Computational Statistics - Project 1 Climate Change

Braden Hanna

February 19, 2021

1 Introduction

Global warming is a topic of grave concern for humanity. Global warming is caused by a phenomenon called the "Green House Effect". Where the sun's heat gets trapped in earths atmosphere because of increased levels of pollutants. In this project, the "Rise in Temp" dataset by Vageesha Budanur [1] is used to analyze common climate predictor variable's effect on global temperature from 1983-2008. The variables contained in the dataset are: The observation year, the observation month, the multivariate El Nino Souther nOscillation index (MEI), the atmospheric concentrations of carbon dioxide, the atmospheric concentrations of methane, the atmospheric concentrations of nitrous oxide, atmospheric concentrations of trichlorofluoromethane(CFC-11), the atmospheric concentrations of dichlorodifluoromethane(CFC-12), the total solar irradiance (TSI) in W/m2, the mean stratospheric aerosol optical depth at 550 nm, and the difference in degrees Celsius between the average global temperature in that period and a reference value.

2 R Script

```r
#Braden Hanna
#2/13/21
#CS 467 - Project 1

library(ggplot2)
library(zoo)
library(dplyr)
library(gridExtra)

#load csv into a dataframe
weather <- read.csv("ClimateData.csv", header=T)

#convert months and years to a date for plotting next part
```

```r
weather$date <- as.Date(paste0(weather$Year, "-", weather$Month, "-01"))

#create scatter for each dependent variable against date, smooth for regression
scatter <- ggplot(weather, aes(date, MEI))
scatter + geom_point() + geom_smooth(method = "lm", colour = "Red") +
  labs(x = "Date", y = "MEI")

scatter <- ggplot(weather, aes(date, CO2))
scatter + geom_point() + geom_smooth(method = "lm", colour = "Red") +
  labs(x = "Date", y = "CO2")

scatter <- ggplot(weather, aes(date, CH4))
scatter + geom_point() + geom_smooth(method = "lm", colour = "Red") +
  labs(x = "Date", y = "CH4")

scatter <- ggplot(weather, aes(date, N2O))
scatter + geom_point() + geom_smooth(method = "lm", colour = "Red") +
  labs(x = "Date", y = "N2O")

scatter <- ggplot(weather, aes(date, CFC.11))
scatter + geom_point() + geom_smooth(method = "lm", colour = "Red") +
  labs(x = "Date", y = "CFC-11")

scatter <- ggplot(weather, aes(date, CFC.12))
scatter + geom_point() + geom_smooth(method = "lm", colour = "Red") +
  labs(x = "Date", y = "CFC-12")

scatter <- ggplot(weather, aes(date, TSI))
scatter + geom_point() + geom_smooth(method = "lm", colour = "Red") +
  labs(x = "Date", y = "TSI")

scatter <- ggplot(weather, aes(date, Aerosols))
scatter + geom_point() + geom_smooth(method = "lm", colour = "Red") +
  labs(x = "Date", y = "Aerosols")

scatter <- ggplot(weather, aes(date, Temp))
scatter + geom_point() + geom_smooth(method = "lm", colour = "Red") +
  labs(x = "Date", y = "Temp")

#create subsets for each 5 year interval, save mean for each dependent variable
weather_subset <- subset(weather, Year >= 1983 & Year <= 1987)
firstMeanMEI <- mean(weather_subset$MEI)
firstMeanCO2 <- mean(weather_subset$CO2)
firstMeanCH4 <- mean(weather_subset$CH4)
firstMeanN2O <- mean(weather_subset$N2O)
firstMeanCFC.11 <- mean(weather_subset$CFC.11)
```

```r
firstMeanCFC.12 <- mean(weather_subset$CFC.12)
firstMeanTSI <- mean(weather_subset$TSI)
firstMeanAerosols <- mean(weather_subset$Aerosols)
firstMeanTemp <- mean(weather_subset$Temp)

weather_subset <- subset(weather, Year >= 1988 & Year <= 1992)
secondMeanMEI <- mean(weather_subset$MEI)
secondMeanCO2 <- mean(weather_subset$CO2)
secondMeanCH4 <- mean(weather_subset$CH4)
secondMeanN2O <- mean(weather_subset$N2O)
