UNIVERSITY OF TORONTO

CSC110 Final Project

Reverberations of Climate Change on Earth and Humans

Authors:
Eeshan Narula, Garv Sood, Avnish Pasari, Aakash Vaithyanathan

December 2020

Contents

Chapter 1

Introduction

Our planet is becoming warmer day by day, from the North to the South pole. Human activities have led to an exponential rise in the temperature of the earth, and its impacts have started to prevail. The effects of global warming are appearing right now, glaciers are melting, sea level is rising, people are getting sick, more wildfires are being produced, and carbon emissions are increasing. (1)

The rising global land temperature has a detrimental impact on the ecosystem affecting various other components of the earth. It has a serious effect on vegetation and causes various health problems for humans. Furthermore, an increase in land temperatures across the globe results in the melting of glaciers, permafrost, and ice sheets all of which in turn contribute to an increase in sea levels which affects the coastal inhabitants and results in loss of habitat for fishes, birds, and plants.(2)

Human activities such as the burning of fossil fuels and large-scale deforestation in recent years have led to the release of greenhouse gases on a scale that has never been seen before. The rise of the concentration of greenhouse gases in the air is an indication that we are turning up the dial on global warming. The greenhouse effect (due to increased greenhouse gas emissions) traps the Sun's radiation and re-radiates it. This is causing catastrophic effects by altering ecosystems, raising sea levels, and causing more frequent extreme weather events.(2)

These increasing emissions vary from country to country, continent to continent. Developed countries are taking steps to decrease climate change, but developing countries are emitting more gasses which is why more people are getting sick in those areas. Thus there is a relation between the GDP of a country and the deaths per year.(3)

The question we want to analyze in our project is that **How with time, climate change has affected and will affect the environment and humans? How has the human response to climate change grown over time?**

1. https://www.nasa.gov/audience/forstudents/k-4/stories/nasa-knows/what-is-climate-change-k4.html

2. https://climate.nasa.gov/effects/

3. https://cleanair.camfil.us/2017/10/30/air-pollution-in-developing-countries/

Chapter 2

Datasets

- The first dataset consists of levels of emissions for different greenhouse gases and for different countries from 1990 to 2014. The columns we would be using are country_or_area, year, value and category.

  Source: Kaggle: https://www.kaggle.com/unitednations/international-greenhouse-gas-emissions

  Format: csv

- The second dataset contains land temperature for different countries from 1743-11-01 to 2013-09-01. Although we would only be using data from 1990 to 2013. The columns we would be using are dt(date), AverageTemperature, Country.

  Source: Kaggle: https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data

  Format: csv

- The third dataset contains CO2 emissions (kg per PPP $ of GDP) for different countries from 1960 to 2020, but the data has consistent values from 1990 to 2016, thus we only use columns for 1990 to 2016, and Country Name.

  Source: Worldbank: https://data.worldbank.org/indicator/EN.ATM.CO2E.PP.GD

  Format: csv

- The fourth dataset is made by us, using the Twitter API to fetch the most recent tweets from 31st October to 9th November, for previous 10 days. We also added sentiment scores and sentiment type(neutral, positive or negative) to the dataset. The columns we would be using are tweet_text, senti-score,senti-type and created_at.

  Source: made by us using Twitter API   Format: csv

- The last dataset contains words with their sentiment score. This dataset is used to calculate the sentiment score for the 4th dataset. The dataset only has two columns, which are the words and the sentiment scores, and we are using both of the columns.

  Source: HLT: https://hlt-nlp.fbk.eu/technologies/sentiwords   Format: text

Chapter 3

Computational Overview and Discussion

3.1 Data management and transformations

To manage all these datasets and perform various transformation on them, we made a class called Dataset. A Dataset object contains a dataset, and many methods to mutate the dataset. Following are the methods we can apply to a Dataset object :-

- We can initialise the data by passing in the filepath of the dataset, or a List object containing data. We can also pass a list of data types for each element in the rows of the data, and they would be converted to respective datatype.

- To convert the elements to different datatypes, the class has a function called transform, which takes in a list of data types, and convert each element with respect to the datatype passed.

- To filter the dataset, which only contain specific values for a column in the dataset, the class has a function called filter_by_value and filter_by_function. The filter_by_value function takes in the column and the list of values we want to keep for the column. The filter_by_function takes in the column and a Predicate function; if the function returns true, we keep the row, else we remove the row.

- To remove all the NA values or empty string, the class has a function called remove_na, it does not take any arguments.

- To keep only some columns or delete some of the columns of the dataset, the class has functions select and delete. Both of them take in the list of columns to keep or delete.

