Exploring Weather Trends in Wellington vs. World

Phil Nelson | April 2018 NZST

Summary outline of approach

1. Two SQL queries: One - "results_global_temp.csv" and Two - "results_wellington_nz.csv" download
2. Load packages into R
3. Import csv file data using R
4. Perform an inner_join using R for the two files so time periods match (161 Wgtn v. 266 years Global)
5. Create rolling means for 10,20,30 year periods using "zoo" package using "rollmeans"" in R
6. Plot & consider 10 and 30 year mean line graphs over time for Wellington City and Global in R
7. Calculate summary stats, graph boxplot and histograms to compare 10 & 30 year city and global in R
8. Create geom_smooth regression trendline and contrast with actual annual temperature averages in R
9. Examine correlation between Wellington City and Global temperatures using R and calculate
10. Publish using knitr package and Rmarkdown in R-studio, use latex to create pdf in R

1. SQL queries

Query 1 - "results_wellington_nz.csv" SELECT year, avg_temp FROM city_data WHERE city = 'Wellington' AND country = 'New Zealand';

Query 2 - "results_global_temp.csv" SELECT * FROM global_data;

2. Load Packages into R for the Weather Study

```
#load packages
library(dplyr)
library(ggplot2)
library(zoo)
#install the zoo package to use the rollmean function below
#(n.b.mean from center of period)
```

3. Import CSV files generated from SQL queries using R script

```
#rename imported csv files from Udacity.
wellington_yr_temp <- read.csv('results_wellington_nz.csv')
global_yr_temp <- read.csv('results_global_temp.csv')
```

4. Prepare the data and use "inner_join"" to merge City and Global data

This section of code uses an "inner join" because the Wellington City covers a 161 year period which is smaller than the Global annual temperature average records of a period of 266 years. *New Zealand is one of the "youngest" of the OECD countries so this shorter period of temperature data makes sense - starting to record data 13 years after the signing of the 1840 Treaty of Waitangi https://en.wikipedia.org/wiki/Treaty_of_Waitangi.*

```r
#merge files based on joining the common years that exist in both data sets
# (Wellington data is only for 161 years (young country!!!), global 266 years data)
well_v_global <- inner_join(wellington_yr_temp,global_yr_temp,by="year")
#I manually check some of the data points from global are mapped correctly - YES

#change the Column Name av_temp.x to av_temp.wgtn and av_temp.y to av_temp.global
colnames(well_v_global) <- c('year','av_temp.wgtn','av_temp.global')
```

5. Create rolling mean periods for 10,20 and 30 years using rollmeans from "zoo" package

```r
#create moving average variable columns based on 10 year and 20 year averages
#rollmean example at http://uc-r.github.io/ts_moving_averages#centered-moving-averages

averages <- well_v_global %>%
  mutate(av_well_10yr = rollmean(av_temp.wgtn, k = 10, fill = NA),
         av_well_20yr = rollmean(av_temp.wgtn, k = 20, fill = NA),
         av_well_30yr = rollmean(av_temp.wgtn, k = 30, fill = NA),
         av_glob_10yr = rollmean(av_temp.global, k = 10, fill = NA),
         av_glob_20yr = rollmean(av_temp.global, k = 20, fill = NA),
         av_glob_30yr = rollmean(av_temp.global, k = 30, fill = NA))
```

During some data exploration and visuals I found there was not quite enough variation between 10 year averages and 20 year averages. So went with a 30 year period to compare 10 year temperature rolling averages against 30 years rolling averages.

To understand the periodicity examine the head of the averages dataset

```r
head(averages,20)
```

