Global Analysis of Increase in Land Temperature

Maxwell Solko, Ayush Pandey, Upvandeep Kaur, Andrea Bendayan, Shrestha Acharya, Jihoon Yun

2022-12-07

# **Introduction**

There is evidence that human activities like burning of fossil fuels and cutting down forests have caused soil degradation and have increased carbon emissions. Something that definitely can, and will, be shown is the change of temperature over the years. The increase in carbon emission causes the greenhouse gas effect which helps to trap heat in the atmosphere. The increase in heat in the atmosphere have caused changes in temperatures and weather patterns. The impacts of these changes are drastic: glaciers and ice sheets are shrinking, fresh water ice sheets in rivers and lakes are disappearing at an alarming rate, plant and animal species are being extinguished, geographic ranges are shifting, and plants and trees in temperate latitudes are blooming in early spring. Data science is emerging as a powerful weapon in the fight against climate change. It can be helpful in understanding the climate changes over years. This information can be used by policy makers and leaders to take vital measures to control climate change and its effects.

In context to the above motive, we decided to work on a long-time study of climate trends and land temperatures over various countries, which includes all the major continents.

# **Data Gathering and Preparation**

## Data Gathering

The raw data has been collected from Kaggle. Further, the Kaggle data set is repackaged from a new compilation created by the Berkeley Earth, which is affiliated with Lawrence Berkeley National Laboratory. The Berkeley Earth Surface Temperature Study combines 1.6 billion temperature reports from 16 pre-existing archives. It is nicely packaged and allows for slicing into interesting subsets (for example by country).

Dataset Details:

* dt: provides the date of observation starting from 1750 to 2013.
* AverageTemperature: details about global average land temperature in celsius.
* AverageTemperatureUncertainty: provides global land temperature with 95% confidence interval around the average.
* Country: details about country names.

## Data Cleaning and Preparation

In order to clean and prepare the datasets, we performed the following steps:

* Handling the null values: For the dataset, we analaysed the missing data values in all the variables. 5% of the data had null values in the columns AverageTemperature & AverageTemperatureUncertainty. We decided to remove these data rows as most of them dated around 1750 - 1850.
* Appropriate data type conversion: We converted the data type format of the column “dt” to date to handle the data easily.
* Handling the data for the column “country” - Removed continent names from country column to increase the accuracy.

Below is the summary statistics of the collected dataset:

df <- read.csv('https://raw.githubusercontent.com/msolko/DTSC5301Project/main/GlobalLandTemperaturesByCountry.csv')  
  
#changing dates to date format  
df$dt <- as.Date(df$dt)  
  
#Analyzing the null data  
df\_u<-df %>% filter(is.na(AverageTemperature))  
max\_dt<-max(df\_u$dt)  
min\_dt<-min(df\_u$dt)  
  
#removing NA values  
df <- drop\_na(df)  
  
#checking NA values  
nadt=sum(is.na(df["dt"]))  
natemp=sum(is.na(df["AverageTemperature"]))  
natempunct=sum(is.na(df["AverageTemperatureUncertainty"]))  
nacountry=sum(is.na(df["Country"]))  
  
#nadt  
#natemp  
#natempunct  
#nacountry  
#summary of the dataset  
summary(df)

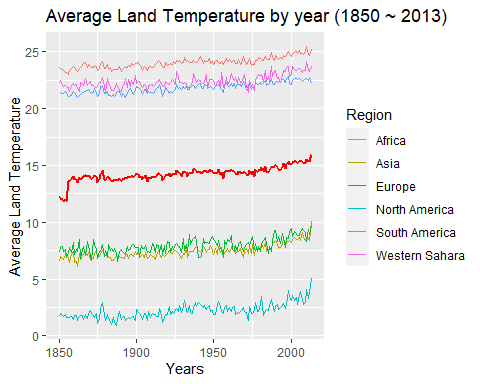
## dt AverageTemperature AverageTemperatureUncertainty  
## Min. :1743-11-01 Min. :-37.66 Min. : 0.052   
## 1st Qu.:1869-11-01 1st Qu.: 10.03 1st Qu.: 0.323   
## Median :1919-08-01 Median : 20.90 Median : 0.571   
## Mean :1913-08-08 Mean : 17.19 Mean : 1.019   
## 3rd Qu.:1966-10-01 3rd Qu.: 25.81 3rd Qu.: 1.207   
## Max. :2013-09-01 Max. : 38.84 Max. :15.003   
## Country   
## Length:544811   
## Class :character   
## Mode :character   
##   
##   
##

