Global Temperature Dataset Project CSC 465 Spring

Prabhakaran Raghavan

Abstract:
To understand the Temperature change over the last 2 centuries. This Dataset has Average temperature data for 7 different countries and its states from the year 1740 in Average Temp in Celsius

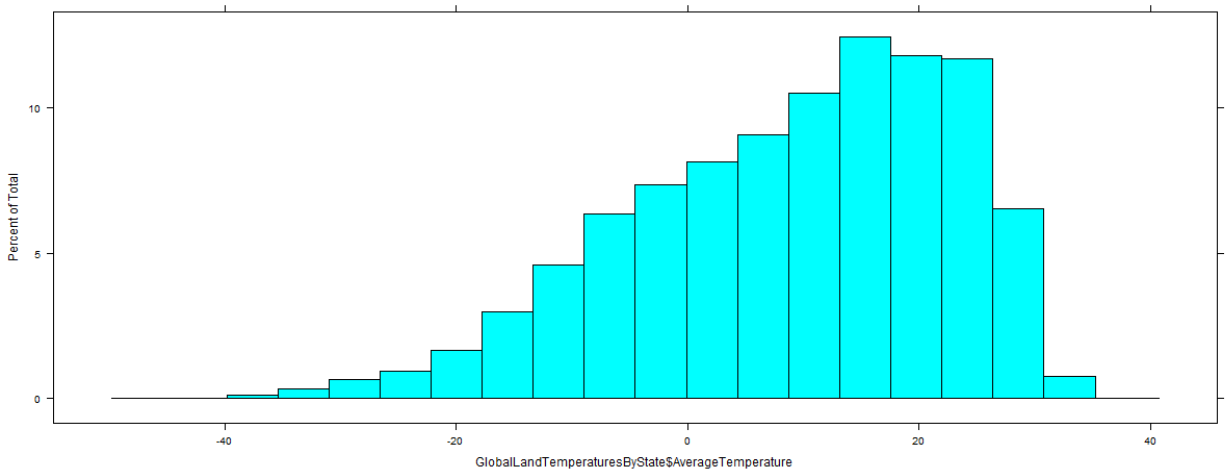
a) Data Exploration:

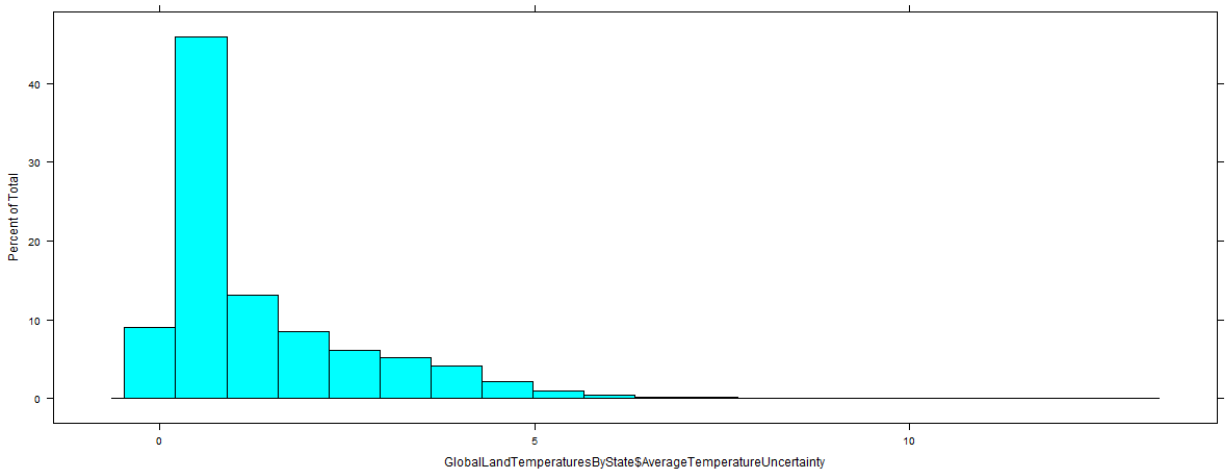
This dataset has 645675 rows and 5 Variables (Date, AverageTemperature, AverageTemperatureUncertainty, State, Country).

AverageTemperature and AverageTemperatureUncertainty is in Celsius, AverageTemperature (Dependent variable) is the focus of our Project for Data visualization. Project main goal is to find out whether the temperature is rising? For which Independent variables are Date,Country and State, while AverageTemperatureUncertainty could be considered as noise.

```
> summary(GlobalLandTemperaturesByState)
       dt             AverageTemperature AverageTemperatureUncertainty
 Min.   :1743-11-01   Min.   :-45.389    Min.   : 0.036
 1st Qu.:1845-12-01   1st Qu.: -0.693    1st Qu.: 0.316
 Median :1902-02-01   Median : 11.199    Median : 0.656
 Mean   :1898-10-01   Mean   :  8.993    Mean   : 1.288
 3rd Qu.:1957-12-01   3rd Qu.: 19.899    3rd Qu.: 1.850
 Max.   :2013-09-01   Max.   : 36.339    Max.   :12.646
                      NA's   :25648      NA's   :25648
    State              Country              Month        Month.String
 Length:645675      Length:645675      Min.   : 1.000   Length:645675
 Class :character   Class :character   1st Qu.: 3.000   Class :character
 Mode  :character   Mode  :character   Median : 6.000   Mode  :character
                                       Mean   : 6.498
                                       3rd Qu.: 9.000
                                       Max.   :12.000

      Year
 Min.   :1743
 1st Qu.:1845
 Median :1902
 Mean   :1898
 3rd Qu.:1957
 Max.   :2013

Variables introduced for the analysis are Month,Month.String and Year.

> describe(GlobalLandTemperaturesByState)
GlobalLandTemperaturesByState
```

```
 8  Variables      645675  Observations
dt
      n missing  unique
 645675       0    3239

lowest : 1743-11-01 1743-12-01 1744-01-01 1744-02-01 1744-03-01
highest: 2013-05-01 2013-06-01 2013-07-01 2013-08-01 2013-09-01
AverageTemperature
   n missing  unique    Info    Mean     .05     .10     .25     .50     .75
 620027   25648  100924       1   8.993 -15.828 -10.354  -0.693  11.199  19.8
99
    .90     .95
 25.659  27.296

lowest : -45.39 -44.74 -44.72 -44.59 -44.45
highest:  35.99  36.01  36.20  36.29  36.34

AverageTemperatureUncertainty
      n missing  unique    Info    Mean     .05     .10     .25     .50     .
75
 620027   25648    8040       1   1.288   0.179   0.216   0.316   0.656   1.8
50
    .90     .95
  3.417   4.169

lowest :  0.036  0.037  0.041  0.042  0.044
highest: 11.914 12.060 12.189 12.352 12.646
State
      n missing  unique
 645675       0     241

lowest : Acre       Adygey     Aga Buryat Alabama    Alagoas
highest: Yaroslavl' Yevrey     Yukon      Yunnan     Zhejiang

Country
      n missing  unique
 645675       0       7

          Australia Brazil Canada China India Russia United States
Frequency     16102  34328  35358 68506 86664 254972        149745
%                 2      5      5    11    13     39            23

Month
      n missing  unique    Info    Mean     .05     .10     .25     .50     .
75
 645675       0      12    0.99   6.498       1       2       3       6
9
    .90     .95
     11      12

              1     2     3     4     5     6     7     8     9    10    11
12
Frequency 53808 53812 53828 53833 53842 53843 53844 53872 53882 53643 53734 5
3734
%             8     8     8     8     8     8     8     8     8     8     8
8
```



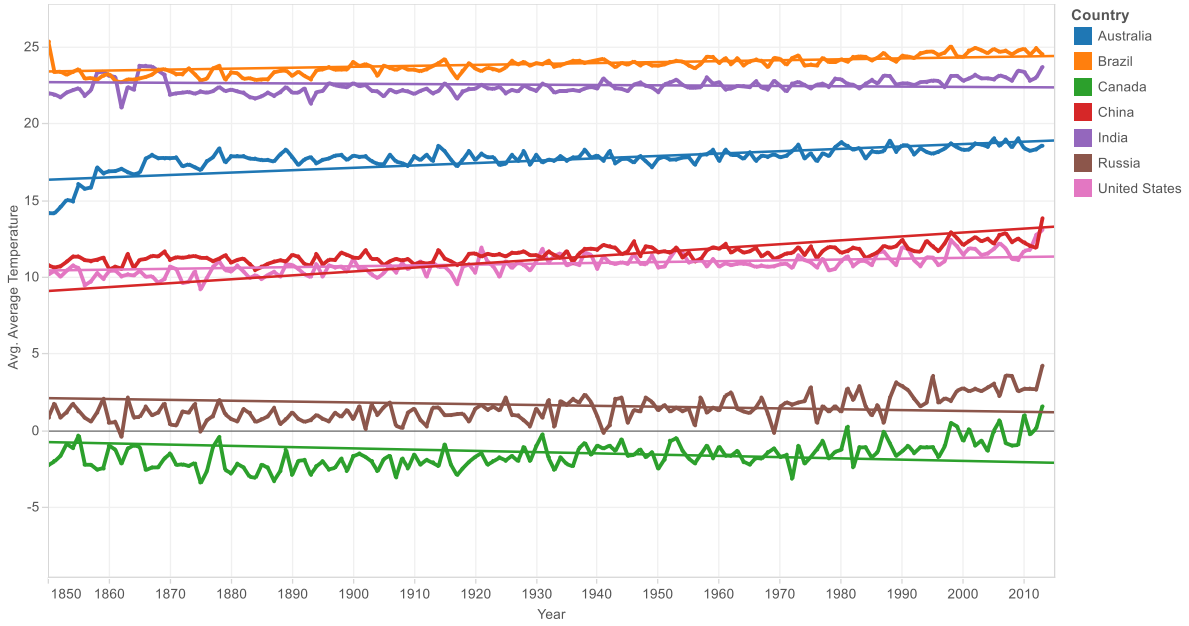


Looking at initial histogram during data exploration the data AverageTemperature is more skewed towards above 10 degree Celsius.

b) Visualization techniques:

The dataset has significant amount of data with 250 years of data. Timeseries Graph was the first thing would like try. Then since it 7 countries (with its State level data) data Geospatial / Choropleth map are Good use for this dataset. After intial analysis the question tried to answer ”Is the temperature rising?”. If so what are the plots would help answer the question.
Initial Plots and Analysis were line graph since it has time series data and so it would enable to find mor e details about each country’s Avg temperature trends.

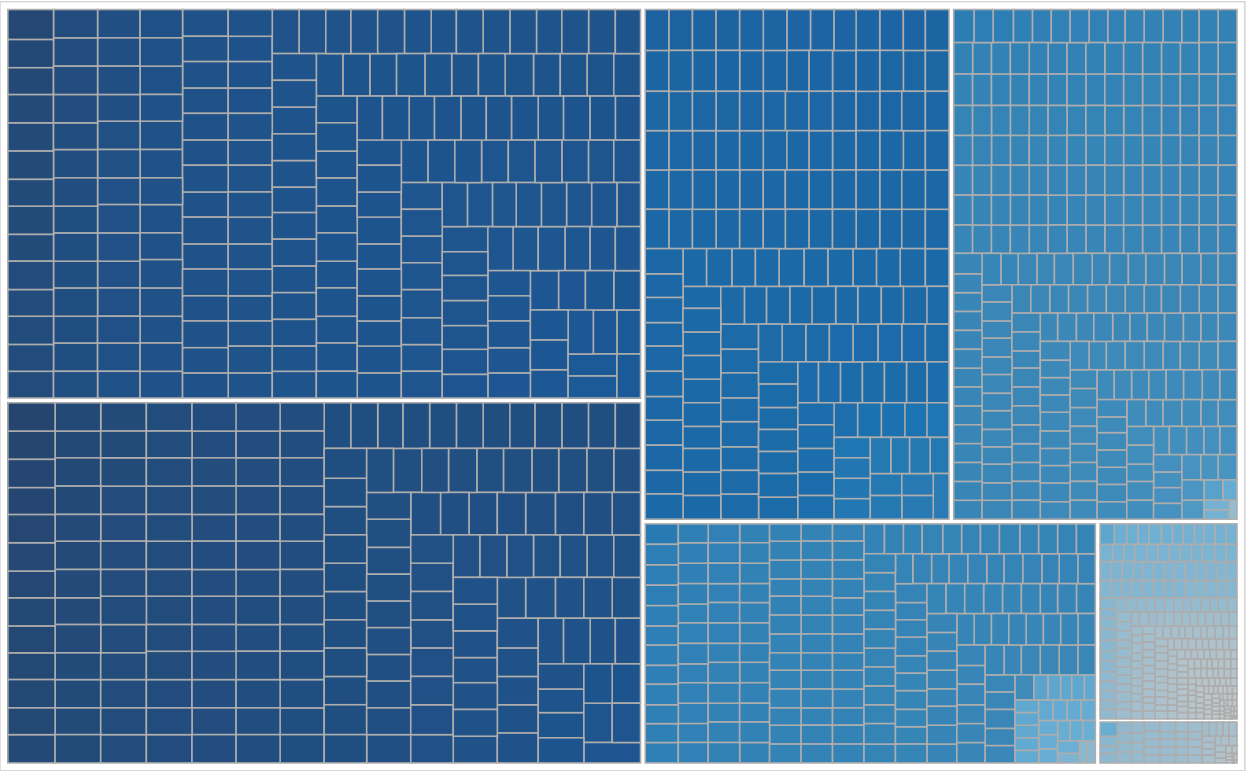
Linegraph from 1850



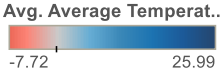
The trend of average of Average Temperature for Year. Color shows details about Country. The view is filtered on Country, which keeps 7 of 7 members.

