CSC110 Project Written Report:
Greed or Need? The Relationship between Money and Climate Change

Jia Hao Choo, Komal Saini

Monday, January 4, 2021

Problem Description and Research Question

This month, Nestlé announced their plan to half their emissions by 2030, and achieve net-zero emissions by 2050 ("Focusing on regenerative agriculture and moving to renewable electricity, Nestlé redoubles efforts to combat climate change"). In 2019, Nike announced multiple "Move to Zero" initiatives, such as powering their facilities with 100% renewable energy by 2050 and reducing their carbon emissions by 30% by 2030 ("What to Know About Nike's Stance on Tackling Climate Change"). Though some of the world's leading corporations have committed to combating climate change, its effects continue to accelerate over time. This got us thinking—are certain countries to blame for the repercussions faced by others on a global scale? In our project, we are interested in understanding the discrepancies in contributions to global climate change within countries based on their position on the income hierarchy. In our project, we are interested in understanding the discrepancies in contributions to global climate change within countries based on their income classification. There are 4 income classifications, as specified by the World Bank: High, Upper-middle, Lower-middle, and Low (Prydz and Wadhwa). Each country is placed into one category based on their GNI per capita, according to the table below (Prydz and Wadhwa):

Classification	GNI per capita
Low Income	$1025 or less
Lower-middle Income	between $1,026 and $3,995
Upper-middle Income	between $3,996 and $12,375
High Income	$12,376 or more

When trying to better understand climate change as a whole, we adopted the Socratic method and tried to dig at its foundation—attempting to determine whether the underlying factor was money. We discovered that the processes that release emissions and contribute to climate change are the very processes that drive economic growth and prosperity. As a result, we are interested in understanding whether the specific industries producing the most emissions are also generating the most revenue; interesting conclusions can be drawn—both from an economic and ethical standpoint—regardless of the result. We have chosen to focus on revenue instead of profit because revenue encompasses the expenses. We want to ensure that we take the expenses into account; they may also be indicative of processes that contribute to climate change. For example, high expenses for maintaining furnaces may be due to the fact that they are used very frequently, therefore contributing to the release of pollutants and toxic emissions into the atmosphere.

As a result, the research question that we are posing is: **In relation to the income classifications, what is the relationship between the revenue generated by different sectors in countries worldwide and their contributions to greenhouse gas emissions?**. The specific sectors that we intend to focus on are industry, manufacturing, and agriculture. Industry refers to the industrial processes, including constructions, electricity, water, gas and mining, while manufacturing refers to the processes of producing a product in factories, such as casting and moulding (World Bank). Agriculture refers to the cultivation of livestock, which includes fishing and farming (World Bank). We have chosen these three sectors because they are the sectors that form the basis of human civilization in the sense that every country leverages these sectors to develop and grow. These three sectors also have the longest history in modern human civilization, as opposed to the technology sector which only began to grow in recent years. In studying these sectors, we can gain a more thorough understanding through using a greater timeline to analyze how emissions from these sectors correlate to the revenue that they generate.

From a social science standpoint, we hope to better understand what motivates our actions as human beings—are we willing to sacrifice our planet's health, and ultimately, the health of future generations, for financial gain? Are growth targets for emissions preventing economic development in low-income countries more than they for countries in other income groups? There are a multitude of questions that our research may reveal answers to. In doing so, we hope to highlight areas where there is potential for change and advancement; perhaps, there are industries that are not generating substantial revenue while also exacerbating climate change.

As students interested in innovation and global problem-solving, both of us are incredibly excited to work through this project and analyze our results. Although our research product may not be directly translated into the real world, we know that the insights that we gain will inspire and inform our future actions.

Dataset Description

We will be using a total of 5 datasets, separated into 3 categories:

0.1 Value (Revenue) of each sector:

There will be 3 different datasets in this category: one for manufacturing, one for industry and one for agriculture. Each dataset will contain the values for the respective sector from the year 1990 to 2016. All datasets will be in csv format. The source of all three datasets are World Bank Data. Below are the specific links to each dataset.

