Global Warming: A Detailed Look at Historical Trends, Variations and Issues of Burden and Responsibilities

W200 Project 2

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

Over the past decades, climate change and global warming have emerged as the greatest challenge that humanity has ever faced. Research has shown that the emission of greenhouse gases (GHG)[1] is a major contributor to the phenomenon, which is 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. In this project, we will (a) show how the trends of GHG emission and global warming have changed over time and whether we can observe any patterns between the two; and (b) explore whether the warming is taking place evenly, both in terms of different geographical locations and times of the year. Finally, we will look at the issues of cause and burden sharing, i.e. which countries have been major contributors in terms of GHG emission and which countries may potentially be most affected.

(II) Data sanity check and preprocessing

The primary data set used in the project is 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, regional, city and even weather station levels, and all these just for land-based data alone) and in different forms of measurement (monthly, annual, ten-year temperature anomalies, etc.). The data is also scattered across the website, with each data point having its own high level summary page and a separate text file providing detailed data. The first challenge was therefore for the team to extract the data for analysis. To this end, we wrote a number of in-house scripts to automatically fetch and preprocess the data to produce the data used in the analysis. This included the time series data itself, as well as other time-invariant data such as latitude and longitude.

In addition, the time series data from Berkeley Earth are temperature anomalies, which are important in climate change studies. However, we are interested in variables related to the absolute temperatures in some exploratory analyses, such as maximum/minimum temperature and temperature difference throughout the year. Therefore, we have also customized our scripts to generate absolute temperature data during the preprocessing step.

As for sanity check, the data from Berkeley Earth is generally of high quality, and the biggest issue is perhaps the presence of missing values which tend to be more common for certain types of data points, namely temperature data from further back in time (e.g. 18th and early 19th century) and for less

[1] Primarily carbon dioxide and methane, and nitrous oxides.

developed areas. Since different subsets of the Berkeley Earth dataset was used for different analysis (e.g. global data for time series analysis, city/hemisphere data for geographical analysis), the general protocol for our team was to perform sanity checks on the subset of data being used before carrying out each set of analysis. The scope of sanity checks typically involved finding the ranges of data where we can get continuous data without missing values, as well as looking for inherent biases for data points (e.g. data points concentrated in only several locations).

Finally, our team also made use of GHG emission and population data from the United Nations. These data are typically of reasonable quality as well, and the main challenge was the limited range of the data (e.g. GHG data only has data from the beginning of the 20th century) and in data matching (the UN data sets often use different conventions in grouping and naming countries/cities which required significant manual effort from the team in matching the country/city names). Despite this effort, ~20 countries within the Greenhouse gas data set remained unmatched with the Berkeley Earth datasets. This discrepancy had no major impact on the analysis.

(III) Results

(a) Initial analysis: How global is global warming?

While there has been wide consensus that global warming is a worldwide issue, we consider it useful to first get a general picture of the phenomenon before delving into the nurances. We first used the Berkeley Earth data[2] to find the most/least heated countries. **Figures 1a and 1b** show the temperature increasing rate for (a) the top ten most heated countries and (b) the ten least heated countries since 1960. Even for those least heated countries, their temperatures have risen at least 0.65 °C/century. Global warming is indeed a worldwide issue.

a. Most heated countries

b. Least heated countries

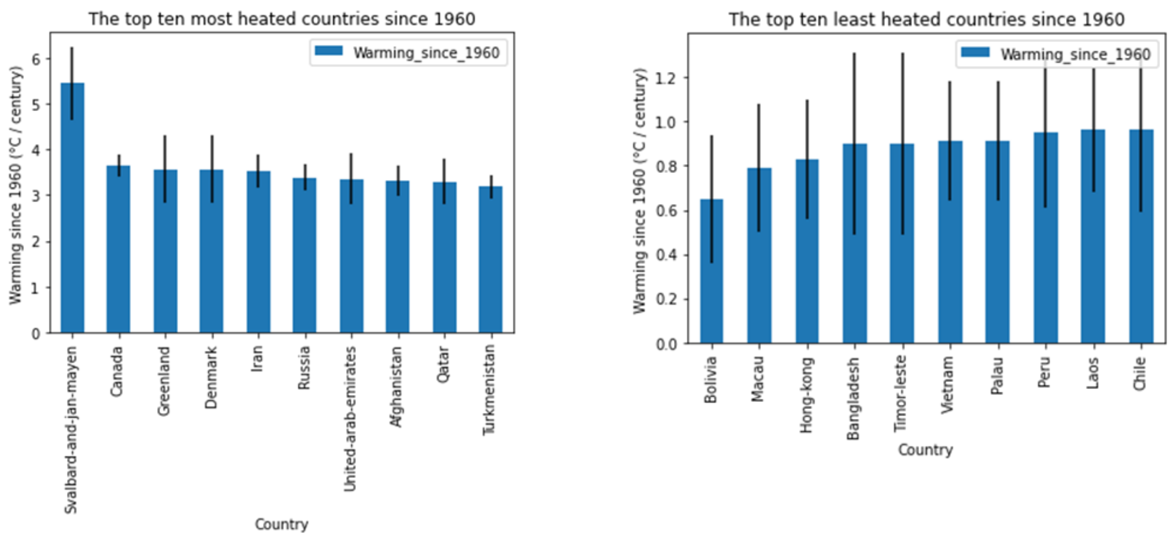


Figure 1. The most/least heated countries since 1960.

[2] As noted above, Berkeley Earth provides a wide range of time-series data at different levels. For this initial exploration, the data was from the summary page for countries, i.e. http://berkeleyearth.lbl.gov/country-list/

(b) Research questions

With the initial exploration completed, we proceeded to look into the nuances of global warming along the following themes:

Theme 1: Have global warming trends evolved over time?

1.1: How do the heating rates change over the year for top-heated countries?

1.2: How does the trend of GHG emission look?

Theme 2: Is global warming taking place evenly?

2.1: Do different times of the year heat up evenly?

2.2: Do all places experience the same level of heating?

2.3: Is there a discernible difference in warming between the two hemispheres?

Theme 3: Which countries are the largest contributors and which countries are most affected?

3.1: Which countries are the largest contributors?

3.2: Which countries are most affected?

(c) Theme 1: Has the global warming trends evolved over time?

1.1 How do the heating rates change over the year for top-heated countries?

We are interested in the heating rates for top-heated countries, which were represented by the slope of the fitting line to the temperature time series. We firstly added fitting lines to the temperature data after 1850 and after 1975, respectively, as shown in **Figures 2a and 2b**.

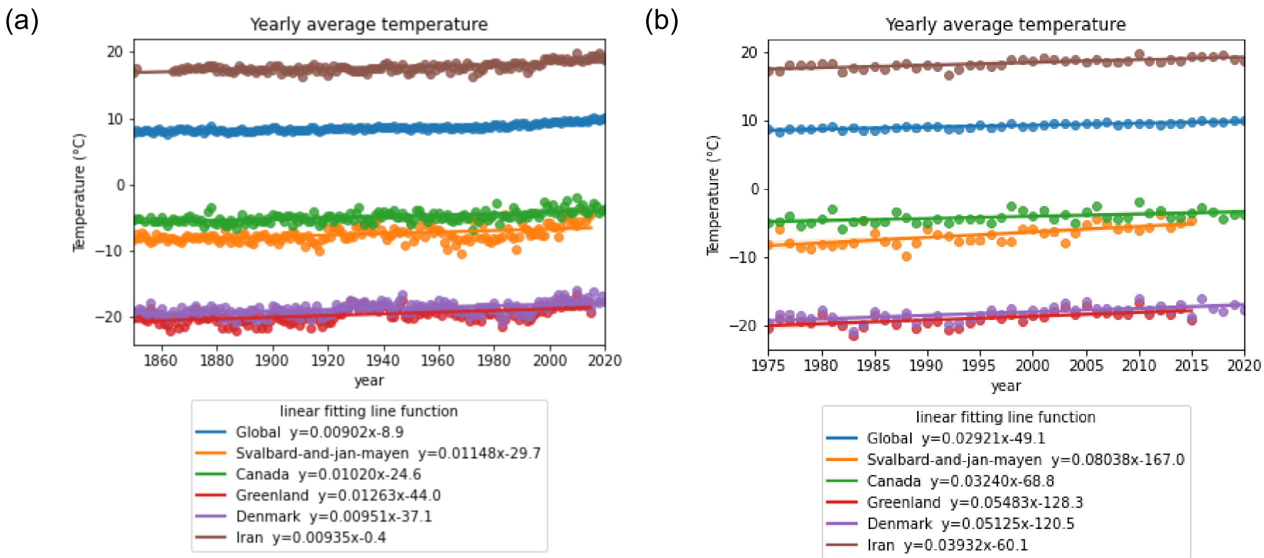


Figure 2. Yearly average temperature after 1850 (a) and after 1975 (b) with linear fitting lines.