Had a treemap for different countries from 1745

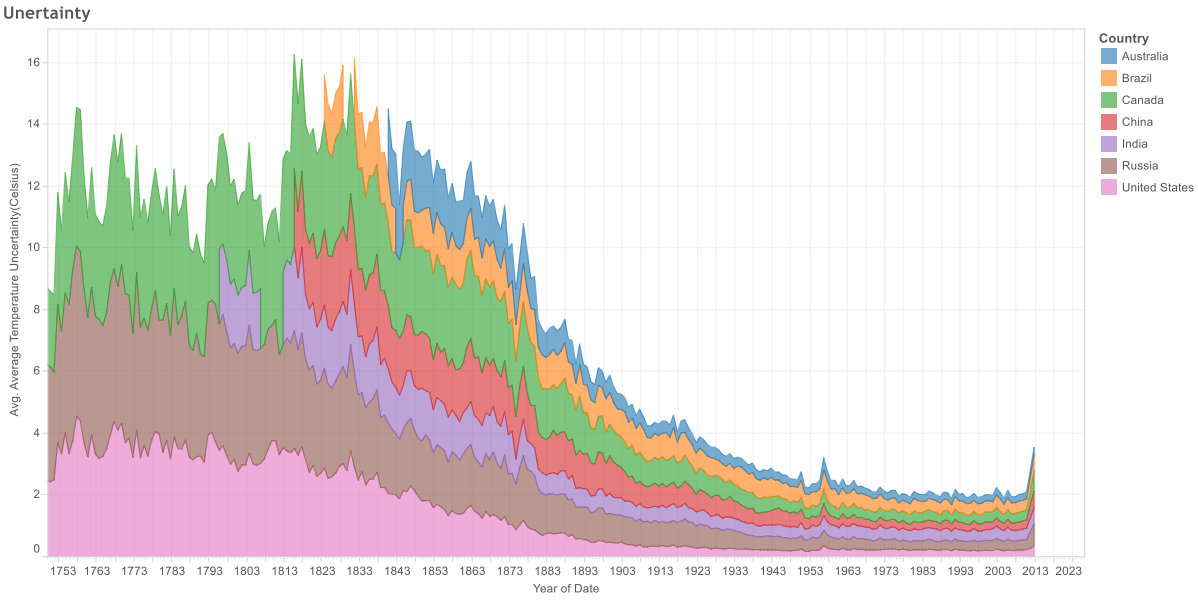
Tree Map for Global temp change



Country and Date Year. Color shows average of Average Temperature. Size shows average of Average Temperature. The marks are labeled by Country and Date Year.



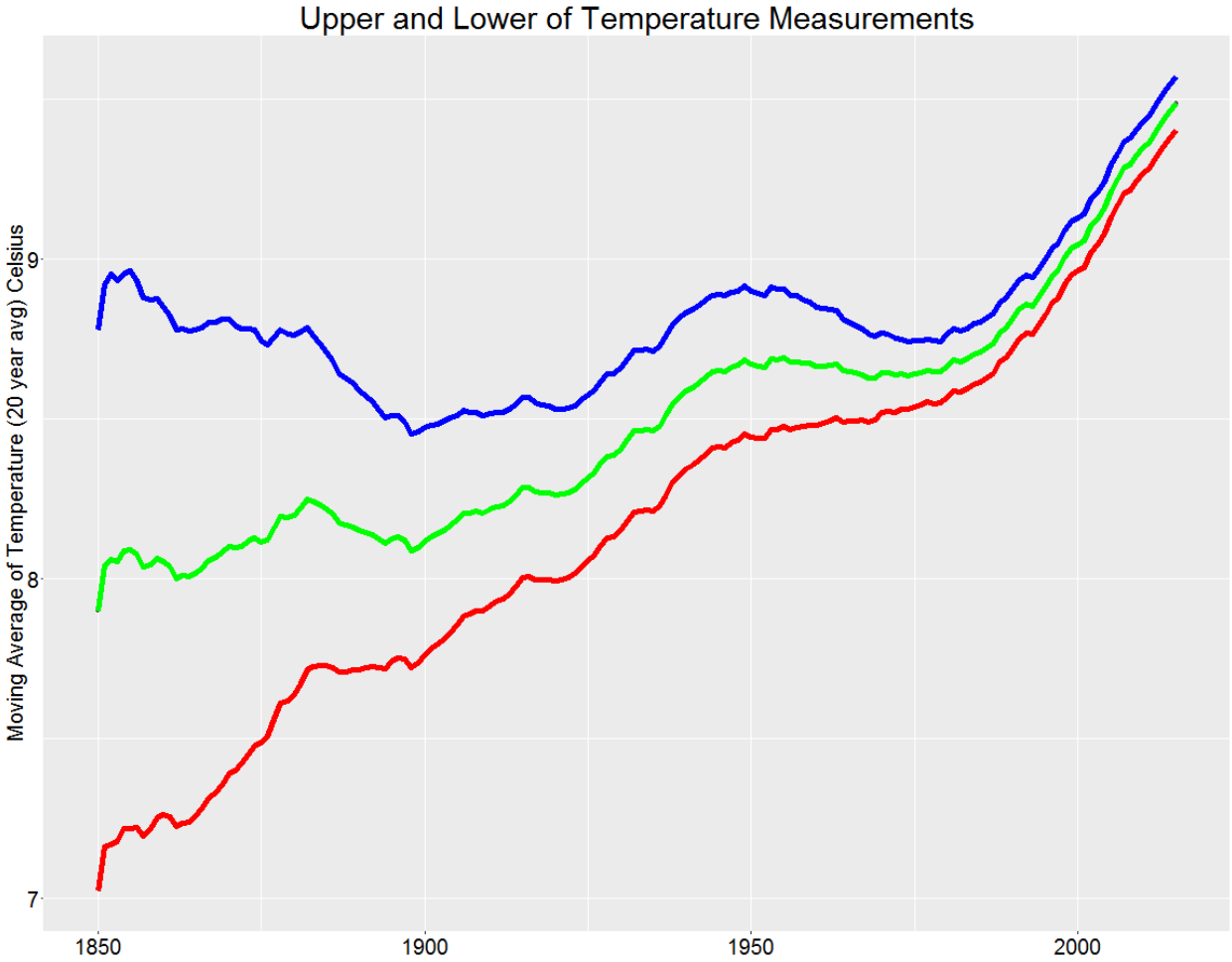
Created a uncertainty linear graph from 1745 to understand the data Average temperature uncertainty and its signficance on the Average temperature.



The plot of average of Average Temperature Uncertainty for Date Year. Color shows details about Country.

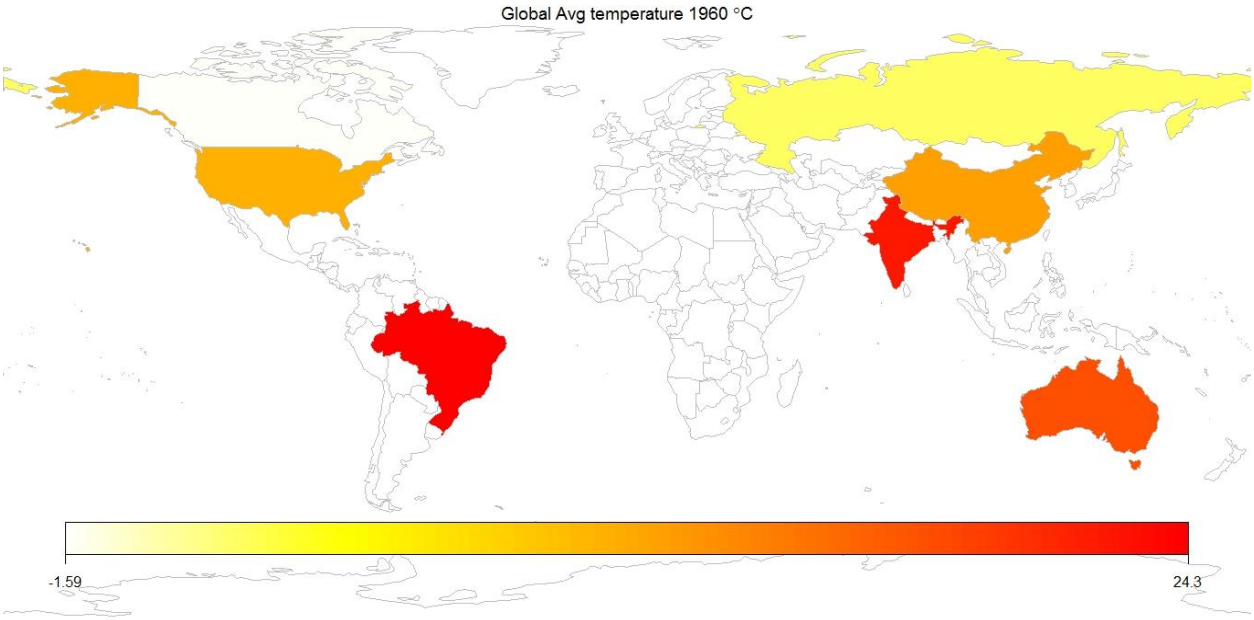
With the Uncertainty graph shows that from 1750 till 1900s there is signficant change in Actual Average Temperature if we add the AverageTemperature Uncertainty.

Approach I:



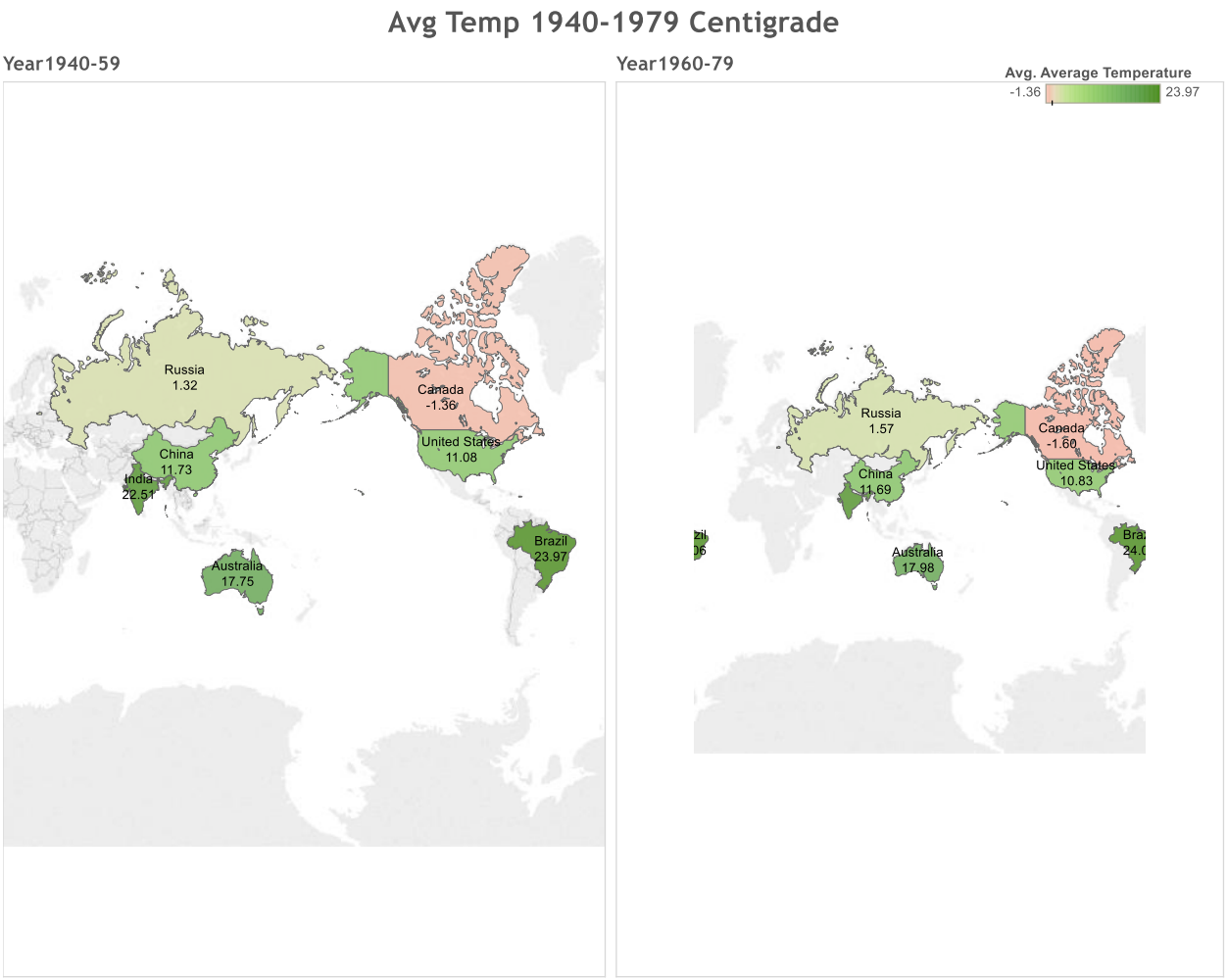
So this made to start deeper analysis of data from 1960s onwards which is of ±.09 celisus degree change in Avg temperature. Here the graph is plotted with green line is the actual AverageTemperature and Red and Blue line is got by adding and subtracting AveragetempUncertainty