Industry: https://data.worldbank.org/indicator/NV.IND.TOTL.CD

Manufacturing: https://data.worldbank.org/indicator/NV.IND.MANF.CD

Agriculture: https://data.worldbank.org/indicator/NV.AGR.TOTL.CD

For all three datasets, there will be 30 columns. The first three columns are Country Name, Country Code and Indicator Name. The Indicator Name is the same in each file, for example in the agriculture dataset, the Indicator Name for all rows will be "Agriculture Value (US$)". The other 27 columns contain the value of each sector (in US$) for the year 1990 to 2016. A sample agriculture data looks like this (for the first 6 columns):

Country Name	Country Code	Indicator Name	1990	1991	1992
Brazil	BRA	Agriculture Value (US$)	31751335800	40927000000	27160437500

In the project, column 1: Country Name and column 3: Indicator Name from all three datasets in this category will not be used.

0.2 Country Metadata

This dataset specifies the Country Code used for each country, its region and its income classification. This dataset is also in csv format, and this is extracted from the same World Bank Data used in the previous type of dataset.

There are 4 columns in this dataset: Country Code, Region, Income Group, Country Name. The country code specified in this dataset will be constant and used across all datasets, and will be what we use to identify different countries, as the country names are represented slightly differently between different datasets (e.g. South Korea is written as “Korea, Rep.” in some datasets, but its code “KOR” is the same across all datasets).

A sample data looks like this:

Country Code	Region	IncomeGroup	Country Name
AFG	South Asia	Low Income	Afghanistan

In this project, column 2: Region will be omitted.

0.3 Emissions Data

There will only be one collective dataset containing the greenhouse gas emission data for all 3 sectors. This dataset is also in csv format. The source of this dataset is from an organisation called Climate Watch Data. Here is the link to the original datafile: https://www.climatewatchdata.org/data-explorer/historical-emissions

There are 31 columns in this dataset. The first 4 columns contain the basic information: Country Name, Country Code, Sector, and Unit. The Unit column is the same for each row of data, as the greenhouse gas emissions are measured in $MtCO_2e$ (metric tonnes of carbon dioxide equivalents). The sector column contains one of the three possible information: “Agriculture”, “Manufacturing/Construction”, or “Industrial Processes.” The last 27 columns contain the emissions data from 1990 - 2016, but the data is arranged in reverse order. So, 2016 will be the fifth column, and 1990 will be at the last column.

A sample data looks something like this (for the first 6 columns):

Country Name	Country Code	Sector	Unit	2016	2015
Angola	AGO	Industrial Processes	$MtCO_2e$	1.35	1.32

In this project, column 1: Country Name and column 4: Unit will be omitted.

Computational Overview

Each of these sections below represents a distinct module in the program (not including main.py):

1 Read CSV Files

Our first step was to collect data from CSV files and transform it into computable data (list of list for country metadata and mappings for emissions and revenues data) in Python. There are three functions, with each one used to respectively collect each category of data specified in the previous section. At this stage, all rows (except header rows) are collected, regardless of validity. For example, when reading the country_metadata csv, some rows do not contain country data. For instance, one of the rows represents information for ’Heavily indebted poor countries (HIPC)’. Regardless, these instances are still collected, but will be filtered out later in the next section.

2 Initializing and Creating Data Class Instances

We created three data classes: Emissions, Revenue, and Country. The Emissions and Revenue data classes each contain four instance attributes, one being the country code, and the others representing the emissions or revenue data for each of the three sectors. Note that the instance attributes for the three sectors in the Emissions data class have type Optional[list] = None because we realized that not all countries have emissions data for all three sectors in the csv file. Therefore, we decided to make the default value None for those instance attributes to avoid any errors moving forward.

The Country data class contains instance attributes like name, country_code, and income_group and also its sector_emissions and sector_revenues. sector_emissions has Emissions class as its type, and sector_revenues has Revenue class as its type. This allows us to access all the information of a country directly through its Country instance without having to search for its Emissions and Revenue instances separately. Note that sector_emissions and sector_revenues attributes are set to None as default using the Optional typing as well, as not every country in the file has its corresponding Emissions and/or Revenue data as we curated the csv files from different sources, and unfortunately, there are some discrepancies between different sources.

