Exploring Weather Trends

Lesson 3 Project
Report by Devin de Araujo

Outline.

1. Determining the data available for cities in Australia.

```
SELECT *
FROM city_list
WHERE city IN ('Melbourne', 'Sydney', 'Brisbane', 'Canberra', 'Adelaide', 'Perth');
```

2. Extract local City Data

```
SELECT *
FROM city_data
WHERE city IN ('Melbourne', 'Adelaide', 'Brisbane', 'Sydney', 'Perth', 'Canberra');
```

3. Extract the Global Data

```
SELECT *
FROM global_data;
```

4. Assessing the data

It was determined that the available City data contained less observations than the Global data set. I decided to combine the data and use only the available information that was true for both data sets.

5. Combining data

Excel formula - Combine Data into one
=VLOOKUP(A2, global_data!A:B,2,FALSE)

	A	B	C	D	F
	year	city	countr	city_avg_temp	global_avg_temp
21	1841	Melbourne	Australia	13.09	7.69
22	1842	Melbourne	Australia	12.96	8.02
23	1843	Melbourne	Australia	13.34	8.17
24	1844	Melbourne	Australia	12.60	7.65
25	1845	Melbourne	Australia	13.16	7.85

6. Calculate moving averages

Given the range of the City data (MIN 1841, MAX 2013), I decided to calculate moving averages in 10 year periods.

Formula
=ROUND(AVERAGE(D22:D31),2)

	A	B	C	D	E	F	G
1	year	city	country	city_avg_temp	10_city_avg	global_avg_temp	10_glo_a
29	1849	Melbourne	Australia	12.30		7.98	
30	1850	Melbourne	Australia	12.99		7.90	
31	1851	Melbourne	Australia	12.97	12.94	8.18	8.04
32	1852	Melbourne	Australia	12.72	12.91	8.10	8.05

7. Plotting the Data

I used RStudio since I am already somewhat familiar with the program and syntax.

Plot 1

How do the yearly averages for Melbourne, Australia compare to the 10 year moving average?

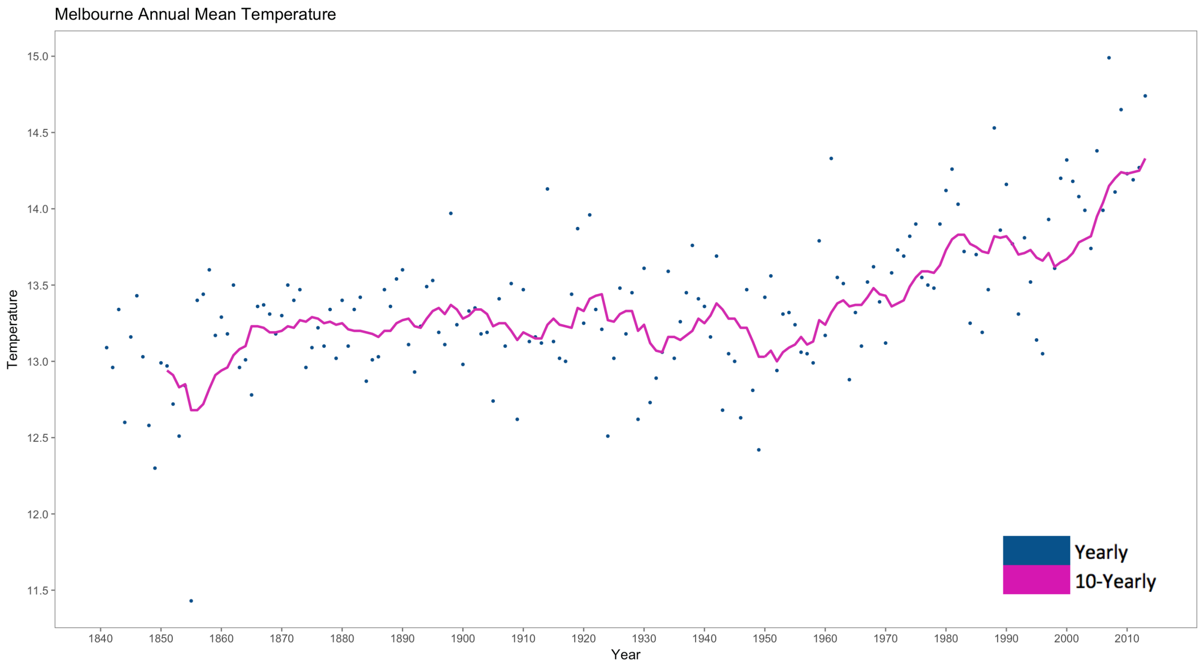
A dot plot was chosen for the yearly averages since a line graph would be too hard to read. The dot plot provides a broad indication of the yearly temperatures while the moving average is a solid indication of the overall trend.