Table 1 shows the slopes in Figure 2. The slopes representing the heating rates in the second column are much higher than those in the first column, indicating the temperature increase has been largely accelerated over the last 45 years. **Figure 3** is shown as one example to exhibit this trend.

	linear fitting slope after 1850	linear fitting slope after 1975
Global	0.009024	0.029208
Svalbard-and-jan-mayen	0.011481	0.080377
Canada	0.010198	0.032397
Greenland	0.012633	0.054833
Denmark	0.009515	0.051250
Iran	0.009351	0.039317

Table 1. Slope values for fitting lines from Figure 2

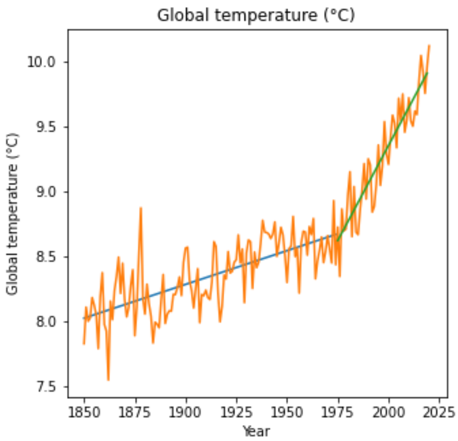


Figure 3. Yearly global temperature with fitting lines

1.2 How does the trend of GHG emission look?

Globally, greenhouse gas emissions have risen every year since 1850 (**see Figure 4)** - in 2018, an estimated 48,601 metric tons of carbon dioxide equivalent were emitted globally. Burning fossil fuels are the largest contributor to greenhouse gases emissions, which, in turn, is the largest contributor to climate change.

Greenhouse gas emissions rose steadily from 1850 until the mid-1900's at which point global emissions reached an inflection point. Around 1935, the rate of global emissions increased considerably. **Figure 5** depicts this sharp uptick in emissions which corresponds to the sharp uptick in warming rates in subsequent decades.

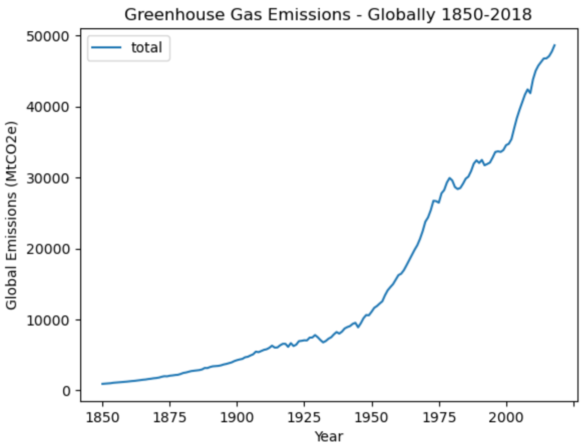


Figure 4. Global GHG Emissions Trend

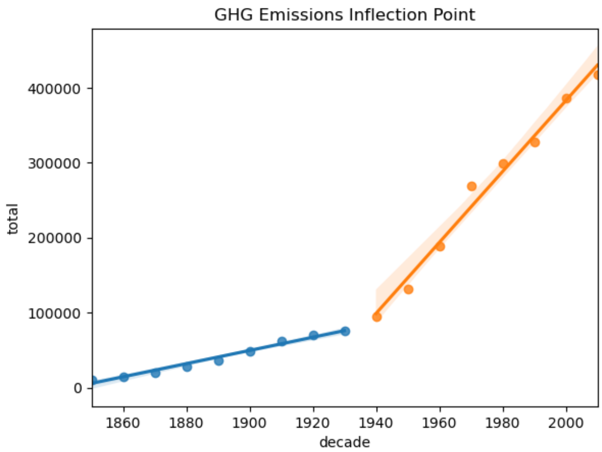


Figure 5. GHG Trend by Decade

(d) Theme 2: Is global warming taking place evenly?

For this theme, we are interested in understanding whether there are variations in the patterns of global warming for (a) different seasons/months of the year; and (b) for different geographical locations (hemisphere, latitude, longitude) as well as neighborhood environment (presence of water bodies).

2.1:Do different times of the year heat up evenly?

For this question, we made use of the two datasets for the northern and southern hemispheres[3], and observed the heating patterns for different seasons[4]. Analysis was done by grouping the monthly temperature-rise data into seasons and calculating the 10-year running average[5]. **Figures 5a and 5b** show the results.

a. Northern hemisphere b. Southern hemisphere

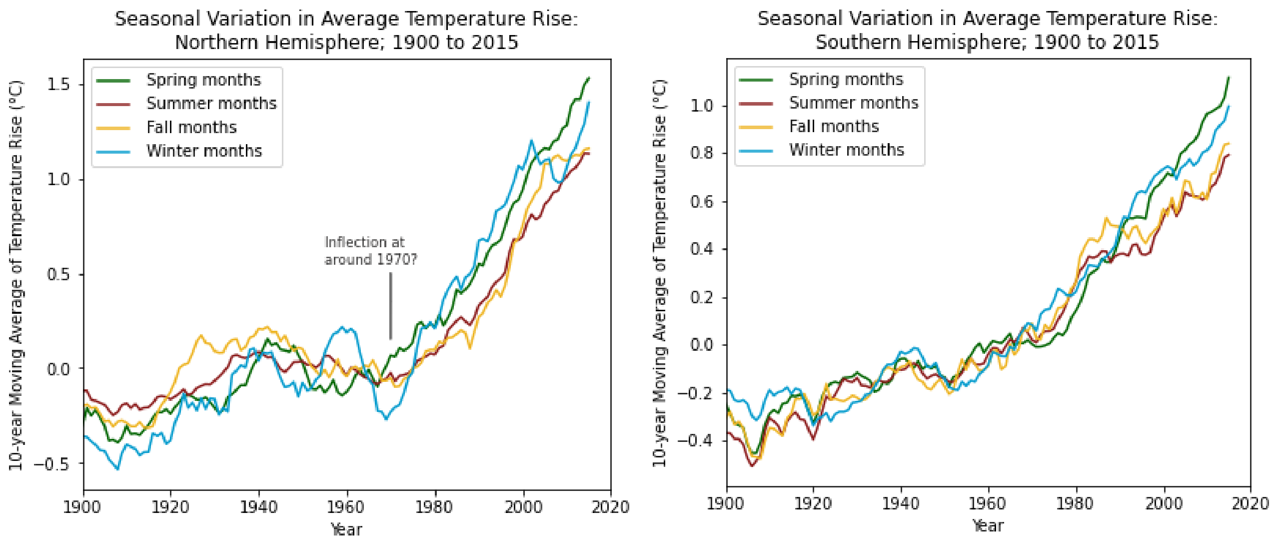


Figure 5. Seasonal variation in temperature rise for the two hemispheres; 1900 - 2015

From Figures 5a and 5b above, it can be seen that there are visible differences between different seasons, with months in winter (coldest season) and spring (2nd coldest season for the northern hemisphere) warming up the most. This will be a recurring observation - **colder places/months tend to heat up more**. Interestingly, the northern hemisphere graph also seems to have an inflection point in the 1970s when heating rates picked up, echoing the trend observed in the analysis for sub-question

---

3 That is, http://berkeleyearth.lbl.gov/regions/northern-hemisphere and http://berkeleyearth.lbl.gov/regions/southern-hemisphere.

4 We adopted the common meteorological definition for the four seasons. Specifically this means: (vide: https://www.ncei.noaa.gov/news/meteorological-versus-astronomical-seasons)

Winter:	Dec - Feb	*(northern hemisphere)*	&	Jun - Aug	*(southern hemisphere)*
Spring:	Mar - May	*(northern hemisphere)*	&	Sep - Nov	*(southern hemisphere)*
Summer:	Jun - Aug	*(northern hemisphere)*	&	Dec - Jan	*(southern hemisphere)*
Fall:	Sep - Nov	*(northern hemisphere)*	&	Mar - May	*(southern hemisphere)*

The meteorological definition is used as it is mainly defined in a way that, for each hemisphere, winter is the 3 coldest months of the year while summer is the 3 hottest months. This fits our purpose as we are primarily interested in temperature variations. That said, before adopting this definition, an analysis was done on the average monthly temperatures for the period Jan 1951-Dec 1980 for each hemisphere to verify that for our dataset, the 3 coldest and hottest months are indeed the ones set out above.

5 This follows the tradition adopted for most Berkeley Earth headline statistics by using 10-year averages, which has the benefit of reducing the distraction from noise.

This trend, however, is not as observable for the southern hemisphere[6], providing a first hint that **the two hemispheres could be experiencing different heating patterns**.

2.2: Do all places experience the same level of heating?

For this question, we compared the temperature rise of cities[7] against three variables, namely (a) latitude; (b) longitude; and (c) presence or amount of water in the neighbourhood of a location[8]. As opposed to previous analysis which relied on global/hemisphere/country level data, this analysis made use of city level data to avoid issues posed by geographically large countries.

(a) Latitude

Figures 6a and 6b below show the results of scatter plots showing all 3 523 cities. Based on the graphs,for the northern hemisphere, the warming rate seems to be higher the further north we go. Considering that cities higher in the north are generally colder[9], the recurring observation that **colder places/months tend to heat up more can be observed here again**. **The same cannot be said for the south**, however, as the warming rates seem to stay within 0.5°C to 2.5°C irrespective of latitude.