Initial Cholorpeth was per year graph



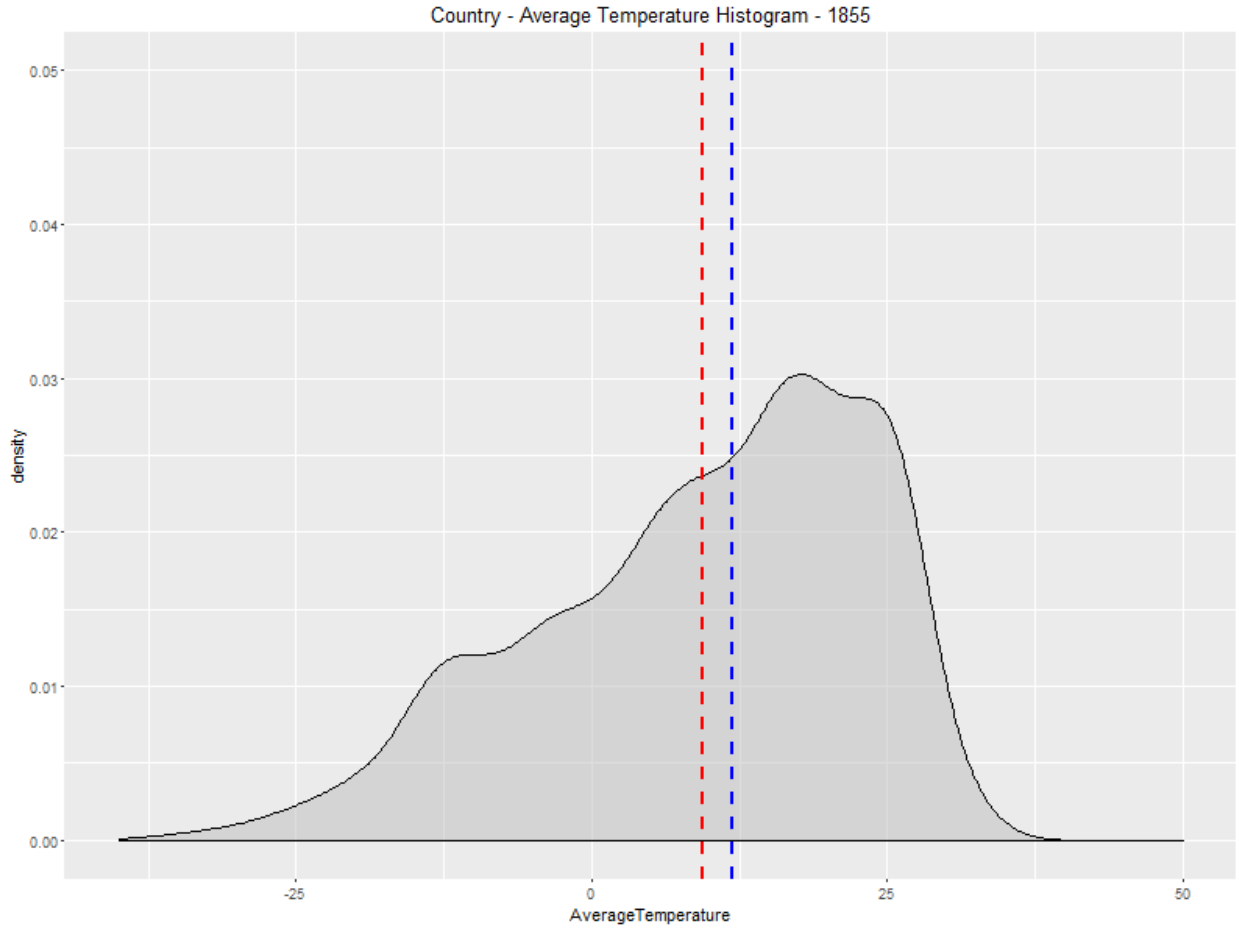
Here choropleth map to present the average temp for a year for different countries, here since only 7 Countries are there most of the Map looks not filled.
So which does not give the full picture of how the avg temperature changed over the years.

So decided to have Choropleth map for Decade to show the Average temperature for 7 countries. This is show how the AverageTemperature has increased for different countries from 1940 to the present 20 Years.



Here also it was too may maps and color used were divergent which did not give clear message.

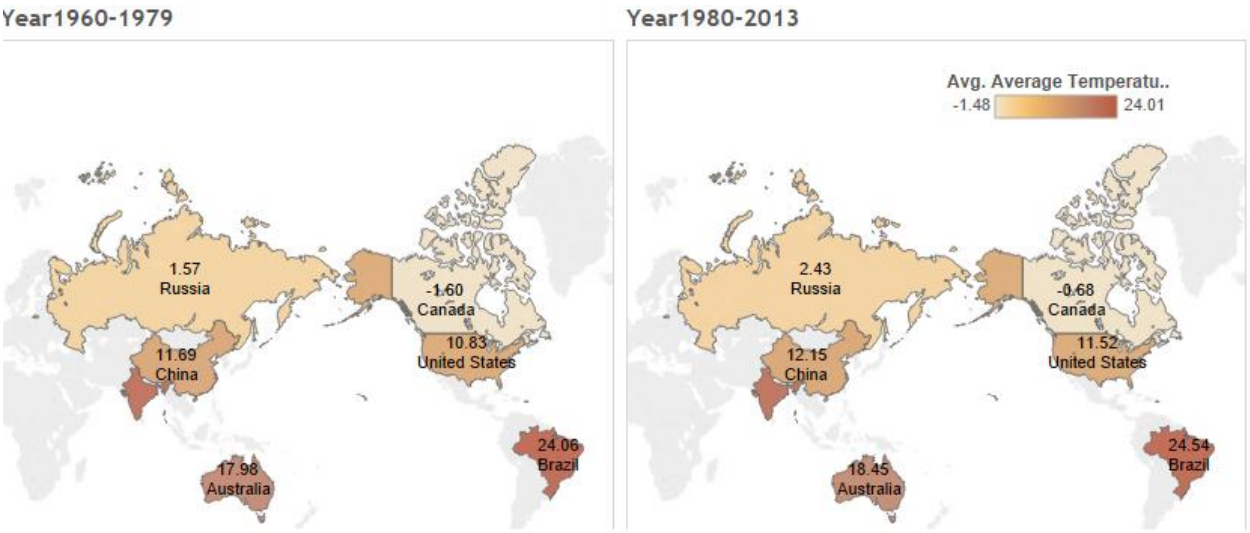
Here below I had to explore find the Density of Avg temperature throughout the year using visually to show the change



The above graph show the change in Avg temperature over the the years and it show the higher density of higher average temperature from 1960s.

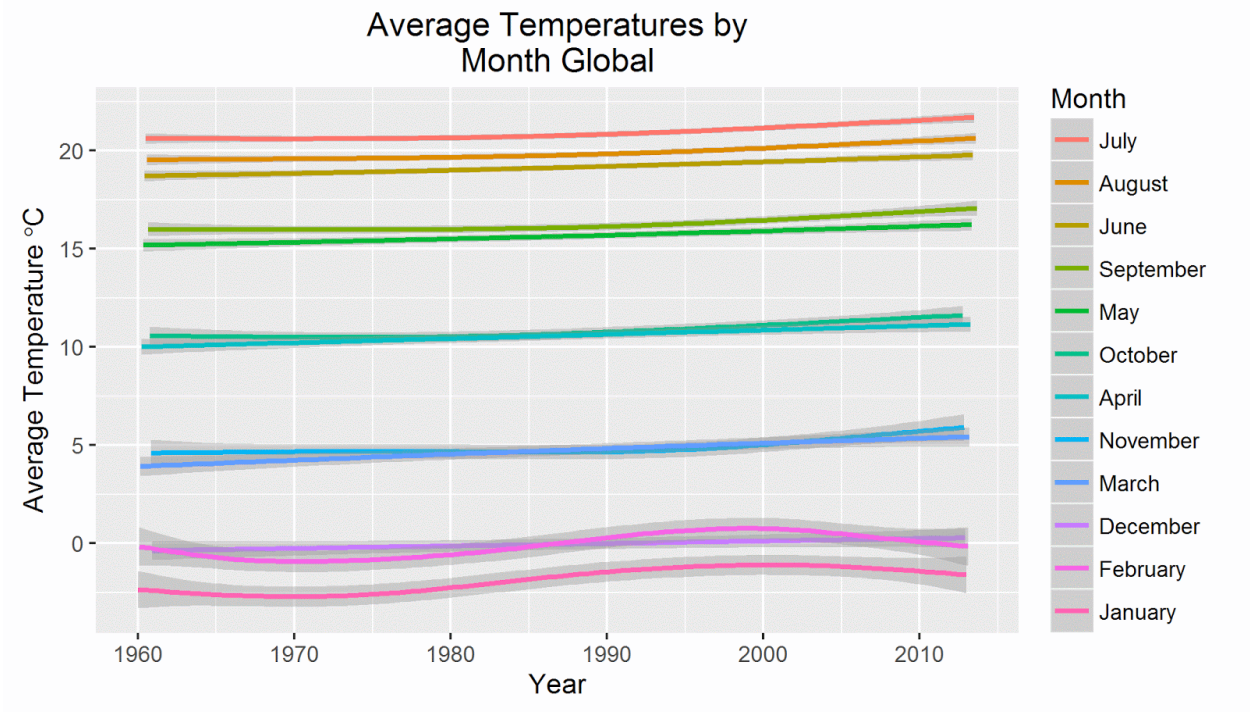
Approach II

Now decided to make the choropleth for 20 years intead of 10 years and which show clearly analysis cou ntry level Average temperature fared from 1960. Tried the Choropleth graph
With country Avg temp from 20 years from 1960 – 79 and 1980 – 2013.

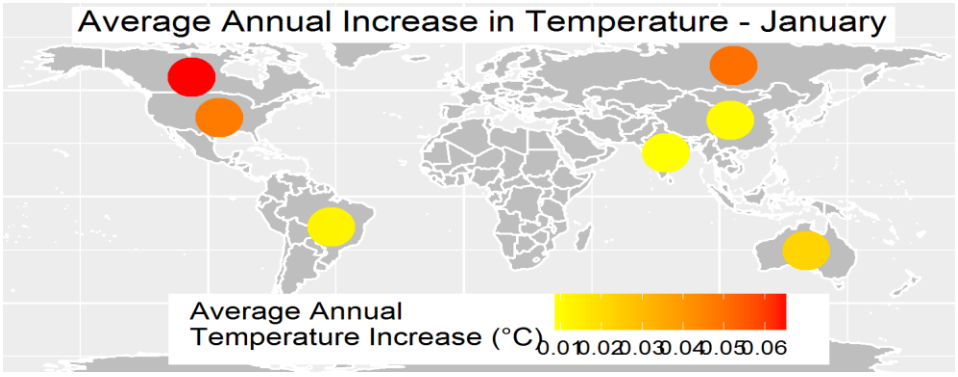


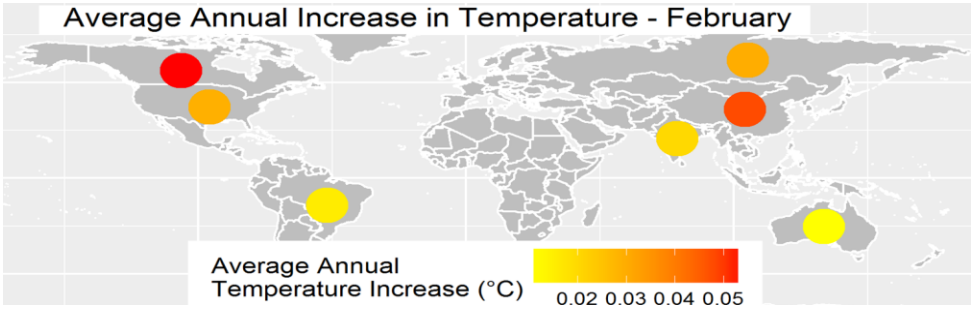
Here we could clearly see the Avg temperature has increased for all 7 countries. We are seeing signficant temperature change for Canada,Russia and USA. The color used is sequential with single color saturation. It is done using Tableau Dashboard.

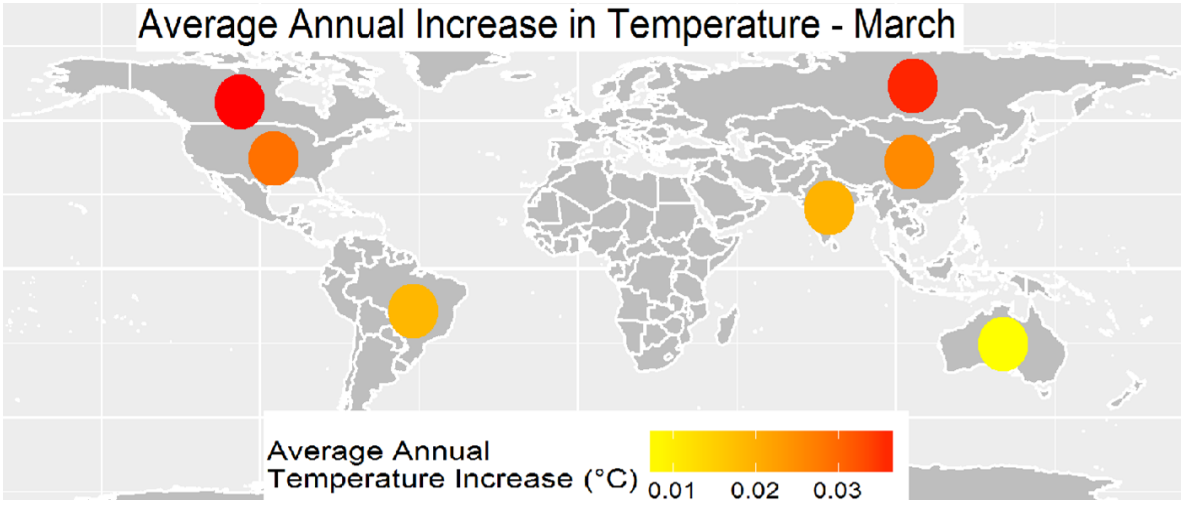
Now we know the temperature are increasing annualy , wanted to create visualization to have monthly change for the each countries.