# **Analysis**

We analyzed the trends of average land temperature for the list of continents available in the dataset. Below are the details around it.

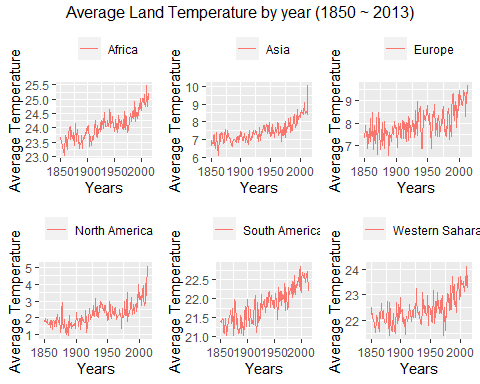
#Separating dataframe region wise  
rdf <- df %>% filter(Country == "Europe"|Country == "Africa"| Country == "North America" | Country == "South America" | Country == "Asia" | Country == "Western Sahara")  
  
#changing columnname  
colnames(rdf)[4] <- "Region"  
#print(unique(rdf$Region))  
  
  
# deleting regions from Country Column  
  
cdf<-subset(df, Country != "Europe" & Country != "Africa" & Country != "North America" & Country != "South America" & Country != "Asia" & Country != "Western Sahara")  
#print(unique(cdf$Country))

# rdf - group by year  
rdf\_year <- rdf %>%   
 group\_by(year = substr(.$dt,1,4), Region) %>%   
 summarise(  
 AverageTemperature\_by\_Region = mean(AverageTemperature)  
 )   
  
# Convert year type (chr > numeric)  
rdf\_year$year <- as.numeric(rdf\_year$year)  
  
# filter - Asia and Western  
rdf\_year\_asia\_western <- rdf\_year %>%   
 filter(Region %in% c('Asia', 'Western Sahara'))   
  
# filter - Asia   
rdf\_year\_asia <- rdf\_year %>%   
 filter(Region %in% c('Asia'))   
  
# filter - Western   
rdf\_year\_western <- rdf\_year %>%   
 filter(Region %in% c('Western Sahara'))   
  
#   
rdf\_mean\_year <- rdf\_year %>%   
 filter(year >= 1850) %>%   
 group\_by(year) %>%   
 summarise(  
 mean\_temperature = mean(AverageTemperature\_by\_Region)  
 )   
  
######################################################  
# Plots all regions with average temperature of all regions  
######################################################  
  
plot\_all\_mean <- rdf\_year %>%   
 filter(year >= 1850) %>%   
 ggplot() +  
 geom\_line(aes(x = year, y = AverageTemperature\_by\_Region, color = Region)) +   
 geom\_line(data = rdf\_mean\_year, aes(x = year, y = mean\_temperature), color = 'red', size=1) +  
 ggtitle("Average Land Temperature by year (1850 ~ 2013)") +  
 xlab("Years") +   
 ylab("Average Land Temperature")  
  
# plot all regions with the average temperature)  
plot\_all\_mean



We can easily understand and compare the different temperature ranges for different continents using the above visualizations. Africa, Western Sahara and South America are way hotter than Europe, Asia, and North America as expected. The “red bold line” projects the mean value of temperatures observed in these 6 continents. The mean is approximately around 15° Celsius. We can also infer that Asia and Europe have mild land temperatures, while North America has the colded recorded land temperatures.

