W200 Project 2 Proposal
Fall 2021, Section 10, Group 1
Nan Li, Justin To, Tiffany Smalley
GitHub Repo: https://github.com/UC-Berkeley-I-School/Project2_Li_Smalley_To.git

(I) Project Summary

Overview

In this project, we will show how temperature is increasing globally under climate change and analyze the effect of greenhouse gas emissions on climate change. Research has long shown that greenhouse gases (primarily Carbon Dioxide and Methane, and Nitrous Oxide) contribute directly to global warming and climate change - a looming threat to life as we know it on earth.

While it is widely recognized that climate change is a global phenomenon, we believe that through data analysis, we will be able to gain a better understanding of the nuances. For example, data can be used to show how the warming trend has changed over time, whether the warming is taking place evenly across the globe and throughout the year. Temperature and location data can also be merged with socio-demographic data to yield insights on the cause and impact of climate change.

Questions for analysis

With the above in mind, our team has come up with the following primary questions for our analysis. As we proceed with the analysis, we envisage that the results may give rise to new questions or modification of existing ones.

Primary Questions:

- What are the top heating countries/cities (heating rates, increase per year)?
- How do the rate/year increases compare between the top heating countries/areas and globally?
- Do countries/areas heat up evenly, and do different times of the year heat up evenly?
- Is there a discernible difference between warming in the northern and southern hemispheres?
- How might greenhouse gas emissions mirror global warming in these countries/cities?

- Can greenhouse gas data provide insights on the variation in temperature rise across regions/countries/cities?

As time permits, we also intend to take into account datasets on socio-demographic variables (e.g. population size) to provide further insights on the effects of global warming. For example, address the following questions:

- What are the largest countries/cities that will see a temperature rise beyond livable range (e.g. 40 degrees Celsius)

Final deliverable

The final report will cover brief analyses of the data to answer the questions posed. We will include a summary, our research questions, data sources, methods, description of our analysis, results, supporting visuals, challenges, considerations, and a conclusion with suggestions for further research.

(II) Data Sets

Primary data set

The primary data set used in the project will be the time series data of Berkeley Earth. As one would quickly notice, the data set is very broad, with temperature data provided at different levels (e.g. data at global, country, state/province, city and even weather station levels, all these just for land-based data alone) and in different forms of measurement (e.g. monthly, annual, 5-year, 10-year and 20-year averages). There is considerable variation in data quality for different types of data, e.g. NaN values being much more common for 18th and early 19th century data and for developing countries.

While this may change as we proceed with the analysis, our current plan is to focus on the analysis of the following data:

- **Data level**: primarily data at global, country and city levels, with occasional use of northern/southern hemisphere data
- **Temperature measurements:** primarily 10-year averages, though monthly averages in some cases
- **Time horizon**: primarily data from 1850 and beyond (which matches the time series data available for GHG emissions), with particular focus on the 20th century

Supplemental data sets

Our current plan is to incorporate the use of the following data sets which will provide additional dimensions to our analysis:

- **Greenhouse gases**: Historical GHG Emissions data (specifically the PIK PRIMAP data set) for 216 countries, regions and country groups was sourced from Climate Watch, an online platform managed by World Resources Institute.
- **Population data** sourced from UN Data Explorer: specifically the Demographic Statistics Database from the UN Statistics Division.

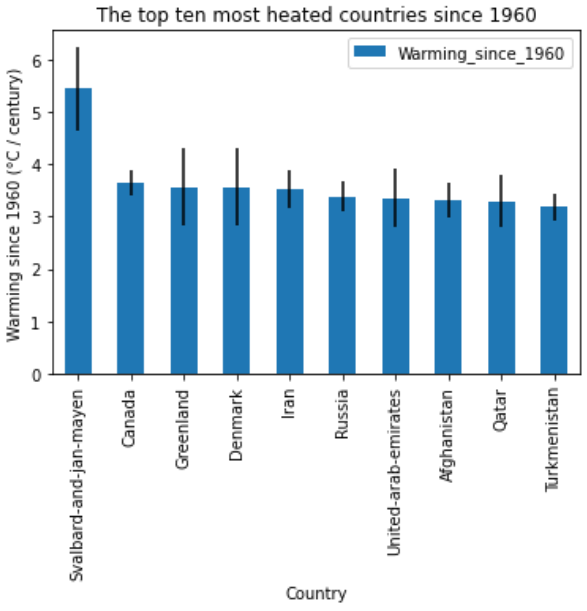
(III) Initial Findings

1. How global is global warming?

To get the general picture of global warming. We first used the data from the Berkeley Earth data webpage (http://berkeleyearth.lbl.gov/country-list/) to find the most/least heated countries.

Figures 1a and 1b show the temperature increasing rate for the top ten most heated countries and the ten least heated countries since 1960, respectively. Even for those least heated countries, their temperatures have risen at least 0.65 °C/century, which means global warming is happening worldwide.