Then, we created three functions to create a mapping of Country instances that represent each country in the csv file keyed by their country codes. First, we created a function that initializes all emissions data keyed by the country code. A similar function was also created to initialize revenue data. Any missing information in a particular year will be represented by an empty string in the list. After having initialized all Emissions and Revenue instances, we initialized all Country instances keyed by the country codes. In the init_countries function, we first created a Country instance for all valid countries (at this step, we have included an if statement to filter out instances in the csv that are not countries by checking whether there is an income group specified to the country or not). Then, we used two for loops to initialize sector_emissions and sector_revenues instance attributes in each Country instance that we already created. This does not cause any error because we set the default data types of those attributes to None using the Optional typing earlier.

At this point, we have a complete mapping of all valid Country instances keyed by the country code.

3 Computations

There are three parts to this module: filtering countries according to their income group, finding the average of emissions data of each sector every year from 1990 to 2016, and finding the average revenue of each sector.

Filtering out the countries according to their income group is easier since we just access the instance attributes of each Country instance in the mapping using a loop to return only those that match the income group that is specified as the function argument.

Finding the emissions average was more complicated as we cannot just use a for loop to loop over each Country instance; not all Country instances have a corresponding emission value for each sector. Therefore, there are two helper functions that we created to filter, collect all valid instances and calculate the average for a particular year, and then use the "main" function to append it back to the main accumulator.

Finding the revenue average is slightly simpler because all Country instances have corresponding revenue for each sector. However, there is still missing information. For example, while all Country instances have corresponding revenue data for each sector, there is no guarantee that the revenue data in each sector for every year is complete. Therefore, we use if statements to deal with any missing information for a particular year.

In both parts that find the averages, we use statistics.mean methods instead of having to use sum() and len() functions, which complicate the code.

At this point, we have the average revenue and emissions data for every year from 1990 to 2016 for each sector of each of the four income groups. This is represented using mappings, with each list of average revenue/emission keyed by the name of the sector.

4 Pandas Dataframe and Plotting

In the first part of the module display_graph, we created a function that returns a pandas data frame from the mapping containing the averages ("pandas.DataFrame.from_dict"). We created a data frame for each income group that looks like this (we only show that first row here, excluding the header, but there are a total of 27 rows with each row representing the data for every year from 1990 to 2016):

Year	Agriculture Emission	Manufacturing Emission	Industry Emission	Agriculture Revenue	Manufacturing Revenue	Industry Revenue
1990	12.39000	8.44000	0.412821	1381441372.64	463163497.556	792257427.250

We have chosen to use pandas data frame to structure the data because it would be a more efficient way to access the data to plot (as we can just specify which columns to plot against) rather than having to plot data from a mapping dict. It is also an easy way to visualize the data.

In the second and last part of this module, we created a function plot_graph that plots a bar chart using plotly. We decided to use plotly to plot a grouped bar chart so we can plot the revenue and the emission data for a particular year side by side. This will allow comparison and visual observation of the trend of how revenue and emissions data develop over the years ("Bar Charts"). We use the offsetgroup argument for python.graph_objects.Bar method to put the two bars side by side by setting one bar to an offsetgroup of 0 and another to an offsetgroup of 1 (Bengtsson). Also, since the emission data and the revenue data have very different orders of magnitude (the revenue data can get to billions while emission data does not exceed 100), we decided to use plotly.subplots to create a secondary y-axis, so that we have two axes that represent the scale of emissions value and revenue respectively ("Multiple Axes").