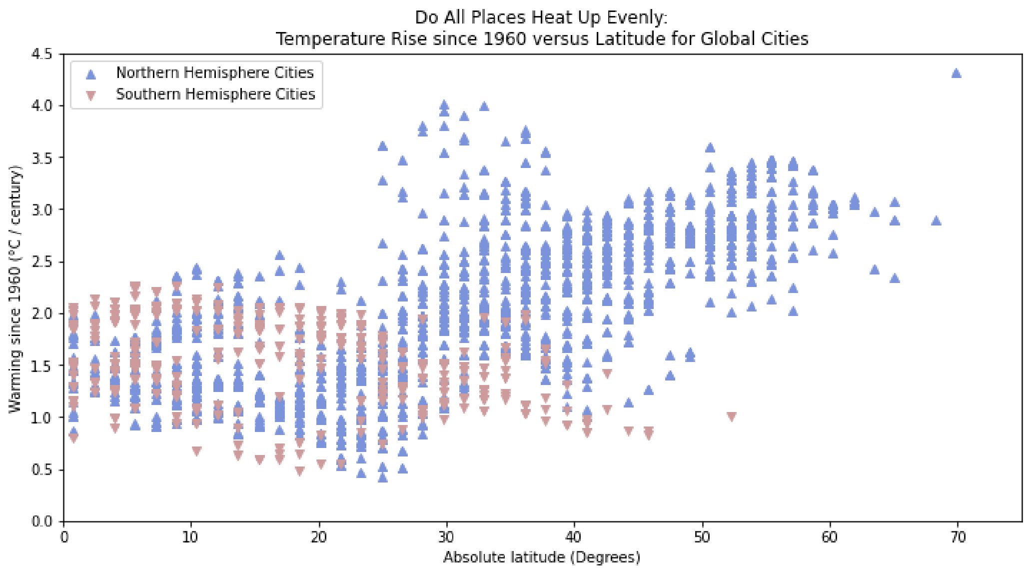


Figure 6a. Temperature variation against city latitude (both hemispheres)

[6] This can of course be due to biases within the southern hemisphere data, for example, the data points could be skewed in terms of the distribution in latitude and longitude. But as we proceed with the analysis on the city-based data under sub-questions 2.2 and 2.3, this appears not to be the case.

[7] This included a total of 3 523 cities. In the project proposal, we looked only at the 100 largest cities identified by Berkeley Earth (vide: http://berkeleyearth.lbl.gov/city-list/). However, this made it much more difficult to identify clear patterns. For example, of these 100 cities only 19 are from the southern hemisphere. When we expand the data set to the 3 523 cities (identified by going to the alphabetical listing page for each alphabet, e.g. cities starting with letter 'A' can be found at: http://berkeleyearth.lbl.gov/city-list/A), the data proved to be much more nuanced, e.g. there are now 616 cities with latitudes below 0.

[8] Rationale for choosing these three variables:
Latitude: when we eyeballed the data, it appeared that cities higher up in the north generally experienced higher rates of heating.
Longitude: helps contextualize findings from latitude, given that location is just a factor of latitude and longitude
Amount of water: we always had the implicit assumption that water helps mitigate temperature changes (due to its large heat capacity, for example). Berkeley Earth conveniently provides data on this for every city level data point.

[9] This can be easily verified from the data. Supplementary plots showing this can be provided upon request.

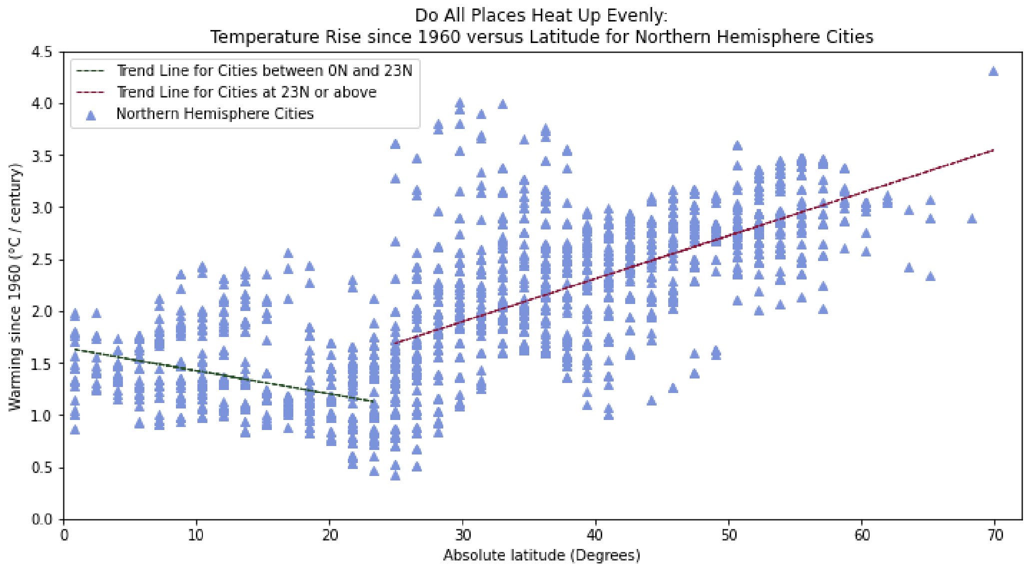


Figure 6b. Temperature variation against latitude for cities in northern hemisphere[10]

(b) Longitude

From Figures 6a and 6b above, there appears to be some outlier groups for the northern hemisphere cities (at latitudes of around (a) 8°N to 20°N; and (b) 23°N to 40°N). We therefore attempted to see if combining a longitudinal grouping (i.e. by defining a "natural time zone" categorization[11]) will help shed light on where these data points are. **Figure 7** below shows that these points are in fact found in the "natural time zones" of "UTC+3" and "UTC+4" (i.e. longitudinal values of 37.5° - 52.5° and 52.5° - 67.5°)

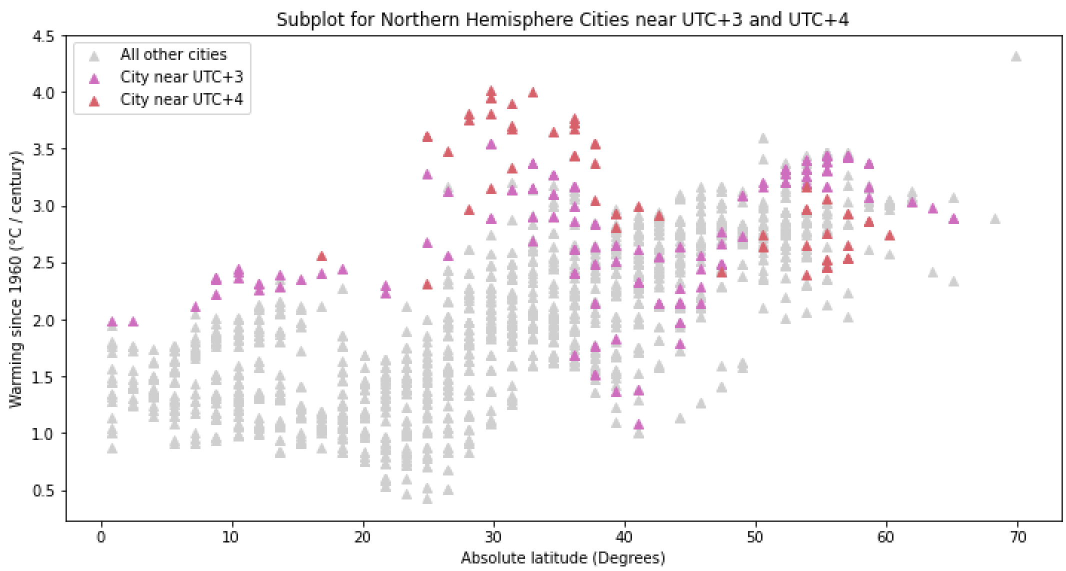


Figure 7. Subplot highlighting cities with "natural time zones" in 'UTC+3' and 'UTC+4' [12]

---

[10] Supplemental graph for just the southern hemisphere at Appendix.

[11] Put simply, we divided the full longitudinal range of -180° to +180° into 25 sections, with a span of 15° each for the 23 sections counting from the GMT line, corresponding to something akin to UTC-11 to UTC+0 to UTC+11. The last two sections have a span of 7.5° each, similar to UTC-12 and UTC+12. This is not unlike how real timezones work, though for real timezones, the categorization will depend not only on the longitudinal value of a location, but also any special choices made by the government of that location. For our study, we are only interested in the longitudinal location, so we call our categorization the "natural time zone".

[12] A full set of subplots for all the "natural time zones" found in the northern hemisphere can be found in the Appendix. While the outliers are found predominantly in "UTC+3" and "UTC+4", points of other "natural time zones" also show some degree of clustering.

And if we take a look at the countries having these outlier cities, we can see that there are only 15 of them, containing mostly countries in the Middle East and near Arabia (**Figures 8a and 8b**).

