



Earth Surface Temperature

IE6600 – Computation and Visualization for Analytics Final Report

Group Number 06

Neeraj Kulkarni - 002770750

Atharva Uplenchwar - 002990536

Holin lee - 002664643

Harshada Terwadkar - 002702245

Manthan Gadhikar - 002991668

Parag Jain - 002984373

Table of Contents

PART 1 - INTRODUCTION AND RESEARCH QUESTIONS.....	2
PART 2 - SUMMARY OF RESULTS	2
PART 3 - DATA SOURCES	3
PART 4 – RESULTS AND METHODS.....	3
DATA CLEANING	3
VISUALIZING DATA USING PYTHON.....	4
<i>Global Temperature Trend Analysis Using Matplotlib:</i>	4
<i>Top US states analysis:</i>	5
<i>Top 10 countries analysis:</i>	6
<i>Forecasting using Linear Regression:</i>	7
VISUALIZING DATA IN TABLEAU	9
PART 5 - LIMITATIONS AND FUTURE WORK.....	13

PART 1 - Introduction and Research Questions

Climate change is said to be one the biggest threats to human society and is also one of the concerns that can't be retaliated back. There can be various reasons causing climate change, like natural factors, human factors, and any other external factors. This Cause has an evident example like low snowfall in the state of Massachusetts in 2022-2023. Increase in earth's surface temperature is a huge reason for the same. This is also one of the reasons why this made us investigate deeper through data and know the reason for the increase in earth's surface temperature.

The research questions for this investigation are as follows:

- How has Earth's surface temperature changed over 150 years, and what are the main drivers of this change?
- What are the main causes of the difference in surface temperature. Is it a natural phenomenon? or is it because of humans?
- What does the future hold when it comes to climate change? How much will the temperature increase by 2050?
- Where is it being affected more? Is there a particular region for the same?

To answer the above question, historical data of average temperature is taken and based on the posed questions, data analysis is performed and visualized using multiple libraries like Matplotlib in Python and Tableau visualizations. Similarly, to find the drivers causing climate change, a dataset that has attributes like change in temperature due to specific human or natural activities along the specific period is used.

Tools like python and excel have been utilized to clean the data and analyze the data on earth's surface temperature. To visualize the plots, tableau, and python libraries like matplotlib is utilized. The results of this analysis will help us understand the depth of concerns surrounding climate change and looking ahead, why it is very important to take the rights steps to conserve nature.

PART 2 - Summary of Results

Post analysis, the data from 2 legitimate sources, there was a significant relation between temperature rise and increase in the level of greenhouse gas emissions. Natural factors like volcanic activity, solar activity, orbital activity did not have much effect on temperature changes, while human factors like land use, aerosol emissions, and finally greenhouse gas emission combined have a clear relation to the temperature increase over the years.

The data also showed the temperature trend in 195 countries and their respective states. The trend gave us a clear upward increase of temperature across the world. From the analysis, a conclusive insight can be inferred that the countries in the northern hemisphere are undergoing maximum temperature changes than other countries.

Forecasting the data up to 2050 gave a 3 degree rise in temperature than temperature in 1850 about 20 degrees Celsius in 2050, indicating that with such a rise in temperature there can be huge natural consequences. Furthermore, a separate analysis was performed for United states which gave top 10 states that underwent significant temperature differences. Similarly, an analysis for all 195 countries gave the top 5 countries that underwent significant temperature differences.

PART 3 - Data Sources

Berkley Earth and Intergovernmental Panel on Climate Change (IPCC) are the 2 main data sources that are used. The Berkley earths dataset was sourced from Kaggle. The data shows temperature trend data from 1740 – 2013 which was broken down into by 195 countries and their respective states. The dataset has approximately 600000 entries as it includes monthly temperature data from 1st of January 1740 to 1st of December 2013. The temperature data is in degree Celsius.

The 2nd data was sourced from IPCC. This data set provided attributes like temperature changes due to natural factors like Solar activity, Volcanic activity, etc. And temperature changes due to Human factors like Greenhouse gas emissions, land use, aerosol emissions, etc. The data is presented in temperature values, indicating the degree to which each factor contributes to temperature change.

Analyzing both datasets gave great insights regarding the temperature trend for 150 years. And the main driving factors for climate change.