- We can see the starting of the dataset - using the head function in the class, which by default will pretty print the first five rows of the dataset. We can also pass an argument nrows to it specifying the number of rows to print.

- To get all the unique values of a column, the class has a function called unique which takes in the column index for which we want the unique values, and return a list with all the

unique values for that column.

- To calculate the average for a column in the dataset, the class has a function called calc_avg_col, which takes in the column index for which we want to find the average, and returns the average.

- To extract a column and return it, the class has a function called extract_column, which takes in the argument column index.

- To add a column to dataset the class has a function called push

- To calculate the average for different values of a column (for example to calculate average emissions for different countries), the class has a function called calculate_average.

There also some other global functions for data handling :-

- group_by_values: It returns a list of Dataset objects, where each object is a sub data of the dataset object we pass in, and represents data for a specific value of the column we pass in to the function. For instance if we want separate datasets for each of the country in a dataset then we would use this function.

- group_by_function: It does the same job as group_by_values, but takes in a function which adds a new column to the dataset and groups by that column.

The other functions are helper functions for all of these functions. The code for the Dataset class and all the functions can be found in data_manager.py

3.2 Change in Average Land Temperature over time

3.2.1 Introduction to the problem

First we would be looking at the Average land temperature dataset. The dataset consists of the average land temperature for Australia, Canada, Russia, India, United States, China, and Brazil.

From this dataset we would analyze :-

- The percentage difference in the Land Temperature for Brazil, India, China, and Australia, which are in the top 10 most populated countries.

- The difference in the average temperature of United states, Canada and the rest of the countries in the dataset (which are Australia, Russia, India, China, and Brazil).

3.2.2 Dataset transformations

We would make use of two helper functions (get_avg_by_year and get_avg_by_country) to perform various data transformations on our dataset (filtered_land_temp.csv file). These two helper functions help us extract the relevant data from the csv file and store them in appropriate format making it easier for us to access and perform various computations on them. The above-mentioned helper functions make use of a Dataset class to store, handle and transform data.

- The get_avg_by_year() function takes the various countries from the csv file and returns a List[List] where each inner list contains the yearly average land temperature values from 1990 to 2013 inclusive.

- The get_avg_by_country() returns a dictionary where each country (key) is mapped to its yearly average land temperature value from 1990 to 2013 (value).

3.2.3 Visualizations

For representing our data, we make use of plotly's bar graphs and mathplotlib's pie chart.

- In our plot_data_usa_can_all() function, we use a bar graph to show the varying global land temperature values between Canada, United States and the Rest of the countries (India, Brazil, China and Australia, Russia) between 1990 and 2013. Plotly's bar graph allows us to simultaneously compare the changes between the land temperatures over the years and how these trends have been changing between various countries.

- In our plot_data_greater_average() function we compare the land temperature between countries whose land temperatures exceeds the global average temperature between 1990 and 2013 (Australia, Brazil, China, and India). We use a pie chart to show the varying percentage composition of various countries' contribution to rising land temperature. This way we can better understand the ratio of land temperature difference among these countries.

3.2.4 Discussion
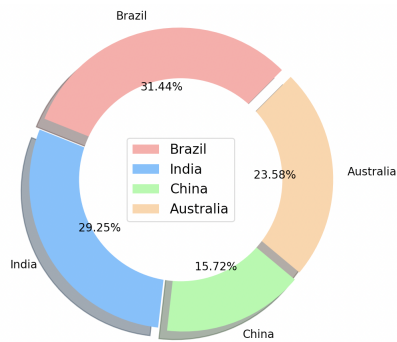
Our objective is to analyze how has the global land temperature changed over the years and understand why it is crucial to take necessary actions to reduce its effect.

Our pie chart graph gives us a representation of the varying land temperatures between Australia, Brazil, China, and India from 1990 to 2013. The graph gives us a percentage composition of the land temperature among these countries.

From the pie chart we can analyze that countries like Brazil (still developing) and India (highly populated) contribute greatly to the rising global land temperatures. This rise in trend can be traced to the population and the development of the country.

From the bar graph, we can visualize and observe how the land temperature changes between United States, Canada, and various other countries over a period of 24 years from 1990 to 2013. We can observe that the land temperatures of these countries show a rise and fall trend from 1990 to 2010. However, we notice a rising trend in land temperatures over the subsequent years.