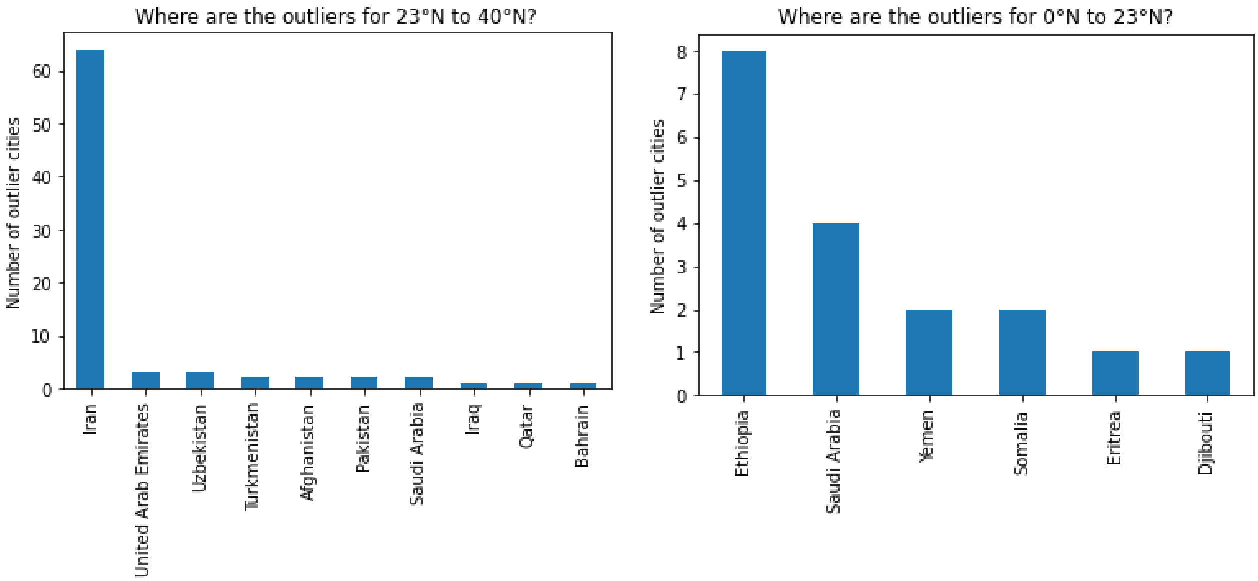


Figure 8. Countries with outlier cities near (a) 23°N to 40°N and (b) 8° to 20°N

A further attempt was made to plot all the cities on a world map in **Figure 9** below, which provides a clear picture of the trends noted in previous sections: (a) cities in the northern hemisphere generally heat up more (redder in color) while those in the southern hemisphere show more uniform heating rates; (b) warming rates typically increases as one goes up north; and (c) the existence of outliers for the 15 countries noted above (in red and orange color as opposed to yellow).

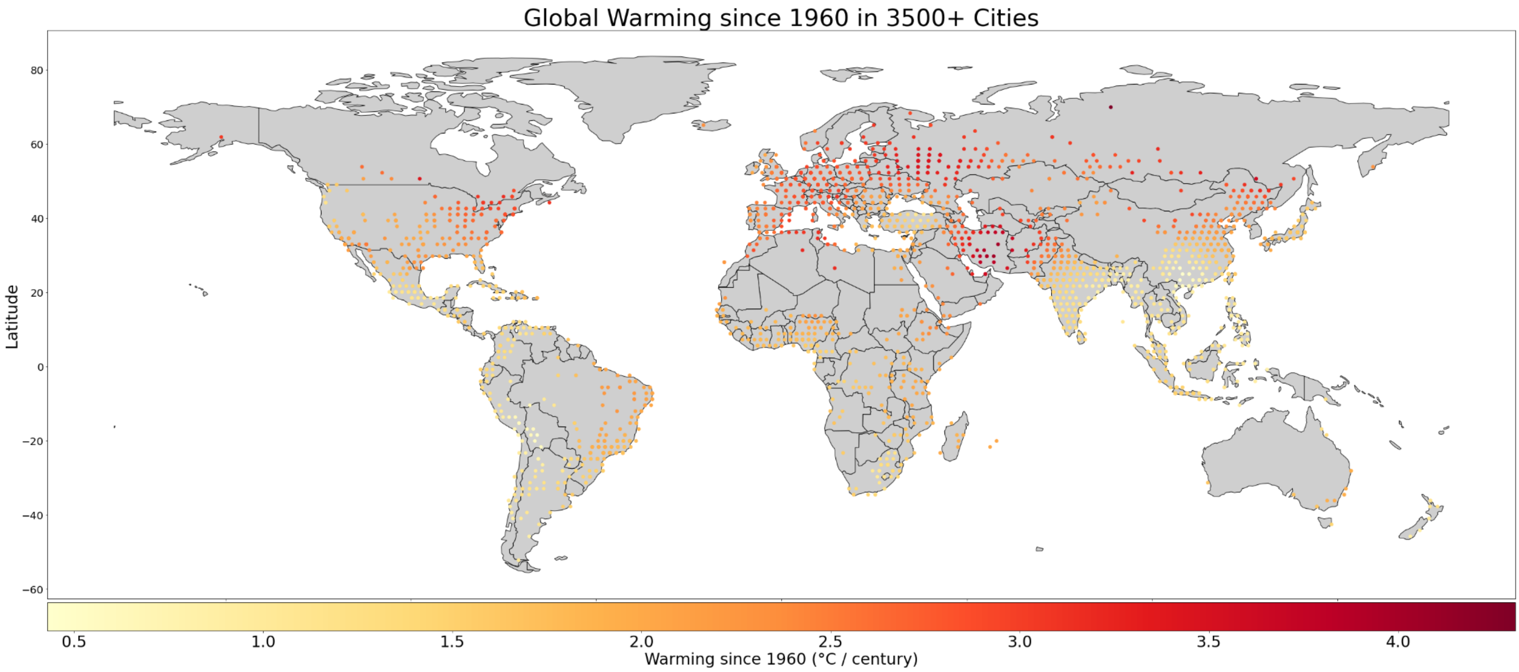


Figure 9. Subplot highlighting cities with “natural time zones” in ‘UTC+3’ and ‘UTC+4’

(c) Presence or amount of water near city

Contrary to what one may expect, the presence of water does not seem to have any noticeable effects. **Figure 10** below shows the result of a plot of warming rates versus water presence.

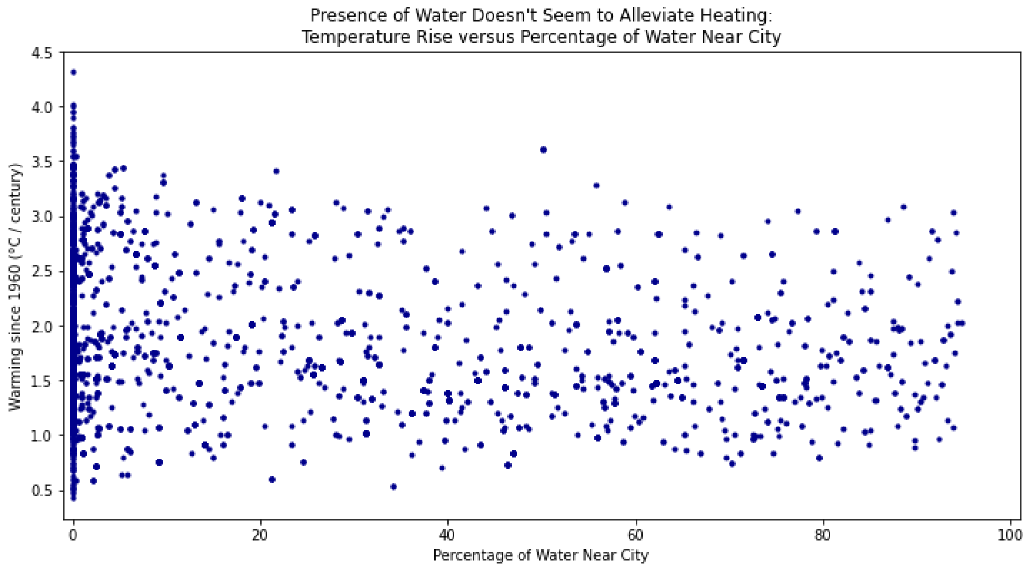


Figure 10. Warming rates versus water presence near city

2.3:Is there a discernible difference in warming between the two hemispheres?

As observed from Figures 5, 6 and 9 above. The two hemispheres do appear to exhibit different heating patterns. In general, the northern hemisphere has been heating up more and showing more variation. The southern hemisphere, in contrast, exhibited lower rates and a greater homogeneity.

Observation from Theme 2

Based on the analysis for sub-questions 2.1 to 2.3 above, the key takeaway is that, while global warming is a worldwide phenomenon, the pattern of heating is not uniform. Variation is observed between the heating rates and patterns for the hemispheres, while different countries and different times of the year experience a varying degree of regional warming. In general, colder seasons and colder places tend to be affected more, though major outliers can also be found.