PART 4 – Results and Methods

Data Cleaning

The temperature readings were inaccurate because mercury thermometers were used in the early phases of data gathering. But as technology has developed over time, more advanced measurement techniques were used, producing temperature readings that are more accurate. As a result of this a subset of the data from 1850 – 2013 was taken.

Even beginning from 1850s there were irregularities in the data. The way the data set was initially structured was that there was an average temperature of a country for every month of the year. For

some of the countries especially in the Southern Hemisphere there were missing values for some of the months. We used the forward fill method to get rid of all the null value fields.

Next, for the initial trend analysis, we did not require the month wise average temperature, so we were able to transform the data in such a way that we obtained a yearly average temperature for every country ranging from 1850-2013. Post this process, all the outliers were identified and removed as their quantity was inconsequential.

Subsequently, we wanted to find the countries that were most affected by global warming in the past 150 years. To achieve this, we reduced the multivalued attribute of country name to an atomic value, and we got the temperature rise for each country by subtracting the temperature reading in 1850 from the temperature reading in 2013. This dataset was then used to find the top 5 countries that had faced a significant temperature rise.

A similar approach was incorporated to find the 10 states in US which were most affected global warming.

Steps taken to clean the data:

- Removing duplicate data
- Checking missing values
- Finding missing values and forward filling them
- Identifying outliers and removing them as the number of outliers were very less.

Data Preparation:

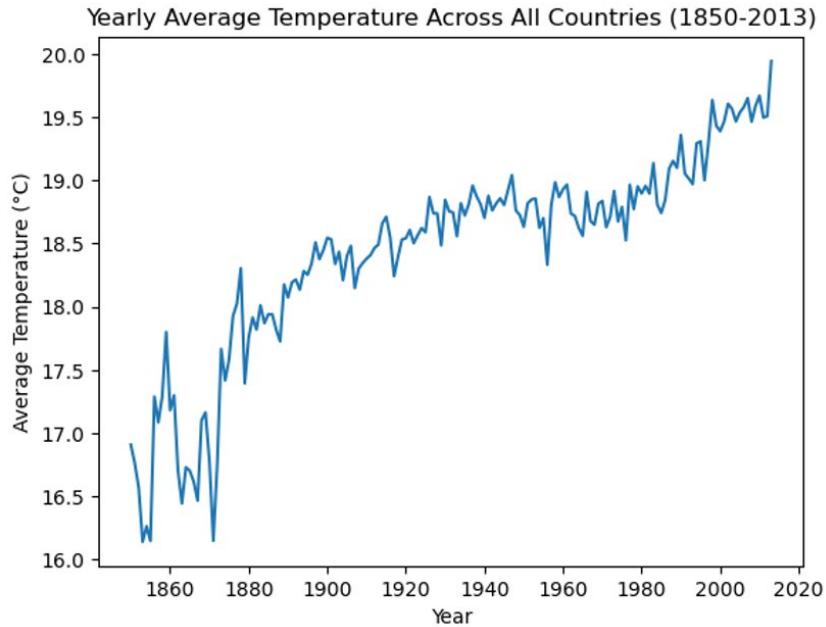
- Grouped dataset by country name and state name.
- Grouped dataset by Year.
- Converted dates into date time function.

Visualizing data using Python.

Global Temperature Trend Analysis Using Matplotlib:

Matplotlib library was used to plot the line plot. The line plot gave a temperature trend over 150 years. We saw the following results.

- Increase of 2 degrees since pre-industrial times (1850-1900).
- Significant rise in temperature after 1995 and steep rise in 20th century.
- Industrial revolution has been a major impact for the rise in temperature.



Top US states analysis:

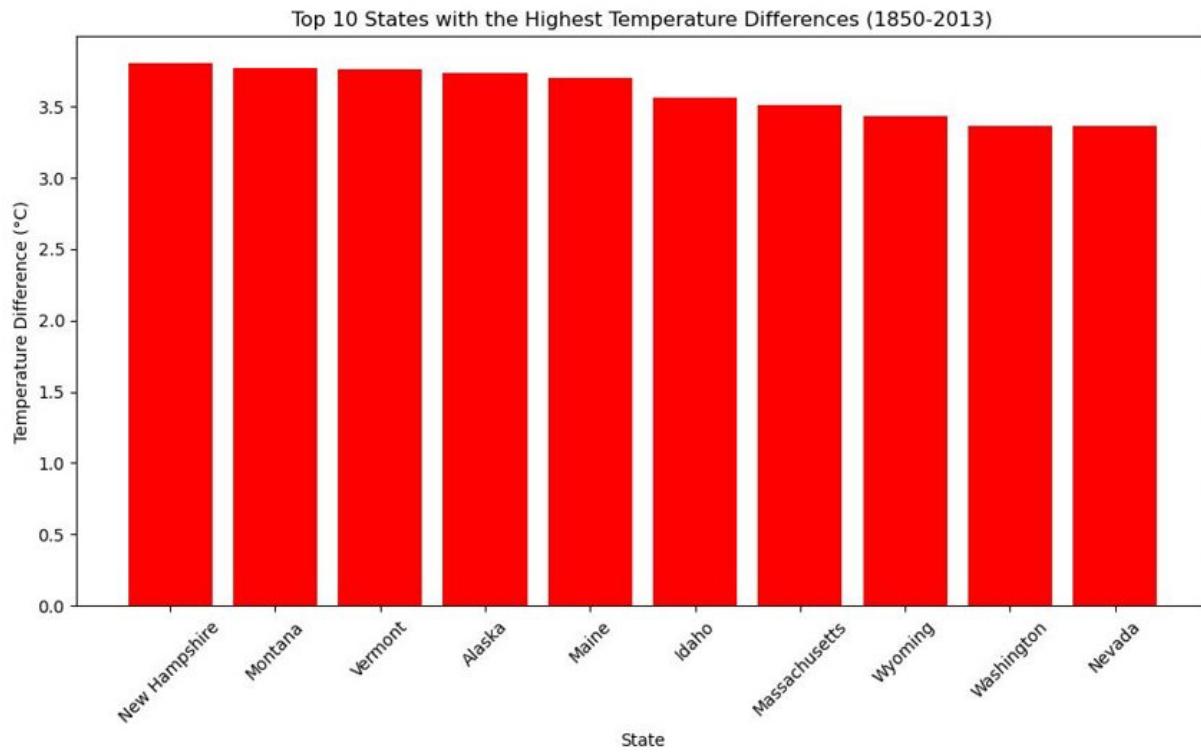
Using the global land temperatures by state dataset from the Berkley earth site. This analysis was focused on to analyze data specifically of 50 states of The United States. Grouping by state made sure that data from 1850 to 2013 for each state was considered. Temperature difference was calculated for each state by differencing last month's data to that of the first month. Using matplotlib library, the top 10 states temperature difference were plotted.

```
# Filter the DataFrame for years 1850 and 2013
df_1850 = df[df['Year'] == 1850]
df_2013 = df[df['Year'] == 2013]

# Merge the two DataFrames on the 'Country' column
merged_df = pd.merge(df_1850, df_2013, on='Country', suffixes=('_1850', '_2013'))

# Calculate the temperature difference
merged_df['Temp_Difference 1850-2013'] = merged_df['Avg_Temperature_2013'] - merged_df['Avg_Temperature_1850']

# Create a new DataFrame with only the desired columns
result_df = merged_df[['Country', 'Temp_Difference 1850-2013']]
```