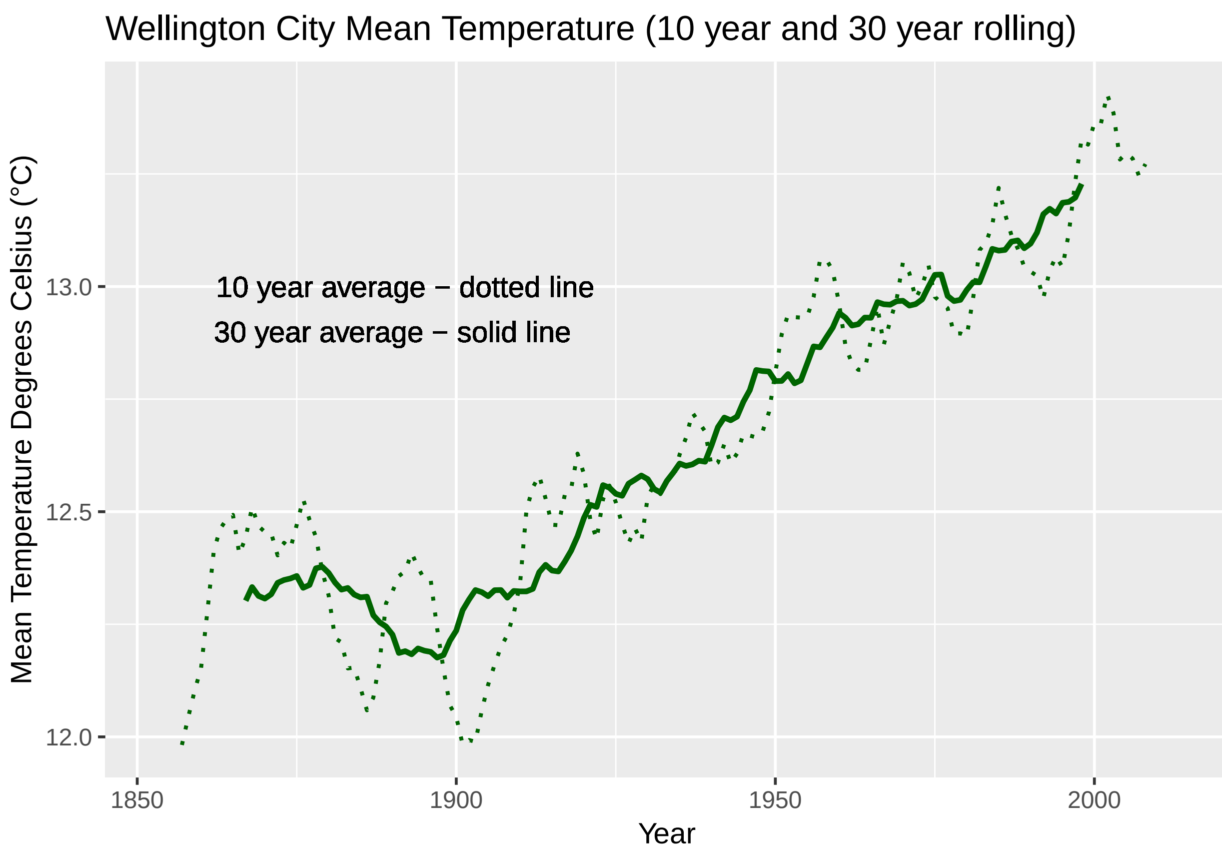
```
##    year av_temp.wgtn av_temp.global av_well_10yr av_well_20yr av_well_30yr
## 1  1853        11.21           8.04           NA           NA           NA
## 2  1854        11.99           8.21           NA           NA           NA
## 3  1855        12.09           8.11           NA           NA           NA
## 4  1856        11.64           8.00           NA           NA           NA
## 5  1857        11.49           7.76       11.982           NA           NA
## 6  1858        11.51           8.10       12.044           NA           NA
## 7  1859        12.21           8.25       12.103           NA           NA
## 8  1860        12.32           7.96       12.152           NA           NA
## 9  1861        12.91           7.85       12.279           NA           NA
## 10 1862        12.45           7.56       12.413      12.2125           NA
## 11 1863        11.83           8.11       12.462      12.2770           NA
## 12 1864        12.58           7.98       12.483      12.2855           NA
## 13 1865        12.58           8.18       12.493      12.3035           NA
## 14 1866        12.91           8.29       12.406      12.3655           NA
## 15 1867        12.83           8.44       12.443      12.4080     12.30233
## 16 1868        12.00           8.25       12.510      12.4465     12.33267
## 17 1869        12.42           8.43       12.468      12.4505     12.31300
## 18 1870        12.42           8.20       12.455      12.4820     12.30700
## 19 1871        12.04           8.12       12.452      12.4670     12.31667
## 20 1872        12.82           8.19       12.403      12.4625     12.34200
##    av_glob_10yr av_glob_20yr av_glob_30yr
## 1            NA           NA           NA
## 2            NA           NA           NA
## 3            NA           NA           NA
```

```
## 4            NA            NA            NA
## 5         7.984            NA            NA
## 6         7.991            NA            NA
## 7         7.968            NA            NA
## 8         7.975            NA            NA
## 9         8.004            NA            NA
## 10        8.072        8.1015            NA
## 11        8.087        8.1170            NA
## 12        8.105        8.1280            NA
## 13        8.129        8.1155            NA
## 14        8.156        8.1195            NA
## 15        8.219        8.1585      8.160333
## 16        8.243        8.1950      8.158333
## 17        8.288        8.1910      8.143667
## 18        8.256        8.1990      8.137333
## 19        8.235        8.2200      8.135667
## 20        8.245        8.2485      8.140667
```