secondMeanCFC.11 <- mean(weather_subset$CFC.11)
secondMeanCFC.12 <- mean(weather_subset$CFC.12)
secondMeanTSI <- mean(weather_subset$TSI)
secondMeanAerosols <- mean(weather_subset$Aerosols)
secondMeanTemp <- mean(weather_subset$Temp)

weather_subset <- subset(weather, Year >= 1993 & Year <= 1997)
thirdMeanMEI <- mean(weather_subset$MEI)
thirdMeanCO2 <- mean(weather_subset$CO2)
thirdMeanCH4 <- mean(weather_subset$CH4)
thirdMeanN2O <- mean(weather_subset$N2O)
thirdMeanCFC.11 <- mean(weather_subset$CFC.11)
thirdMeanCFC.12 <- mean(weather_subset$CFC.12)
thirdMeanTSI <- mean(weather_subset$TSI)
thirdMeanAerosols <- mean(weather_subset$Aerosols)
thirdMeanTemp <- mean(weather_subset$Temp)

weather_subset <- subset(weather, Year >= 1998 & Year <= 2002)
fourthMeanMEI <- mean(weather_subset$MEI)
fourthMeanCO2 <- mean(weather_subset$CO2)
fourthMeanCH4 <- mean(weather_subset$CH4)
fourthMeanN2O <- mean(weather_subset$N2O)
fourthMeanCFC.11 <- mean(weather_subset$CFC.11)
fourthMeanCFC.12 <- mean(weather_subset$CFC.12)
fourthMeanTSI <- mean(weather_subset$TSI)
fourthMeanAerosols <- mean(weather_subset$Aerosols)
fourthMeanTemp <- mean(weather_subset$Temp)

weather_subset <- subset(weather, Year >= 2003 & Year <= 2008)
fifthMeanMEI <- mean(weather_subset$MEI)
fifthMeanCO2 <- mean(weather_subset$CO2)
fifthMeanCH4 <- mean(weather_subset$CH4)
fifthMeanN2O <- mean(weather_subset$N2O)
fifthMeanCFC.11 <- mean(weather_subset$CFC.11)
fifthMeanCFC.12 <- mean(weather_subset$CFC.12)
fifthMeanTSI <- mean(weather_subset$TSI)
```

```
fifthMeanAerosols <- mean(weather_subset$Aerosols)
fifthMeanTemp <- mean(weather_subset$Temp)

#create a new dataframe to hold these mean values
df <- data.frame(DateRange=character(), meanMEI=double(), meanCO2=double(),
                 meanCH4=double(),meanN2O=double(), meanCFC.11=double(),
                 meanCFC.12=double(), meanTSI=double(),
                 meanAerosols=double(), meanTemp = double(),
                 stringsAsFactors=FALSE)

#add new rows to the dataframe, each row is the means for each dependent
#variable in that 5 year period
new_row <- data.frame("1983-1987", firstMeanMEI, firstMeanCO2, firstMeanCH4,
                      firstMeanN2O, firstMeanCFC.11, firstMeanCFC.12,
                      firstMeanTSI, firstMeanAerosols, firstMeanTemp,
                      row.names = "new_row")
names(new_row) <- names(df)
df <- rbind(df, new_row)

new_row <- data.frame("1988-1992", secondMeanMEI, secondMeanCO2, secondMeanCH4,
                      secondMeanN2O, secondMeanCFC.11, secondMeanCFC.12,
                      secondMeanTSI, secondMeanAerosols, secondMeanTemp,
                      row.names = "new_row")
names(new_row) <- names(df)
df <- rbind(df, new_row)

new_row <- data.frame("1993-1997", thirdMeanMEI, thirdMeanCO2, thirdMeanCH4,
                      thirdMeanN2O, thirdMeanCFC.11, thirdMeanCFC.12,
                      thirdMeanTSI, thirdMeanAerosols, thirdMeanTemp,
                      row.names = "new_row")
names(new_row) <- names(df)
df <- rbind(df, new_row)

new_row <- data.frame("1998-2002", fourthMeanMEI, fourthMeanCO2, fourthMeanCH4,
                      fourthMeanN2O, fourthMeanCFC.11, fourthMeanCFC.12,
                      fourthMeanTSI, fourthMeanAerosols, fourthMeanTemp,
                      row.names = "new_row")
names(new_row) <- names(df)
df <- rbind(df, new_row)

new_row <- data.frame("2003-2008", fifthMeanMEI, fifthMeanCO2, fifthMeanCH4,
                      fifthMeanN2O, fifthMeanCFC.11, fifthMeanCFC.12,
                      fifthMeanTSI, fifthMeanAerosols, fifthMeanTemp,
                      row.names = "new_row")
names(new_row) <- names(df)
df <- rbind(df, new_row)
