Temperature Annual Means 1750 - 2015: Moving Averages for Global and Local (Los Angeles) Temperature Data

Extract Data

Write a SQL query to extract the city level data. Export to CSV.

```
    SELECT year, avg_temp
    FROM city_data
    WHERE city = 'Los Angeles' AND country = 'United States'
    ORDER BY year
```

Write a SQL query to extract the global data. Export to CSV.

```
    SELECT *
    FROM global_data
    ORDER BY year
```

Process Data

Load both .csv files in R Studio.

```
# clear working environment
rm(list=ls())

# for the rollmean function
library(zoo)

# for the plot
library(ggplot2)

# read in files extracted by SQL
global_data <- read.csv('global_data.csv')
city_data <- read.csv('city_data.csv')

# check data
head(global_data)
```

```
##   year avg_temp
## 1 1750     8.72
## 2 1751     7.98
## 3 1752     5.78
## 4 1753     8.39
## 5 1754     8.47
## 6 1755     8.36
```

```
head(city_data)
```

```
##   year avg_temp
## 1 1849    15.71
## 2 1850    15.28
## 3 1851    15.53
## 4 1852    15.61
## 5 1853    16.27
## 6 1854    15.74
```

Merge city data and global data into one data frame joining on ‘year’
Pad missing values with ‘NA’

```
# join the global and city data based on year
combined <- merge(x = global_data, y = city_data, by = "year", all = TRUE, suffixes = c(".global",".city"
))
```

Calculate moving averages in different intervals for both global and city temperature data using `rollmeanr` function

```
# compute moving averages of intervals (3, 5, 9, 15) for global temp data
combined$global_mavg03 <- rollmeanr(combined$avg_temp.global, 3, fill = NA)
combined$global_mavg05 <- rollmeanr(combined$avg_temp.global, 5, fill = NA)
combined$global_mavg09 <- rollmeanr(combined$avg_temp.global, 9, fill = NA)
combined$global_mavg15 <- rollmeanr(combined$avg_temp.global, 15, fill = NA)

# compute moving averages of intervals (3, 5, 9, 15) for city temp data
combined$city_mavg03 <- rollmeanr(combined$avg_temp.city, 3, fill = NA)
combined$city_mavg05 <- rollmeanr(combined$avg_temp.city, 5, fill = NA)
combined$city_mavg09 <- rollmeanr(combined$avg_temp.city, 9, fill = NA)
combined$city_mavg15 <- rollmeanr(combined$avg_temp.city, 15, fill = NA)

# attach the data frame
attach(combined)
```

Plot Data

Create a line chart that compares your city's temperatures with the global temperatures.
Check minima and maxima of all temperatures to determine limits for the Y axis
Use similar colour families for each moving average data set: blue for global data, orange for city data
Add title, axis labels, and legend for the plot

```
# check min for the plot axis limit
min(c(global_mavg03, global_mavg05, global_mavg09, global_mavg15), na.rm = TRUE)
```

```
## [1] 6.943333
```

```
min(c(city_mavg03, city_mavg05, city_mavg09, city_mavg15), na.rm = TRUE)
```

```
## [1] 14.95333
```