One of the limitations of the above-mentioned plots is that they only show changing land temperatures up to 2013 and reveals no information about the present trend.
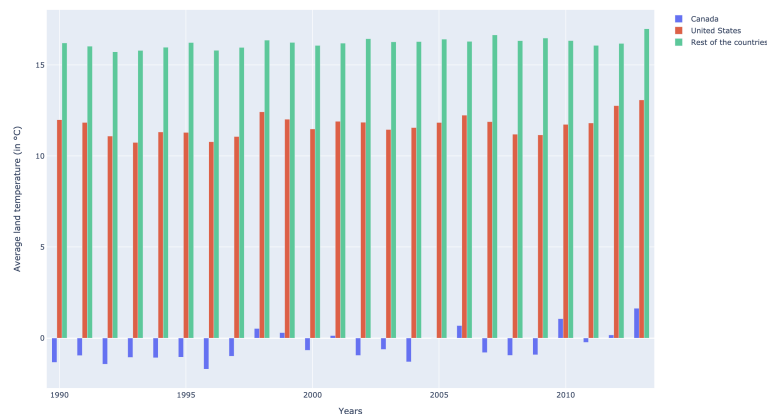
The rise in land temperature has two serious effects across the globe. One of them is damage it causes to vegetation thus causing several health problems for both humans and animals. The other is the rising sea levels caused due to melting of ice sheets, permafrost, and glaciers. The rise in land temperature is mainly caused due to the emissions from the greenhouse gases due to the greenhouse effect.

These land temperatures are positively correlated to the population and negatively correlated to development of a country. However, The United States, for example, contains just five percent of world population, yet contributes a quarter of total CO_2 output (1) and has highest land temperature changes for any developed country.

On January 15, 2020, a report published by NOAA and NASA confirmed that 2010 to 2019 was the hottest decade to be ever recorded. (2). It is imperative for us to recognize the detrimental effects climate change has on the globe.

1. https://www.scientificamerican.com/article/population-growth-climate-change/

2. https://www.nationalgeographic.com/science/2019/12/the-decade-we-finally-woke-up-to-climate-change/

3.3 Greenhouse Gases and their effect over the years

3.3.1 Introduction to the problem

Now, we would analyse data from Greenhouse gases data, and Co2 emissions (PPP per $ GDP) data. The Greenhouse gases data consists of values for emissions for different gasses for different countries from 1990 to 2014. The Co2 emissions (PPP per $ GDP data) data consist of co2 emissions per dollar of gdp of different countries from 1960 to 2020, although the dataset is empty until 1990 and after 2016.

From these datasets we would analyse:-

- Change in the emissions of nitrogen triflouride from 2010 to 2014

- Difference in the Change in the emissions of nitrogen triflouride in USA, Canada and European Union from 1990 to 2014

- Change in the emissions for 3 sparsely populated countries, Austria, Belarus and Bulgaria in the 24 year period (between 1990 and 2014).

- Difference in emissions of hydroflourocarbon in 1990, 2002 and 2014.

- Change in the Carbon dioxide emissions per dollar of gdp form 1990 to 2014 for the all the countries.

- Percentage contribution each country is making to the total emissions of Carbon dioxide per dollar gdp.

3.3.2 Dataset transformations

We have used data transformation techniques in functions, data_by_tags_and_country_or_year, data_by_tags, and tidy_gdp_data, to manipulate our datasets. These functions are used to transform the data so that it can easily be used for subsequent visualizations.

- data_by_tags_country_year: The function splits the dataset by countries and extracts emission values for each of the desired countries. It returns a list containing lists of each country's emission value. This function is applied to Greenhouse gas dataset.

- data_by_tags: The function filters data on the basis of countries and years and returns a list with emission values along with the respective country and year. This function is applied to the Greenhouse gas dataset.

- tidy_gdp_data: The function removes all the columns with years less than 1990 and greater than 2016, and removes all the rows with NA value or empty strings. It then tidies the data statistically by keeping only one column of years, instead of having several rows, one for each year. This function is applied to Co2 emissions (PPP per $ gdp) dataset.

3.3.3 Visualizations

The functions which make visualisations use the numpy, plotly, pandas and matplotlib libraries. We use the numpy library to create a one-dimensional array with numerical data giving information about emissions. Pandas library is used to create a dataframe from the dictionary of lists. This dataframe creates a two-dimensional array, which is the standard way to store data that has two different indexes (rows and columns), and is a valid input to the plotting libraries.

- First we analyze the nitrogen triflouride emissions of all countries from 2010 to 2014 using boxplots from the matplotlib.pyplot library, in the plot_all_countries_peryear function.

- Next, we analyse the change in amount of nitrous oxide emissions in Canada, The United States and European Union using a scatter-plot using matplotlib.pyplot library, in the function plot_scatter_n2o.

- For the third graph, we plot a plotly's interactive visualizations having both the boxplots and scatterplots in a single graph, to show the emissions for 3 sparsely populated countries, Austria, Belarus and Bulgariain the 24 year period (between 1990 and 2014), in the scatterboxplot_sparsely_pop function.