(e) Theme 3: Which countries are the largest contributors and which countries are most affected?

3.1: Which countries are the largest contributors?

Today, just 10 countries are responsible for nearly 60% of greenhouse gas emissions . The United States, China, Russia, Germany, and India have contributed most to greenhouse gas emissions, since 1850 (**Figures 11, 12**).

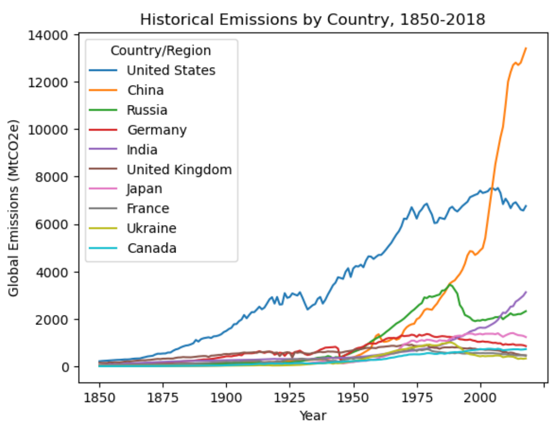


Figure 11. Historical Emissions by Country

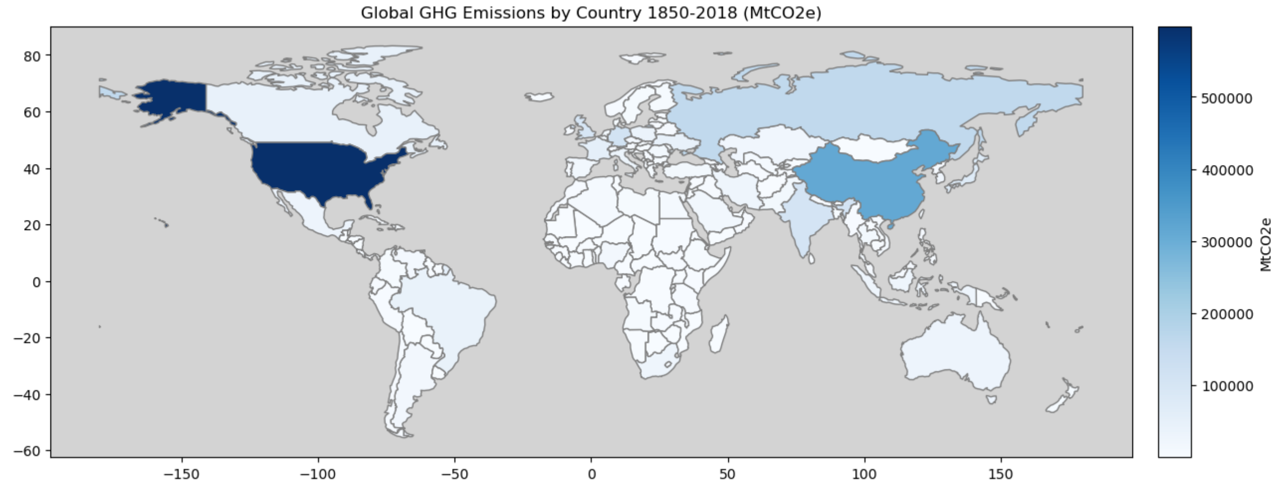


Figure 12. Global GHG Emissions 1850-2018 (MtCO2e)

3.2: Which countries are most affected?

Despite their outsized contributions to the issue of global warming, the highest emission countries have not experienced the same outcome/burden that other, lower emissions countries have faced. **Figure 13** shows a value-by-alpha map of the comparison of global emissions to temperature increase since 1960.

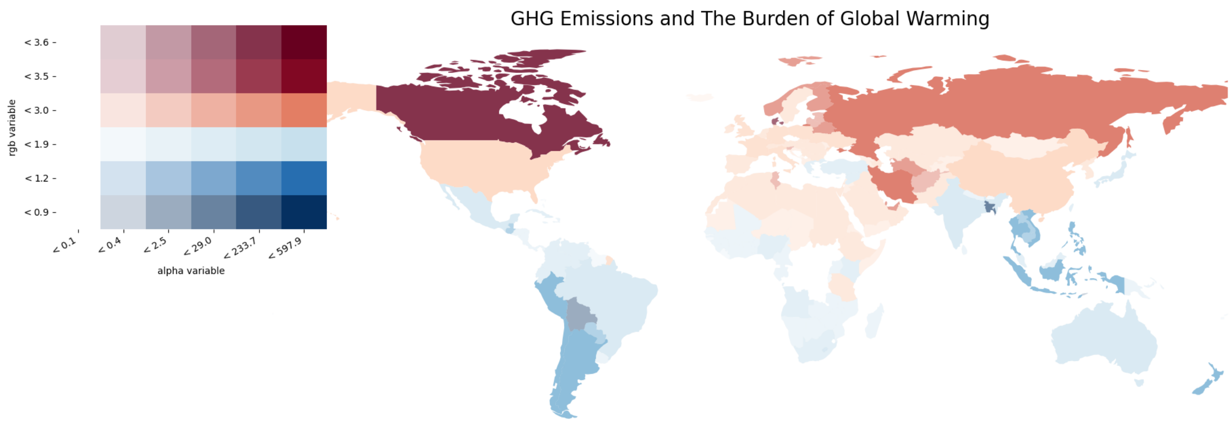


Figure 13. Disproportionate impact of global warming on low emissions countries

Countries with colors toward the top left of the legend have experienced the highest warming trends but have emitted the least amount of greenhouse gases. Countries with colors toward the top right of the legend have experienced the highest warming trends but have also emitted the most greenhouse gases. On the bottom left of the legend are countries that have not experienced warming or emitted large amounts of greenhouse gases, and finally, the colors on the bottom right of the legend represent countries that have emitted large amounts of greenhouse gases but have not seen major warming trends.

A further analysis has also been carried out by merging data from the United Nations World Urbanization Prospects 2018[13] with the Berkeley Earth data. The UN datasets provide an additional

[13] The dataset is available at https://population.un.org/wup/Download/. Specifically, the data from "File 12: Population of Urban Agglomerations with 300,000 Inhabitants or More in 2018, by country, 1950-2035 (thousands)" is used.

layer for us to filter out the cities from Berkeley Earth that are actually large cities (> 300k population in 2018) and look at which of these cities may actually reach uninhabitable temperatures (defined by us as having any month with monthly average temperature above 35°C[14]) if the existing warming rate continues.

As one can observe from **Figure 14** below, which shows a map plot showing the projected location of cities reaching monthly average temperature above 35°C in 2100 assuming the current warming trend continues, a majority of the cities most affected by global warming are concentrated in a few regions.

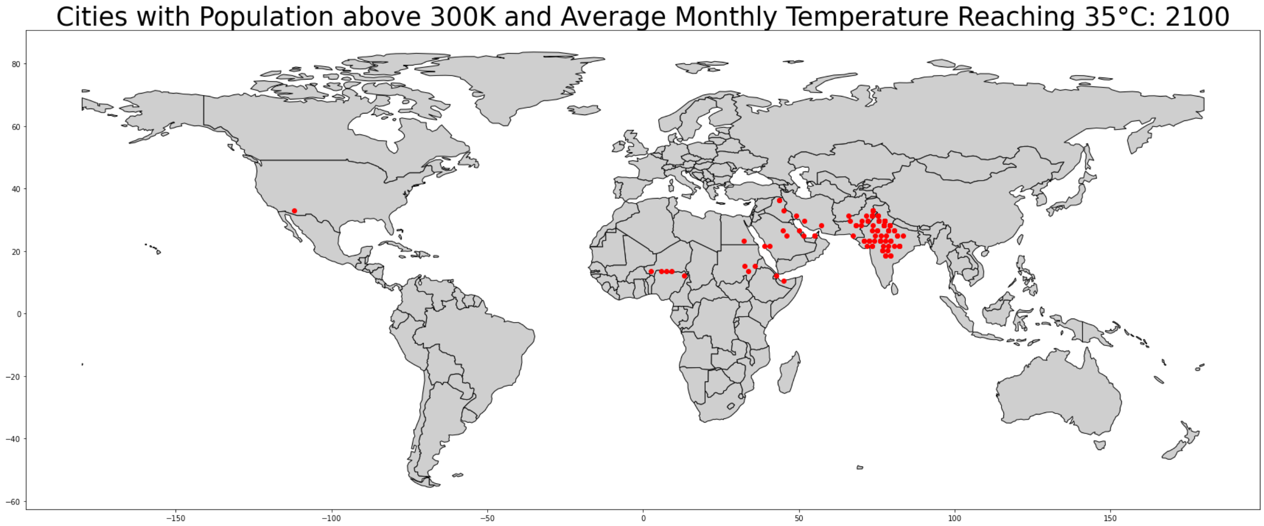


Figure 14. Cities most affected by global warming and their locations[15]

Observation from Theme 3

In summary, we see some of the top emissions countries like the United States, China, and Russia, have only experienced moderate warming trends while countries like Estonia, Norway, Afghanistan, and Turkmenistan have experienced the largest warming trends while emitting much less greenhouse gas. As part of the top 10 ten emissions countries, Canada has emitted 89,322 MtCO2e since 1850, yet is experiencing larger warming trends than the United States which has emitted 1,195,872 MtCO2e- nearly 13 ½ times as much as Canada.