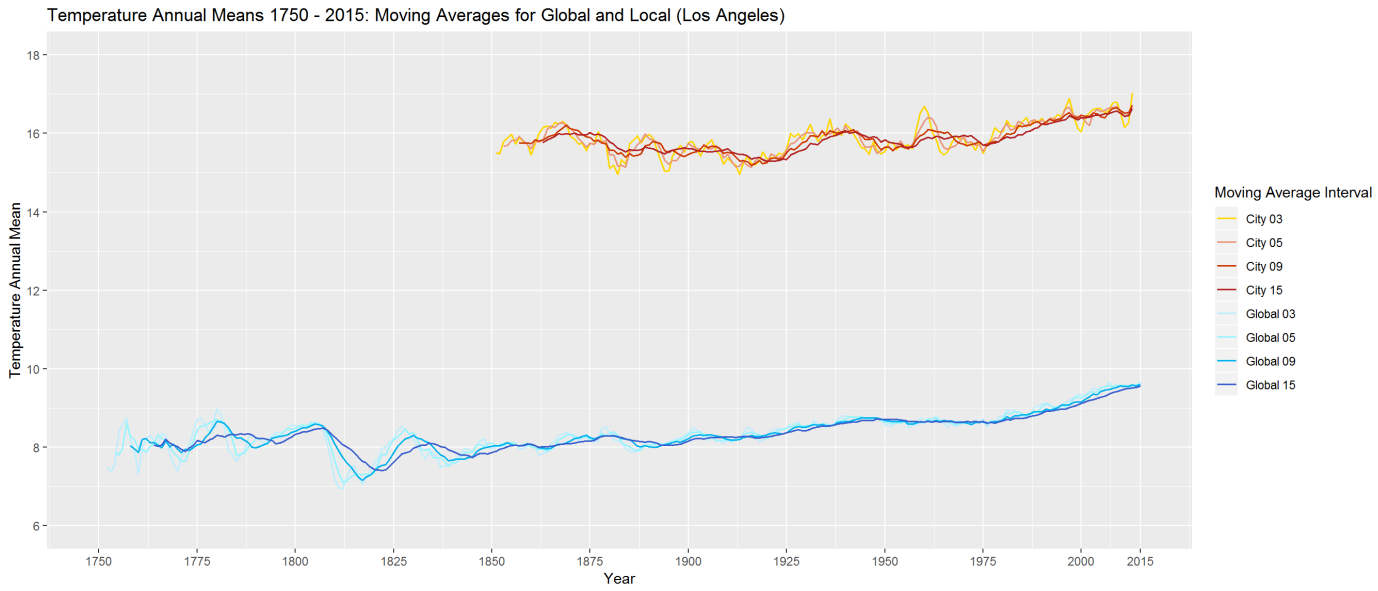
```
# check max for the plot axis limit
max(c(global_mavg03, global_mavg05, global_mavg09, global_mavg15), na.rm = TRUE)
```

```
## [1] 9.67
```

```
max(c(city_mavg03, city_mavg05, city_mavg09, city_mavg15), na.rm = TRUE)
```