- Next, we would use a similar interactive visualizations having both the boxplots and scatterplots in a single graph, to show the difference in emissions of hydroflourocarbon in 1990, 2002 and 2014, in the plot_scatter_boxplot_1990_2014 function.

- Now, we would use a Choropleth map showing the distribution of CO2 emissions (kg per PPP $ of GDP), using pyplot, which would animate from 1990 to 2016, in the plot_emission_per_gdp function.

- The final Visualization is a pie chart, which would show the percentage of CO2 emissions (kg per PPP $ of GDP) contribution made by all the countries with respect to other countries, using plotly, in the dist_2016 function.

3.3.4 Discussion

We want to assess how gas emissions have increased over the years and whether governments have done enough to control emissions.

The boxplot gives us the levels of nitrogen triflouride emissions over a 5-year period (2010 - 2014). Unfortunately the dataset only has data till 2014 so figures of recent years could not be presented. In the graph we can see that the emissions of nitrogen triflouride are falling each year. So, countries all over the world were extremely successful in reducing its emissions levels from the year 2010 to 2014


8
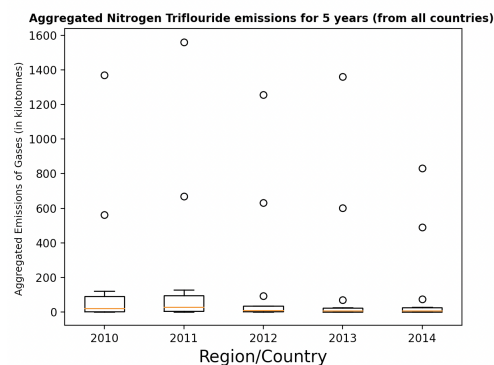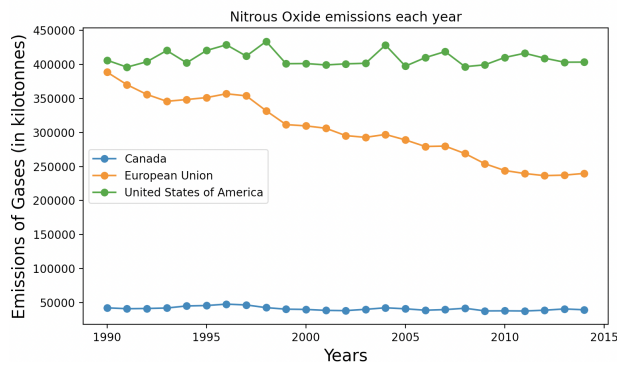
The scatterplot displaying the nitrous oxide emissions by Canada, The United States and the European Union shows that the nitrous oxide emissions in Canada are the lowest out of the 3 regions by a large margin.

This can be accounted to a smaller population and strict rules. Agriculture is the largest contributor towards nitrous oxide emissions. Canada has implemented the Nitrous Oxide Emission Reduction Protocol so that farmers can earn carbon credits if they reduce their nitrous oxide production, thus giving them incentives to reduce pollutants (1). One limitation of this visualization (that could be explored further) is that even though emissions of nitrous oxide by the USA remain the same throughout, we have no information about the remaining fossil fuels so cannot make any comments on America's handling of fossil fuels.

Usually sparsely populated countries have lower pollutant emissions due to strict regulations and less population density. Out of the 3 countries, Austria has the highest carbon dioxide emissions, and Austria has the highest population (2). So it can be concluded that there is a direct relation between CO2 emissions and population density. However, the Austrian government has set up a plan to reduce CO2 emissions by 36% by 2030 (3). Such measures are the need of the hour to reduce climate change.
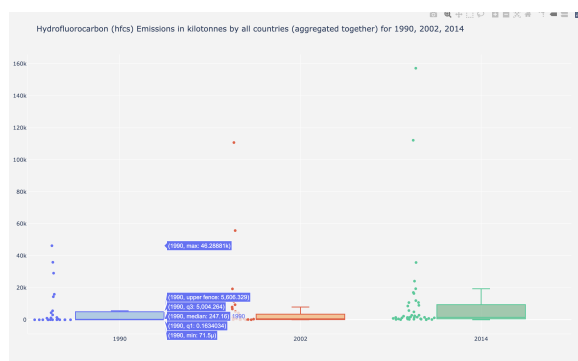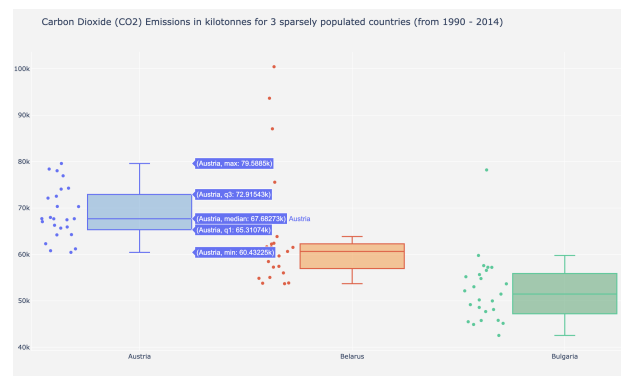
The graph summarizes the emissions of Hydroflourocarbons (HFCs) of all countries at intervals of 12 years (1990, 2002, 2014). We can see that over the decades, the maximum values of emissions are higher and there is an overall rise in pollutants. Hydroflourocarbons (HFCs) are mainly use as refrigerants. These levels have continued to rise in spite of the Montreal Protocol (which aims to phase out such chemicals) the levels increase.