To allow switching between data for different sectors, we created interactive buttons using plotly custom buttons capabilities ("Custom Buttons"). To allow the buttons to work properly, we first wrote codes to plot bar graphs for every sector. However, we set the argument "visible" in the same python.graph_objects.Bar method as earlier to False for the graphs of sectors that we don't want to show until the users have clicked on corresponding buttons. Note that the default graph shows the agriculture sector, so initially, the bar graphs codes for agriculture revenue and emission have their argument "visible" set to True. Then, we use the update_layout method and use its updatemenus argument to create three buttons, labelled "Agriculture", "Manufacturing" and "Industry" respectively ("Layout.updatemenus)". Each button has the capabilities to set the argument "visible" to True or False according to which button is selected. For example, if the "Industry" button is selected, the argument "visible" in the codes for plotting industry emissions and industry revenue will become True, and the codes for other plots will become False.

Lastly, we also use f-string formatting to automatically update the title of the graph automatically according to the input argument of the function plot_graph (Jablonski).

Instructions for running the program

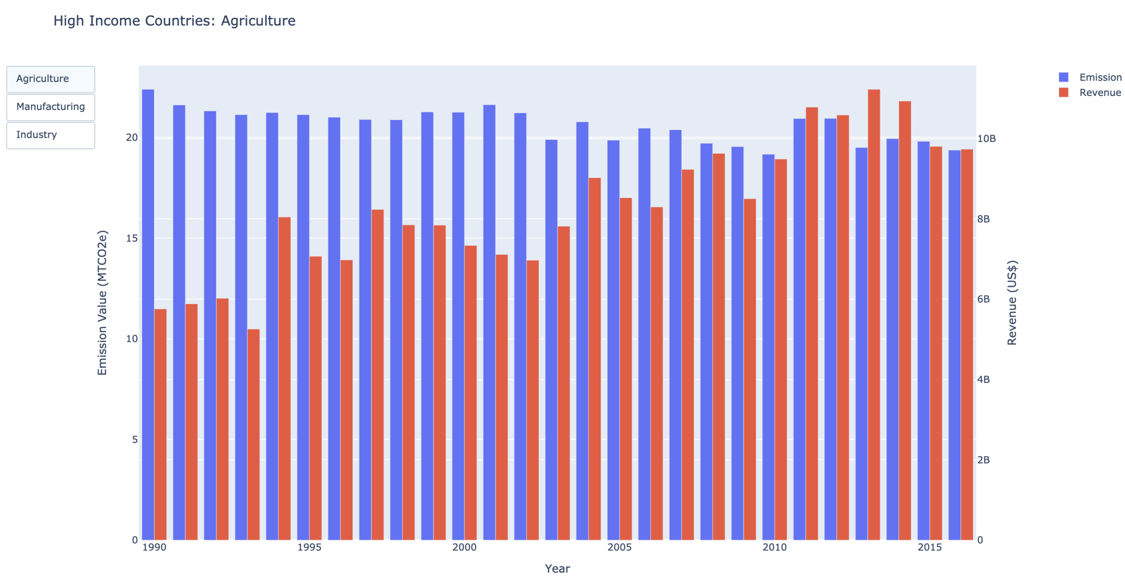
1. Make sure you have downloaded all required Python libraries specified in requirements.txt on your Python IDE.

2. Make sure that you have downloaded every python file and the dataset folder and put them all under the same directory on your computer.