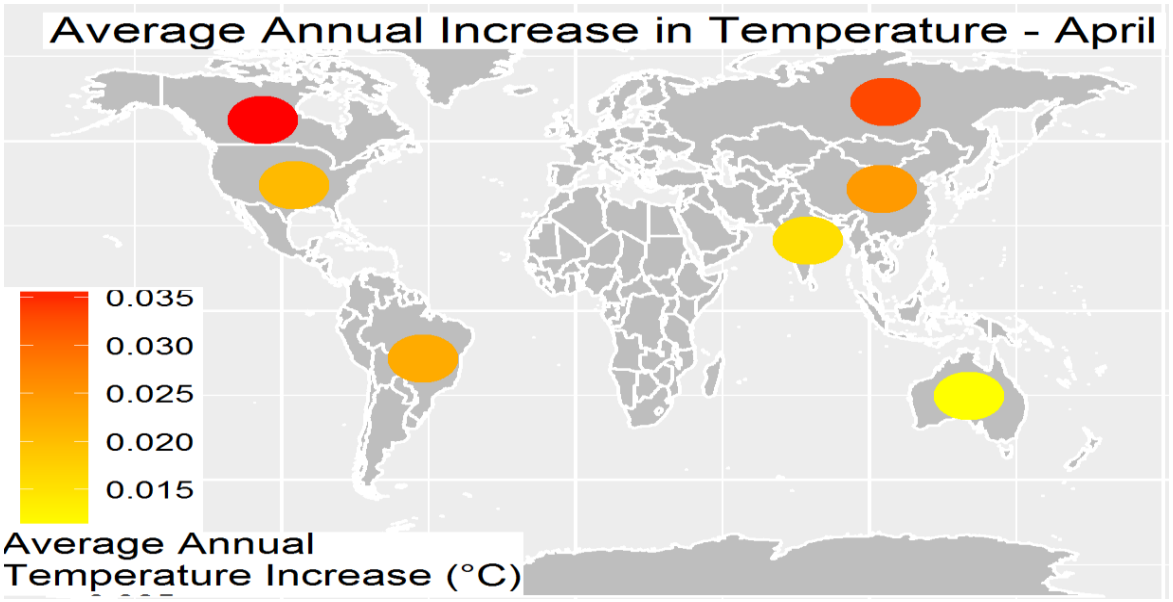


Linear graph clearly show the temperature has increased for every month from 1960.Here the 12 different colors for 12 months .It was build using ggplot.



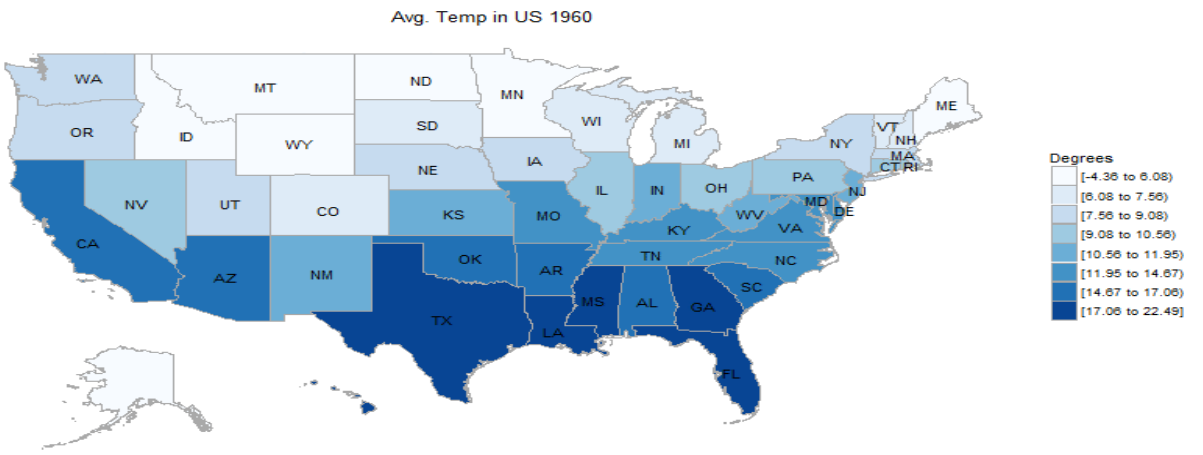






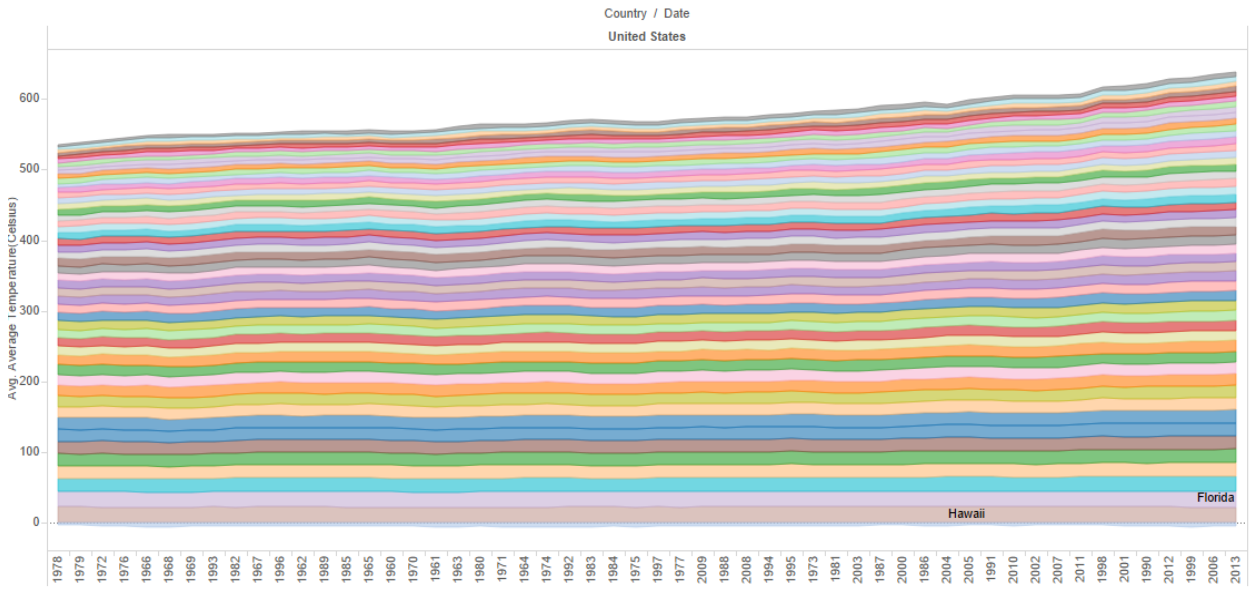
Here it was decided to have world map and point the Monthly Avg temperature Increase per country. Decided to use point to avoid too much color and want show the clear difference between the countries for the month. Colors used are Yellow, orange to red. Red signifies the highest average temp increase while yellow shows the lowest increase. It was done with R with lm function for Avg Temp~ country + month.

This clearly show every month there is change in Average Temperature over the last 50 years.



Created Choropleth map to show the change in Average temperature for US States from 1960

It is animated to show the Average temperature change over the last 50 years. It shows the temperature changes in the northern states. Used 8 color sequential to show the difference in temperature across the 50 states.



Here this graph shows the 50 states (50 different colors) average temperature has increased over the last years. Lowest temperature was in 1978 and highest in 2013.

Approach III

c)Refined Techniques

There were few things has been suggested for Choropleth Decades Avg. temp change for the countries , for the US states animation is for the Average temperature from 1960 it was not providing Visually change in Avg temp for the US States. Also for the US states Cummulative Avg temperature with 50 different color was not reaching the audience. So decided to those Plots.

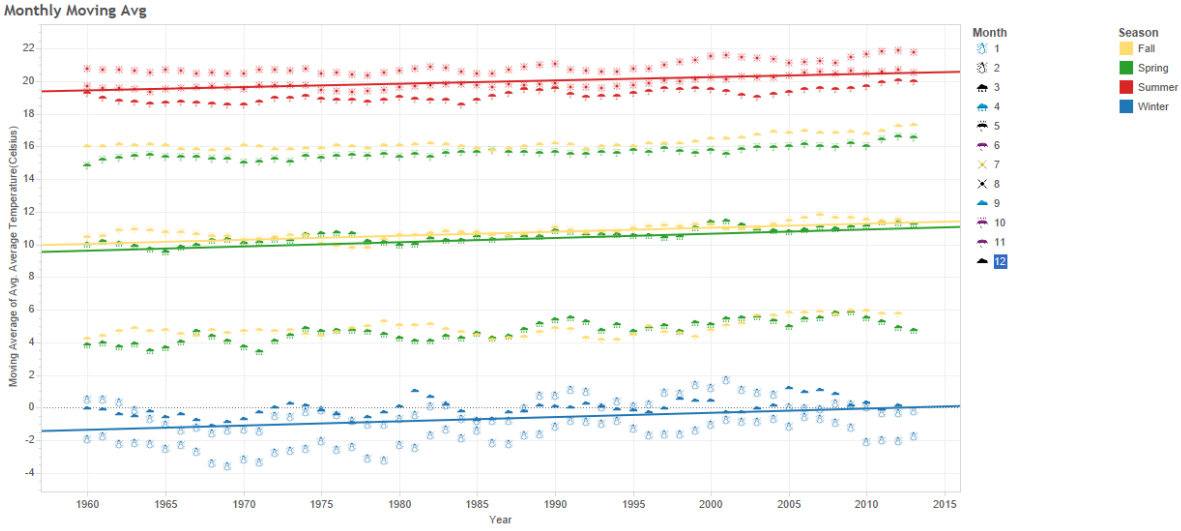
Made the changes necessary in the choropleth map has to be of same time period and also color for the Avg temperature is Divergent instead of Sequential

Map based on Longitude (generated) and Latitude (generated). Color shows average of Average Temperature. The marks are labeled by Country and average of Average Temperature. The data is filtered on Year, which ranges from 1993 to 2002. The view is filtered on Country, which keeps 7 of 7 members.

Yr2003-2012

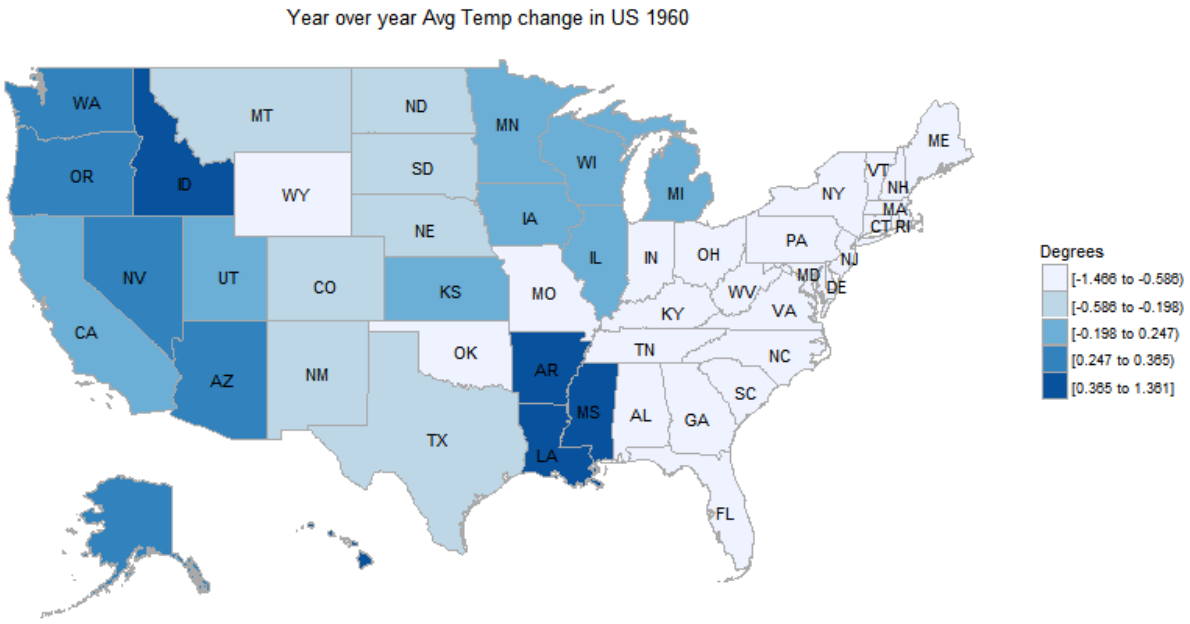
Map based on Longitude (generated) and Latitude (generated). Color shows average of Average Temperature. The marks are labeled by Country and average of Average Tempera
The data is filtered on Year, which ranges from 2003 to 2012. The view is filtered on Country, which keeps 7 of 7 members.

Next suggestion is to make the AverageTemp Monthly linear graph with seasoned coloring .
So is the changed plot.



