Project 1: Exploring Weather Trends
BY: Khan Akbar

Outline

- Purpose
- Tools and Methods
- Data
- Analysis
- Results

Purpose

Describing the similarities and differences between global temperature trends, and temperature trends in the closest big city

Tools and Methods

I wrote SQL queries for extracting the city and global data from a database and saved them as CSV files. The report was generated by RMarkdown. I used the libraries reader, dplyr, ggplot, and caret to manipulate the data, and create line charts

Project Outline:

1. I used SQL to query the database on the Udacity website. My SQL code is as follows:

```
### Extract the data
- city list
    SELECT *
    FROM city_list
    ORDER BY country, city

- city data
    SELECT *
    FROM city_data
    WHERE city IN ('San Jose', 'Los Angeles','San Francisco') AND country = 'United States'

- global data
    SELECT *
    FROM global_data
```

Analysis

1) I imported the data set CSV file into R studio
2) I created the time from 1849 to 2013 for each data set because it is easy fair for comparison
3) I created 10 years of moving average and generate a line chart to avoid shoving the line in NA Graph we removing the NA and missing value.
4) I generated the following line charts.

- ➔ 7 year moving average in Global
- ➔ 7 year moving average in San Jose
- ➔ 7 year moving average in Global versus San Jose
- ➔ 7 year moving average in Los Angeles
- ➔ 7 year moving average in Taipei
- ➔ 7 year moving average in the world, San Jose, Los Angeles, and Taipei

Results:

1) Compared to the 7-year moving average temperature in my current city, San Jose with in the world from 1849 to 2013

- The temperature trends in the world and San Jose have gradually become hotter and hotter, especially from 1975 to 2013.
- The temperature in San Jose has not been consistently going up. Before 1975, the temperature has been dramatically up and down. The temperature in the world has been consistently going up, especially from 1975 to 2013.
- The average temperature in San Jose is hotter than in the world, around 6 degrees difference.
- In the world, the year is highly and positively correlated with average temperature than in San Jose.

1) Compared to the 7-year moving average temperature among the world, San Jose, Los Angeles, and Taipei from 1849 to 2013.

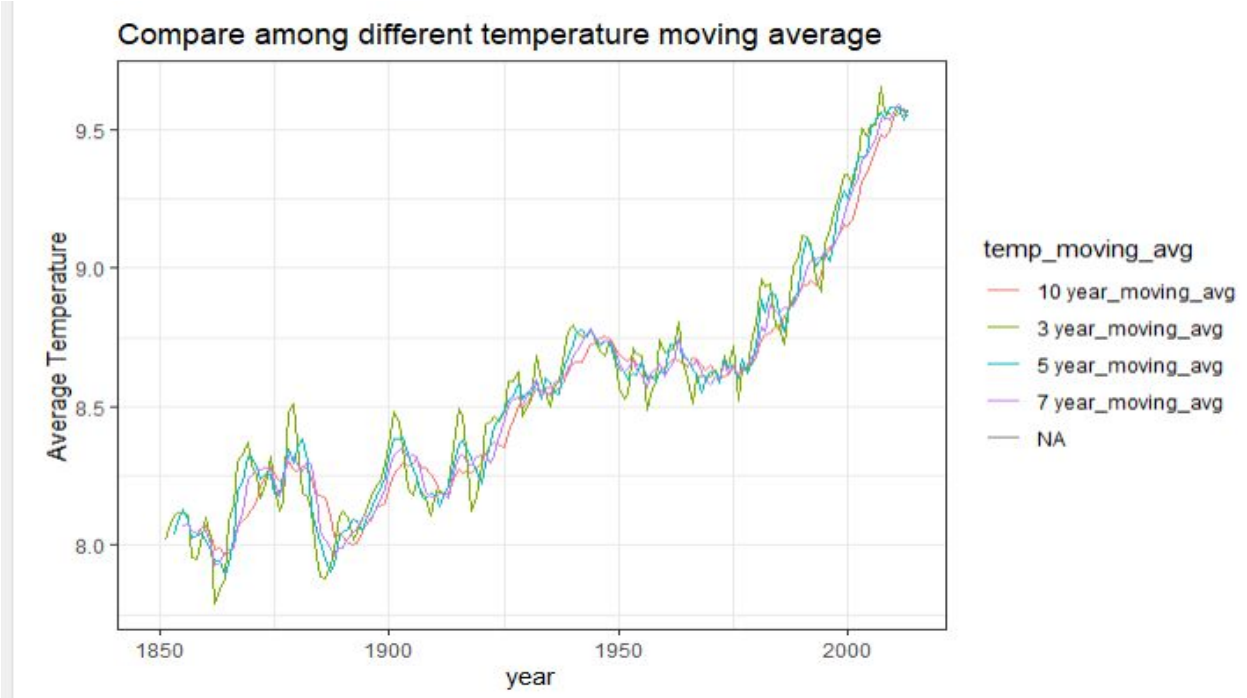
- The temperature in Taipei has been the hottest over time compared to the world, San Jose and Los Angeles.
- The temperatures in San Jose and in Los Angeles have similar trends over the years even though the temperature in Los Angeles has been always higher than in San Jose over the years.
- The temperatures in San Jose, Los Angeles, and Taipei are all higher than in the world.
- The world has the highest correlation between year and average temperature compared to San Jose, Los Angeles, and Taipei.