```

```r
#plot the means of each period from 1983-2008
scatter <- ggplot(df, aes(DateRange, meanMEI))
scatter + geom_point()

scatter <- ggplot(df, aes(DateRange, meanCO2))
scatter + geom_point()

scatter <- ggplot(df, aes(DateRange, meanCH4))
scatter + geom_point()

scatter <- ggplot(df, aes(DateRange, meanN2O))
scatter + geom_point()

scatter <- ggplot(df, aes(DateRange, meanCFC.11))
scatter + geom_point()

scatter <- ggplot(df, aes(DateRange, meanCFC.12))
scatter + geom_point()

scatter <- ggplot(df, aes(DateRange, meanTSI))
scatter + geom_point()

scatter <- ggplot(df, aes(DateRange, meanAerosols))
scatter + geom_point()

scatter <- ggplot(df, aes(DateRange, meanTemp))
scatter + geom_point()

#find outliers and apply transformation if needed
boxplot(weather$MEI, ylab = "MEI")
summary(weather$MEI)
#create dataframe to remove items from, do not want to ruin original dataset
MEIdata <- weather[,c(1,3)]
#set cut off benchmark, values will be removed if not within
bench <- .8305 + 1.5*IQR(MEIdata$MEI)
#remove outliers, 5 found
MEIdata <- filter(MEIdata, MEI < bench)

#no outliers
boxplot(weather$CO2, ylab = "CO2")
#no outliers
boxplot(weather$CH4, ylab = "CH4")
#no outliers
boxplot(weather$N2O, ylab = "N2O")
```

```
#outliers found
boxplot(weather$CFC.11, ylab = "CFC-11")
summary(weather$CFC.11)
CFC.11data <- weather[,c(1,7)]
bench <- 246.3 - 1.5*IQR(CFC.11data$CFC.11)
#remove outliers, 29 found
CFC.11data <- filter(CFC.11data, CFC.11 > bench)

#outliers found
boxplot(weather$CFC.12, ylab = "CFC-12")
summary(weather$CFC.12)
CFC.12data <- weather[,c(1,8)]
bench <- 472.4 - 1.5*IQR(CFC.12data$CFC.12)
#remove outliers, 17 found
CFC.12data <- filter(CFC.12data, CFC.12 > bench)

#no outliers
boxplot(weather$TSI, ylab = "TSI")
#outliers found
boxplot(weather$Aerosols, ylab = "Aerosols")
summary(weather$Aerosols)
Aerosolsdata <- weather[,c(1,10)]
bench <- .0126 + 1.5*IQR(Aerosolsdata$Aerosols)
#remove outliers, 262 found?
Aerosolsdata <- filter(Aerosolsdata, Aerosols > bench)

#no outliers
boxplot(weather$Temp, ylab = "Temp")

#find regression equation, create a new plot if outliers were found and removed
scatter <- ggplot(MEIdata, aes(Year, MEI))
scatter + geom_point() + geom_smooth(method = "lm", colour = "Red") +
  labs(x = "Year", y = "MEI")
regressionMEI <- lm(MEI ~ Year, data=MEIdata)
print(regressionMEI)

#find min, max, median, 1Q and 3Q
summary(MEIdata$MEI)

regressionCO2 <- lm(CO2 ~ Year, data=weather)
print(regressionCO2)

summary(weather$CO2)

regressionCH4 <- lm(CH4 ~ Year, data=weather)
print(regressionCH4)
```

```r
summary(weather$CH4)

regressionN2O <- lm(N2O ~ Year, data=weather)
print(regressionN2O)

summary(weather$N2O)

scatter <- ggplot(CFC.11data, aes(Year, CFC.11))
scatter + geom_point() + geom_smooth(method = "lm", colour = "Red") +
  labs(x = "Year", y = "CFC-11")
regressionCFC.11 <- lm(CFC.11 ~ Year, data=CFC.11data)
print(regressionCFC.11)

summary(CFC.11data$CFC.11)

scatter <- ggplot(CFC.12data, aes(Year, CFC.12))
scatter + geom_point() + geom_smooth(method = "lm", colour = "Red") +
  labs(x = "Year", y = "CFC-12")
regressionCFC.12 <- lm(CFC.12 ~ Year, data=CFC.12data)
print(regressionCFC.12)

summary(CFC.11data$CFC.11)

regressionTSI <- lm(TSI ~ Year, data=weather)
print(regressionTSI)

summary(weather$TSI)

scatter <- ggplot(Aerosolsdata, aes(Year, Aerosols))
scatter + geom_point() + geom_smooth(method = "lm", colour = "Red") +
  labs(x = "Year", y = "Aerosols")
regressionAerosols <- lm(Aerosols ~ Year, data=Aerosolsdata)
print(regressionAerosols)

summary(Aerosolsdata$Aerosols)

regressionTemp <- lm(Temp ~ Year, data=weather)
print(regressionTemp)

summary(weather$Temp)
```