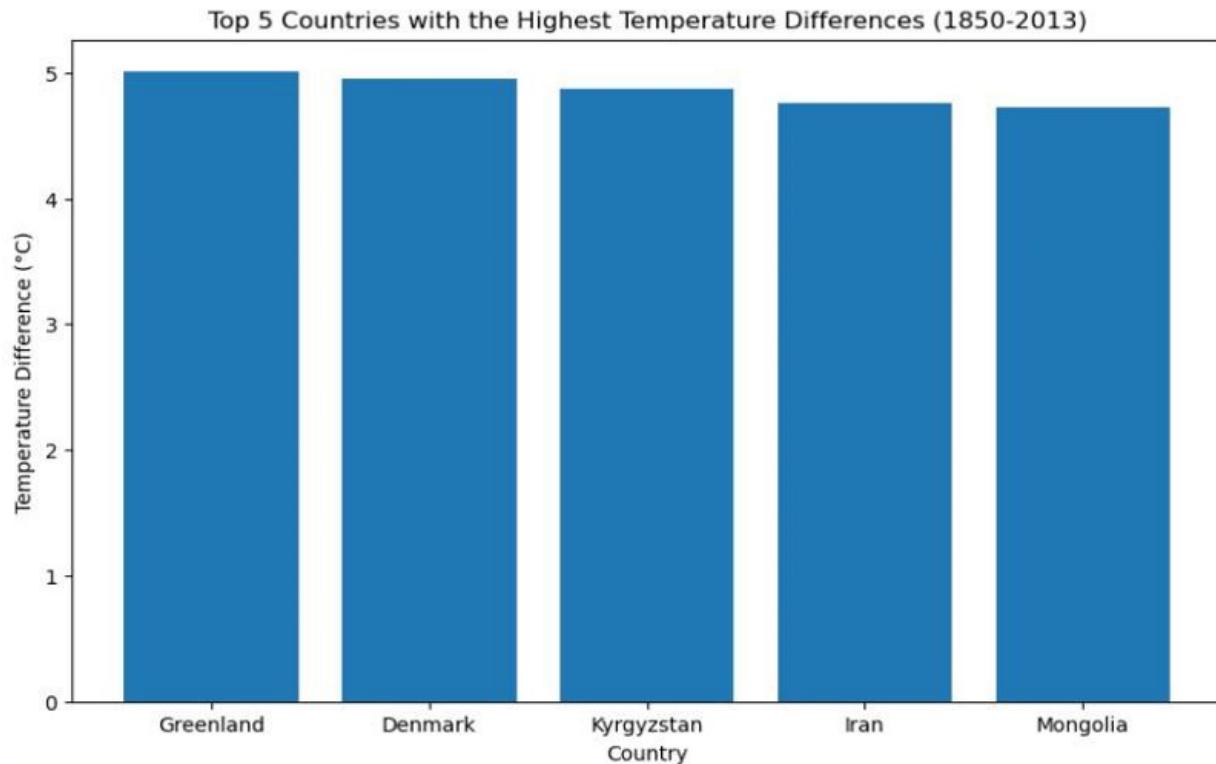


Insights of the above analysis:

- New Hampshire is the state that experienced the most temperature difference since 1850.
- Northern part of USA particularly is very much affected by the climate change over the years.
- Alaska to be also one of the top states to undergo maximum temperature difference.

Top 10 countries analysis:

Like the 2nd analysis, which was done by states for USA, analysis for temperature difference for all 195 countries was performed and top 5 countries that underwent significant temperature differences were plotted.



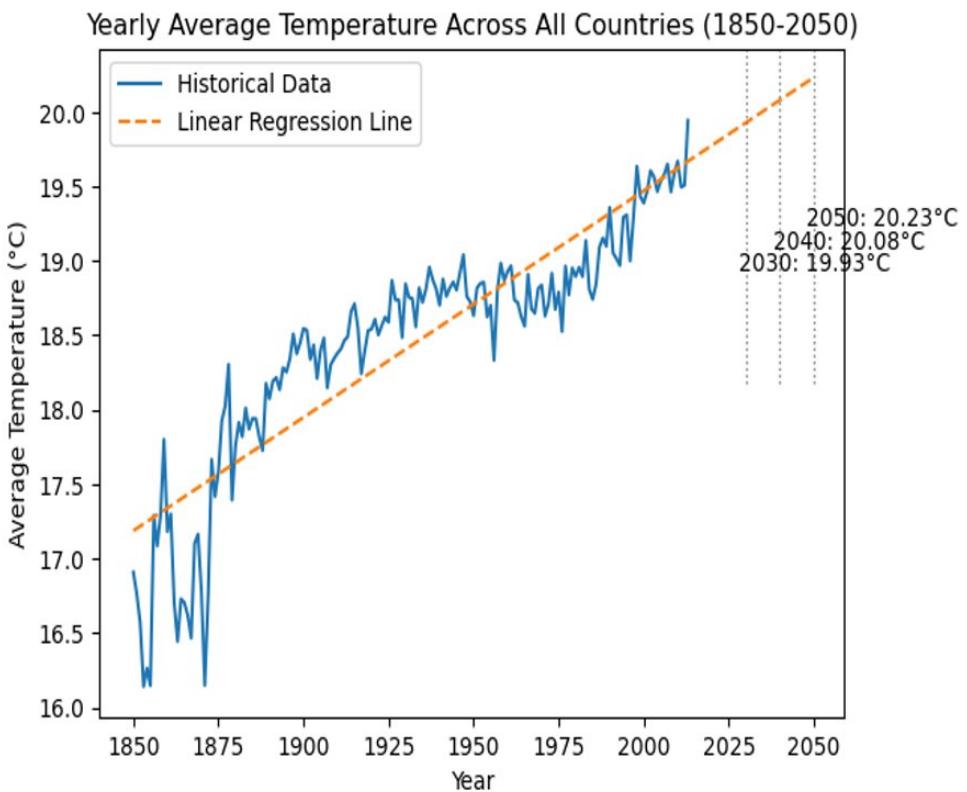
Insights of the above analysis:

- Greenland experienced a significant temperature rise.
- Supporting the 2nd analysis, here also the countries in the northern hemisphere underwent a higher temperature change than those in other regions.

Forecasting using Linear Regression:

Using the large volume of data, temperature data was forecasted for the next 30 years i.e., up to 2050. This was done by liner regression method. Creating a liner regression model helped to establish the relation between attributes like year and temperature. The model successfully predicted the temperature based on the historic temperatures.

Analysis shows that the average temperature will reach 19.93 degrees Celsius by 2030. 2050 will be the hottest year on record, 3 degrees warmer than 1850. A temperature rise of 3 degrees is associated with many natural disasters and loss of coastline, loss of vegetation due to dry air, etc.

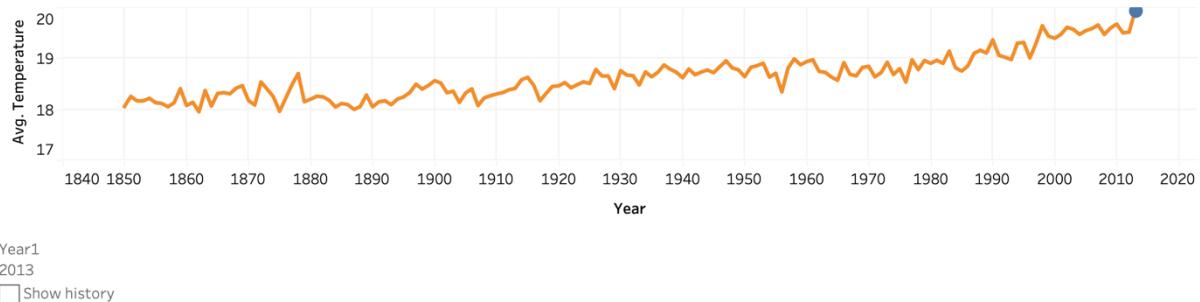


Visualizing data in Tableau

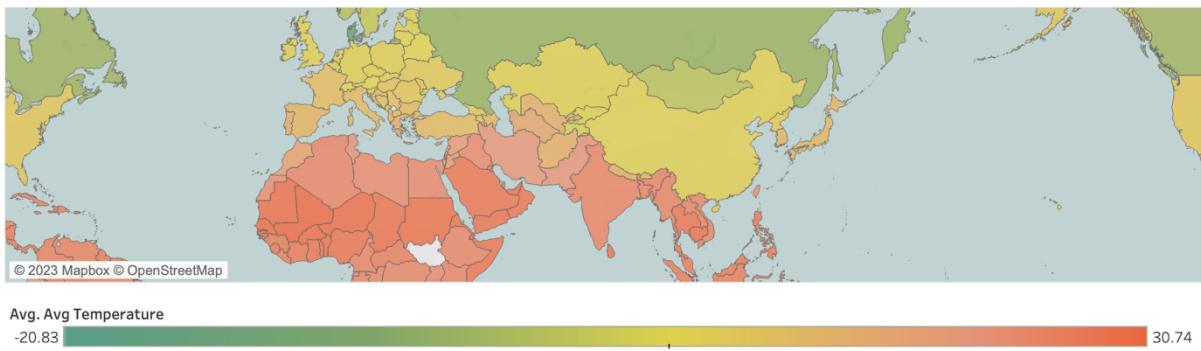
Temperature Variation in the last 150 years

Global Temperature Gradient	US Temperature Heat Map	Monthly Average	Temperature Variations caused by Natural Activities	Temperature Variations Caused by Human Activities
-----------------------------	-------------------------	-----------------	---	---