Considerations

Appropriate breaks were set on the X and Y axes. These were determined on the X axis by the 10 year moving average and on the Y axis, by half degrees.

Observation 1

The data suggests that on average, local temperatures in Melbourne have followed an increasing trend.



R Script

```
ggplot(data = subset(weather, city == 'Melbourne'), aes(x = year, y = city_avg_temp)) +
  geom_point(size = .8, colour = '#08538b') +
  geom_line(aes(y = X10_city_avg), colour = '#d617b1', size = 1) +
  scale_x_continuous(breaks = seq(1840, 2013, 10)) +
  scale_y_continuous(breaks = seq(0,16, 0.5)) +
  labs(title = "Melbourne Annual Mean Temperature",
     x = "Year",
     y = "Temperature") +
  theme_few() + scale_colour_tableau()
```

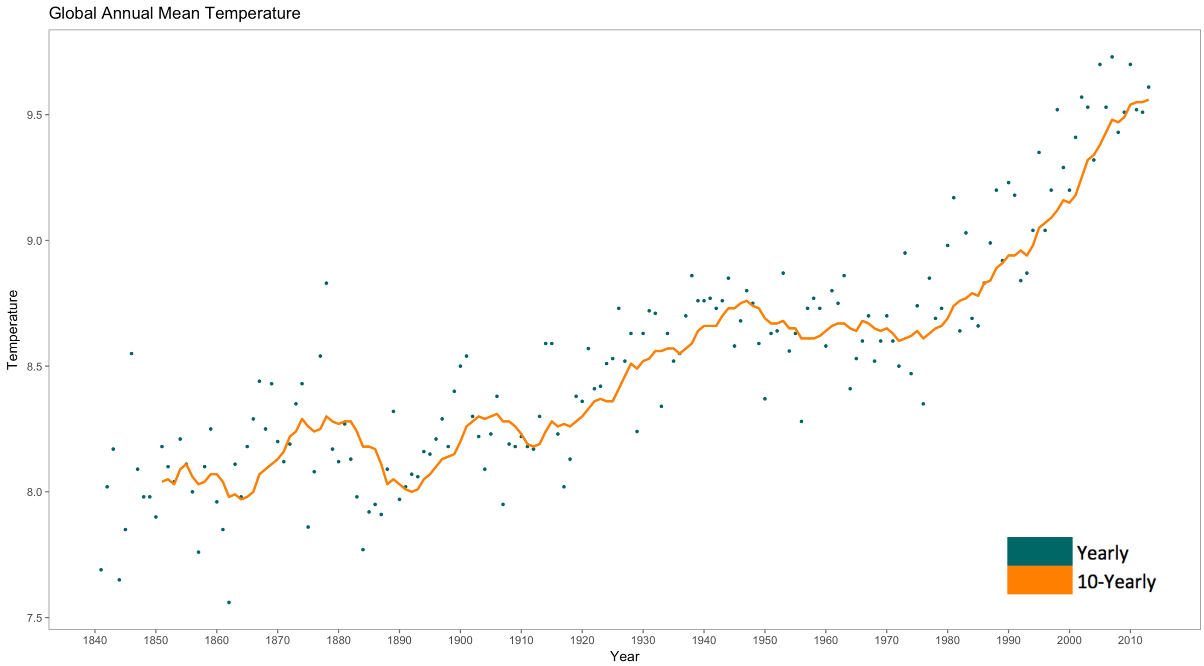
Plot 2

What do yearly averages suggest is happening to global temperature?