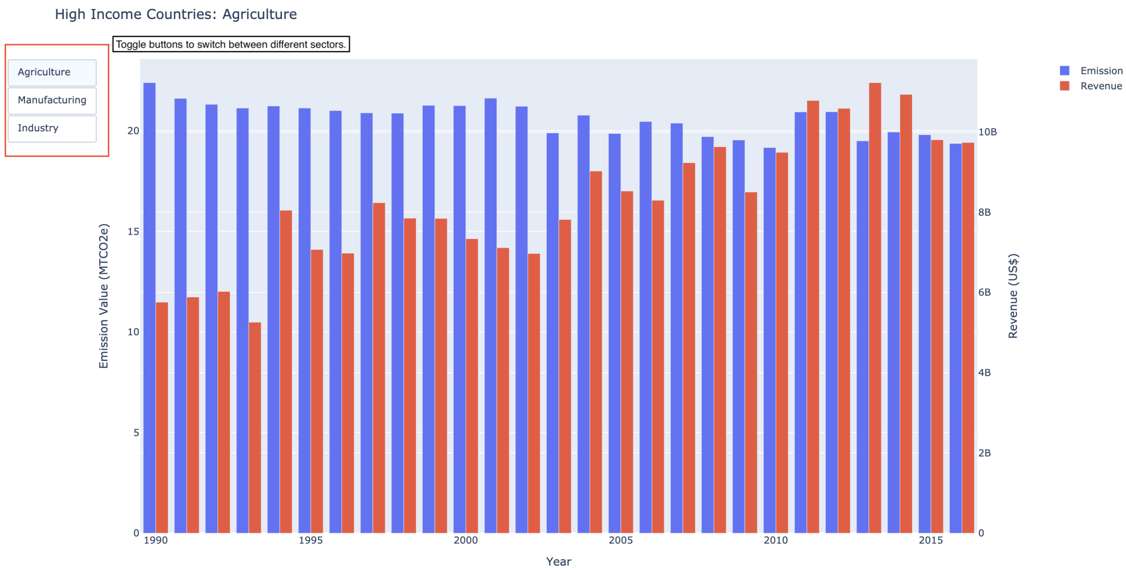
    - The original datasets can obtained from the following sources:
        - Emissions Data (for all three sectors) - www.climatewatchdata.org/data-explorer/historical-emissions?historical-emissions-data-sources=cait
        - Agriculture Value - data.worldbank.org/indicator/NV.AGR.TOTL.CD
        - Industry Value - data.worldbank.org/indicator/NV.IND.TOTL.CD
        - Manufacturing Value - data.worldbank.org/indicator/NV.IND.MANF.CD.
        - The country metadata data can obtained from any of the above three WorldBank sources. If you download any of the three datasets in csv format, you will get a folder containing three items. One of those is the actual Value dataset, and one of the remaining two files will be the country metadata dataset.

3. Make that main folder source root in PyCharm (or whatever IDE you can use to run Python), and run the main.py module.

4. You should see a bar graph titled "High Income Countries: Agriculture" as default (screenshot below)



5. If you wish to see the graph for other sectors in High Income Countries, click on any one of the buttons corresponding to the sector on the left of the bar chart (screenshot below).



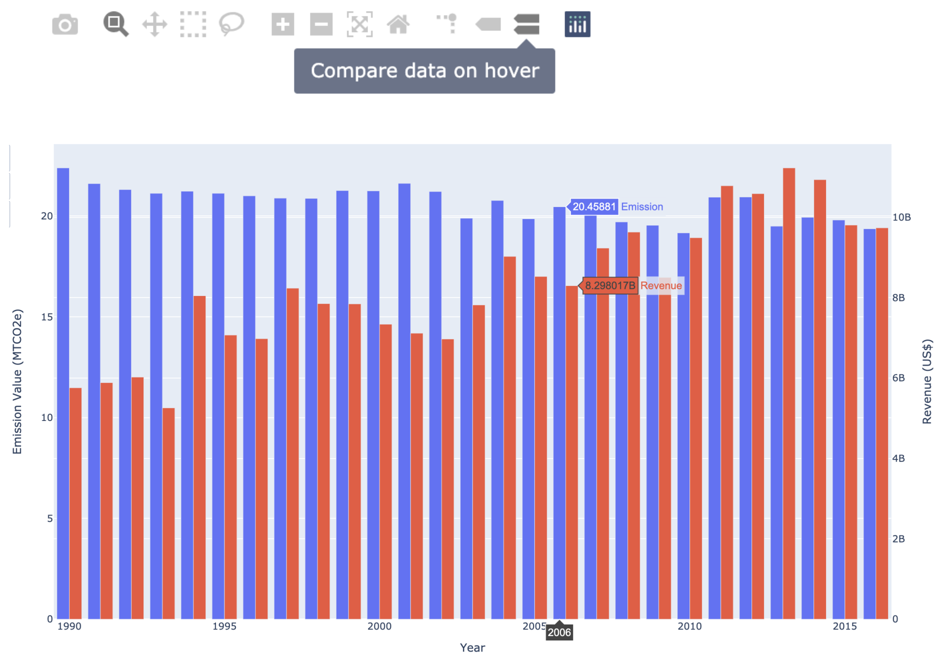
6. If you wish to see the graphs for other income groups (Upper middle income, lower middle income or low income), you will need to go to main.py to uncomment the specific line of code calling the plot_graph function (see screenshot below). For example, if you want to see the graph for Low Income countries, uncomment "plot_graph(low_income_df, 'Low Income')".

```
#################################################################################
# Editable Part - Feel free to uncomment and comment the functions below.
#################################################################################

#  Plotting - default graph is for High Income countries
plot_graph(high_income_df, 'High Income')

#   If you want to see the graph for upper middle income countries,
#   uncomment the code below.
# plot_graph(upper_middle_income_df, 'Upper Middle Income')

#   If you want to see the graph for lower middle income countries,
#   uncomment the code below.
# plot_graph(lower_middle_income_df, 'Lower Middle Income')

#   If you want to see the graph for low income countries,
#   uncomment the code below.
# plot_graph(low_income_df, 'Low Income')
```