3 Graphs and Analysis

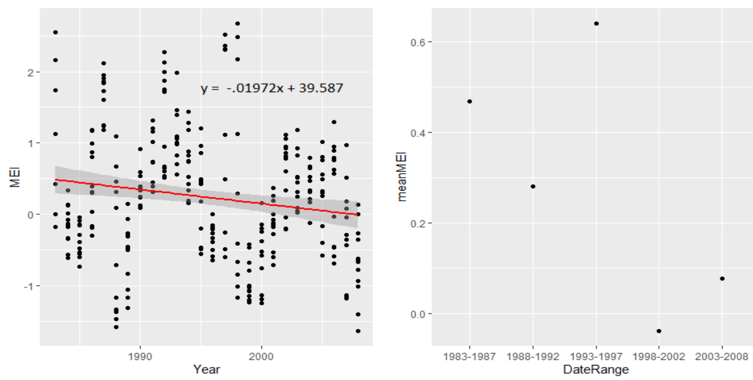

Figure 1: Multivariate El Nino Southern Oscillation index (MEI) from 1983-2008. Left graph shows each month's MEI with regression. Right graph shows every 5 years' mean MEI

The MEI varied drastically over the time frame. It was either really high or really low depending on if the year was an El Nino or La Nina. There were 5 outliers found in the data. Once removed, it was found to have a slight negative trend over time. After removing outliers, the minimum value was -1.635, the maximum was 2.673, the median was .1920, the first quartile was -.4085 and the third quartile was .806.