(a) Most heated countries



(b) Least heated countries

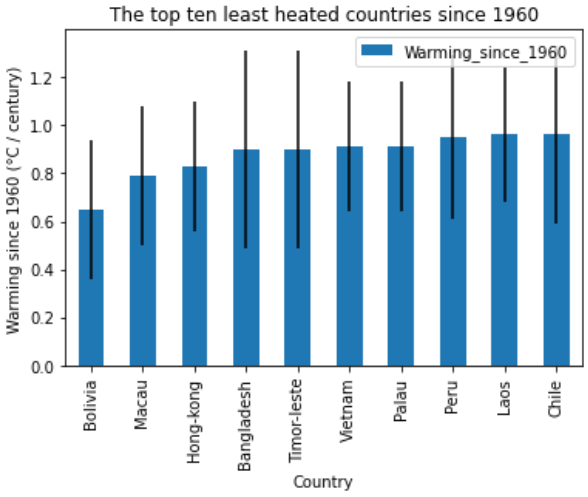


Figure 1. The most/least heated countries since 1960.

2. But do all places experience the same level of heating?

Next, we used the city level data to check for any initial patterns as to whether all places show similar levels of heating. Figure 2(a) below shows the findings of plotting the temperature rise (using 10-year averages) of the 100 major cities (as defined by Berkeley Earth at http://berkeleyearth.lbl.gov/city-list/) against the absolute value of the city's latitude.

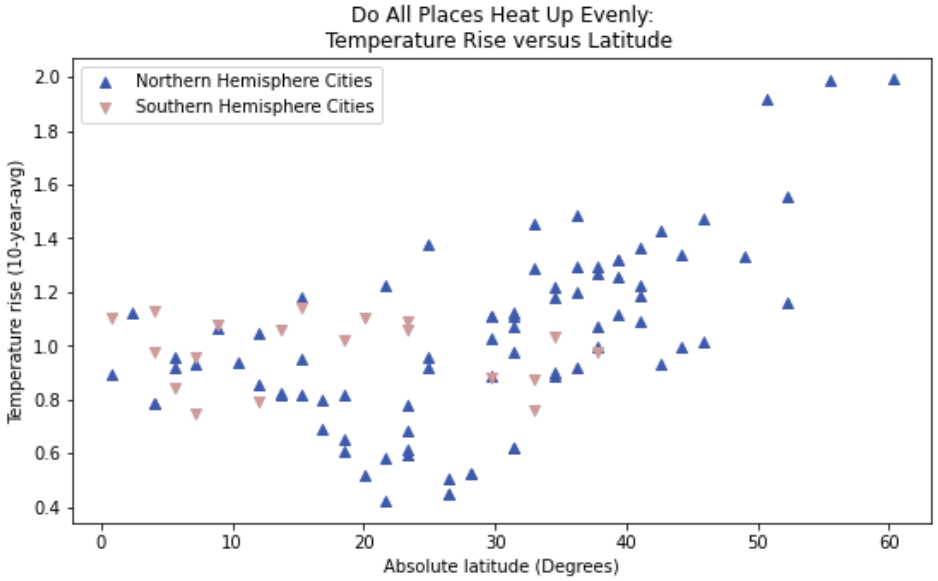


Figure 2. Scatter plot of temperature rise against latitude for 100 major cities

The initial observation is that the northern and southern hemispheres appear to show different heating patterns. Cities in the northern hemisphere seem to be heating up more, with the increase larger for places higher up in the north. It also seems that the above observation mostly applies to cities outside the tropical area (i.e. absolute latitude less than ~23°N).

3. And is the level of heating similar throughout the year?

Next, we used the global level data to see if the heating pattern of different months looks similar. Figure 3 below shows the time series data for the winter and summer months, defined respectively to mean Dec - Jan and Jun - Aug. It appears that colder months seem to heat up more than summer months[1]. A recurring theme seems to be showing up already - colder places/months seem to experience more heating?

[1] In the actual analysis, we will likely switch to the use of northern and southern hemisphere data for this part.

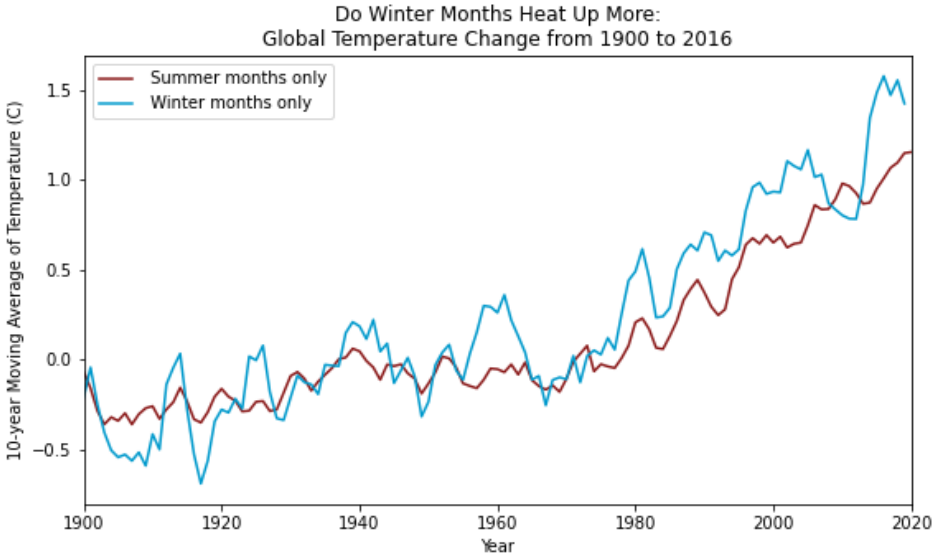


Figure 3. Change of global temperature with separate lines for winter and summer months