The same plotting method as above was used to determine this.

Observation 2

Over time, Global temperatures have continually increased. After a fairlystable period between 1950 - 1980, this increase accelerated rather drastically.



R Script

```
ggplot(data = subset(weather, city == 'Melbourne'), aes(x = year, y = global_avg_temp)) +
  geom_point(size = .8, colour = '#006666') +
  geom_line(aes(y = X10_glo_avg), colour = '#ff8100', size = 1) +
  scale_x_continuous(breaks = seq(1840, 2013, 10)) +
  scale_y_continuous(breaks = seq(0,10, 0.5)) +
  labs(title = "Global Annual Mean Temperature",
     x = "Year",
     y = "Temperature") +
  theme_few() + scale_colour_tableau()
```

Plot 3

How do temperatures in Melbourne city compare to Global temperatures?

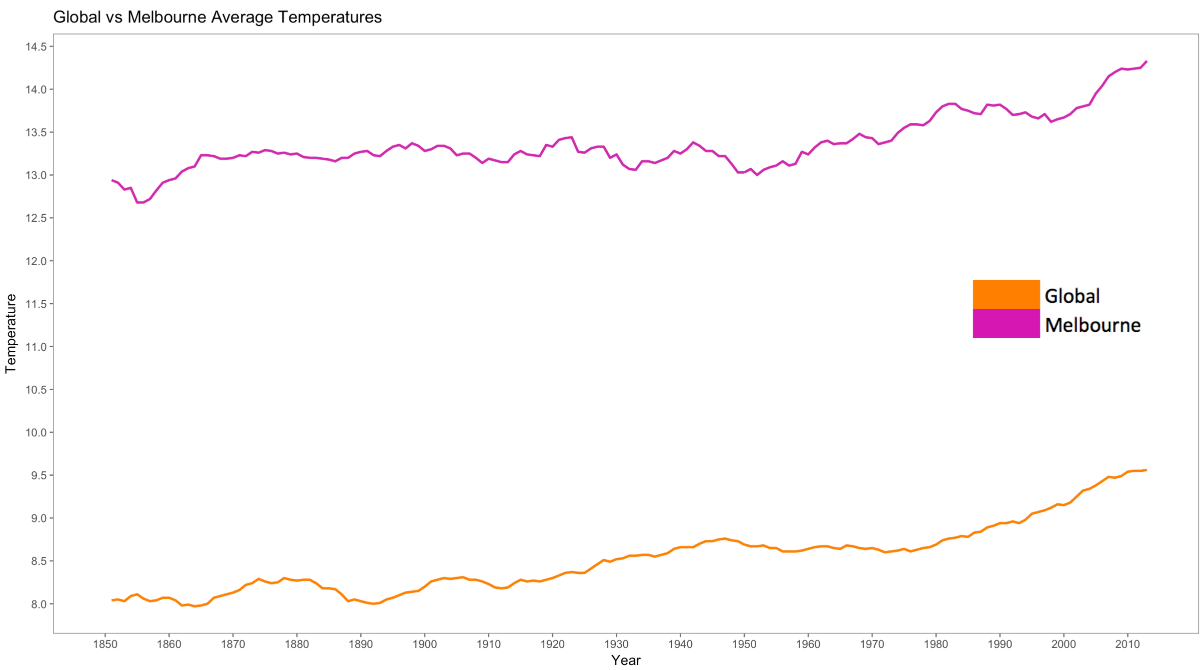
Observation 3

When plotted together, we can see that both Melbourne and Global temperatures have increased over time. The data suggests that changes in global temperature are reflected by changes in temperature in Melbourne.

On average, temperatures in Melbourne are well above Global temperatures. This has been fairly consistent over time, however, the trend is more dramatic on the global scale.

	MIN	MEDIAN	MAX
Global	7.97	8.55	9.56
Melbourne	10.74	15.94	20.29

Overall, the world is getting hotter. This increase has been more dramatic since the 1970s. The temperatures in Melbourne are succinct with what is happening globally.