6. Plot Wellington and Global 10 year and 30 year averages and compare

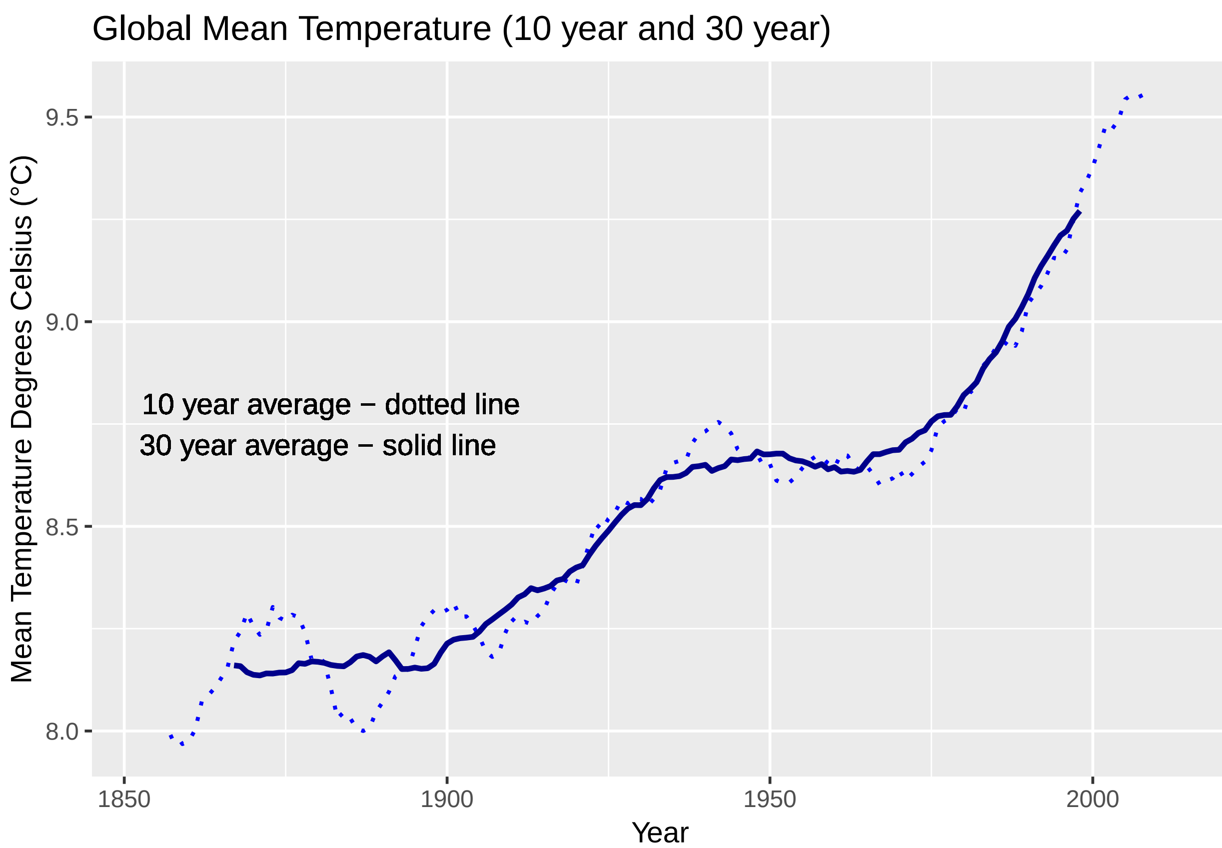
The code and plot below shows Wellington City 10 year and 30 year temperature averages.

```r
#Line plot of Average Temperatures in Wellington (10 and 30 year averages)
ggplot(averages,aes(x=year,y=av_temp.wgtn)) +
  geom_line(aes(x=year,y=av_well_10yr),color = "dark green",size=0.7,linetype="dotted") +
  geom_line(aes(x=year,y=av_well_30yr),color ="dark green",size=1) +
  #geom_point(alpha = 0.4, color = "dark green") +
  labs(title="Wellington City Mean Temperature (10 year and 30 year rolling)",
       x = "Year", y = "Mean Temperature Degrees Celsius (°C)") +
  geom_text(aes(x = 1892, y = 13, label = "10 year average - dotted line"))+
  geom_text(aes(x = 1890, y = 12.9, label = "30 year average - solid line"))
```



The code and plot below shows the Global 10 year and 30 year temperature averages.

```
#Line plot of Average Temperatures Global (10 and 30 year averages)
ggplot(averages,aes(x=year,y=av_temp.global)) +
  geom_line(aes(x=year,y=av_glob_10yr),color="blue",linetype="dotted",size=0.7) +
  geom_line(aes(x=year,y=av_glob_30yr),color="dark blue",size=1) +
  labs(title="Global Mean Temperature (10 year and 30 year)",
       x = "Year", y = "Mean Temperature Degrees Celsius (°C)") +
  geom_text(aes(x = 1882, y = 8.8, label = "10 year average - dotted line"))+
  geom_text(aes(x = 1880, y = 8.7, label = "30 year average - solid line"))
```