4. How does the trend of GHG emission look?

Initial explorations of the greenhouse (GHG) gas data revealed a steep, upward trend from 1850 to 2018 in the volume of global greenhouse gas emissions (Figure 4). Delving deeper into this upward trend, we identified the top ten GTG emissions countries to observe this general upward trend by country. (Figure 5) In our analysis, we'll plan to compare this data to the temperature variation data. In addition, we calculated the percentage change in GHG emissions between different time frames- we also plan to compare that to our temperature variation data.

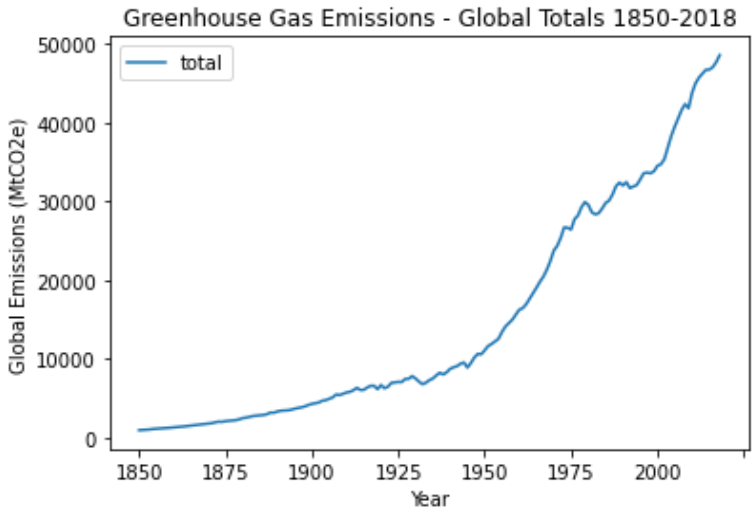


Figure 4. Upward GHG emissions trend 1850-present

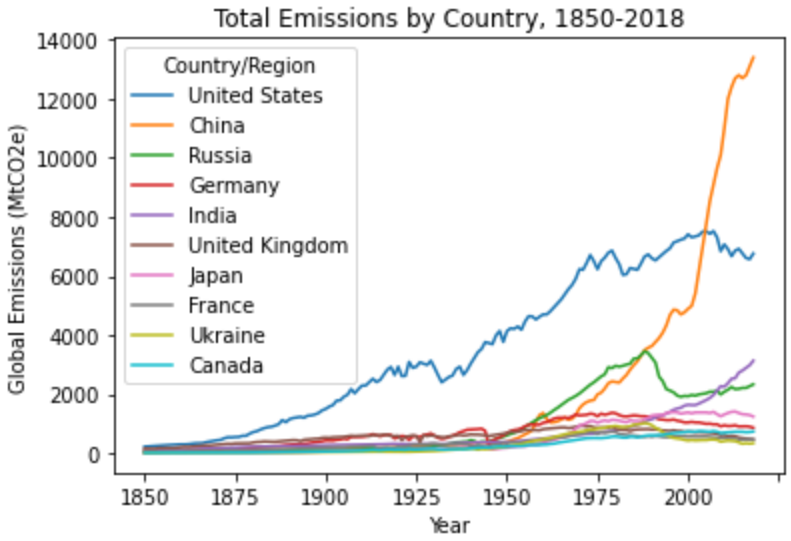


Figure 5. Upward GHG emissions trend, top 10 countries 1850-present

(IV) Challenges

From the initial exploration we have done so far, some of the challenges we envisage are:

1. **Data extraction:** data from Berkeley Earth are scattered across its website, e.g. high-level statistics are available in the global, country and city pages. As for the actual temperature data points, latitude/longitude, baseline temperature and other data of a country/city, they are all stored in separate text files (example), which need to be preprocessed and cleaned to make into a proper dataset.

2. **Data merging:** the data set from Berkeley Earth features data from multiple levels across a very long time horizon (1750 to present). As noted above, the quality of data for different levels, across different regions and in different time periods exhibits significant variation. This could present difficulties when we try to merge and compare the results of analyses performed using different subsets of data from Berkeley Earth, and will only be complicated when we mix in the GHG emissions and socio-demographic data from the UN which each has different geographical and time coverage.

3. **Data visualization:** since much of the analysis we plan to do will be about the different patterns of global warming and the effects for different regions/countries/cities, it could be difficult to visualize the findings using typical graphs (e.g. Figure 2 above). We plan to pick up and utilize geopandas which will require additional time commitment from the team.