From the Choropleth map we can observe that the Carbon emissions (kg perPPP \$ of GDP), are decreasing every where with respect to their GDP. Even though the the emissions per dollar gdp are decreasing, the fact is that the total emissions are increasing. Although, because of the pandemic, record drop in the emissions was seen (4).

CO2 emissions (kg per PPP \$ of GDP) by country
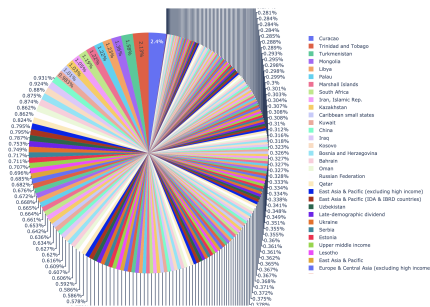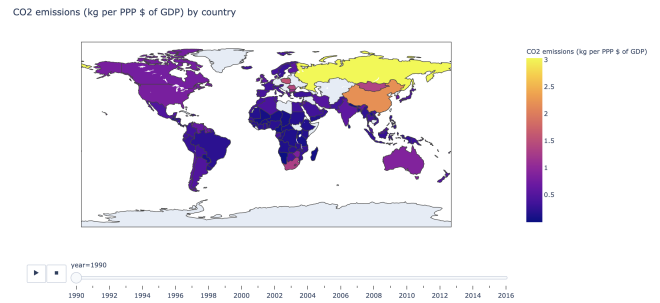
The pie chart shows the percentage of CO2 emissions (kg per PPP \$ of GDP) contribution made by all the countries. It can be seen that the country with the highest percentage is Curacao. It does not have a high gdp but is producing immense Co2 emissions with respect to its gdp. Although the highest Co2 emissions are made by USA (5)

1. https://farmingfirst.org/2016/06/nitrous-oxide-emission-reduction-protocol-2/

2. https://www.worldometers.info/population/europe/

3. https://www.themayor.eu/en/austria-plans-for-36-percent-co2-emissions-reduction-by-2030

4. https://www.sciencedaily.com/releases/2020/10/201014082806.htm

5. https://www.economicshelp.org/blog/10296/economics/top-co2-polluters-highest-per-capita/

3.4 Human Reaction to Climate Change

3.4.1 Introduction to the problem

We would use data from twitter, to analyse the sentiments of people about climate change. The dataset is made by us using the Twitter API. The dataset contains tweets with hashtags about climate change, for 10 days from 31st October to 9th of November. We can re-run the program to get latest dates (31st October to 9th of November was the time when US elections were held) and we extracted the data at that time.

To analyse the sentiments, we would be using the Sentiworld dataset (5th dataset in the dataset description), which has score of -1 to 1 for a large number of words.

From this dataset we would analyse:-

- Change in the sentiment scores in 10 days time (also the time during US elections and some days after elections)

- Top 10 words which occurred the most in the tweets

3.4.2 Dataset transformations

- First we used the Twitter API's search function to search for 9 different hashtags related to climate change. The search function only allows us to search for a 100 tweets or a single page at a time, so to prevent the usage of the for-loop for each page, we used the API's function called Cursor. We conducted a search with every hashtag as the query, and filtered out the retweets. All the Reference used to write the code are in the Introductory docstring itself of get_data.py The function search_for_query is written in the python file, which when run, generates a dataset, to the path argument given to it.

- Now, to add sentiment score to the tweet text, and the type of sentiment (neutral, positive or negative), we used the words and their scores in the Sentiworld dataset. The function takes in the dataset we have created and creates a new dataset with 2 extra columns (which are senti-score and senti-type).