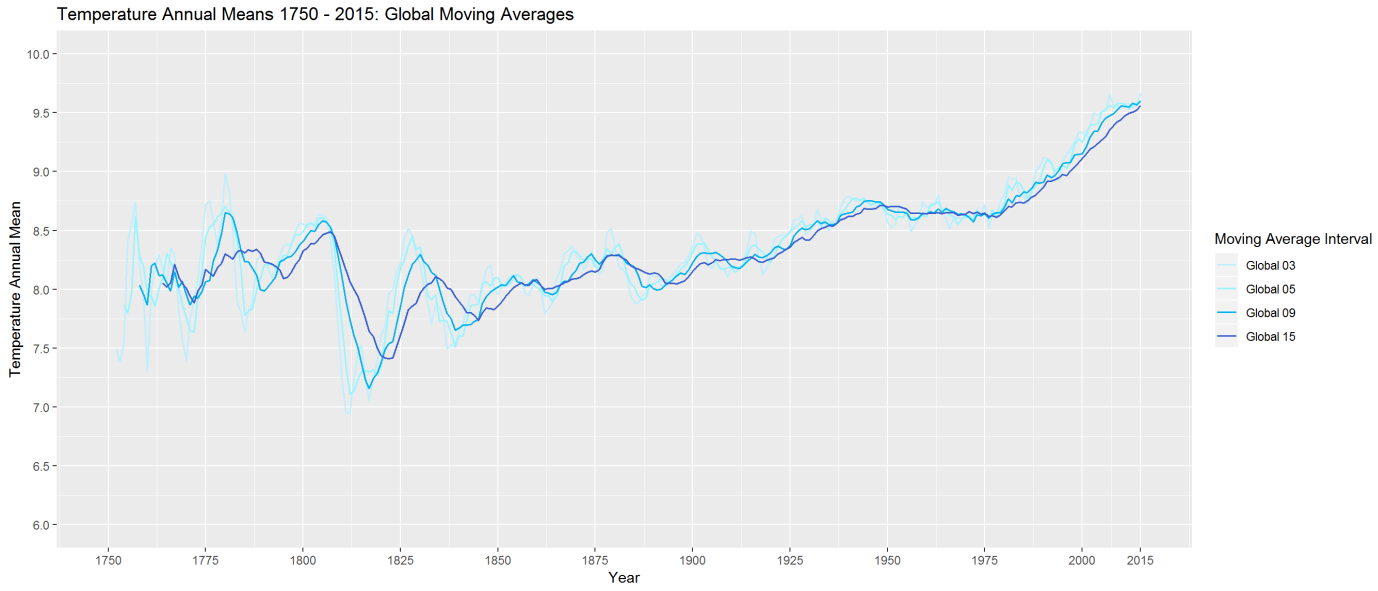
```
## [1] 17.02667
```

```
# create the plot using ggplot
ggplot(data=combined, aes(x=year)) +
  geom_line(aes(y=global_mavg03, colour = "Global 03"), size=0.6) +
  geom_line(aes(y=global_mavg05, colour = "Global 05"), size=0.6) +
  geom_line(aes(y=global_mavg09, colour = "Global 09"), size=0.6) +
  geom_line(aes(y=global_mavg15, colour = "Global 15"), size=0.6) +
  geom_line(aes(y=city_mavg03, colour = "City 03"), size=0.6) +
  geom_line(aes(y=city_mavg05, colour = "City 05"), size=0.6) +
  geom_line(aes(y=city_mavg09, colour = "City 09"), size=0.6) +
  geom_line(aes(y=city_mavg15, colour = "City 15"), size=0.6) +
  scale_color_manual(values=c("gold1", "darksalmon", "orangered3", "firebrick",
                              "lightblue1", "cadetblue1", "deepskyblue2", "royalblue3"),
                              name="Moving Average Interval") +
  scale_x_continuous(breaks=c(1750, 1775, 1800, 1825, 1850, 1875, 1900, 1925, 1950, 1975, 2000, 2015)) +
  scale_y_continuous(limits=c(6, 18), breaks=c(6, 8, 10, 12, 14, 16, 18)) +
  labs(x = "Year", y = "Temperature Annual Mean") +
  ggtitle("Temperature Annual Means 1750 - 2015: Moving Averages for Global and Local (Los Angeles)")
```