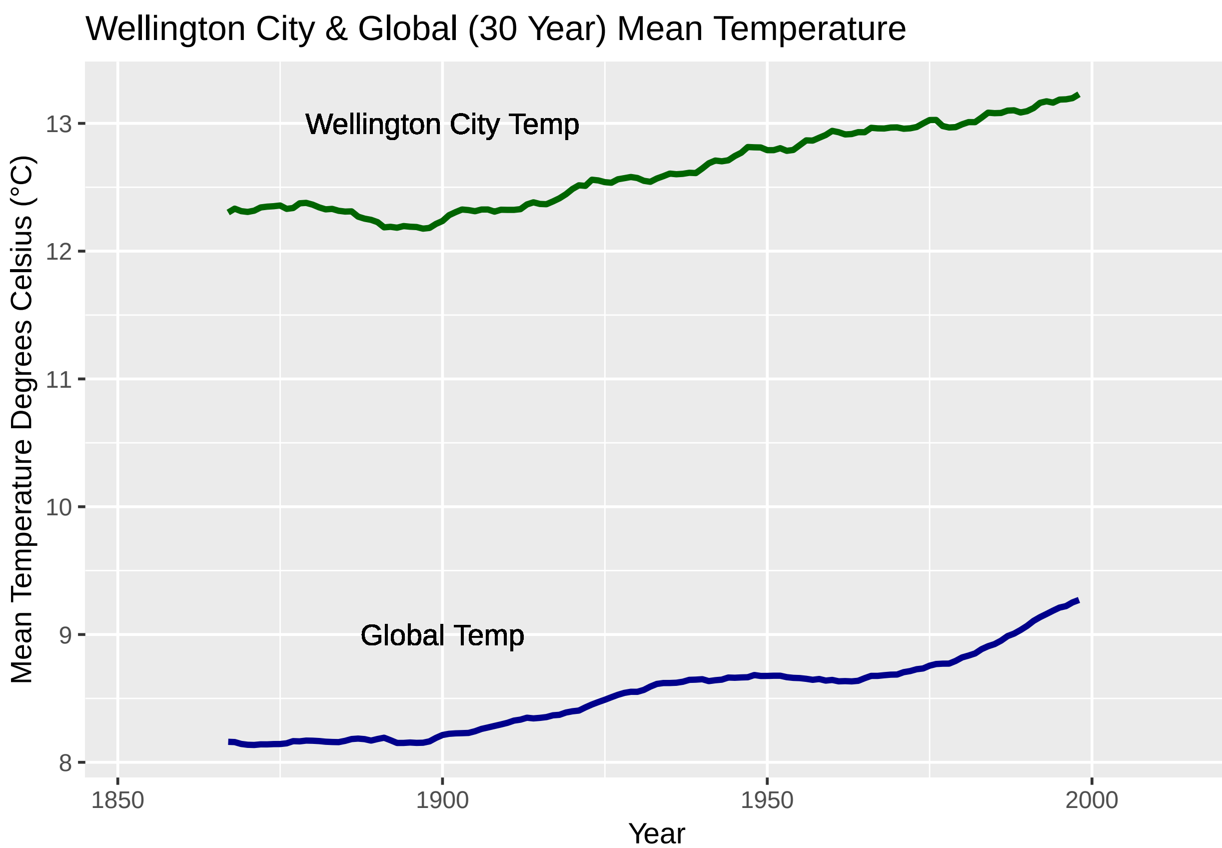


Some considerations regarding variability over different time periods and what it means

In Chapter 12 "A Climate of Healthy Skepticism" of Nate Silver's book "The Signal and the Noise" https://www.amazon.com/Signal-Noise-Many-Predictions-Fail-but/dp/0143125087 which I have read ;-) makes some interesting observations including bayesian probability estimates that global warming is occurring even if we did observe no temperature increase over a 10 year period is still reaonably high for an initial 95% prior probability. The other useful insight is that over shorter periods average temperature is expected to more variation than over longer time period averages.

There is high temperature variability across most time domains and the trend line signal only emerges over from the noise over long periods of time (several decades). So I decided to use a 30 year average in the comparison graphs below.

```
#Show actual temperatures versus 10 year and 30 year rolling averages
ggplot(averages,aes(x=year,y=av_temp.wgtn)) +
  #geom_point(alpha = 0.4, color = "dark green") +
  #geom_line(aes(x=year,y=av_well_10yr),color="dark green",linetype="dotted",size=0.75) +
  geom_line(aes(x=year,y=av_well_30yr),color="dark green",size=1.1) +
  #geom_point(aes(x=year,y=av_temp.global),alpha = 0.4,color ="blue") +
  #geom_line(aes(x=year,y=av_glob_10yr),color="blue",linetype="dotted",size=0.75) +
  geom_line(aes(x=year,y=av_glob_30yr),color="dark blue",size=1.1) +
  labs(title="Wellington City & Global (30 Year) Mean Temperature", x = "Year",
       y = "Mean Temperature Degrees Celsius (°C)") +
  geom_text(aes(x = 1900, y = 13, label = "Wellington City Temp")) +
  geom_text(aes(x = 1900, y = 9, label = "Global Temp"))
```