Global Temperature Timeline



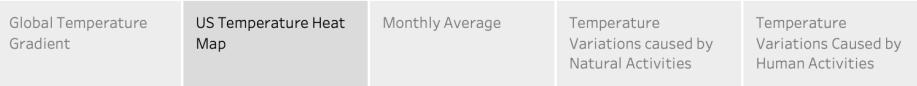
Temperature Map - 2013



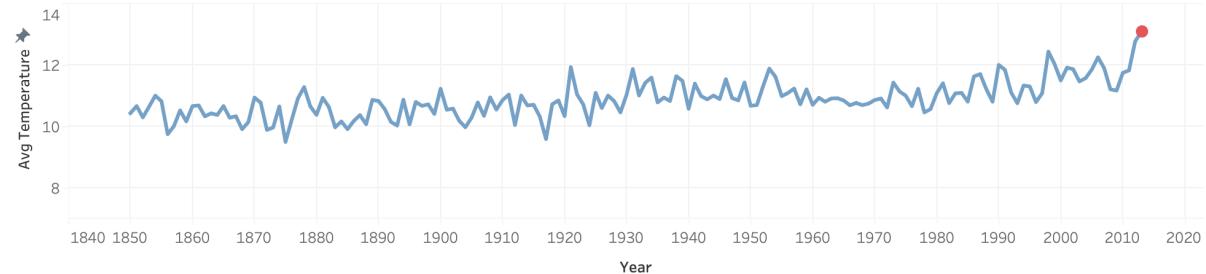
From the above visualizations we can conclude that the global land temperature has risen by approximately 2 degrees Celsius from the beginning of 1850 to 2013. If we were to include the ocean temperatures as well in this data, the global average temperature rise would be even more.

Even though a 2-to-3-degree rise doesn't seem like much, it has had a significant and far-reaching impact on the planet as well as human society. As it is apparent from the above visualizations the countries that are most affected by global warming are the ones that are away from the equator. This is mostly because these countries are usually sensitive to climate changes as they face a wide range of temperatures throughout the year. Due to global warming, this range of temperature has drastically reduced.

Temperature Variation in the last 150 years



US Temperature Timeline - 2013

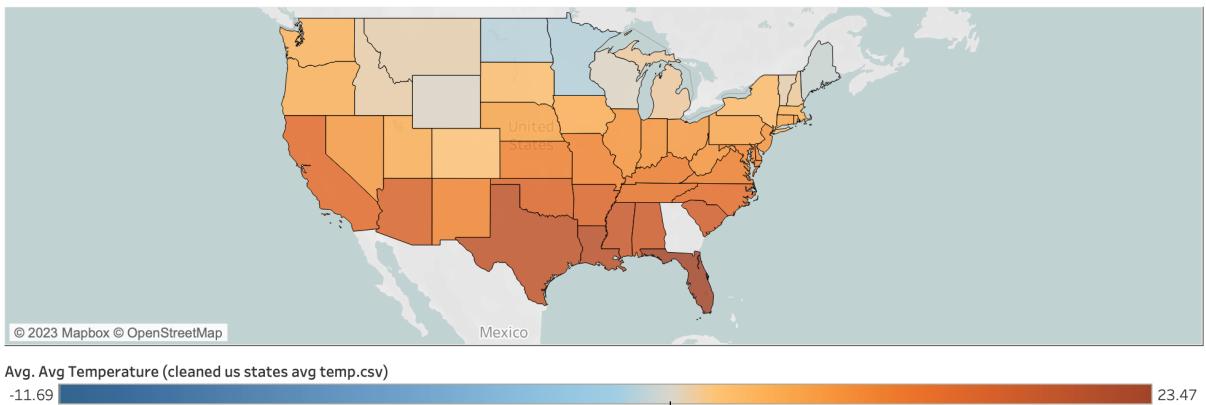


Year (cleaned us states avg temp.csv)