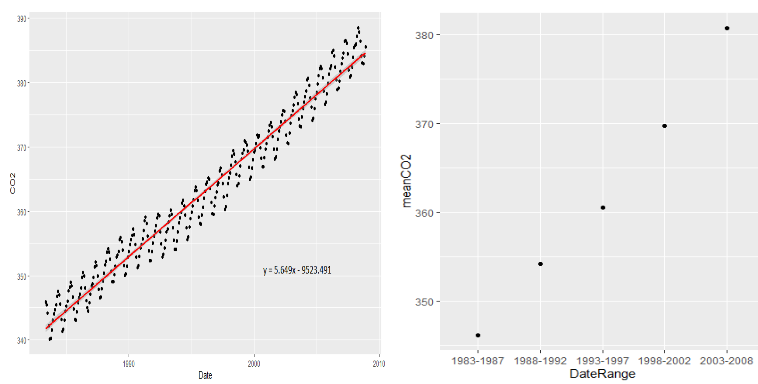

Figure 2: Atmospheric concentration of CO2 from 1983-2008. Left graph shows each month's CO2 with regression. Right graph shows every 5 years' mean CO2

The atmospheric CO2 measurements from 1983-2008 followed a consistent positive trend. There were no outliers. The minimum value was 340.2, the maximum was 388.5, the median was 361.7, the first quartile was 353.0, and the third quartile was 373.5. Atmospheric CO2 has a positive correlation with global temperature.

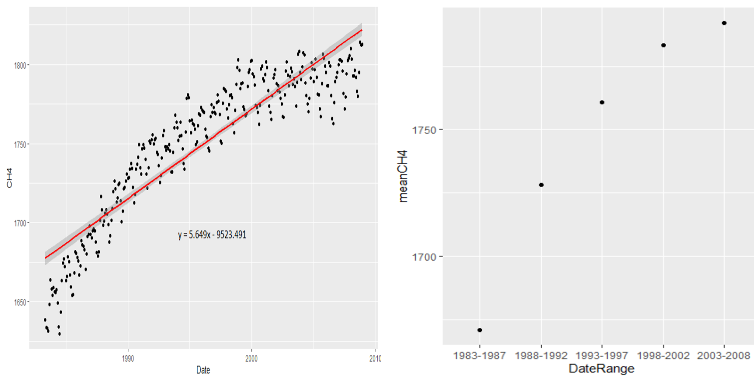

Figure 3: Atmospheric concentration of CH4 from 1983-2008. Left graph shows each month's CH4 with regression. Right graph shows every 5 years' mean CH4

The atmospheric CH4 measurements from 1983-2008 followed a consistent positive trend. There were no outliers. The minimum value was 1630, the maximum was 1814, the median was 1764, the first quartile was 1722, and the third quartile was 1787. Atmospheric CH4 has a positive correlation with global temperature.

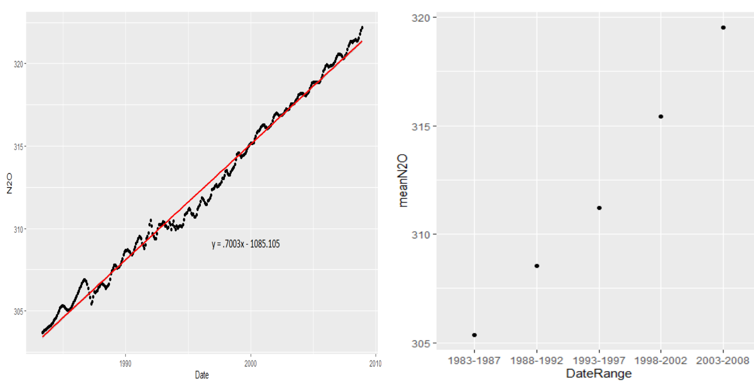

Figure 4: Atmospheric concentration of N2O from 1983-2008. Left graph shows each month's N2O with regression. Right graph shows every 5 years' mean N2O

The atmospheric N2O measurements from 1983-2008 followed a consistent positive trend. There were no outliers. The minimum value was 303.7, the maximum was 322.2, the median was 311.5, the first quartile was 308.1, and the third quartile was 317.0. Atmospheric NO2 has a positive correlation with global temperature.