  The type is determined by the score. If the score is greater than .5 then the tweet is positive, if it is less than -.25 then negative or else neutral. The unbalanced type distribution is due to the fact that the dataset is biased towards positive words, and to eliminate this bias we have chosen the values of .5 and -.25. The code for the sentiment score and type calculations is in the analyze_sentiments.py file.

- To use the data for Visualization, and modifying it, Dataset class is used.

3.4.3 Visualizations

- The first visualization is a line graph made with matplotlib.pylot library to show the average score of each day from 31st October to 9th November. For calculating the average we use the Dataset objects calculate_average function. The function which plots this graph is plot_sentiments, available in analyze_data.py.

- The second visualization is a bar graph showing the top 10 occurring words in the dataset. A dict of words mapped to their respective word count, is made with in the helper function word_count, and then a sorted tuple is out of the dictionary which is used to plot a count plot/ bar plot using matplotlib.pyplot library. Both the helper funtion (word_count) and the plotting function (plot_top_10) are also in the file analyze_data.py.

3.4.4 Discussion

From the dataset we want to analyze the human behaviour, and activities on climate change topic.

Today environment and climate change is one of the main issues and people are becoming more and more aware about it. They are posting about it to share their views. In Britain, a majority of people said that the environment and climate change are one of the 3 most important issues. (1).
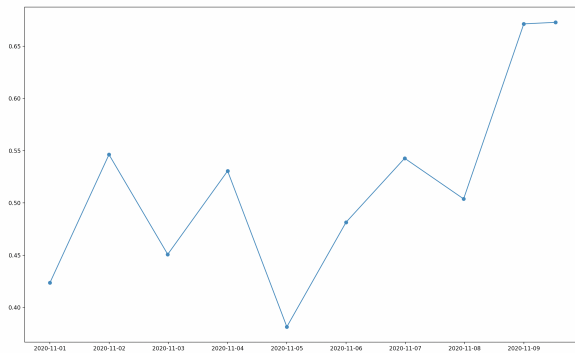
This can be seen in the line plot where the sentiment score is increasing on average, and is positive, which means people are motivated to protect the environment and are willing to act.
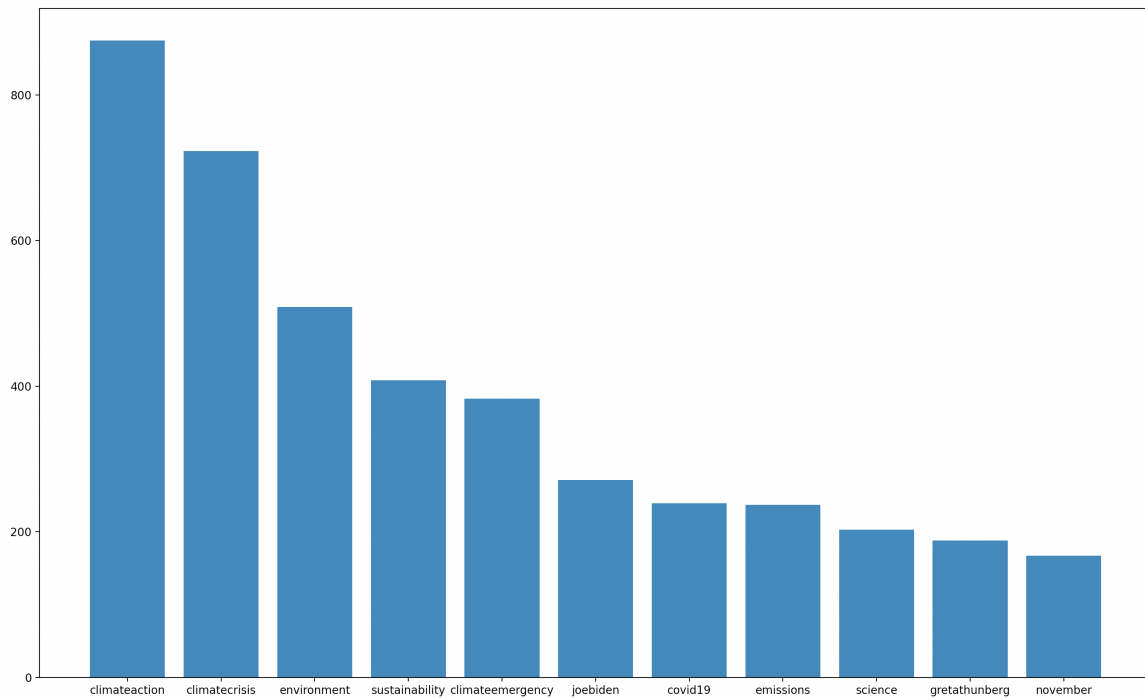
Even from the bar graph we can see that the top 10 words contain more positive words like sustainability, climate action, and Greta Hunberg who is a huge pro environment protester. The list also has the president elect Joe Biden, who is a huge supporter of environmental protection. He has very strong plans for helping environment like plans to make electricity carbon free by 2050. (2)