7. Similar to the High Income Countries graphs, feel free to toggle between different sectors using the buttons in any other income groups' graphs.

8. In any graph, you can zoom in and zoom out, or download the plot as png file using a series a buttons on the top right corner of the graph (screenshot below). To reset axis, double click on the graph.

9. You can also hover over a bar to see the specific value of the data. If you select "Compare data on hover" button (screenshot below), you can even compare the emission and revenue data of a particular year side-by-side by hovering over a bar.



10. Have fun exploring! Remember to NOT edit any other module except the main.py module otherwise you may run into some issues running the program.

Discussion

The question we aimed to answer with our computation exploration was: In relation to the income classifications, what is the relationship between the revenue generated by different sectors in countries worldwide and their contributions to greenhouse gas emissions? Though our findings do not provide a clear answer to our question, there are interesting trends and conclusions we can draw.

Low Income Countries

In the manufacturing graph, it was quite surprising to see that high emissions and low revenue in the 1990s transformed into quite the opposite: high revenue generation and low emission rates. Perhaps, this sudden change can be equated to the integration of sustainable technology—but this is merely our guess. Another trend can be noted in the industry graph; overtime, revenue began to soar, drastically changing the ratio of emissions to revenue. In the agriculture graph, the release of emissions seems to grow consistently, while revenue soars overtime. *Overall Conclusion:* Decrease in emissions in manufacturing sector as revenue increases over time. The other two sectors see a positive correlation between revenue and emissions.

High Income Countries

From the manufacturing and agriculture sector graphs, a slight decrease in emissions over time can be observed. For industry sector, there is an inconsistent trend; the rate of emissions rises and falls over time. Throughout all sectors, revenue generation subtly increases overtime. *Overall Conclusion:* Subtle decrease in emissions in two sectors as revenue increases over time. The emissions of industry sector does not seem to have any strong correlation with the revenue.

Upper Middle Income Countries

From the manufacturing and industry sector graphs, an increase in emissions over time can be observed. For agriculture, the rate of emissions remains quite stagnant over time. Throughout all sectors, revenue generation increases immensely over time, with a sharp increase noted beginning around 2005. *Overall Conclusion:* Increased emissions as revenue increases over time in two sectors. The correlation between rate of emissions and revenue in agriculture is weak.

Lower Middle Income Countries
As the emissions released by each of the sectors—agriculture, manufacturing, and industry—increased over time, the revenue generated by them also increased over time. Although the increase in both variables—emissions and revenue—was not proportionate per sector, the overall trend was the fact that there was, indeed, an increase. *Overall Conclusion:* Positive correlation between rate of emissions and revenue in all three sectors.