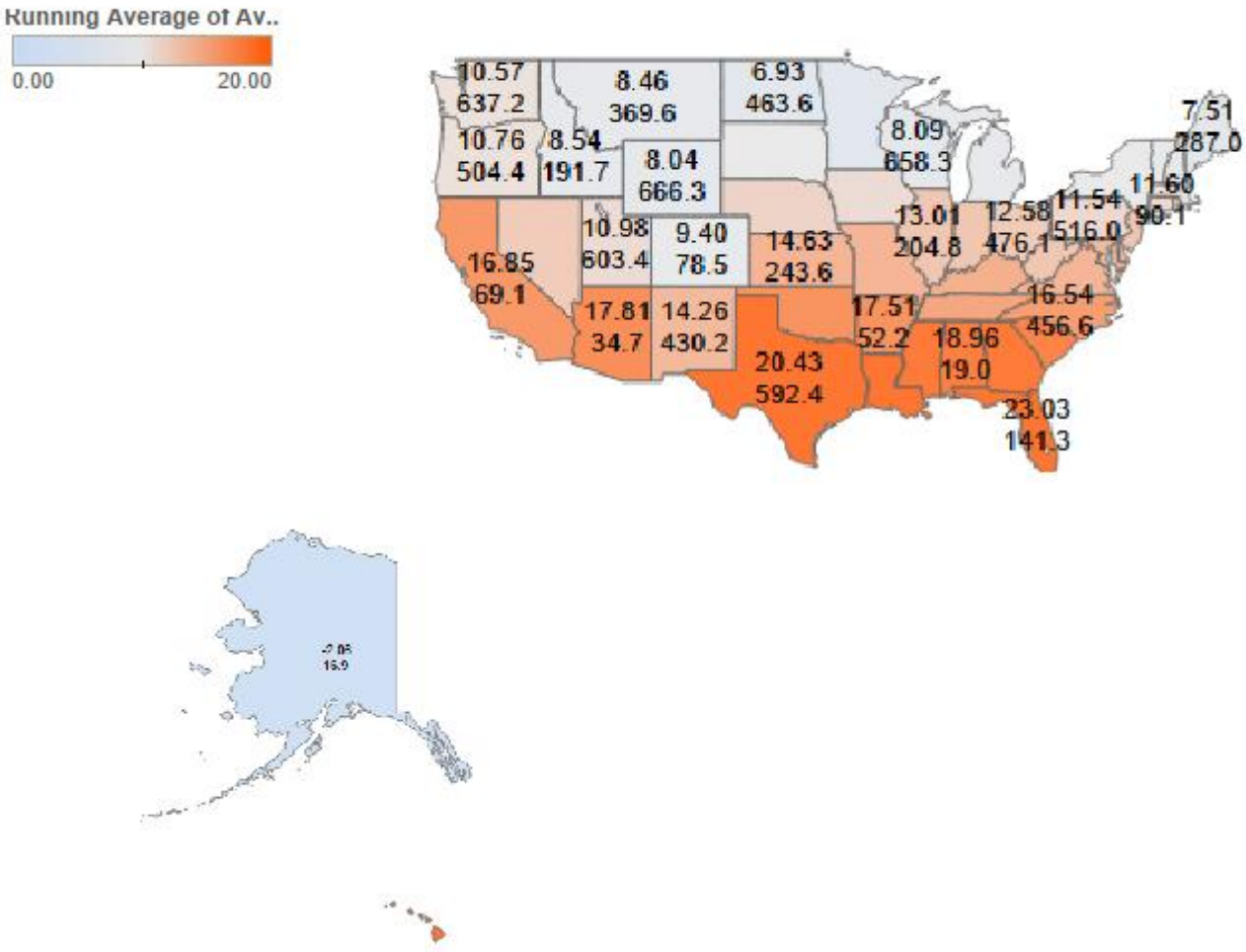
The plot of Moving Average of Avg. Average Temperature for Year. Color shows details about Season. Shape shows details about Month. The view is filtered on Year, which ranges from 1960 to 2013.

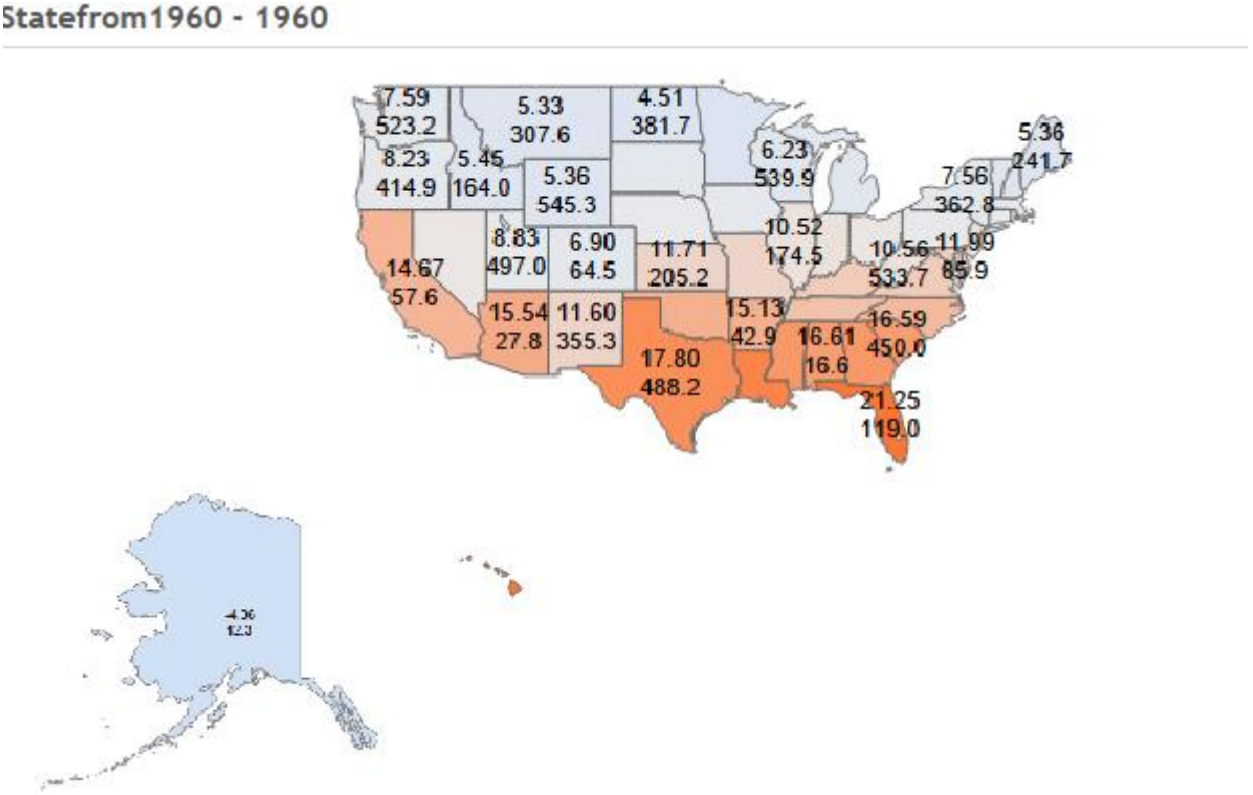
One more thing changed was the US State Animation was changed to show the Year over year change in Avg temperature for different states.



Modified the Cummulative US 50 state Avg temperature from line graph to a choropleth to show running total with diveregent colors for Avg temperature for the States.

Statefrom1960 - 2013





c) Analysis and results

With the help of various graphs and plots did the analysis to answer the question "Is the temperature rising?"

Initial linear graphs have shown the global Average temperature is trending higher. Graph with Average Temperature uncertainties shows that 1700s and 1800s has significant impact on the actual Average temperature. So decided to do analysis from 1960s to find the answer the question "Is the temperature rising?".

With the help of choropleth map we could really see from 1992s to 2012, the Avg. temperature for 7 the countries has increased.

Russia 0.51 Celsius ,Canada 0.36 Celsius, U.S.A 0.27 Celsius

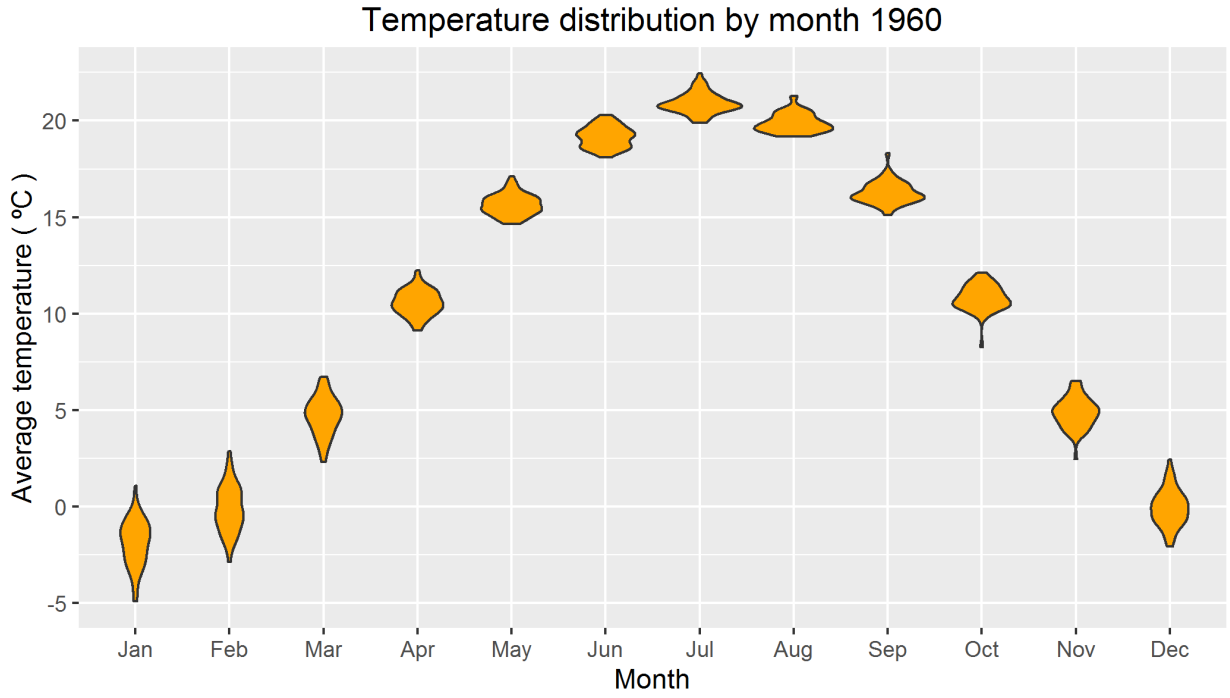
With the help of monthly Average temperature choropleth map it shows every month the temper ature is increasing, significant increase is for the months of January, February and December. Also it shows Canada, Russia and United states are increasing more. It also shows Northern Hem isphere Countries are increasing the Average temperature since 1993s .

With the help of the density plot for the Average temperature also confirms that over the years, te mperature has increased. Density has increased more from 1960s.

Choropleth map for US 50 states shows the Year over Year change in temperature from 1960s. Here for Map US States Avg temp Visually we could see changes are mostly in every region of US. Average Temperature. In 1990s the Average temp on the East coast states have increased. In 1960s the Northern states Avg temperature has increased. In 1980s Western and South west ern states Avg temperature trend increased

In US the during1960 Total 50 states had cumulative total average temperature was 545 Celsius, now in 2013 Total cumulative 666 Celsius. It clearly show the average temperature of most of the 50 states has increasing trend.

Here is a violin plot with Monthly distribution of temperature in 1960



So with the data analysis and plots used we are good to say Average temperature is increasing ov er last 50 years. This data shows that temperature is rising. So we answer the question “Is the tem perature rising?” we could say “YES”. It is a fact it is trending higher. Have not seen data which show that temperature is going down.

d) Appendix

Prab:
I have attached the R code and Tableau Code in drop box. I have learned many things in this course, R ggplot2 , choropleth. Tried few of the packages googleVis,zoo and Js on formats. Tableau was used for creating Dashboard and even tried to publish the Work to public.tableau.com.
With this course now, I would be able to generate plot with proper data visualization using proper color and correct plots.