2013

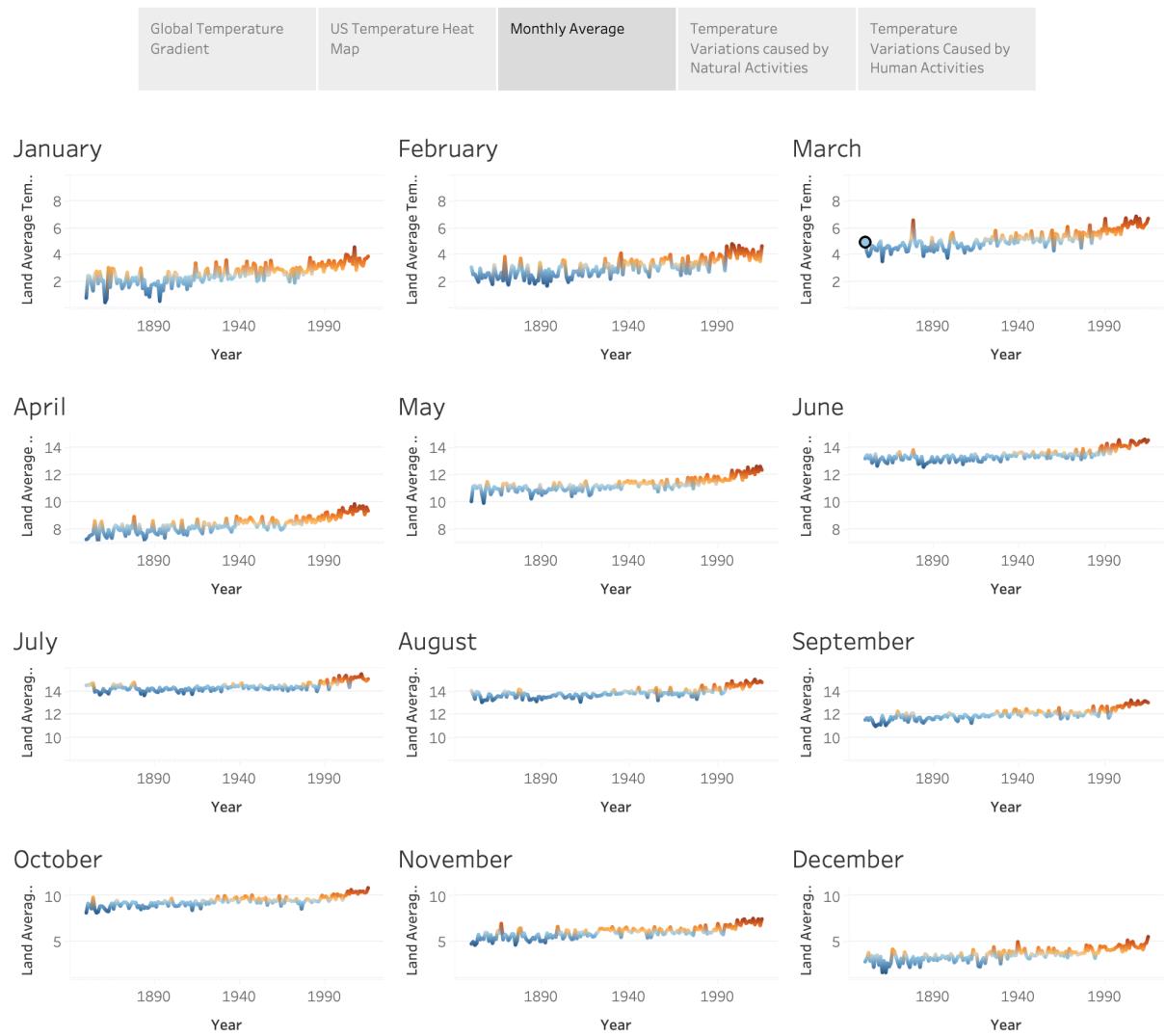
Show history

US Heat Map - 2013



The United States has also witnessed a lot of temperature changes due to global warming. A lot of the northern states that usually had an average temperature of below 6-7 degrees Celsius have now reached average temperatures of above 8 degrees. Alaska has witnessed a 4 degree rise in the past decades.