#######################################################  
# Plot using grid.arrange  
#######################################################  
  
# define regions vector  
v\_regions <- unique(rdf$Region)  
  
# plots all regions by using grid.arrange  
l\_plot <- lapply(v\_regions, function(x) {  
 p <- rdf\_year %>%   
 filter(year >= 1850 & Region == x) %>%   
 ggplot(aes(x = year, y = AverageTemperature\_by\_Region, color = Region)) +  
 geom\_line() +   
 theme(  
 legend.position = "top",  
 legend.title = element\_blank()  
 ) +  
 xlab("Years") +   
 ylab("Average Temperature")  
})  
  
# show plots  
n <- length(l\_plot)  
do.call("grid.arrange", c(l\_plot, ncol=3, top="Average Land Temperature by year (1850 ~ 2013)"))



Furthermore, the above visualization focuses on the change in land temperature in various continents from 1850 to 2013. The land temperature in all the continents have increased significantly and linearly, reflecting a positive correlation between years and an increase in temperature. There is an average increase of 2° to 4° Celsius for all the continents. Below is the detailed analysis for all the continents:

1. South America, Western Sahara, Europe, Africa: - For all these continents, the increase in land temperature is gradual over the years i.e., the temperature is increasing slowly and gradually between the decades. Apart from Western Sahara, the countries in South America, Europe and Africa have a mix power grid that includes fossil fuels and renewable energy. The countries in these 3 regions are also smaller and require less energy consumption to function.
2. Asia, North America: For these two continents, there is a gradual increase of land temperatures until the beginning of 2000’s however, after that there is a sudden temperature influx of approximately 1°Celsius. North America and Asia have countries with developed economies that are highly reliant on fossil fuels to power up their energy grids.

# **Conclusions**

For the last 163 years we can conclude that the land temperature averages have been increasing steadily across all the regions in the world. The industrial revolution which began in the early 1800’s has increased the quality of life for humanity in both developed and developing countries. However, the increase in quality of life has been sustained by the indiscriminate use of burning fossil fuels to power electric grids all over the world. Countries in North America and Asia have seen the steepest increase in temperature in the last 20 years. The increase in temperature in the last 20 years is caused by the increase in demand for energy. North America and Asia are the two regions that produce the most products and services, therefore they require more energy that can be obtained easily, cheaply, and efficiently, which is obtained by burning fossil fuels such as coal or oil. The carbon emissions caused by the burning of fossil fuels provokes the greenhouse effect phenomena and traps heats within the atmosphere, causing the rapid and sharp increase in temperature. On the contrary, countries in South America, Europe and Africa that are more likely to incorporate renewable sources of energy, have a more modest increase in temperature throughout the last 20 years. Countries in Europe, South America, and Africa also have smaller economies and populations compared to North America and Asia; therefore, they do not necessarily need to rely in burning fossil fuels for additional energy needs. In conclusion, data science can be a powerful tool to study, manage and mitigate the negative effects of climate change.

## Possible Bias

For our bias, even though the data visualization clearly shows a positive linear relationship between time in years and an increase in average temperature, we cannot confirm that carbon emissions caused by the burning of fossil fuels is the main cause to the increase in temperature. The data set started around the same time when the industrial revolution started to occur. There is no temperature data previously to the industrial revolution, therefore we cannot confirm with certainty the cause of the increase in temperature throughout the years. Our bias attributed the increase in temperature to the burning of fossil fuels, however since there is no previous data there is a possibility that this can be a natural process that planet Earth experiences. Even if the data did go back that far it still could just be a natural cycle of warming and cooling in the earth. Correlation doesn’t equal causation. To better make these connections, bringing in data about carbon usage or other activities to show their growths as well would help. It is still important to realize our biases and bring in fair data, not just data that supports our view.

### Resources:

<https://www.worldometers.info/co2-emissions/co2-emissions-by-country/>

<https://ourworldindata.org/low-carbon-electricity-by-country>