Source code of R

```
library(ggplot2) # Data visualization
#library(readr) # CSV file I/O, e.g. the read_csv function
library(animation)
library(ggplot2)
library(maptools)
library(rgeos)
library(Cairo)
library(ggmap)
library(scales)
library(RColorBrewer)
library(choroplethrMaps)
library(choroplethr)
library(maps)
library(data.table)
library(dplyr)
library(statebins)
library(SGP)
library(rgdal)
library(rworldmap)
ani.options(convert = 'convert.exe')

GlobalLandTemperaturesByState <- fread("c:/CSC465/project/GlobalLandTemperatu
resByState.csv")

#city <- read.csv("../input/GlobalLandTemperaturesByCity.csv")

GlobalLandTemperaturesByState$year <- substr(GlobalLandTemperaturesByState$dt
,1,4)
allYears <- unique(GlobalLandTemperaturesByState$year)

USData<-GlobalLandTemperaturesByState[GlobalLandTemperaturesByState$Count
ry=='United States']
head(USData)
USData1 <- aggregate(AverageTemperature~year+State,USData,mean)
head(USData1)
library(stringi)
USData1$State<- as.character(USData1$State)
USData1$State[USData1$State=="Georgia (State)"]<-"Georgia"
USData1$State<- as.factor(USData1$State)
USData5<-data.table(USData1)
USData5[, yoy := c(NA, diff(USData5$AverageTemperature)), by = "State"]

typeof(USData5$AverageTemperature)
```

```
USData5$yoy=as.numeric(USData5$yoy)
head(USData5)

datm <- melt(USData1, 'year',
             USData1.name = 'region',
             value.name = 'State'
             # USData1.name = 'AverageTemperature',
             #  value1.name = 'AverageTemperature'
)
head(datm)
#,n=12500)
choropleths = list()
for (i in 2:ncol(datm)) {
  df          = datm[, c(1, i)]
  colnames(df) = c("region", "value")
  title       = paste0("US Temp: ", colnames(USData1)[i])
  choropleths[[i-1]] = state_choropleth(df, title=title)
}

library(choroplethr)
library(choroplethrMaps)
library(sqldf)
library(tcltk)
head(USData)
USData4<-
  sqldf("
select
        year,State
        ,avg(AverageTemperature) as value
        from USData

        group by
        State,year
        ")
USData6<-
  sqldf("
        select
        year,State
        ,yoy as value
        from USData5
        where year >1959
        group by
        State,year
        ")

#USData %>%
```

```
#  select(year,AverageTemperature,State) %>%
#  group_by(year,State) %>%
#  summarise(value=mean(AverageTemperature))-> USData4

USData4$State<- as.character(USData4$State)
USData4$State[USData4$State=="Georgia (State)"]<-"Georgia"
USData4$State<- as.factor(USData4$State)

USData6$State<- as.character(USData6$State)
USData6$State[USData6$State=="Georgia (State)"]<-"Georgia"
USData6$State<- as.factor(USData6$State)


colnames(USData4)[2]<- "region"
USData4$region<-tolower(USData4$region)
head(USData)
USData4 <- na.omit(USData4)

colnames(USData6)[2]<- "region"
USData6$region<-tolower(USData6$region)
head(USData)
USData6 <- na.omit(USData6)


head(USData4)
dev.off()
print(state_choropleth(USData4[USData4$year==1900],
                       title="Land Temperature 1900",
                       num_colors = 8,
                       legend="Degrees"),reference_map=TRUE)



saveGIF({
  for (thisYear in 1960:2013 ) {

    thisYearCountry <- USData4[USData4$year == thisYear, ]
    thisYearCountry <- na.omit(thisYearCountry)
    print(state_choropleth(thisYearCountry,
                           title=paste("Avg. Temp in US", thisYear),
                           num_colors = 8,
                           #brewer_pal="YlOrRd",
                           legend="Degrees"),reference_map=TRUE)
  }
```

```
}, interval = 0.3, movie.name = "tempDensity3.gif", ani.width = 800, ani.height = 600
)


saveGIF({
  for (thisYear in 1960:2013 ) {

    thisYearCountry <- USData6[USData6$year == thisYear, ]
    thisYearCountry <- na.omit(thisYearCountry)
    print(state_choropleth(thisYearCountry,
                           title=paste("Year over year Avg Temp change in US", thisYear),
                           num_colors = 8,
                           #brewer_pal="YlOrRd",
                           legend="Degrees"),reference_map=TRUE)
  }

}, interval = 0.3, movie.name = "tempYoy1.gif", ani.width = 800, ani.height = 600)


saveGIF({

 for (thisYear in 1960:2013 ) {
  thisYearCountry <- USData4[USData4$year == thisYear, ]
  thisYearCountry <- na.omit(thisYearCountry)
  #stateData$value=thisYearCountry$value[!sapply(thisYearCountry, is.null)]
  choro = StateChoropleth$new(na.omit(thisYearCountry))
  choro$title = paste("Avg. Temp in US", thisYear)
  choro$ggplot_scale = scale_fill_brewer(name="Avg. temp Degrees"~degree~C, pal
ette="Purples", drop=FALSE)
  choro$render()
  }

}, interval = 0.3, movie.name = "tempDensity2.gif", ani.width = 800, ani.height = 600
)

#saveGIF({
setwd('C:/CSC465/Project/images')
  for (thisYear in 1960:2013 ) {
    thisYearCountry <- USData4[USData4$year == thisYear, ]
    thisYearCountry <- na.omit(thisYearCountry)
    if (thisYear >= 1960) {name = paste('US', thisYear,'plot.png', sep='')}
    #if (i < 10) {name = paste('000',i,'plot.png',sep='')}

    #if (i < 100 && i >= 10) {name = paste('00',i,'plot.png', sep='')}
    choro = StateChoropleth$new(na.omit(thisYearCountry))
    choro$title = paste("Avg. Temp in US", thisYear)
```

```
   choro$ggplot_scale = scale_fill_brewer(name="Avg. temp Degrees"~degree~C, pa
lette="Purples", drop=FALSE)
   choro$render()
   #$render()
   if (thisYear >= 1960) {name = paste('US', thisYear,'plot60.png', sep='')}
   png(name)

   dev.off()
}


#install.packages('animation', repos = 'http://yihui.name/xran')
#library(animation)

data(thisYearCountry.regions)
head(thisYearCountry)
saveGIF({
  for (thisYear in allYears) {

   thisYearCountry <- GlobalLandTemperaturesByState[GlobalLandTemperaturesBy
State$year == thisYear, ]

   m <- ggplot(thisYearCountry, aes(x=AverageTemperature))
   m <- m + ggtitle(paste("Country - Average Temperature Histogram -", thisYear))
   m <- m + geom_density(alpha=.5, fill = "gray")
   m <- m + xlim(c(-40, 50))
   m <- m + ylim(c(0, .05))
   m <- m + geom_vline(aes(xintercept=mean(AverageTemperature, na.rm=T)),   # Ig
nore NA values for mean
                       color="red", linetype="dashed", size=1)
   m <- m + geom_vline(aes(xintercept=median(AverageTemperature, na.rm=T)),   #
Ignore NA values for mean
                       color="blue", linetype="dashed", size=1)

   print(m)

  }

}, interval = 0.6, movie.name = "tempDensity6.gif", ani.width = 800, ani.height = 600
)
```