Some observations from the plots above:

1. In general temperature appears to be increasing for both Wellington City and Global data
2. Wellington is approximately 4 degrees hotter than the Global average temperature
3. There is more average temperature fluctuation in Wellington City than for Global data under all time periods
4. There is a greater overall increase in Global temperature than in Wellington City (cause unknown to author area for more research)
5. Even with a 30 year rolling average period there are still some brief periods where there is a reduction in temperature for individual cities this is less likely for the global data

7. Calculate summary stats, graph boxplot summary stats

For points 3 and 4 see summary statistics for Wellington City:

```
summary(averages$av_well_30yr)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##   12.18   12.33   12.58   12.63   12.95   13.23      29
```

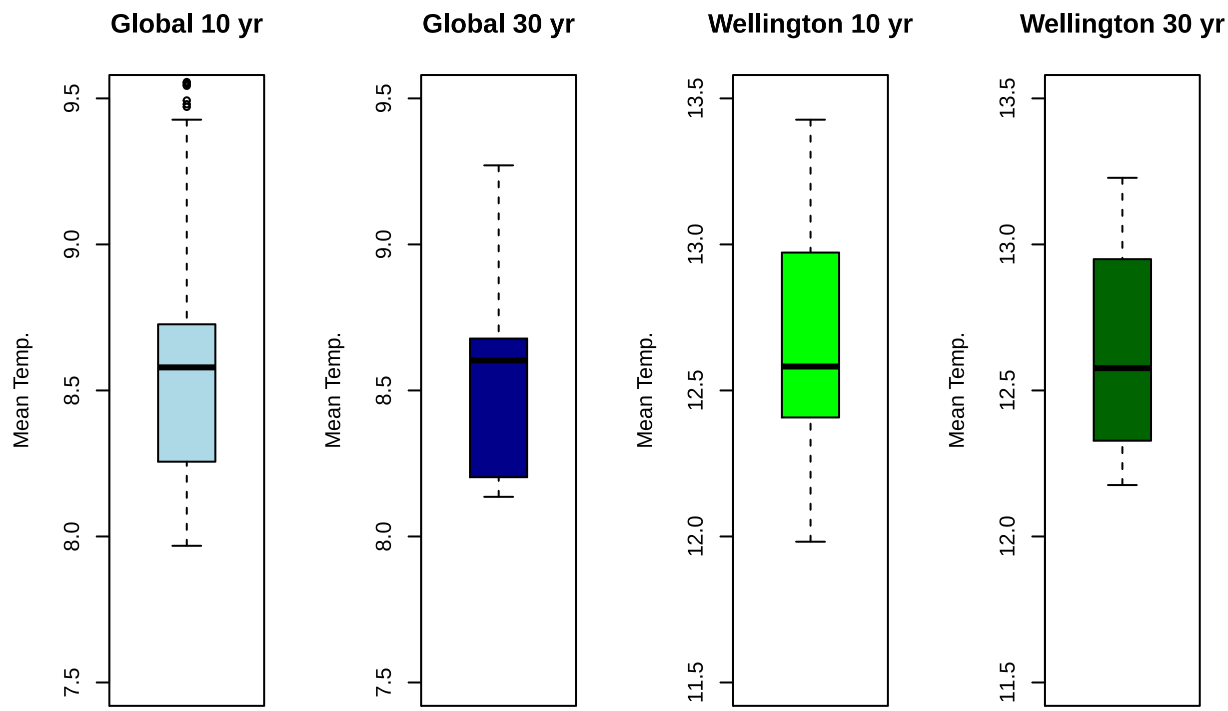
And the summary statistics for Global Data:

```
summary(averages$av_glob_30yr)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##   8.136   8.208   8.603   8.524   8.678   9.271      29
```

Generate boxplots for Wellington and Global 10 year and 30 years:

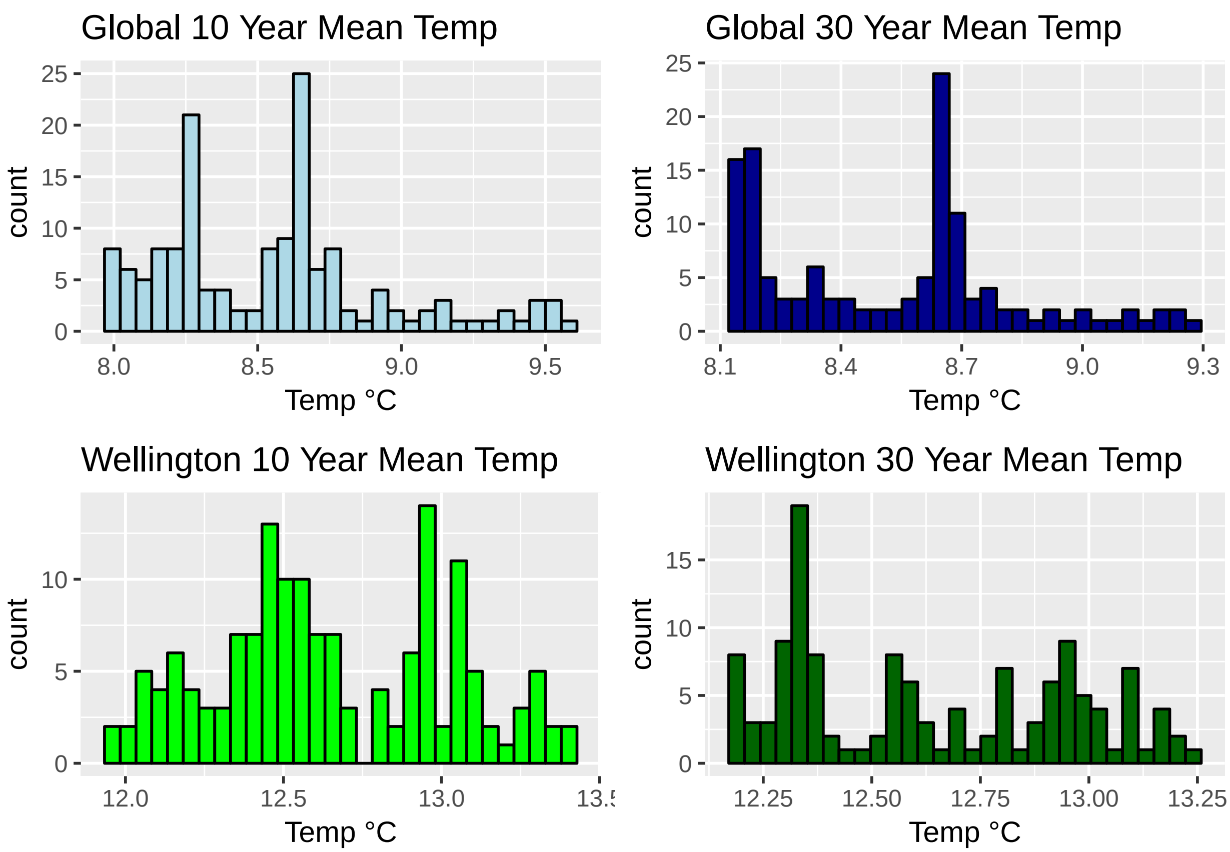
```r
par(mfrow=c(1,4))
boxplot(averages$av_glob_10yr,main="Global 10 yr",ylab = "Mean Temp.",
        ylim=c(7.5,9.5),col="light blue")
boxplot(averages$av_glob_30yr,main="Global 30 yr",ylab = "Mean Temp.",
        ylim=c(7.5,9.5),col="dark blue")
boxplot(averages$av_well_10yr,main="Wellington 10 yr",ylab = "Mean Temp.",
        ylim=c(11.5,13.5),col="green")
boxplot(averages$av_well_30yr,main="Wellington 30 yr",ylab = "Mean Temp.",
        ylim=c(11.5,13.5),col="dark green")
```



As shown in the time series the 10 year average range is greater than the 30 year range. Interesting to note that the boxplots might indicate some skew but it is unclear what the nature of the distributions may be. Some histograms with time bands may provide more insight.

```r
#Install libraries to create a grid for ggplot2
library(grid)
library(gridExtra)

#allocate each plot to a variable
g30 <- ggplot(averages,aes(x=av_glob_30yr)) +
  geom_histogram(bins = 30,fill="dark blue",color = "black")+
  labs(title="Global 30 Year Mean Temp",x="Temp °C")

g10 <- ggplot(averages,aes(x=av_glob_10yr)) +
  geom_histogram(bins = 30,fill="light blue",color = "black")+
  labs(title="Global 10 Year Mean Temp",x="Temp °C")
```

```
w30 <- ggplot(averages,aes(x=av_well_30yr)) +
  geom_histogram(bins = 30,fill="dark green",color = "black")+
  labs(title="Wellington 30 Year Mean Temp",x="Temp °C")

w10 <- ggplot(averages,aes(x=av_well_10yr)) +
  geom_histogram(bins = 30,fill="green",color = "black")+
  labs(title="Wellington 10 Year Mean Temp",x="Temp °C")

grid.arrange(g10, g30, w10, w30, ncol = 2)
```



The bimodal nature of 30 year global distribution suggests the possibility of two different modes or states these probably correspond with two plateaus (i.e. ~1870-1900 and ~1940-1960) in the 30 year average data.