Global warming is a global issue, however the inequity between high emissions countries and countries that must bear the burden of these actions has grown.

[14] A threshold of 35°C may seem low at first glance. However, it has to be remembered that this is going to be the average temperature. If we look at the relevant Wikipedia page (https://en.wikipedia.org/wiki/List_of_cities_by_average_temperature), few countries actually have average temperatures above 30°C even in the hottest month. If we pick Assab, a city with average temperature in July at 35.0°C, the average daily high temperature in July is already at 41.2°C. The same goes for Bosaso, with average temperature in June and August at 35.6°C, corresponding to average daily highs of 41.0°C and 40.0°C.

[15] In preparing this plot, we have also produced a series of similar plots showing the situation from around 1960 (the Baseline Scenario), in 2020, 2040, 2060 and 2080. A picture in GIF format showing these map plots in succession is available at https://drive.google.com/file/d/1iXdOt8LntDn22yJe4WF5u-POsPZgXh1c, accessed by anyone from the UC Berkeley group.

Appendix

Additional Results of the Data Exploration

Purpose

This Appendix presents other findings of our exploratory analysis which have not been included in the main report.

Additional Findings

(i) How do the temperatures change over the year for top-heated countries?

In the initial analysis section, we identified the top 10 most heated countries and the 10 least heated countries since 1960. As a follow-up, we conducted a detailed analysis of the time series data for the top 5 most heated countries and compared their results with global temperature data. Our research focused on the overall temperature variation, yearly temperature variation, ten-year average temperature, and heating rate.

Figure a.1 shows the temperature difference between the maximum and minimum temperatures for the top 5 most heated countries and the whole world. It was demonstrated that the temperature differences in the top heated countries are much higher than the global temperature difference, which is above 30 °C.

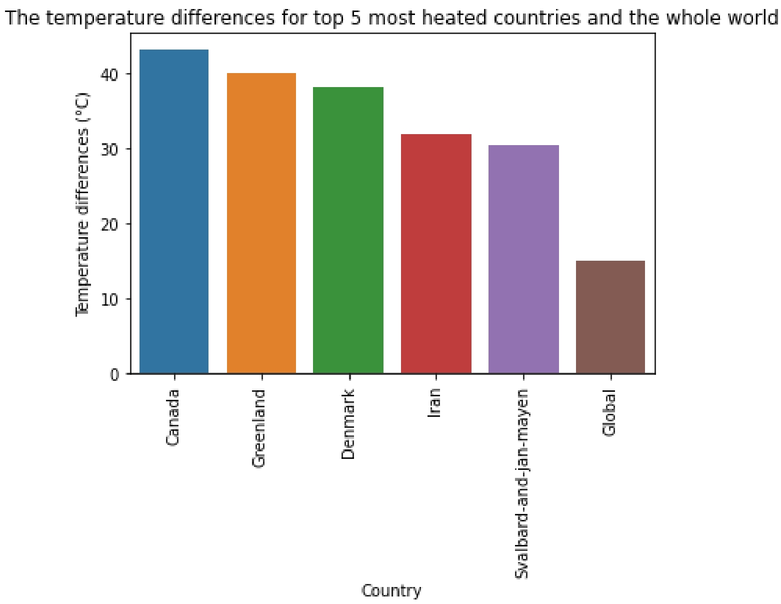


Figure a.1. The temperature differences for the top 5 most heated countries and the whole world.

Figures a.2 and a.3 show the yearly/ten-year average temperature, and yearly/ten-year averaged temperature difference (maximum-minimum temperatures) for the five countries and the global area. The average temperatures all show an increasing trend over the year. And it was demonstrated that the global temperature difference decreased monotonically over the year. While all of the country data maintain the decreasing trend, the trend is less distinct. A significant fluctuation can be observed for several countries, such as Iran and Svalbard-and-jan-mayen.

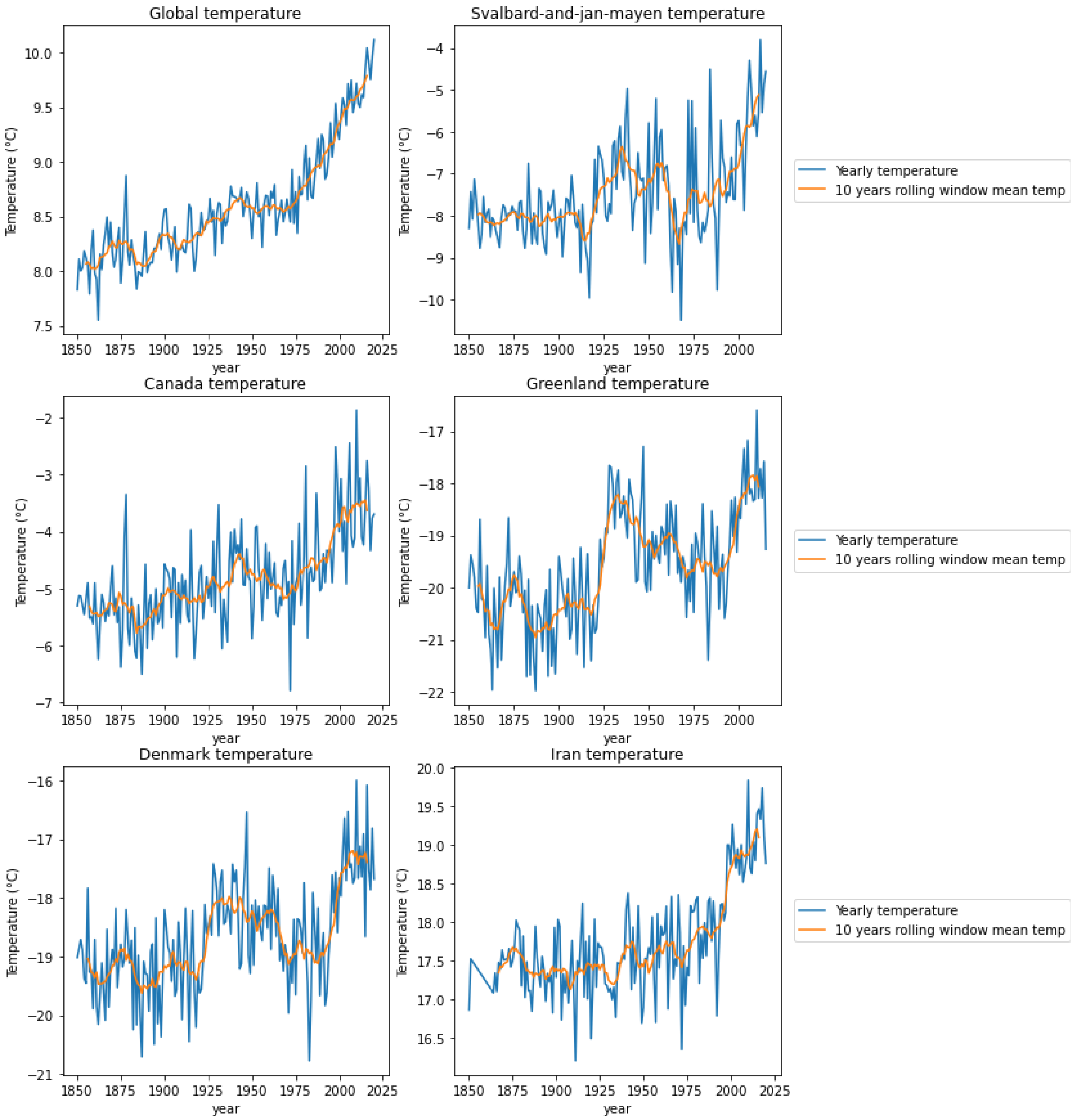
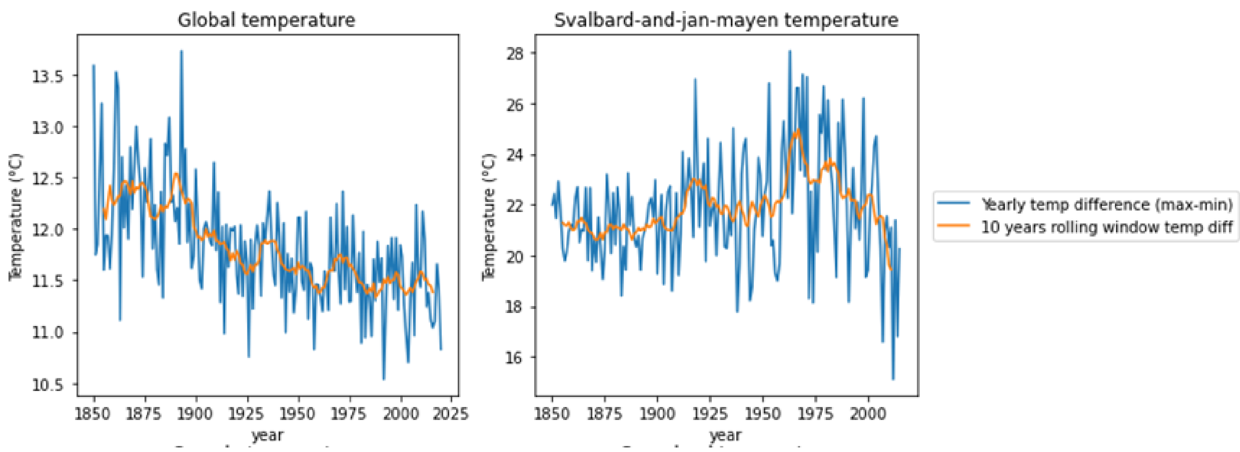