World temp change from 1960 R source code

```
library(ggplot2)
library(maptools)
```

```
library(rgeos)
library(Cairo)
library(ggmap)
library(scales)
library(RColorBrewer)
library(choroplethrMaps)
library(choroplethr)
library(maps)
library(data.table)
library(dplyr)
library(statebins)
library(SGP)
library(rgdal)
library(rworldmap)
typeof(GlobalLandTemperaturesByState$AverageTemperature)
GlobalLandTemperaturesByState <- fread("c:/CSC465/project/GlobalLandTemperaturesByState.csv")
GlobalLandTemperaturesByState1 <- fread("c:/CSC465/project/GlobalLandTemperaturesByState1.csv")

head(GlobalLandTemperaturesByState)
GlobalLandTemperaturesByState$dt<-as.Date(GlobalLandTemperaturesByState$dt,"%Y-%m-%d")
GlobalLandTemperaturesByState$Month<-as.numeric(format(GlobalLandTemperaturesByState$dt,"%m"))
GlobalLandTemperaturesByState$Month.String<-format(GlobalLandTemperaturesByState$dt,"%B")
GlobalLandTemperaturesByState$Year<-as.numeric(format(GlobalLandTemperaturesByState$dt,"%Y"))

GlobalLandTemperaturesByState1$dt<-as.Date(GlobalLandTemperaturesByState1$dt,"%Y-%m-%d")
GlobalLandTemperaturesByState1$Month<-as.numeric(format(GlobalLandTemperaturesByState1$dt,"%m"))
GlobalLandTemperaturesByState1$Month.String<-format(GlobalLandTemperaturesByState1$dt,"%B")
GlobalLandTemperaturesByState1$Year<-as.numeric(format(GlobalLandTemperaturesByState1$dt,"%Y"))
head(GlobalLandTemperaturesByState1)
data11<-GlobalLandTemperaturesByState1[GlobalLandTemperaturesByState1$Year>1960]
head(data11)
GlobalLandTemperaturesByState.recent.1960 <- GlobalLandTemperaturesByState[GlobalLandTemperaturesByState$Year==1960,]
GlobalLandTemperaturesByState.recent.1960 <- aggregate(AverageTemperature~Year+Country,GlobalLandTemperaturesByState.recent,mean)
```

```
head(GlobalLandTemperaturesByState.recent.1960)
#join data to a map
gtdMap.1960 <- joinCountryData2Map( GlobalLandTemperaturesByState.recent.196
0,
                    nameJoinColumn="Country",
                    joinCode="NAME" )

mapDevice('x11') #create a world shaped window

#plot the map
mapCountryData( gtdMap.1960,
          nameColumnToPlot='AverageTemperature',
          catMethod='fixedWidth',
          numCats=100,
          mapTitle="Global Avg temperature 1960"~degree*C)


GlobalLandTemperaturesByState.recent1.1960 <- GlobalLandTemperaturesByState[
GlobalLandTemperaturesByState$Year==1960,]
head(GlobalLandTemperaturesByState.recent1.1960)
GlobalLandTemperaturesByState.recent1.1960 <- aggregate(GlobalLandTemperature
sByState.recent1.1960$AverageTemperatureUncertainty~GlobalLandTemperaturesB
yState.recent1.1960$Year+GlobalLandTemperaturesByState.recent1.1960$Country,
GlobalLandTemperaturesByState.recent1.1960,mean)

head(GlobalLandTemperaturesByState.recent1.1960)
#join data to a map
gtdMap1.1960 <- joinCountryData2Map( GlobalLandTemperaturesByState.recent1.1
960,
                    nameJoinColumn="GlobalLandTemperaturesByState.recent1.196
0$Country",
                    joinCode="NAME" )

mapDevice('x11') #create a world shaped window

#plot the map
mapCountryData( gtdMap1.1960,
          addLegend = TRUE, borderCol = "grey",
          nameColumnToPlot='GlobalLandTemperaturesByState.recent1.1960$Aver
ageTemperatureUncertainty',
          catMethod='fixedWidth',
          numCats=100,
          mapTitle="Global Avg.Uncertainty temperature 1960")
```

```
GlobalLandTemperaturesByState.recent1.1980 <- GlobalLandTemperaturesByState[
GlobalLandTemperaturesByState$Year==1980,]
head(GlobalLandTemperaturesByState.recent1.1980)
GlobalLandTemperaturesByState.recent1.1980 <- aggregate(GlobalLandTemperature
sByState.recent1.1980$AverageTemperature~GlobalLandTemperaturesByState.recen
t1.1980$Year+GlobalLandTemperaturesByState.recent1.1980$Country,GlobalLandT
emperaturesByState.recent1.1980,mean)

head(GlobalLandTemperaturesByState.recent1.1980)
#join data to a map
gtdMap1.1980 <- joinCountryData2Map( GlobalLandTemperaturesByState.recent1.1
980,
                        nameJoinColumn="GlobalLandTemperaturesByState.recent1.
1980$Country",
                        joinCode="NAME" )

mapDevice('x11') #create a world shaped window

#plot the map
mapCountryData( gtdMap1.1980,
            addLegend = TRUE, borderCol = "grey",
            nameColumnToPlot='GlobalLandTemperaturesByState.recent1.1980$Aver
ageTemperature',
            catMethod='fixedWidth',
            numCats=100,
            mapTitle="Global Avg. temperature 1980"~degree*C)


library(scatterplot3d)
# create column indicating point color
GlobalLandTemperaturesByState.recent1.1960$pcolor[GlobalLandTemperaturesByS
tate.recent1.1960$Country=="India"] <- "brown"
GlobalLandTemperaturesByState.recent$pcolor[GlobalLandTemperaturesByState.rec
ent$Country=="Brazil"] <- "yellow"
GlobalLandTemperaturesByState.recent$pcolor[GlobalLandTemperaturesByState.rec
ent$Country=="United States"] <- "blue"
GlobalLandTemperaturesByState.recent$pcolor[GlobalLandTemperaturesByState.rec
ent$Country=="Russia"] <- "orange"
GlobalLandTemperaturesByState.recent$pcolor[GlobalLandTemperaturesByState.rec
ent$Country=="China"] <- "red"
GlobalLandTemperaturesByState.recent$pcolor[GlobalLandTemperaturesByState.rec
ent$Country=="Australia"] <- "black"
GlobalLandTemperaturesByState.recent$pcolor[GlobalLandTemperaturesByState.rec
ent$Country=="Canada"] <- "green"

with(GlobalLandTemperaturesByState.recent1.1960, {
```

```
  s3d <- scatterplot3d(Country,Year,AverageTemperature,        # x y and z axis
                       color=pcolor, pch=19,        # circle color indicates no. of cylinders
                       type="h", lty.hplot=2,       # lines to the horizontal plane
                       scale.y=.75,                 # scale y axis (reduce by 25%)
                       main="3-D Scatterplot Example 4",
                       zlab="Avg temperature"~degree*C,
                       ylab="Year",
                       ylab="Country",
                      # ylim=c(1963, 2013),
                       zlim=c(-10, 30),
                       ylim=c("India", "Brazil", "United States","Russia","China","Australia",
"Canada"), fill=c("brown", "yellow", "blue","orange","red","black","green"))
  s3d.coords <- s3d$xyz.convert(Country,Year,AverageTemperature)
  text(s3d.coords$x, s3d.coords$y,     # x and y coordinates
       labels=row.names(GlobalLandTemperaturesByState.recent),       # text to plot
       pos=4, cex=.5)                 # shrink text 50% and place to right of points)
  # add the legend
  legend("topleft", inset=.05,      # location and inset
         bty="n", cex=.5,              # suppress legend box, shrink text 50%
         title="Global Temperature",
         c("India", "Brazil", "United States","Russia","China","Australia","Canada"), fill
=c("brown", "yellow", "blue","orange","red","black","green"))
})
USData<-GlobalLandTemperaturesByState.recent.1960[GlobalLandTemperaturesBy
State.recent.1960$Country=="United States"]

head(USData)
library(lattice)
library(gclus)
library(car)
scatterplotMatrix(~AverageTemperature+Month,data=USData,
      main="US Avg temp month Scatter Plot 1960")
GlobalLandTemperaturesByState[,scaled.temperature:=(ScaledTemperature=scale(A
verageTemperature)), by=.(Country,Month)]
gc(reset=TRUE)

ggplot(GlobalLandTemperaturesByState,aes(x=dt,y=scaled.temperature))+
  stat_bin_2d(bins=100)+scale_fill_gradient(low="lightblue",high="red")+geom_smo
oth(colour="purple")+
  ggtitle("Centered and Scaled Temperatures Over Time")

summary(GlobalLandTemperaturesByState.recent.1960$AverageTemperature)
ggplot(GlobalLandTemperaturesByState.recent.1960,aes(x=AverageTemperature))+g
eom_density()
library(randomForest)
library(fitdistrplus)
```

```
descdist(GlobalLandTemperaturesByState.recent.1960$AverageTemperature,graph=
TRUE)
ggplot(GlobalLandTemperaturesByState.recent.1960[Month==1],aes(x=AverageTem
perature))+geom_density()

ggplot(GlobalLandTemperaturesByState.recent.1960[Month==5],aes(x=AverageTem
perature))+geom_density()
len=length(GlobalLandTemperaturesByState.recent.1960$Country)
len
coeff=numeric(len )
coeff

#res=persp(x=GlobalLandTemperaturesByState.recent$Country, y=GlobalLandTemp
eraturesByState.recent$Year, z=GlobalLandTemperaturesByState.recent$AverageTe
mperature)
#mypoints = trans3d(GlobalLandTemperaturesByState.recent$Country,GlobalLandT
emperaturesByState.recent$Year,GlobalLandTemperaturesByState.recent$AverageT
emperature, pmat=res)
#points(mypoints, pch=1, lwd=2, col="red")


install.packages("devtools")  # so we can install from github
library("devtools")
install_github("ropensci/plotly")  # plotly is part of ropensci
library(plotly)

py <- plotly(username="r_user_guide", key="mw5isa4yqp")  # open plotly connectio
n

# Generate data
library(reshape2) # for melt
GLT=data.frame(GlobalLandTemperaturesByState.recent$Year,GlobalLandTempera
turesByState.recent$Country,GlobalLandTemperaturesByState.recent$AverageTemp
erature)
head(GLT)
volcano3d <- melt(GLT,id=c("GlobalLandTemperaturesByState.recent.Year"))
names(volcano3d) <- c("Year", "AverageTemperature", "Country")
head(volcano3d)
# Basic plot
head(GlobalLandTemperaturesByState.recent)
v <- ggplot(volcano3d, aes(GlobalLandTemperaturesByState.recent$Year, GlobalLan
dTemperaturesByState.recent$AverageTemperature ))
v   + stat_contour()
v <- ggplot(volcano3d, aes(Year, AverageTemperature,Country ))
v   + stat_contour() + geom_tile(aes(fill = AverageTemperature))
#+
```

```
#geom_tile(aes(fill = GlobalLandTemperaturesByState.recent$Country))
py$ggplotly()

st1=toupper(state.name)
st1
head(GlobalLandTemperaturesByState)
Data1 <- subset(GlobalLandTemperaturesByState1, toupper(State) %in% c(st1))


  Data1<- subset( Data1,            Data1$Year == '1960')
Data1

data3=aggregate(Data1$AverageTemperature, by=list(Data1$Year,Data1$State), FU
N=mean)[3]
head(Data1)
Data2 <-na.omit(subset(Data1, Year>1960))
setkey(Data2$Year,Data2$State)
Data2 %>%
  #group_by(Year) %>%
  ummarise(Temp = mean(AverageTemperature)) ->Data2
library(sqldf)
head(Data1)

data4=sqldf("
select
     State,Year
     ,avg(AverageTemperature) as Avgtemp
     from Data1

     group by
     State,Year
     ")

head(data4,99)
typeof(data4$Year)

Data1960 <-subset(data4, Year=='1960')
head(Data1960)
#where Year == '1960'
value=data4$Avgtemp[!sapply(data4$Avgtemp, is.null)]
region=data4$State[!sapply(data4$State, is.null)]
region
value

region=tolower(region)
#region=sapply(region, capwords)
```

```
#region1=capwords(region1)
stateData1960=data.frame(region,value)
nclr <- 8 # number of bins
min <- -30 # theoretical minimum
max <- 30 # theoretical maximum
breaks <- (max - min) / nclr
library(RColorBrewer)
library(classInt)

plotclr <- brewer.pal(nclr, "Oranges")
plotvar <- stateData1960$value
class <- classIntervals(plotvar,
                        nclr,
                        style = "fixed",
                        fixedBreaks = seq(min, max, breaks))
colcode <- findColours(class,
                       plotclr)

library(OIdata)

map("state", # base
    col = "gray80",
    fill = TRUE,
    lty = 0)
map("state", # data
    col = colcode,
    fill = TRUE,
    lty = 0,
    add = TRUE)
map("state", # border
    col = "gray",
    lwd = 1.4,
    lty = 1,
    add = TRUE)
legend("bottomright", # position
       legend = names(attr(colcode, "table")),
       title = "Percent",
       fill = attr(colcode, "palette"),
       cex = 0.56,
       bty = "n") # border


plotclr <- brewer.pal(nclr, "Oranges")
plotvar <- St
class <- classIntervals(plotvar,
                        nclr,
```

```
                    style = "fixed",
                    fixedBreaks = seq(min, max, breaks))
colcode <- findColours(class,
                   plotclr)
NAColor <- "gray80"
plotclr <- c(plotclr, NAColor)
choro = StateChoropleth$new(na.omit(stateData1960))
choro$title = "1960 Average Temp by State"
choro$ggplot_scale = scale_fill_brewer(name="Avg Temp", palette=2, drop=FALSE
)
choro$render()

install.packages("ggplot2")
library(ggplot2)
install.packages("maps")
library(maps)
install.packages("mapproj")
#library(mapproj)
#install.packages("spatstat")
library(spatstat)

theme_set(theme_bw(base_size = 8))
options(scipen = 20)

MyPalette <- colorRampPalette(c(hsv(0, 1, 1), hsv(7/12, 1, 1)))

### Map ###
StateMapData <- map_data("state")
head(StateMapData)

### Some Invented Data ###

IndependentVariable1 <- c("1900-1930", "1931-1960", "1961-2013")
IndependentVariable2 <- c("-5-10", "11-15", "16-20", "21+")

# Here is one way to "stack" lots of copies of the shapefile dataframe on top of each o
ther:
# This needs to be done, because (as far as I know) ggplot2 needs to have the state na
mes and polygon coordinates
# for each level of the faceting variables.

TallData <- expand.grid(1:nrow(StateMapData), IndependentVariable1, Independent
Variable2)
TallData <- data.frame(StateMapData[TallData[, 1], ], TallData)
colnames(TallData)[8:9] <- c("IndependentVariable1", "IndependentVariable2")
```

```
# Some random dependent variable we want to plot in color:
TallData$State_IV1_IV2 <- paste(TallData$region, TallData$IndependentVariable1,
TallData$IndependentVariable2)
RandomVariable <- runif(length(unique(TallData$State_IV1_IV2)))
TallData$DependentVariable <- by(RandomVariable, unique(TallData$State_IV1_IV
2), mean)[TallData$State_IV1_IV2]

### Plot ###

MapPlot <- ggplot(TallData,
                  aes(x = long, y = lat, group = group, fill = DependentVariable))
MapPlot <- MapPlot + geom_polygon()
MapPlot <- MapPlot + coord_map(project="albers", at0 = 45.5, lat1 = 29.5)  # Chang
es the projection to something other than Mercator.
MapPlot <- MapPlot + scale_x_continuous(breaks = NA, expand.grid = c(0, 0)) +
  scale_y_continuous(breaks = NA) +
  opts(
    panel.grid.major = theme_blank(),
    panel.grid.minor = theme_blank(),
    panel.background = theme_blank(),
    panel.border = theme_blank(),
    expand.grid = c(0, 0),
    axis.ticks = theme_blank(),
    legend.position = "none",
    legend.box = "horizontal",
    title = "Here is my title",
    legend.key.size = unit(2/3, "lines"))
MapPlot <- MapPlot + xlab(NULL) + ylab(NULL)
MapPlot <- MapPlot + geom_path(fill = "transparent", colour = "BLACK", alpha = I(
2/3), lwd = I(1/10))
MapPlot <- MapPlot + scale_fill_gradientn("Some/nRandom/nVariable", legend = F
ALSE,
                                  colours = MyPalette(100))

# This does the "faceting":
MapPlot <- MapPlot + facet_grid(IndependentVariable2 ~ IndependentVariable1)

# print(MapPlot)

ggsave(plot = MapPlot, "YOUR DIRECTORY HERE.png", h = 8.5, w = 11)

library(RJSONIO)
library(googleVis)
#data11<-GlobalLandTemperaturesByState1[GlobalLandTemperaturesByState1$Yea
r>1960]
#head(data11)
```

```
#GlobalLandTemperaturesByState.YrAvg <- GlobalLandTemperaturesByState[Glob
alLandTemperaturesByState$Year==1960,]
GlobalLandTemperaturesByState.YrAvg <- aggregate(AverageTemperature~Year+C
ountry,GlobalLandTemperaturesByState,mean)
GlobalLandTemperaturesByState.YrAvgUn <- aggregate(AverageTemperatureUncer
tainty~Year+Country,GlobalLandTemperaturesByState,mean)
head(GlobalLandTemperaturesByState.YrAvg)
head(GlobalLandTemperaturesByState.YrAvgUn)
YrAvgUn<-merge(GlobalLandTemperaturesByState.YrAvg,GlobalLandTemperature
sByState.YrAvgUn)
head(YrAvgUn)
vis1<-gvisMotionChart(YrAvgUn,idvar='Country',timevar='Year')
plot(vis1)

Geo=gvisGeoChart(YrAvgUn, locationvar="Country",
                 colorvar="AverageTemperature",
                 options=list(projection="kavrayskiy-vii"))
plot(Geo)
countryunique <- unique(GlobalLandTemperaturesByState$Country)
countryuniqueMonth <- unique( GlobalLandTemperaturesByState$Month)
countryunique1<-merge(countryunique,countryuniqueMonth)
head(countryuniqueMonth)
head(GlobalLandTemperaturesByState)
head(countryunique1,999)
colnames(countryunique1)<-c("Country", "Month")
#countryunique1<-countryunique1[order("Month"),]
#sort(countryunique1$Month)

#countrylatlong <- data.frame([unique(GlobalLandTemperaturesByState$Country),un
ique(GlobalLandTemperaturesByState$Year)
#head(GlobalLandTemperaturesByState)
  data(wrld_simpl)
head(wrld_simpl)
wrld_simpl@data$id <- wrld_simpl@data$NAME
wrld <- fortify(wrld_simpl, region="id")
wrld <- subset(wrld, id != "Antarctica")
head(wrld,999)
countrylatlong1 <- fread("c:/CSC465/project/countrylatlong.csv")
head(countrylatlong1)


head(countrylatlong1)
countrylatlong<-merge(countryunique1,countrylatlong1,by.x="Country",by.y="Count
ry")
head(countrylatlong,99)
head(meta.country,99)
```

```
meta.country<-unique(countrylatlong)
meta.country<-data.table(meta.country)
meta.country<-unique(countrylatlong[,c(1:4),with=FALSE],by=c("Country","Month"
,"Latitude","Longitude"))
setkey(meta.country,Country,Month)

meta.country.length<-length(meta.country$Country)
meta.country$intercept.coef<-numeric(meta.country.length)
meta.country$year.coef<-numeric(meta.country.length)
typeof(meta.country$Country)
typeof(meta.country$Month)
typeof(dt1$Country)
typeof(dt1$Month)
typeof(i)
meta.country$Country <- as.double(meta.country$Country)
dt1$Country <- as.integer(dt1$Country)

#Create a data table for faster subsetting. Data before 1880 is rejected (uncertainty is t
oo high)
dt1 <- as.data.table(na.omit(subset(GlobalLandTemperaturesByState,Year>1960)))
#dt1 <- merge(dt1,countrylatlong,by.x="Country",by.y="Country")
data.table(dt1)
setkey(dt1,Country,Month)
head(dt1.subset)
head(meta.country)
#This loop will fill in the columns of the meta.city table.
for(i in 1:meta.country.length){
  dt1.subset<-dt1[list(meta.country$Country[i],meta.country$Month[i]),]
  lmfit<-with(dt1.subset,lm.fit(x=cbind(1,Year),y=AverageTemperature))
  meta.country$intercept.coef[i]<-lmfit$coefficients[1]
  meta.country$year.coef[i]<-lmfit$coefficients[2]
}
#ggplot()+borders("world",colour="white",fill="grey")+
#  theme(panel.background=element_rect(fill = "gray93"))+
  #geom_map(data=subset(meta.country, Month==1),aes((mapid=Country),map=met
a.country$Country,colour=year.coef),size=3)+
  #geom_segment(data=subset(meta.country, Month==1),aes(xend = Longitude      +
delta_Longitude,    yend = Latitude + delta_Latitude,colour=year.coef))
#  scale_colour_gradient(low="yellow",high ="red")+
#  ggtitle("Average Annual Increase in Temperature - January")+
#  labs(colour='Average Annual \nTemperature Increase (°C)')+xlab("Longitude")+yl
ab("Latitude")


Jan<-ggplot()+borders("world",colour="white",fill="grey")+
  theme(panel.background=element_rect(fill = "gray93"))+
```

```
  geom_point(data=subset(meta.country, Month==1),aes(x=Longitude,y=Latitude,col
our=year.coef),size=8)+
  scale_colour_gradient(low="yellow",high ="red")+
  ggtitle("Average Annual Increase in Temperature - January")+theme(legend.positio
n = "bottom")+
  labs(colour='Average Annual \nTemperature Increase (°C)')+xlab("Longitude")+yla
b("Latitude")
ggsave(plot = Jan, "C:/CSC465/Project/images/JanTempIncrease.png", h = 4, w = 6)
Feb<-ggplot()+borders("world",colour="white",fill="grey")+
  theme(panel.background=element_rect(fill = "gray93"))+
  geom_point(data=subset(meta.country, Month==2),aes(x=Longitude,y=Latitude,col
our=year.coef),size=8)+
  scale_colour_gradient(low="yellow",high ="red")+
  ggtitle("Average Annual Increase in Temperature - February")+theme(legend.positi
on = "bottom")+
  labs(colour='Average Annual \nTemperature Increase (°C)')+xlab("Longitude")+yla
b("Latitude")
ggsave(plot = Feb, "C:/CSC465/Project/images/FebTempIncrease.png", h = 4, w = 7)

Mar<-ggplot()+borders("world",colour="white",fill="grey")+
  theme(panel.background=element_rect(fill = "gray93"))+
  geom_point(data=subset(meta.country, Month==3),aes(x=Longitude,y=Latitude,col
our=year.coef),size=8)+
  scale_colour_gradient(low="yellow",high ="red")+theme(legend.position = "bottom
")+
  ggtitle("Average Annual Increase in Temperature - March")+
  labs(colour='Average Annual \nTemperature Increase (°C)')+xlab("Longitude")+yla
b("Latitude")
ggsave(plot = Mar, "C:/CSC465/Project/images/MarTempIncrease.png", h = 4, w = 7
)

Apr<-ggplot()+borders("world",colour="white",fill="grey")+
  theme(panel.background=element_rect(fill = "gray93"))+
  geom_point(data=subset(meta.country, Month==4),aes(x=Longitude,y=Latitude,col
our=year.coef),size=8)+
  scale_colour_gradient(low="yellow",high ="red")+
  ggtitle("Average Annual Increase in Temperature - April")+theme(legend.position =
"right")+
  labs(colour='Average Annual \nTemperature Increase (°C)')+xlab("Longitude")+yla
b("Latitude")
ggsave(plot = Apr, "C:/CSC465/Project/images/AprTempIncrease.png", h = 4, w = 7)

May<-ggplot()+borders("world",colour="white",fill="grey")+
  theme(panel.background=element_rect(fill = "gray93"))+
  geom_point(data=subset(meta.country, Month==5),aes(x=Longitude,y=Latitude,col
our=year.coef),size=8)+
```

```
  scale_colour_gradient(low="yellow",high ="red")+
  ggtitle("Average Annual Increase in Temperature - May")+ theme(legend.position =
"right")+
  labs(colour='Average Annual \nTemperature Increase (°C)')+xlab("Longitude")+yla
b("Latitude")
ggsave(plot = May, "C:/CSC465/Project/images/MayTempIncrease.png", h = 4, w =
7)

Jun<-ggplot()+borders("world",colour="white",fill="grey")+
  theme(panel.background=element_rect(fill = "gray93"))+
  geom_point(data=subset(meta.country, Month==6),aes(x=Longitude,y=Latitude,col
our=year.coef),size=8)+
  scale_colour_gradient(low="yellow",high ="red")+theme(legend.position = "right")
+
  ggtitle("Average Annual Increase in Temperature - June")+
  labs(colour='Average Annual \nTemperature Increase (°C)')+xlab("Longitude")+yla
b("Latitude")
ggsave(plot = Jun, "C:/CSC465/Project/images/JunTempIncrease.png", h = 4, w = 7)

Jul<-ggplot()+borders("world",colour="white",fill="grey")+
  theme(panel.background=element_rect(fill = "gray93"))+
  geom_point(data=subset(meta.country, Month==7),aes(x=Longitude,y=Latitude,col
our=year.coef),size=8)+
  scale_colour_gradient(low="yellow",high ="red")+
  ggtitle("Average Annual Increase in Temperature - July")+theme(legend.position =
"right")+
  labs(colour='Average Annual \nTemperature Increase (°C)')+xlab("Longitude")+yla
b("Latitude")
ggsave(plot = Jul, "C:/CSC465/Project/images/JulTempIncrease.png", h = 4, w = 7)

Aug<-ggplot()+borders("world",colour="white",fill="grey")+
  theme(panel.background=element_rect(fill = "gray93"))+
  geom_point(data=subset(meta.country, Month==8),aes(x=Longitude,y=Latitude,col
our=year.coef),size=8)+
  scale_colour_gradient(low="yellow",high ="red")+
  ggtitle("Average Annual Increase in Temperature - August")+theme(legend.position
= "right")+
  labs(colour='Average Annual \nTemperature Increase (°C)')+xlab("Longitude")+yla
b("Latitude")
ggsave(plot = Aug, "C:/CSC465/Project/images/AugTempIncrease.png", h = 4, w = 7
)

Sep<-ggplot()+borders("world",colour="white",fill="grey")+
  theme(panel.background=element_rect(fill = "gray93"))+
  geom_point(data=subset(meta.country, Month==9),aes(x=Longitude,y=Latitude,col
our=year.coef),size=8)+
```

```
  scale_colour_gradient(low="yellow",high ="red")+
  ggtitle("Average Annual Increase in Temperature - September")+theme(legend.posit
ion = "bottom")+
  labs(colour='Average Annual \nTemperature Increase (°C)')+xlab("Longitude")+yla
b("Latitude")
ggsave(plot = Sep, "C:/CSC465/Project/images/SepTempIncrease.png", h = 4, w = 7)

Oct<-ggplot()+borders("world",colour="white",fill="grey")+
  theme(panel.background=element_rect(fill = "gray93"))+
  geom_point(data=subset(meta.country, Month==10),aes(x=Longitude,y=Latitude,co
lour=year.coef),size=8)+
  scale_colour_gradient(low="yellow",high ="red")+
  ggtitle("Average Annual Increase in Temperature - October")+
  labs(colour='Average Annual \nTemperature Increase (°C)')+xlab("Longitude")+yla
b("Latitude")
ggsave(plot = Oct, "C:/CSC465/Project/images/OctTempIncrease.png", h = 4, w =7)

Nov<-ggplot()+borders("world",colour="white",fill="grey")+
  theme(panel.background=element_rect(fill = "gray93"))+
  geom_point(data=subset(meta.country, Month==11),aes(x=Longitude,y=Latitude,co
lour=year.coef),size=8)+
  scale_colour_gradient(low="yellow",high ="red")+
  ggtitle("Average Annual Increase in Temperature - November")+ theme(legend.posi
tion = "bottom")+
  labs(colour='Average Annual \nTemperature Increase (°C)')+xlab("Longitude")+yla
b("Latitude")
ggsave(plot = Nov, "C:/CSC465/Project/images/NovTempIncrease.png", h = 4, w = 7
)

Dec<-ggplot()+borders("world",colour="white",fill="grey")+
  theme(panel.background=element_rect(fill = "gray93"))+
  geom_point(data=subset(meta.country, Month==12),aes(x=Longitude,y=Latitude,co
lour=year.coef),size=8)+
  scale_colour_gradient(low="yellow",high ="red")+
  ggtitle("Average Annual Increase in Temperature - December")+theme(legend.posit
ion = "bottom")+
  labs(colour='Average Annual \nTemperature Increase (°C)')+xlab("Longitude")+yla
b("Latitude")
ggsave(plot = Dec, "C:/CSC465/Project/images/DecTempIncrease.png", h = 4, w = 7)
```