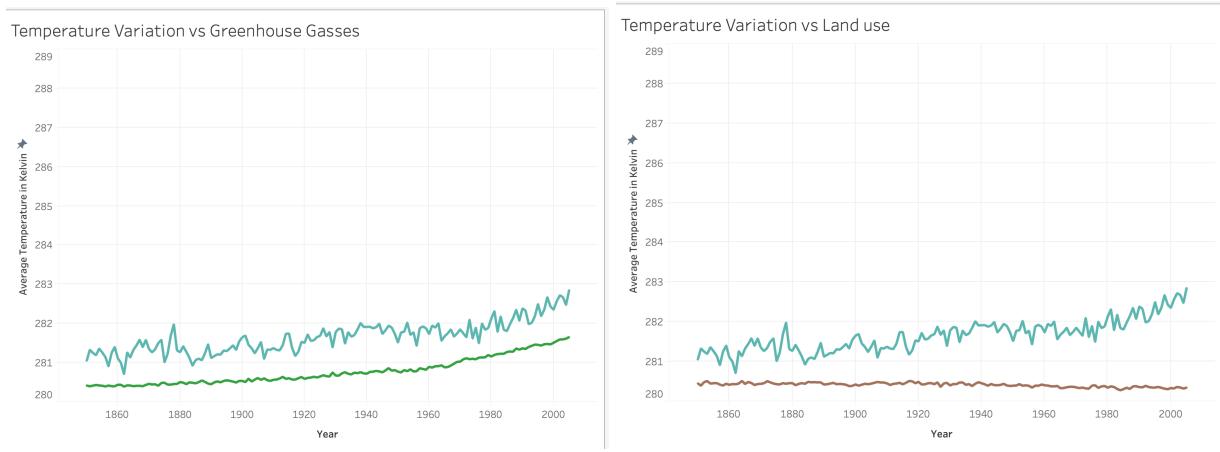
Temperature Variation in the last 150 years

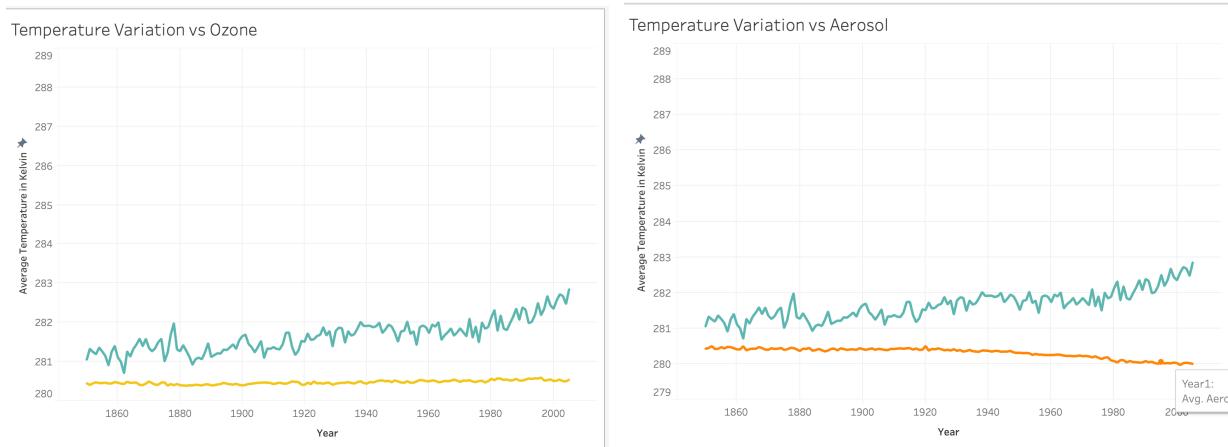


This is a monthly plot of global average temperatures over the last 150 years. Winter months such as December January February have experienced an average 2-degree rise. It can also be seen from the above plot that the temperatures increase steadily from January to June and then start declining till December. As expected, the winter months experience a wide range of temperatures in several countries situated away from the equator but in these past decades the ranges have vastly reduced.



To summarize the above visualization, we can say that none of the natural activities that have occurred over the past 150 years have had a significant impact in the average rise of global temperature.





It is apparent from the above plots that increase in the amount of greenhouse gasses is one of the major reasons that has caused the warming of the earth's surface.

PART 5 - Limitations and Future Work

- The initial data that we're considering i.e., the temperature readings from 1850 to 1900 are not very accurate at that point of time. The apparatus used to measure the average temperature varied greatly based on the country. But to the data can still be used to get an approximate idea of the temperature at that time.
- Forecasting using linear regressions is not completely reliable as it does not consider seasonality and stationarity which might lead to inaccuracies in predictions.
- To increase the scope of this project to accurately determine the overall temperature difference caused by global warming, ocean temperatures over the past 150 years should also be included.
- To portray the consequences of the rise in global temperatures, data that show the rise in sea level over the decades can also be utilized.
- A better prediction model can be incorporated that utilizes a combination of seasonality, stationarity, and trend analysis. This will result in more accurate predictions.