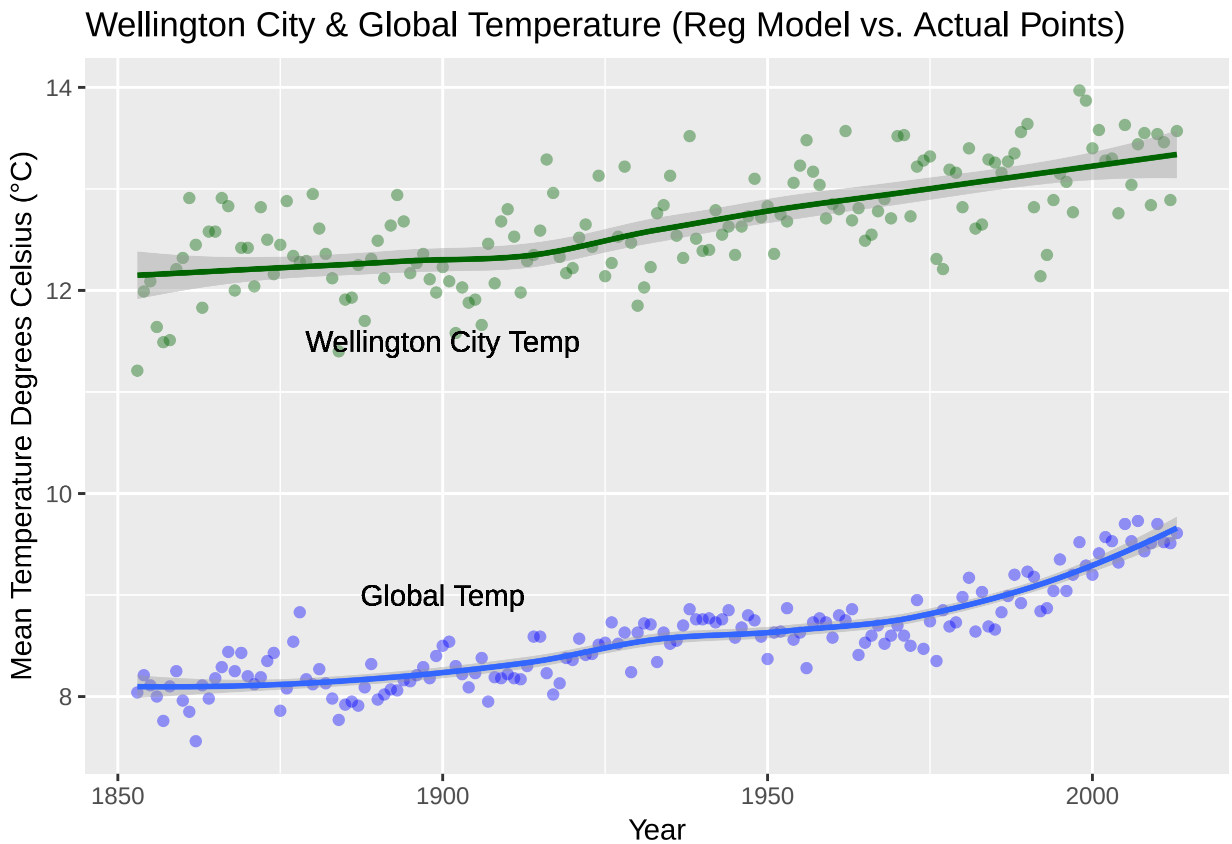
The reason for these modes or plateaus is unknown to me. It is unclear if this is long cycle periodic or random distribution or some other effect. I just googled this: https://en.wikipedia.org/wiki/Global_warming_hiatus. It is I think coincidence (or maybe method in my madness?) that I picked a 30 year average period for my analysis which was determined iteratively using "rollmeans" as shown in section 5 above !!!

It would take more study to determine the reasons for the local variation of Wellington from the global data (perhaps Island or sea/coastal moderating climate or regional effect?). Wellington is only one of many city datapoints.