Figure a.2. Averaged temperature by areas



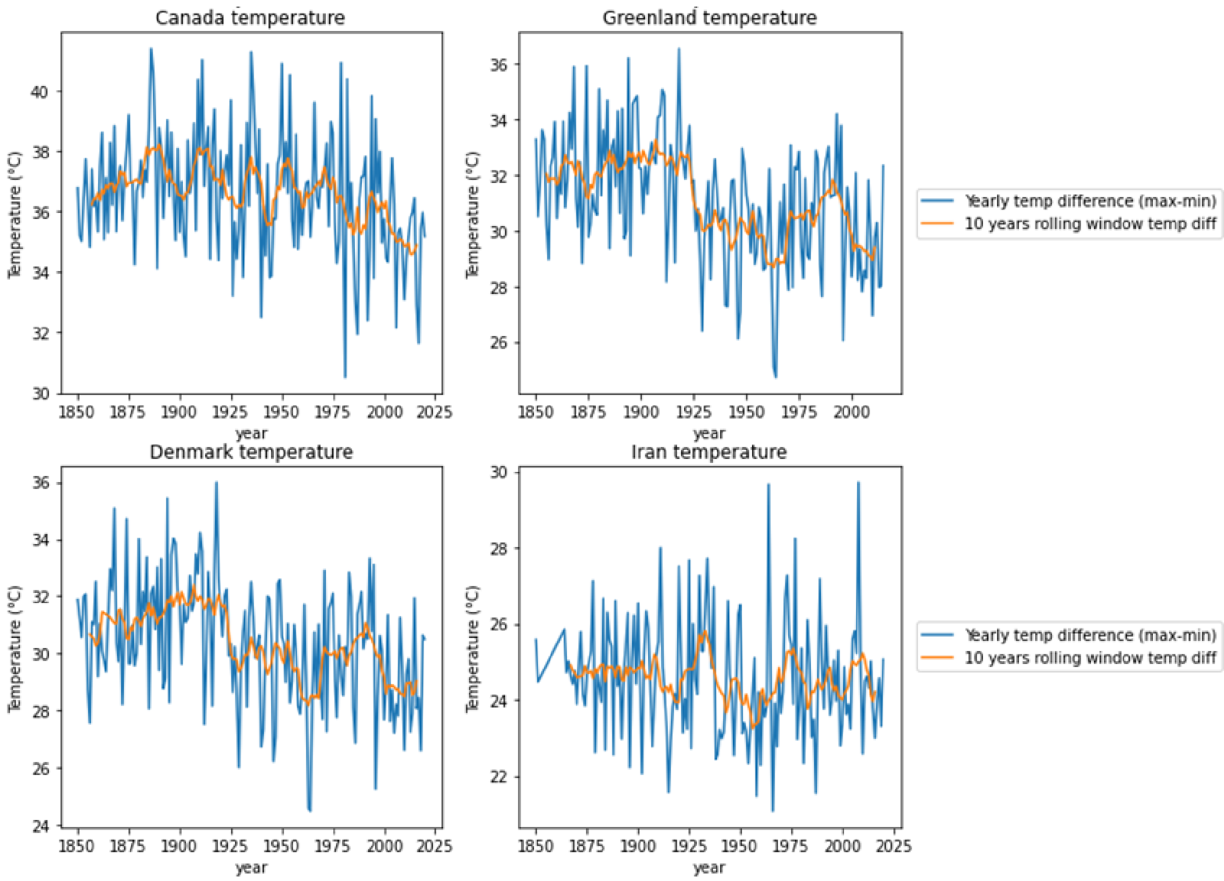


Figure a.3. The temperature difference by areas

(ii) Additional plots for the temperature data for the two hemispheres

Under part (a) of the analysis for sub-question 2.2, we looked at the temperature rise versus latitude plots which seem to show a different heating pattern for the two hemispheres. Here, we provide an additional graph (**Figure a.4**) showing just the data points for the southern hemisphere and in the same scale as the plots in Figures 6a and 6b. This additional plot helps visualize how the southern hemisphere seems to show no visible variation in temperature rise level relative to changes in latitude.

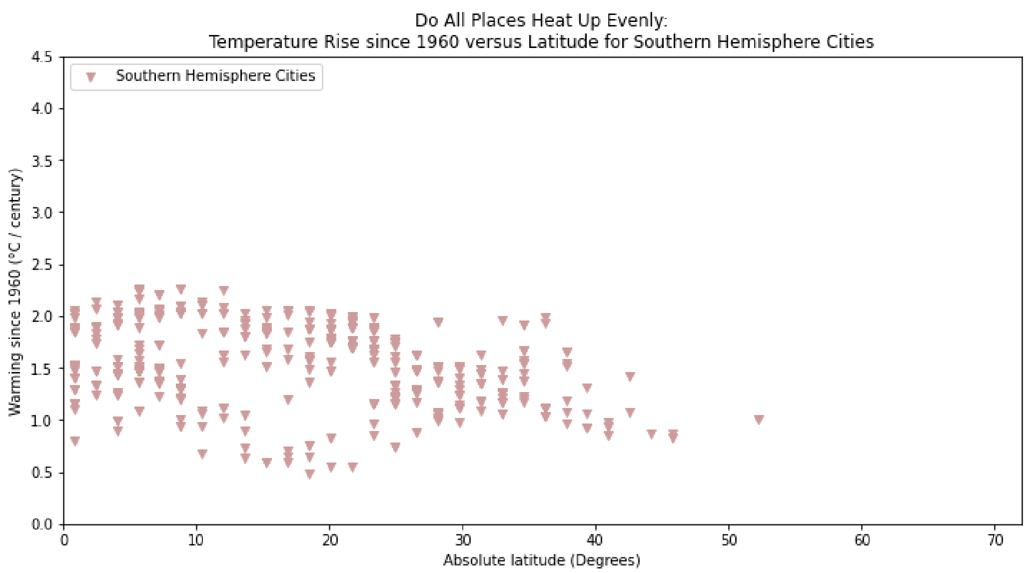


Figure a.4. Temperature variation against latitude for cities in southern hemisphere *(ii) Additional plots for the temperature data for different "natural time zones"*

Under part (b) of the analysis for sub-question 2.2, we made use of the longitude data to identify the location of the outlier data points from the northern hemisphere. As mentioned in the footnote for Figure 7, the plots for the different "natural time zones" also show varying degrees of clustering. These additional plots (Figure a.5 and Figure a.6) are provided here for the interested reader.