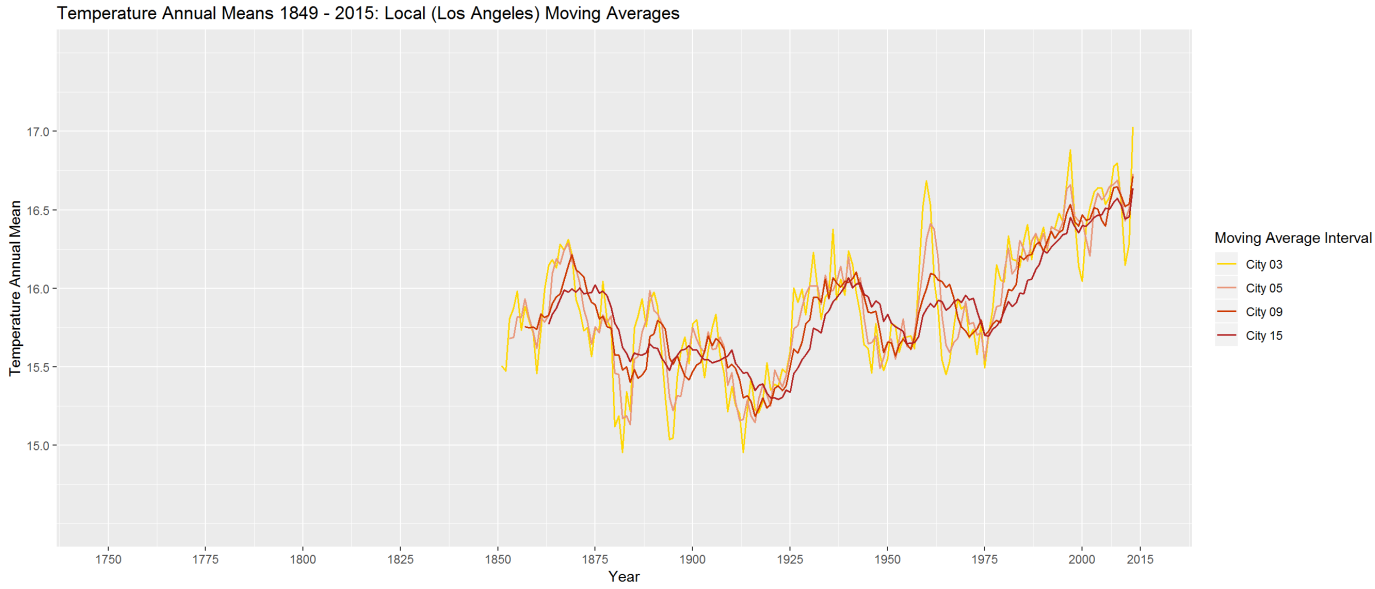
Zoom in on global temperatures only

```
ggplot(data=combined, aes(x=year)) +
  geom_line(aes(y=global_mavg03, colour = "Global 03"), size=0.6) +
  geom_line(aes(y=global_mavg05, colour = "Global 05"), size=0.6) +
  geom_line(aes(y=global_mavg09, colour = "Global 09"), size=0.6) +
  geom_line(aes(y=global_mavg15, colour = "Global 15"), size=0.6) +
  scale_color_manual(values=c("lightblue1", "cadetblue1", "deepskyblue2", "royalblue3"),
                     name="Moving Average Interval") +
  scale_x_continuous(breaks=c(1750, 1775, 1800, 1825, 1850, 1875, 1900, 1925, 1950, 1975, 2000, 2015)) +
  scale_y_continuous(limits=c(6, 10), breaks=c(6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10)) +
  labs(x = "Year", y = "Temperature Annual Mean") +
  ggtitle("Temperature Annual Means 1750 - 2015: Global Moving Averages")
```



Zoom in on local (Los Angeles) temperatures only

```
ggplot(data=combined, aes(x=year)) +
  geom_line(aes(y=city_mavg03, colour = "City 03"), size=0.6) +
  geom_line(aes(y=city_mavg05, colour = "City 05"), size=0.6) +
  geom_line(aes(y=city_mavg09, colour = "City 09"), size=0.6) +
  geom_line(aes(y=city_mavg15, colour = "City 15"), size=0.6) +
  scale_color_manual(values=c("gold1", "darksalmon", "orangered3", "firebrick"),
                     name="Moving Average Interval") +
  scale_x_continuous(breaks=c(1750, 1775, 1800, 1825, 1850, 1875, 1900, 1925, 1950, 1975, 2000, 2015)) +
  scale_y_continuous(limits=c(14.5, 17.5), breaks=c(15, 15.5, 16, 16.5, 17)) +
  labs(x = "Year", y = "Temperature Annual Mean") +
  ggtitle("Temperature Annual Means 1849 - 2015: Local (Los Angeles) Moving Averages")
```



Knit the .rmd file to html then print file to PDF

Observations

- There is higher variability for moving averages calculated with shorter intervals. Moving average interval of 3 for both global (light blue) and city (yellow) show more movement.
- As moving average interval increases (the darkest lines - dark blue and red for interval of 15), the movement becomes smoother.
- Local temperature data for Los Angeles is much warmer than the global average, for all years with data.
- Variability is relatively high until approximately 1900, after which there is a general trend of rising temperatures. This is especially visible from 1975 - 2015 on the individual global and local plots.
- Variability is higher for the local data compared to global data - likely since the global average is aggregated from multiple locations, and the increased sample size will lower the variance.

References

https://www.rdocumentation.org/packages/zoo/versions/1.8-3/topics/rollmean
http://zevross.com/blog/2014/08/04/beautiful-plotting-in-r-a-ggplot2-cheatsheet-3/
http://www.stat.columbia.edu/%7Etzheng/files/Rcolor.pdf
https://rmarkdown.rstudio.com/authoring_basics.html