Final Conclusion: One trend that was common to each income classification and each sector, whether subtle or sharp was an increase in revenue. Also, in all income groups considered, there are at least 2 sectors that see a positive correlation between revenue and rate of emission. Therefore, on the whole, we can conclude to a certain extent that the effects of climate change are being exacerbated by our greed. However, this conclusion is limited in the sense that there are still several weak correlations between emissions and revenue.

Challenges & Obstacles

We encountered some obstacles due to gaps in our datasets; for example, in the country metadata, some country codes were not affiliated with a country name or region. We also faced some issues with converting from data classes to pandas data frames, so we decided to store most of our data in dictionaries instead. In efforts to make our final product more interactive, we added features such as buttons to allow the user to exchange the data depending on which income classification of countries they wanted to analyze. Implementing interactive buttons to switch between different data was difficult at first as both of us had never worked with Plotly extensively. Also, plotting dataframes instances using plotly was tricky at first, as we had to make sure that we structured our dataframes in a certain way that is compatible with plotly. Therefore, we had to do some trials and errors multiple times to identify the best structure that works.

Further Exploration

One possible exploration topic would be to identify why there is a weak correlation between emissions and revenue in some sectors, as described above. It may be possible to look into the trend of technological advancement and rate of emission and see whether an increase in technological advancement results in a lower rate of emission. Also, we can analyze more industries such as transportation and waste, to gain an all-encompassing perspective. Additional research into which specific processes drive emission rates (and/or revenue) per industry, which countries in each classification contribute most to emissions, and how we can leverage emerging technologies (such as artificial intelligence and cellular agriculture) to reduce emissions is recommended to both pinpoint root causes of the issue, along with potential solutions.

Works Cited

"Bar Charts." Plotly, plotly.com/python/bar-charts/.

Bengtsson, Fredrik. "Stacked and Grouped Bar Charts Using Plotly (Python)." DEV Community, DEV Community, 11 Aug. 2020, dev.to/fronkan/stacked-and-grouped-bar-charts-using-plotly-python-a4p.

"Climate Watch." Climate Watch - Data for Climate Action, www.climatewatchdata.org/data-explorer/historical-emissions?historical-emissions-data-sources=cait.

"Custom Buttons." Plotly, plotly.com/python/custom-buttons/.

"DataFrame." DataFrame - Pandas 1.1.4 Documentation, 2020, pandas.pydata.org/pandas-docs/stable/reference/frame.html.

"Focusing on Regenerative Agriculture and Moving to Renewable Electricity, Nestlé Redoubles Efforts to Combat Climate Change." Nestlé Global, 3 Dec. 2020, www.nestle.com/media/pressreleases/allpressreleases/nestle-efforts-combat-climate-change.

Jablonski, Joanna. "Python 3's f-Strings: An Improved String Formatting Syntax (Guide)." Real Python, realpython.com/python-f-strings/.

"Layout.updatemenus." Plotly, plotly.com/python/reference/layout/updatemenus/.

"Multiple Axes." Plotly, plotly.com/python/multiple-axes/.

"Pandas.DataFrame.from_dict." Pandas.DataFrame.from_dict - Pandas 1.1.5 Documentation, pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.from_dict.html.

Prydz, Espen Beer, and Divyanshi Wadhwa. "Classifying Countries by Income." WDI - Classifying Countries by Income, 9 Sept. 2019, datatopics.worldbank.org/world-development-indicators/stories/the-classification-of-countries-by-income.html.

"What to Know About Nike's Stance on Tackling Climate Change." Nike News, 19 Sept. 2019, news.nike.com/news/nike-move-to-zero-climate-change-initiative.

World Bank. "Agriculture, Forestry, and Fishing, Value Added (Current US$)." The World Bank, 2020, data.worldbank.org/indicator/NV.AGR.TOTL.CD.

World Bank. "Industry (Including Construction), Value Added (Current US$)." The World Bank, 2020, data.worldbank.org/indicator/NV.IND.TOTL.CD.

World Bank. "Manufacturing, Value Added (Current US$)." The World Bank, 2020, data.worldbank.org/indicator/NV.IND.MANF.CD.