1. https://www.theguardian.com/environment/2019/jun/05/greta-thunberg-effect-public-concern-over-environment-reaches-record-high

2. https://www.bbc.com/news/science-environment-54858638

3.5 Predicting Future Emissions and Land Temperature

3.5.1 Introduction to the problem

We would be predicting the Co2 emissions and Average land temperature for 2015 2016 and 2017 which means 3 year ahead of what we have data for.

3.5.2 Predicting future values

To predict future values we would be using machine learning techniques to calculate a regression line. We would be using Gradient decent to calculate the weights for a line of best fit, that is linear regression by gradient decent.

A linear regression mode is a straight line which best describes relation between variables. The model can be trained in two ways - with maths or machine learning technique called gradient decent. We would use gradient decent and matrices to find the line of best fit.

The equation of the line can be denoted by $y = mx + b$, where y is the variable we want to predict and x predictor variable used to predict variable y. m is the slope and b is the bias or intercept of that line.

In machine learning terminology this slope m is denoted by θ_1 and is called the weight, and the intercept b is called the bias and denote by θ_0. we want to compute these parameters of the line, which would make it fit to the data.

The algorithm follows the following steps:-

1. first we initialize our weights and biases randomly.

2. then we compute the cost of those weights and biases, which is the average squared difference in the predicted and actual values of our dataset. Our aim is to reduce this cost, by changing the values of the weights and biases

3. This is done by the use of gradient decent formula, where we compute the partial derivative of the cost function which respect to θ_1 to modify θ_1 and with respect to θ_0 to modify θ_0. The value we just calculate multiplied by alpha, is the number we would change our weights and bias by.Here alpha is the learning rate which prevents everything from being to large, and keeps everything in bound.

4. we repeat step 2 to 3 multiple times, to get a line of best fit. the number of times we repeat step 2 to step 3 is called the number of epochs.

Note: A very low learning rate can take the model too many epochs to learn, and too high value of learning rate wont let the model learn, and instead can make it worse with time.
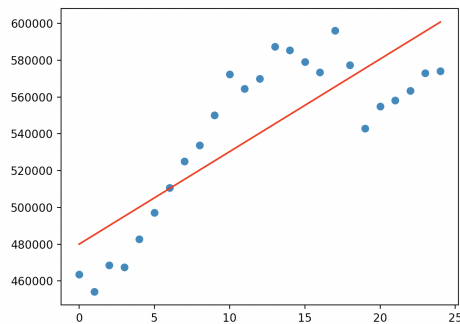
The References to where we took algorithm from is in the code file itself (Linear_regression.py)

3.5.3   Data transformation

- data_by_tags function is used to get Co2 emissions for Canada

- get_avg_by_year() function is used to get average land temperature for Canada
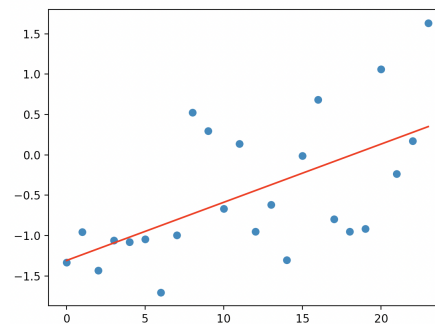
'

3.5.4   Visualizations and Discussion

Canada's future Co2 emissions predictions

From the predicted line we can see that the emissions increased in the 3 year ahead in 2015, 2016 and 2017. It is a fact that they indeed did increase (1) and according to our model the emissions increased by approximately 10 thousand kilotons.

Canada's future average land temperature predictions

From the prediction line we can see that the land temperature would increase in the future. From the model, if we compute the predicted values we observe that the temperature would increase by approximately .23 in the future.

1. https://www.canada.ca/en/environment-climate-change/services/environmental-indicators/greenhouse-gas-emissions.html

Chapter 4

Instructions

4.1   Datasets

All the datasets will be uploaded to Markus. Two of the datasets are helper datasets, and other 4 are for visualizations. All the datasets must be stored in a directory called datasets.