Data import the data in R Studio

```{r message=TRUE, warning=TRUE, paged.print=TRUE}
## 1) Reading datasets
library(readr)
library(dplyr)

# a. Global
global <- read_csv("C:/Users/khana/OneDrive/Desktop/UDACITY project/Udacity-DAND-Weather-Trends-master (1)/Udacity-DAND-Weather-Trends-master/data/global.csv")

# Make the time period the same like San Jose and Los Angeles
global2 <- global %>%
  filter(year >= 1849 & year <= 2013)

# b. My current Location: San Jose
cities <- read_csv("C:/Users/khana/OneDrive/Desktop/UDACITY project/Udacity-DAND-Weather-Trends-master (1)/Udacity-DAND-Weather-Trends-master/data/cities.csv")
sj <- cities %>%
  filter(city == 'San Jose') %>%
  select(year,avg_temp)

# c. The city I stayed before the Bay: Los Angeles
la <- cities %>%
  filter(city == 'Los Angeles') %>%
  select(year,avg_temp)

# d. My hometown: Taipei
tpi <- read_csv("C:/Users/khana/OneDrive/Desktop/UDACITY project/Udacity-DAND-Weather-Trends-master (1)/Udacity-DAND-Weather-Trends-master/data/tpi.csv") %>%
  filter(year >= 1849 & year <= 2013) %>%
  select(year,avg_temp)

```

Calculating moving average

```{r}
################################### 2) Create a function for calculating moving average ##############################3
move <- function(data,loc,y){
    mov_year = data.frame()
    locat = data.frame()
    year = data.frame()
    movavg = data.frame()
    for(i in 1:length(data$avg_temp)){
        if (i < y){
            x <- NA
            z <- NA
        }
            else{
                x <- data %>%
                summarize(movmean = mean(avg_temp[(i-(y-1)):i]))
                z <- paste(y,'year_moving_avg')
                }
        movavg = rbind(movavg,data.frame(as.numeric(x)))
        year = rbind(year,z)
        locat = rbind(locat,loc)
    }
    names(locat) <- 'location'
    names(movavg) <- 'mov_avg'
    names(year) = 'temp_moving_avg'
    mov_year <- cbind(data,movavg,year,locat)
    return(mov_year)
}

globalall <- rbind(move(global2,'Global',3),move(global2,'Global',5),move(global2,'Global',7),move(global2,'Global',10))

library(ggplot2)
ggplot(data=globalall, aes(x=year, y=mov_avg, color=temp_moving_avg))+geom_line()+
  ylab("Average Temperature")+theme_bw()+ggtitle("Compare among different temperature moving average")
```



Removing 1st to 6th year (NAs)

```{r}
# Remove the missing value and data prepreation by remob=ving  NAs
global7 <- move(global2,'Global',7) %>%
    filter(year >= 1855)
head(global7)

sj7 <- move(sj,'San Jose',7) %>%
  filter(year >= 1855)
head(sj7)

glo_sj <- rbind(global7,sj7)

la7 <- move(la,'Los Angeles',7) %>%
  filter(year >= 1855)
head(la7)

tpi7 <- move(tpi,'Taipei',7) %>%
  filter(year >= 1855)
head(tpi7)

glob_city <- rbind(global7,sj7,la7,tpi7)
```

data.frame 6 x 5 | data.frame 6 x 5 | data.frame 6 x 5 | **data.frame 6 x 5**

	year <dbl>	avg_temp <dbl>	mov_avg <dbl>	temp_moving_avg <chr>	location <chr>
1	1855	21.89	21.94714	7 year_moving_avg	Taipei
2	1856	21.54	21.92714	7 year_moving_avg	Taipei
3	1857	21.94	21.93571	7 year_moving_avg	Taipei
4	1858	21.67	21.91000	7 year_moving_avg	Taipei
5	1859	21.87	21.90714	7 year_moving_avg	Taipei
6	1860	21.70	21.83286	7 year_moving_avg	Taipei

6 rows

Compare weather trends in global and cities

```r
## 3) Data Visualization
library(ggplot2)

ggplot(data=global7, aes(x=year, y=mov_avg, color=temp_moving_avg))+geom_line()+
    ylab("Average Temperature")+theme_bw()+ggtitle("7 year Moving Average Temperature in Global")

ggplot(data=sj7, aes(x=year, y=mov_avg, color=temp_moving_avg))+geom_line()+
  ylab("Average Temperature")+theme_bw()+ggtitle("7 year Moving Average Temperature in San Jose")

ggplot(data=glo_sj, aes(x=year, y=mov_avg, color=location))+geom_line()+
  ylab("Average Temperature")+theme_bw()+ggtitle("7 year Moving Average Temperature Global V.S. San Jose")

ggplot(data=la7, aes(x=year, y=mov_avg, color=temp_moving_avg))+geom_line()+
  ylab("Average Temperature")+theme_bw()+ggtitle("7 year Moving Average Temperature in Los Angeles")

ggplot(data=tpi7, aes(x=year, y=mov_avg, color=temp_moving_avg))+geom_line()+
  ylab("Average Temperature")+theme_bw()+ggtitle("7 year Moving Average Temperature in Taipei")

ggplot(data=glob_city, aes(x=year, y=mov_avg, color=location))+geom_line()+
  ylab("Average Temperature")+theme_bw()+ggtitle("7 year Moving Average Temperature")
```