3) Joo’s

```
library(ggplot2)
library(dplyr)
library(tidyr)
```

```
library(lubridate)

movingAverage <- function(x, n=1, centered=FALSE) {

  if (centered) {
    before <- floor  ((n-1)/2)
    after  <- ceiling((n-1)/2)
  } else {
    before <- n-1
    after  <- 0
  }

  # Track the sum and count of number of non-NA items
  s     <- rep(0, length(x))
  count <- rep(0, length(x))

  # Add the centered data
  new <- x
  # Add to count list wherever there isn't a
  count <- count + !is.na(new)
  # Now replace NA_s with 0_s and add to total
  new[is.na(new)] <- 0
  s <- s + new

  # Add the data from before
  i <- 1
  while (i <= before) {
    # This is the vector with offset values to add
    new   <- c(rep(NA, i), x[1:(length(x)-i)])

    count <- count + !is.na(new)
    new[is.na(new)] <- 0
    s <- s + new

    i <- i+1
  }

  # Add the data from after
  i <- 1
  while (i <= after) {
    # This is the vector with offset values to add
    new   <- c(x[(i+1):length(x)], rep(NA, i))

    count <- count + !is.na(new)
    new[is.na(new)] <- 0
    s <- s + new
```

```
  i <- i+1
 }

 # return sum divided by count
 s/count
}

temperature = read.csv("C:/Users/Admin/Documents/data visualization/project/GlobalTemperatures.csv")
temperature$dt = as.Date(temperature$dt)


temperature %>%
 #filter(Country=="United States")  %>%
 separate(col = dt, into = c("Year", "Month", "Day"), convert = TRUE) ->temperature
temperature<-na.omit(temperature)

#cData %>%
# filter(State!="Hawaii" & State!="Alaska") -> cData1
# Remove na's
#cData1 = na.omit(cData1)

temperature %>%
 #filter(Year>1850) %>%
 group_by(Year) %>%
 summarise(Temp = mean(LandAverageTemperature), TempUncertainity = mean(LandAverageTemperatureUncertainty),
           TempUncertainityUpper = mean(LandAverageTemperature) +  mean(LandAverageTemperatureUncertainty),
           TempUncertainityLower = mean(LandAverageTemperature) -  mean(LandAverageTemperatureUncertainty)
           ) ->cData2


cData2$movingAverage = temperature$Temp
cData2$movingAverage = movingAverage(cData2$Temp, 20)

cData2$movingAverageUpper = temperature$Temp
cData2$movingAverageUpper = movingAverage(cData2$TempUncertainityUpper, 20)

cData2$movingAverageLower = temperature$Temp
cData2$movingAverageLower = movingAverage(cData2$TempUncertainityLower, 20)
```

```
ggplot(cData2,
       aes(Year, movingAverage)) + ggtitle("Upper and Lower of Temperature Measure
ments") +
  geom_point(alpha=0.5) +
  geom_line(color="Green",size=2) +
  geom_line(data=cData2,aes(Year,movingAverageUpper),color="Blue",size=2) +
  geom_line(data=cData2,aes(Year,movingAverageLower),color="Red",size=2) +
  labs(y = "Moving Average of Temperature (20 year avg) Celsius") +
  theme(axis.text.x = element_text(colour = 'black', angle = 0, size = 20, hjust = 0.5, vj
ust = 0.5),axis.title.x=element_blank()) +
  theme(axis.text.y = element_text(colour = 'black', size = 20), axis.title.y = element_t
ext(size = 20, hjust = 0.5, vjust = 0.2)) +
    theme(title = element_text(size=24)) +
  legend(2000,9.5, lty=c(1,1), lwd=c(2.5,2.5),col=c("blue","red"))
```

4)

```
library(dplyr)
library(tidyr)
library(ggplot2)
library(lattice)

cData = read.csv("C:/Users/Admin/Documents/data visualization/project/GlobalLand
TemperaturesByState.csv")

cData %>%
  #filter(Country=="United States")  %>%
  separate(col = dt, into = c("Year", "Month", "Day"), convert = TRUE) ->cData
cData<-na.omit(cData)

#cData %>%
# filter(State!="Hawaii" & State!="Alaska") -> cData1
# Remove na's
#cData1 = na.omit(cData1)

cData %>%
  #filter(Year>1850) %>%
  group_by(Year,Country) %>%
  summarise(Temp = mean(AverageTemperatureUncertainty)) ->cData2

ggplot(data=cData2,aes(Year,Temp)) + geom_point() + facet_grid(Country ~ .) +
  aes(colour = Temp) + scale_color_gradient(low="blue", high="red")
```

5)

```
library(dplyr)
library(tidyr)
library(ggplot2)

movingAverage <- function(x, n=1, centered=FALSE) {

  if (centered) {
    before <- floor  ((n-1)/2)
    after  <- ceiling((n-1)/2)
  } else {
    before <- n-1
    after  <- 0
  }

  # Track the sum and count of number of non-NA items
  s     <- rep(0, length(x))
  count <- rep(0, length(x))

  # Add the centered data
  new <- x
  # Add to count list wherever there isn't a
  count <- count + !is.na(new)
  # Now replace NA_s with 0_s and add to total
  new[is.na(new)] <- 0
  s <- s + new

  # Add the data from before
  i <- 1
  while (i <= before) {
    # This is the vector with offset values to add
    new   <- c(rep(NA, i), x[1:(length(x)-i)])

    count <- count + !is.na(new)
    new[is.na(new)] <- 0
    s <- s + new

    i <- i+1
  }

  # Add the data from after
  i <- 1
  while (i <= after) {
    # This is the vector with offset values to add
    new   <- c(x[(i+1):length(x)], rep(NA, i))
```

```
    count <- count + !is.na(new)
    new[is.na(new)] <- 0
    s <- s + new

    i <- i+1
   }

  # return sum divided by count
  s/count
}
cData = read.csv("C:/Users/Admin/Documents/data visualization/project/GlobalTem
peratures.csv")

cData %>%
  #filter(Country=="United States")  %>%
  separate(col = dt, into = c("Year", "Month", "Day"), convert = TRUE) ->cData
cData<-na.omit(cData)

#cData %>%
 # filter(State!="Hawaii" & State!="Alaska") -> cData1
# Remove na's
#cData1 = na.omit(cData1)

cData %>%
  #filter(Year>1850) %>%
  group_by(Year) %>%
  summarise(Temp = mean(LandAverageTemperature), TempUncertainity = mean(La
ndAverageTemperature)) ->cData2

cData2$movingAverage = cData2$Temp
cData2$movingAverage = movingAverage(cData2$Temp, 20)

qplot(Year, movingAverage, data=cData2, main="Average Temperature World-Wide
1850-2013") +  geom_line(size=2) +
  aes(colour = Temp) + scale_color_gradient(low="blue", high="red") +
  theme(axis.text.x = element_text(colour = 'black', angle = 0, size = 20, hjust = 0.5, vj
ust = 0.5),axis.title.x=element_blank()) +
  theme(axis.text.y = element_text(colour = 'black', size = 20), axis.title.y = element_t
ext(size = 20, hjust = 0.5, vjust = 0.2)) +
  labs(y = "Moving Average of Temperature (20 year avg) Celsius") +
  theme(title = element_text(size=24))
```

Tableau code would be attached in the drop box.