R Script

```
ggplot(data = subset(weather, city == 'Melbourne'), aes(x = year)) +
  geom_line(aes(y = X10_city_avg), colour = '#d617b1', size = 1) +
  geom_line(aes(y = X10_glo_avg), colour = '#ff8100', size = 1) +
  scale_x_continuous(limits = c(1850, 2013), breaks = seq(1850, 2013, 10)) +
  scale_y_continuous(breaks = seq(0,20, 0.5)) +
  labs(title = "Global vs Melbourne Average Temperatures",
     y = "Temperature",
     x = "Year") +
 theme_few() + scale_colour_tableau()
```

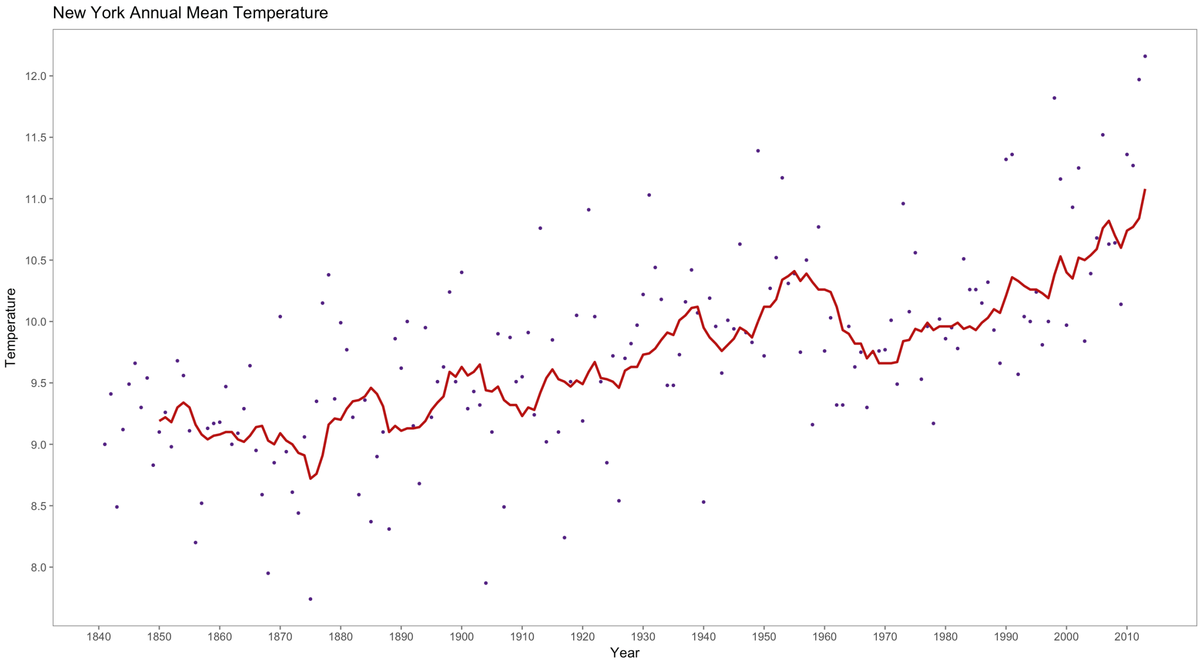
Additional Plots

Plot 4

How do the yearly averages for New York compare to the 10 year moving average?

Observation 4

The data suggests that local average temperatures in New York are continually increasing. The trend is not overly smooth, there was a peak around 1955 followed by a drop around 1970, but overall, temperatures appear to be rising.