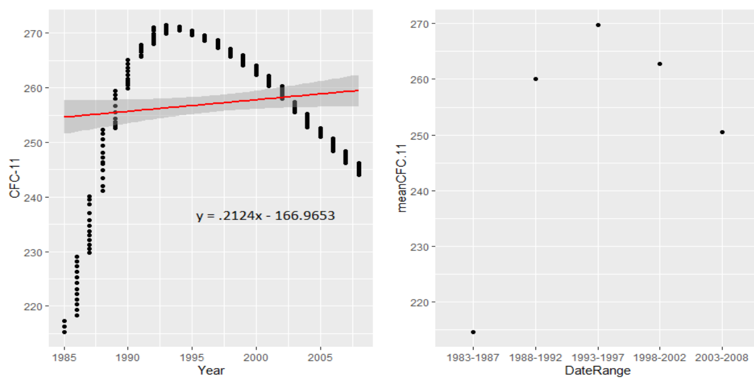

Figure 5: Atmospheric concentration of CFC-11 from 1983-2008. Left graph shows each month's CFC-11 with regression. Right graph shows every 5 years' mean CFC-11

The atmospheric concentration of CFC-11 increased quickly to a maximum around 1992, and then has been following a negative trend since then. There were 29 outliers found. After removing the outliers, the minimum value was 215.3, the maximum was 271.5, the median was 260.4, the first quartile was 250.3, and the third quartile was 267.5.

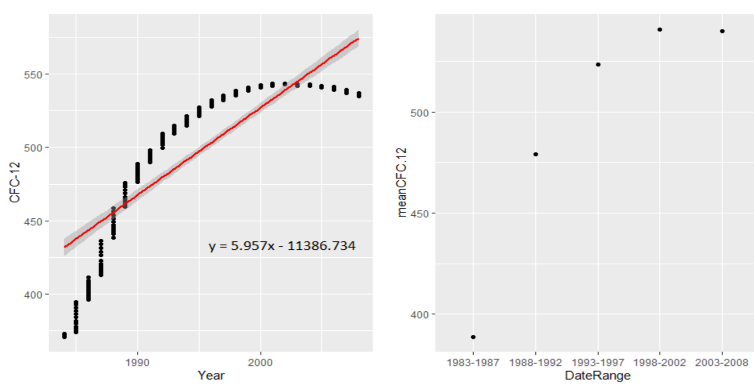

Figure 6: Atmospheric concentration of CFC-12 from 1983-2008. Left graph shows each month's CFC-12 with regression. Right graph shows every 5 years' mean CFC-12

The atmospheric concentration of CFC-12 increased quickly to a maximum around 2000, and then has been following a slight negative trend since then. There were 29 outliers found. After removing the outliers, the minimum value was 350.1, the maximum was 543.8, the median was 528.4, the first quartile was 472.4, and the third quartile was 540.5.

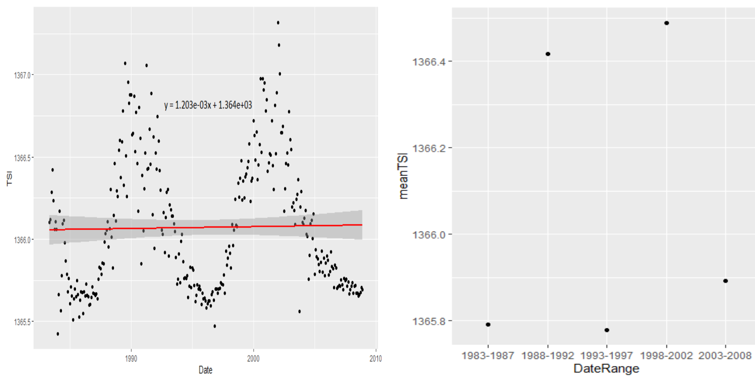

Figure 7: The total solar irradiance (TSI) from 1983-2008. Left graph shows each month's TSI with regression. Right graph shows every 5 years' mean TSI

The total solar irradiance had two major spikes between 1983 and 2008. One around 1990 and another around 2002. There were no outliers. The minimum value was 1365.426, the maximum was 1367.316, the median was 1365.981, the first quartile was 1365.717, and the third quartile was 1366.363.

