Corporate Environmental Impact | Kaggle](https://www.kaggle.com/datasets/mannmann2/corporate-environmental-impact)

**-** [Global CO2 Emissions | Kaggle](https://www.kaggle.com/datasets/prashant808/co2-emissions?select=_COE28282_Emissions2C_Emissions_Intensities2C_and_Emissions_Multipliers.csv)

- [GDP (current US$) | Data (worldbank.org)](https://data.worldbank.org/indicator/NY.GDP.MKTP.CD)

- [Population - The World Factbook (cia.gov)](https://www.cia.gov/the-world-factbook/field/population/country-comparison)

**Rough Breakdown of Tasks:**

**Fiscal Focus** – Arunkumar

**HealthCare Focus**– Steffi, Jack

**Environmental Focus** – Rachel, Robert

**Visualizations – 10 Countries – 2010 - 2018**

1. Environmental – Co2 Emissions
2. Populations of countries
3. GDP
4. Average Life Expectancy
5. Mortality Rate & Child Mortality Rate