R Script

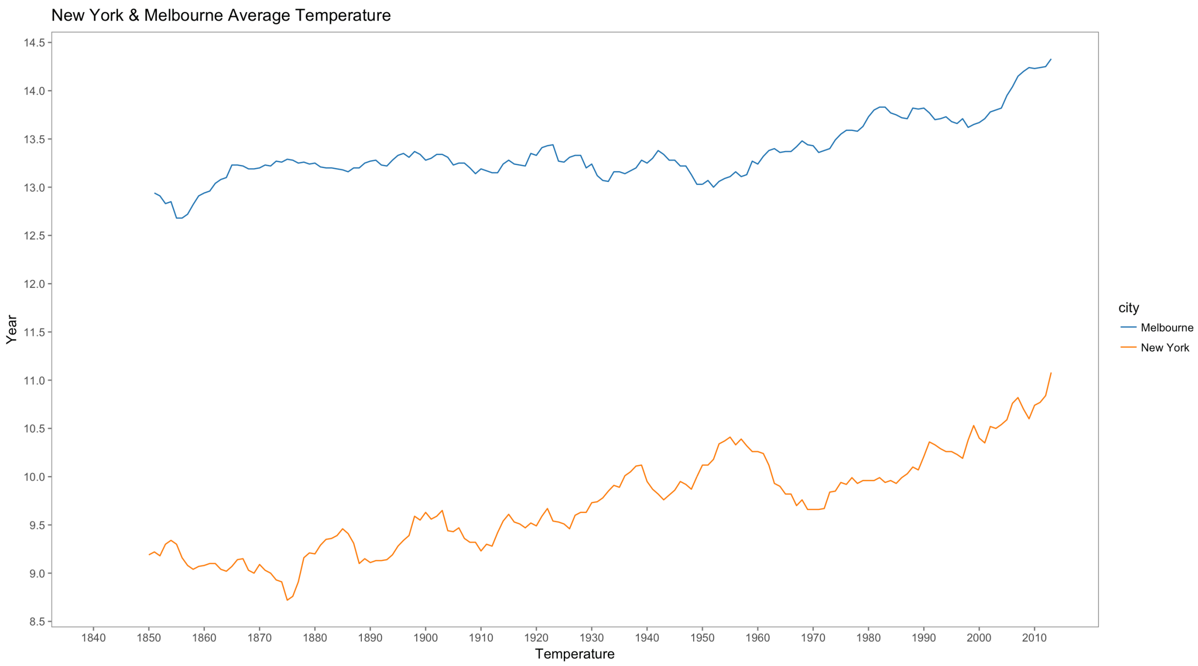
```
ggplot(data = subset(weather, city == 'New York'), aes(x = year, y = city_avg_temp)) +
  geom_point(size = .8, colour = '#562383') +
  geom_line(aes(y = X10_city_avg), colour = '#BA1313', size = 1) +
  scale_x_continuous(breaks = seq(1840, 2013, 10)) +
  scale_y_continuous(breaks = seq(0,16, 0.5)) +
  labs(title = "New York Annual Mean Temperature",
       x = "Year",
       y = "Temperature") +
  theme_few() + scale_colour_tableau()
```

Plot 5

How do the 10 yearly average temperatures for New York compare to Melbourne?

Observation 5

While it is true that both cities have increasing temperatures; those in New York tend to fluctuate more dramatically, while in Melbourne, changes are relatively stable in comparison.



R Script

```
ggplot(data = subset(weather, city == 'Melbourne' | city == 'New York'),
    aes(x = year, y = X10_city_avg, group = city, colour = city)) +
 geom_line() +
 scale_x_continuous(breaks = seq(1840, 2013, 10)) +
 scale_y_continuous(breaks = seq(0,16, 0.5)) +
 labs(title = "New York & Melbourne Average Temperature",
    y = "Year",
    x = "Temperature") +
 theme_few() + scale_colour_tableau()
```

Calculate the Correlation Coefficient

Outline

Using R, fist subset the data to extract only the Melbourne and New York data. Then coerce the matrices into Data Frames using the Tibble package.

OUTPUT

Correlation Coefficient = **0.4084517**

R Script

```
# Subset data, select only city and city_avg_columns
mel <- subset(weather, city == "Melbourne", select = c(city, city_avg_temp))
nyc <- subset(weather, city == "New York", select = c(city, city_avg_temp))

# use tibble to coerce into data frame
mel <- as_tibble(mel)
nyc <- as_tibble(nyc)

# calculate Correlation Coefficient
cor(mel$city_avg_temp, nyc$city_avg_temp)
```

R Packages Used

```
Library(ggplot2)
library(tidyverse)
library(ggthemes)
library(lubridate)
library(RColorBrewer)
```