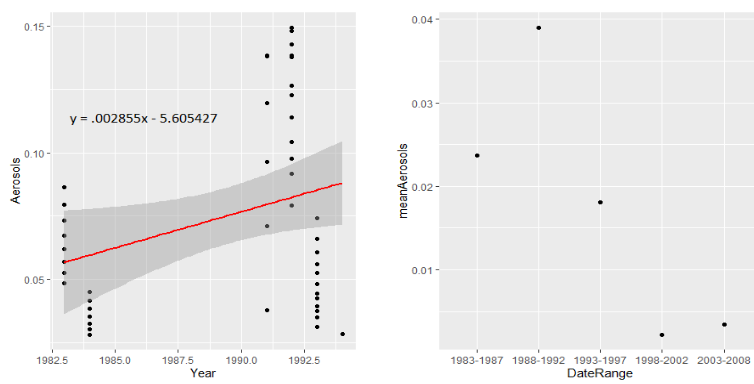

Figure 8: The mean stratospheric aerosol optical depth at 550 nm from 1983-2008. Left graph shows each month's Aerosols with regression. Right graph shows every 5 years' mean Aerosols without removing outliers

The aerosol measurements between 1983-2008 do not offer any value. The values are inconsistent and do not show any particular trend. This is proven by 263 of the 308 entries shown to be outliers. After removing outliers, the minimum value was .0282, the maximum was .1494, the median was .064, the first quartile was .04183, and the third quartile was .10275.

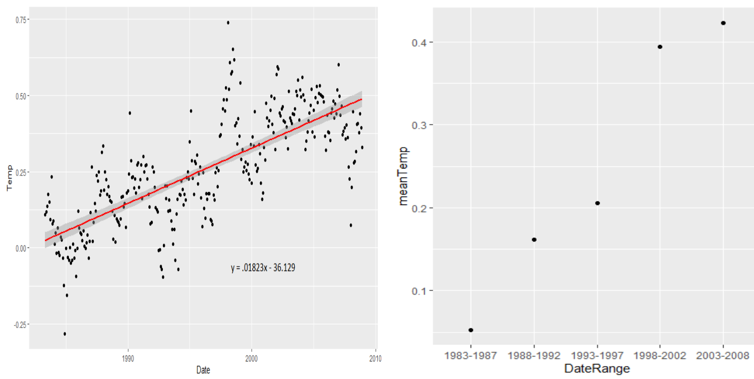

Figure 9: The difference in degrees Celsius between the average global temperature in that period and a reference value from 1983-2008. Left graph shows each month's temperature difference with regression. Right graph shows every 5 years' mean temperature difference.

Global temperature differential followed a positive trend from 1983-2008. There were no outliers. The minimum value was -.282, the maximum was .739, the median was .248, the first quartile was .1217, and the third quartile was .4073.

4 Conclusion

It was found that there are very strong correlations between levels of pollutants in the air and global temperature. CO2, CH4, NO2, CFC-11, and CFC-12 all had positive correlations with rising global temperatures. Elevated levels of these pollutants cause infrared radiation from the sun to be trapped in the earth's atmosphere, warming the earth. The other variables analyzed (MEI, TSI, and Aerosols) did not seem to have any correlation to rising global temperatures.

References

[1] Vageesha Budanur. Rise in temp dataset, Jul 2019.