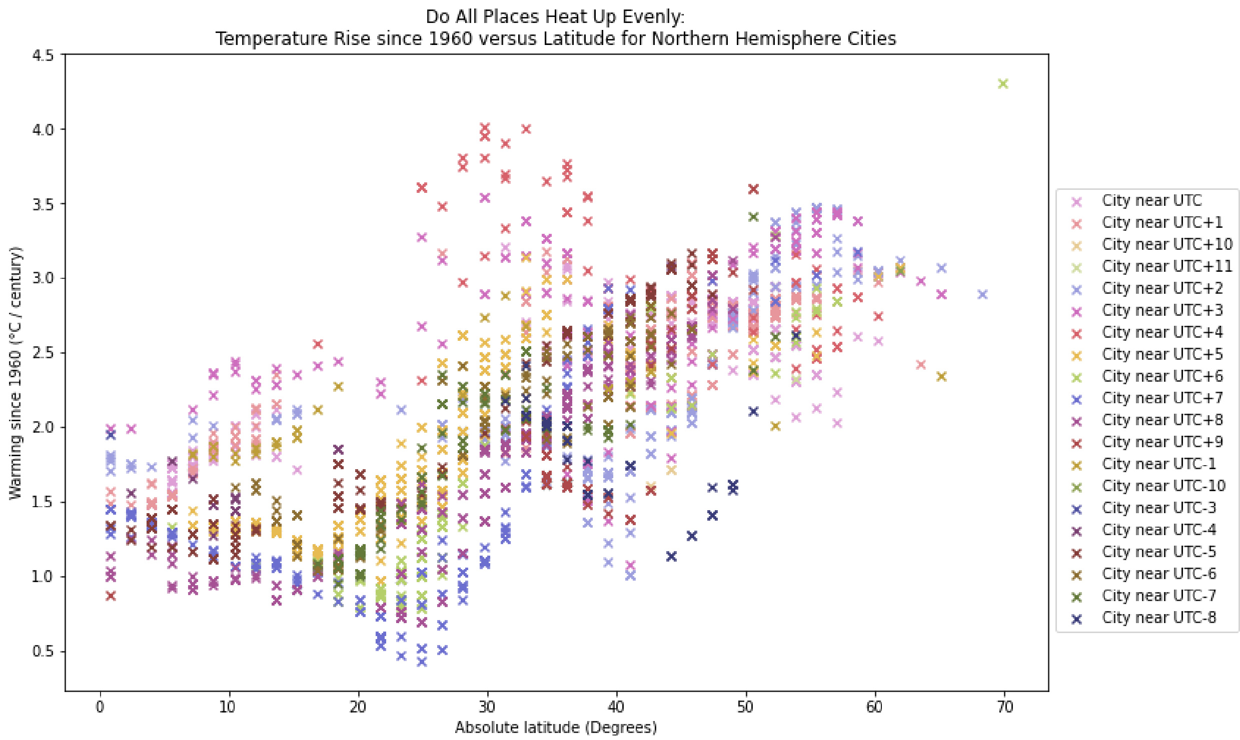
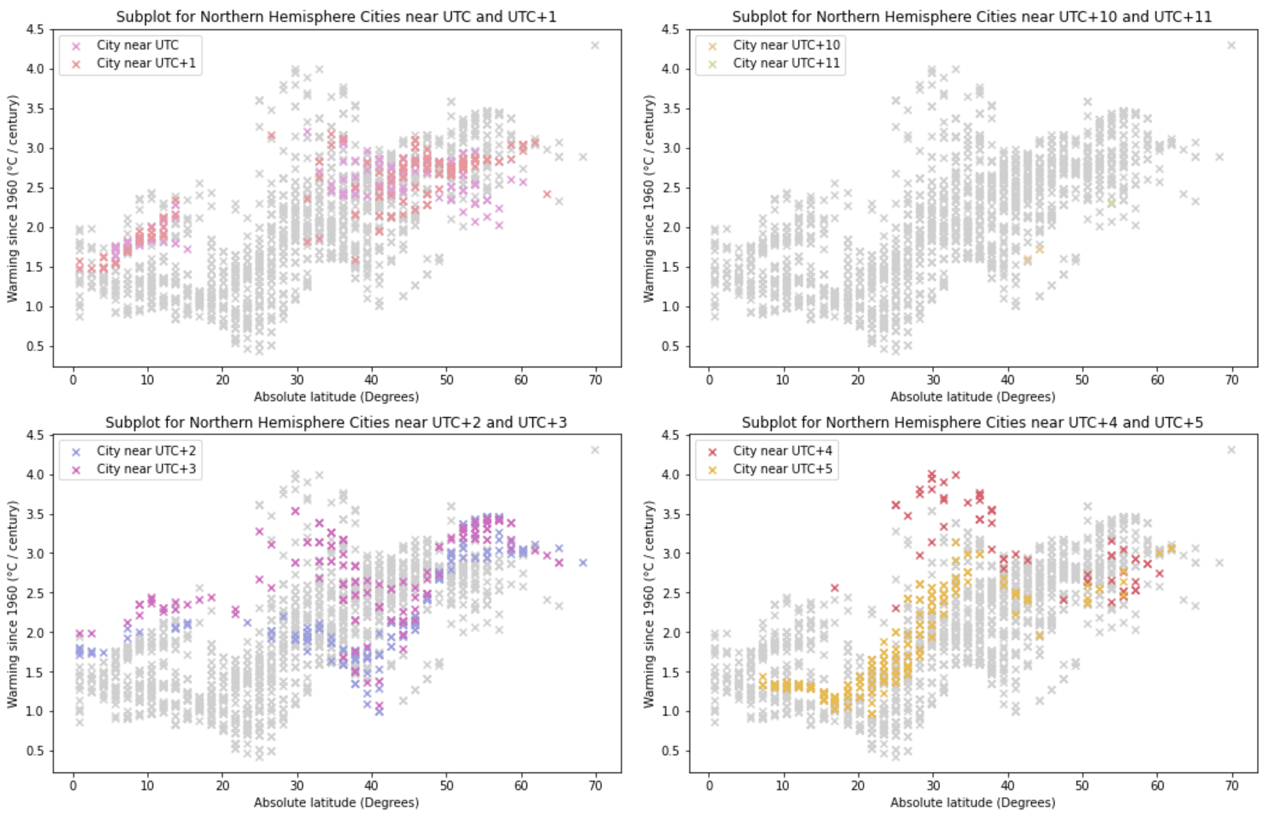


Figure a.5. Overall plot of all "natural time zones"

Note: this is slightly messy to look at but its purpose is to just show the general shape of the data.



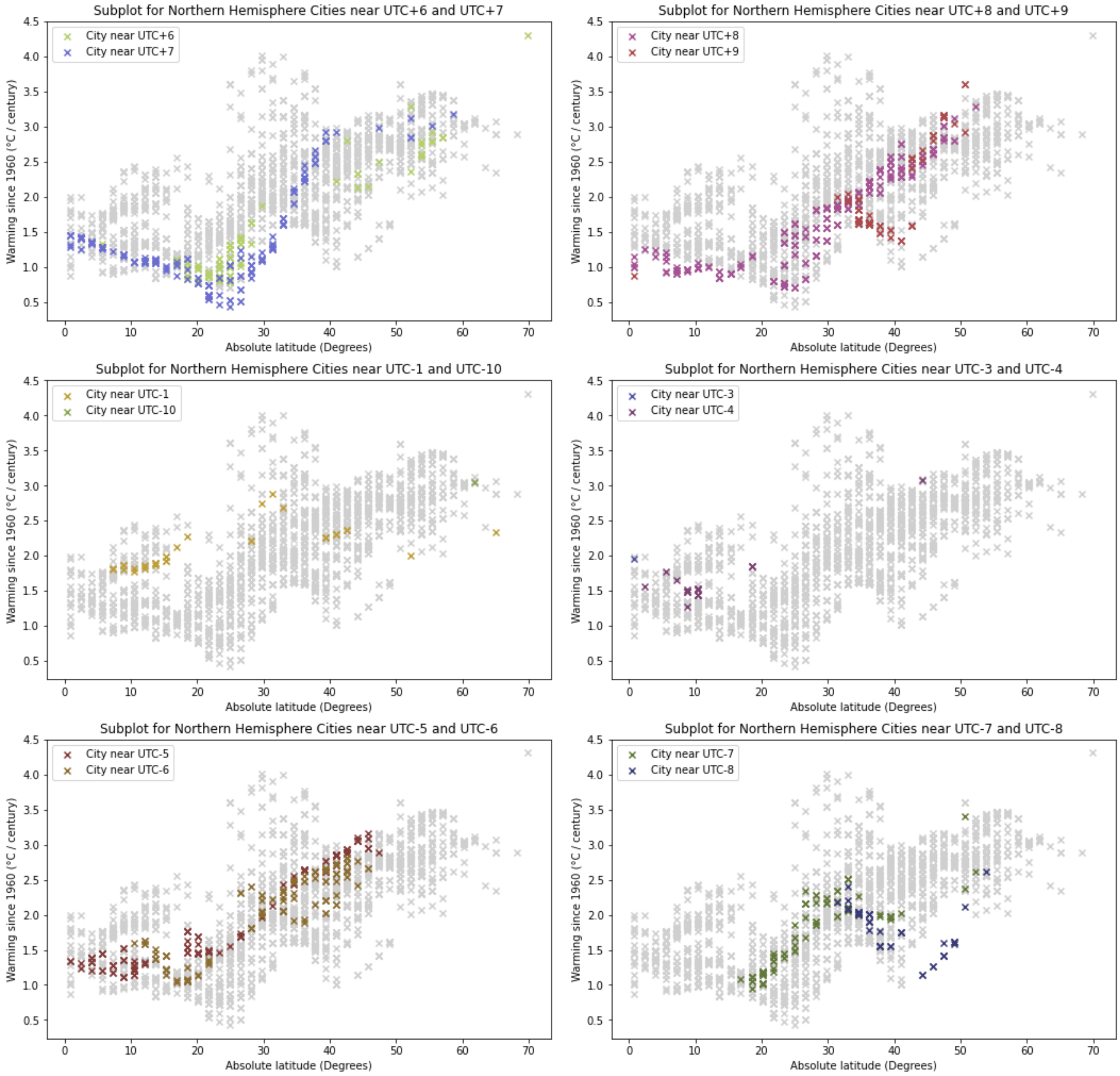


Figure a.6. Subplots each showing data points from two “natural time zones”[16]

[16] Understandably, not all “natural time zones” have data points in the northern hemisphere. The missing ones are UTC-2, UTC-9, UTC-11, UTC-12 and UTC+12. These locations either correspond to oceans (UTC-2) or sparsely populated areas (e.g. North-western tip of Canada for UTC-9, both sides of the Bering Sea for UTC-11, UTC-12 and UTC+12).