The ones for visualizations are :-

- greenhouse_gas_inventory_data_data.csv

- climate-change-sentiment.csv

- emissions_per_capita.csv

- filtered_land_temp.csv (this is the modified version, link to original dataset is in chapter 2 Datasets)

The other two datasets are :-

- SentiWords_1.1.txt

- climate-change.csv (this is the same as climate-change-sentiment, but does not contain the sentiment score and type)

4.2   main.py

When the TA runs the main.py file, a new window will open, with buttons to see different visualizations. Something like this :-
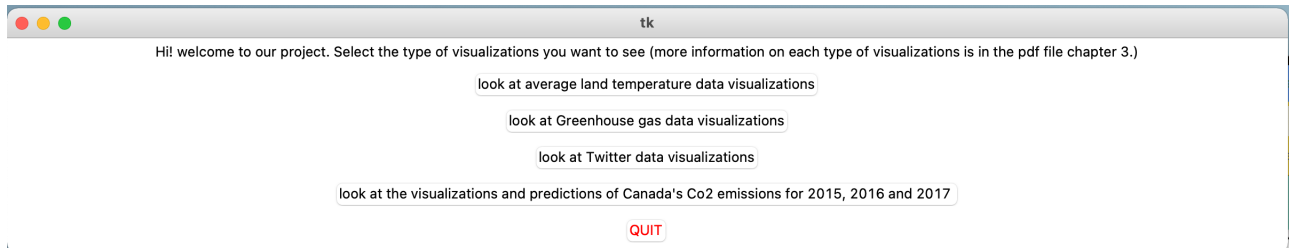
- If the TA selects the first button - all the visualizations related to land temperature, which are in the global_land_temp.py file, would be shown. The TA can shut the window down for those plots, and again come to the interface, and click on a button for different visualisations.

- If the TA selects the second button - all the visualizations related to land temperature, which are in the greenhousegases_project.py file, would be shown. Then can again shut the window down for those plots, and again come to the interface, and click on a button.

- For the third button, similar to the first two button, the TA can see the visualizations of Twitter data, and go back to the interface.

- Similarly, fourth button will show the TA, visualizations for the best fit line and cost function. In addition to this, it will also print the predictions in the console.

  Note: All the visualization, are also shown in this latex/pdf file, for every section's decision subsection of chapter 3 Computational Overview and Discussion

  Some of the visualizations that we are using are interactive, and the description of each one of them is written in chapter 3, Computational Overview and Discussion, under the subsection visualizations of very section of chapter 3.

Chapter 5

References

1. https://www.nasa.gov/audience/forstudents/k-4/stories/nasa-knows/what-is-climate-change-k4.html

2. https://climate.nasa.gov/effects/

3. https://cleanair.camfil.us/2017/10/30/air-pollution-in-developing-countries

4. https://www.kaggle.com/unitednations/international-greenhouse-gas-emissions

5. https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data

6. https://data.worldbank.org/indicator/EN.ATM.CO2E.PP.GD

7. https://hlt-nlp.fbk.eu/technologies/sentiword

8. https://www.scientificamerican.com/article/population-growth-climate-change

9. https://www.nationalgeographic.com/science/2019/12/the-decade-we-finally-woke-up-to-climate-change/

10. https://farmingfirst.org/2016/06/nitrous-oxide-emission-reduction-protocol-2

11. https://www.worldometers.info/population/europe/

12. https://www.themayor.eu/en/austria-plans-for-36-percent-co2-emissions-reduction-by-2030

13. https://www.sciencedaily.com/releases/2020/10/201014082806.html

14. https://www.economicshelp.org/blog/10296/economics/top-co2-polluters-highest-per-capita/

15. https://www.theguardian.com/environment/2019/jun/05/greta-thunberg-effect-public-concern-over-environment-reaches-record-high

16. https://www.bbc.com/news/science-environment-54858638

17. https://www.canada.ca/en/environment-climate-change/services/environmental-indicators/greenhouse-gas-emissions.html

18. https://towardsdatascience.com/linear-regression-using-python-b136c91bf0a2

19. https://matplotlib.org/devdocs/contents.html

20. https://www.programiz.com/python-programming/writing-csv-files

21. https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/overview/tweet-object

22. http://docs.tweepy.org/en/v3.5.0/cursor_tutorial.html

23. http://docs.tweepy.org/en/v3.5.0/api.htmltweepy-api-twitter-api-wrapper

24. https://www.kite.com/python/answers/how-to-color-a-scatter-plot-by-category-using-matplotlib-in-python

25. https://plotly.com/python/box-plots/

26. https://towardsdatascience.com/how-to-create-an-animated-choropleth-map-with-less-than-15-lines-of-code-2ff04921c60b

27. https://numpy.org/doc/

28. https://docs.python.org/3/library/tk.html

29. https://www.tutorialspoint.com/python/tk_text.htm

30. https://www.geeksforgeeks.org/python-grid-method-in-tkinter/