8. Create geom_smooth regression trendline and compare with actual annual temperature means

The plot below shows the measured Wellington vs. Global Temp annual averages in a scatter plot. The smooth line generated with using geom_smooth model from ggplot2 package in R.

```
ggplot(well_v_global,aes(x=year,y=av_temp.wgtn)) +
  geom_point(alpha = 0.4, color = "dark green") + geom_smooth(color="dark green") +
  geom_point(aes(x=year,y=av_temp.global),alpha = 0.4,color ="blue") +
  geom_smooth(aes(x=year,y=av_temp.global)) +
  labs(title="Wellington City & Global Temperature (Reg Model vs. Actual Points)",
       x = "Year", y = "Mean Temperature Degrees Celsius (°C)") +
  geom_text(aes(x = 1900, y = 11.5, label = "Wellington City Temp")) +
  geom_text(aes(x = 1900, y = 9, label = "Global Temp"))
```

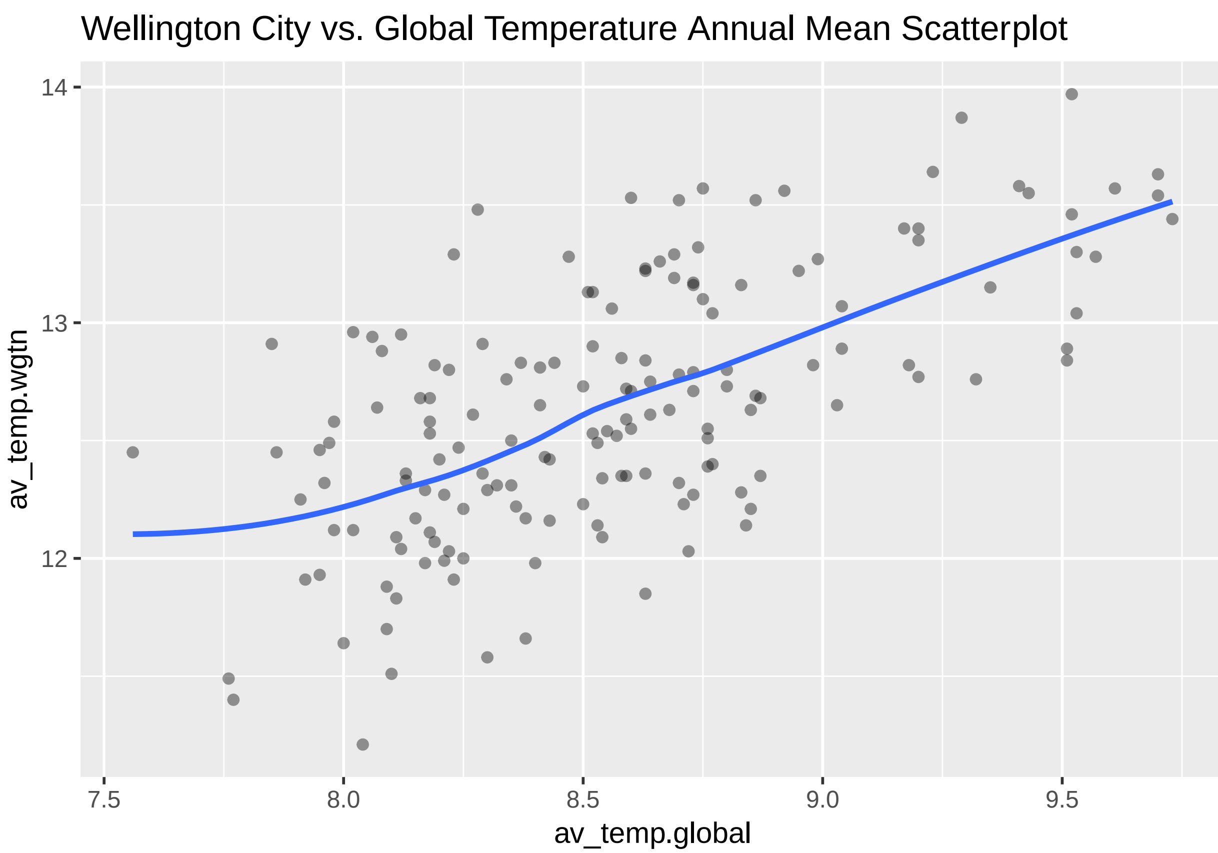


This shows a smooth general trendline with the actual data in general showing the increase in temperature in the model. LOESS smoothing is a non-parametric form of regression that uses a weighted, sliding-window, average to calculate a line of best fit. Within each "window", a weighted average is calculated, and the sliding window passes along the x-axis.

9. Visually examine correlation between Wellington City and Global temperatures and calculate

The geom_smooth regression line shows a line of best fit and the positive slope relationship between global and Wellington City annual avarage temperature over a 161 year period.

```
#have an initial look at the year by year temperature averages scatter between wellington
#and global
ggplot(well_v_global,aes(x=av_temp.global,y=av_temp.wgtn)) +
  geom_point(alpha = 0.4) +
  geom_smooth(se=FALSE) +
  labs(title="Wellington City vs. Global Temperature Annual Mean Scatterplot")
```



Correlation coefficient calculation

```
#calculate correlation between Wellington City and Global temperature averages
cor(well_v_global$av_temp.global,well_v_global$av_temp.wgtn)
```

```
## [1] 0.6395115
```

The correlation of Wellington to Global annual average temperatures is moderately strong.