Western Australia Weather

Christina Dietrich

11/17/2021

Import libraries

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

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(readr)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v tibble  3.1.4     v stringr 1.4.0
## v tidyr   1.1.3     v forcats 0.5.1
## v purrr   0.3.4
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x lubridate::as.difftime() masks base::as.difftime()
## x lubridate::date()        masks base::date()
## x dplyr::filter()          masks stats::filter()
## x lubridate::intersect()   masks base::intersect()
## x dplyr::lag()             masks stats::lag()
## x lubridate::setdiff()     masks base::setdiff()
## x lubridate::union()       masks base::union()
```

```
library(moderndive)
library(skimr)
library(ISLR)
```

Import data

```
au_weather <- read.csv("aus_weather (1).csv")
```

Data Exploration

View the raw data and draw conclusions

```
View(au_weather)
glimpse(au_weather)
```

```
## Rows: 26,543
## Columns: 10
## $ Year        <int> 1944, 1944, 1944, 1944, 1944, 1944, 1944, 1944, 1944, 1944~
## $ Month       <int> 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5~
## $ Day         <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17,~
## $ rainfall_mm <dbl> 0.0, 0.0, 0.0, 4.3, 0.0, 2.5, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0~
## $ min_temp_C  <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
## $ max_temp_C  <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
## $ daily_avg   <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ daily_range <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ uv_MJ_m.m   <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
## $ Season      <int> 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3~
```

Looking at the raw data presented, I determined that I wanted to use the variables of date (season, month, year), “daily_avg”, and “rainfall_mm”. The date will aid in looking at the variables over a certain amount of time and determining if there were any changes. The year spans from 1944 to 2016 which is 72 years. The second variable, “daily_avg”, represents the average temperature each day from 1944 to 2016. It is represented in degrees Celsius and spans from 0 to about 36. The third variable, “rainfall_mm”, is measured in millimeters and is the amount of rain each day from 1944 to 2016. This variable spans from 0 millimeters to 132 millimeters.

Compute Summary Statistics

```
au_weather %>%
  get_correlation(formula = rainfall_mm ~ daily_avg)
```

```
##         cor
## 1 -0.229998
```

This value represents the correlation between rainfall and the average temperature each day. Because the value is -0.229998, it can be seen that the correlation between rainfall and the average temperature is a weaker negative correlation. This value means that if the temperature were to increase by one degree Celsius, the rainfall would decrease by 0.22 millimeters.

```
au_weather %>%
  summarize(mean_rainfall = mean(rainfall_mm),
            median_rainfall = median(rainfall_mm),
            mean_temp = mean(daily_avg),
            median_temp = median(daily_avg))
```

```
##   mean_rainfall median_rainfall mean_temp median_temp
## 1      2.101281               0  18.22445        17.3
```

This table represents the mean and median of the numerical variables I used in my data set. The mean rainfall is 2.102181 millimeters in a day, while the median rainfall is 0 millimeters in one day. The reason that the median is zero is because Australia does not receive much rain a lot of the year. This would mean that a lot of the values are zeros. The mean temperature is 18.22445 degrees Celsius (about 64.8 degrees Fahrenheit), while the median temperature is 17.3 degrees Celsius (about 63.1 degrees Fahrenheit). These values represent Western Australia as dry with temperate weather.

```
# Creating the data set "cat_month_season" transfers month and season into a categorical variabl
e that r understands, as opposed to keeping the month and season as numbers.
cat_month_season <- au_weather %>%
  mutate(month = cut(Month,
    breaks = c(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12),
    labels = c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov", "De
c")
  )) %>%
  mutate(season = cut(Season,
    breaks = c(0, 1, 2, 3, 4),
    labels = c("Winter", "Spring", "Summer", "Fall")
  )) %>%
  slice(1:26543)

cat_month_season %>%
  select(rainfall_mm, daily_avg, season, month) %>%
  skim()
```

Data summary

Name	Piped data
Number of rows	26543
Number of columns	4

Column type frequency:	
factor	2
numeric	2
Group variables	None

Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
season	0	1	FALSE	4	Fal: 6716, Sum: 6655, Win: 6643, Spr: 6529
month	0	1	FALSE	12	May: 2263, Jul: 2263, Aug: 2263, Oct: 2263

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
rainfall_mm	0	1	2.10	5.85	0	0.00	0.0	0.6	132.0	
daily_avg	0	1	18.22	5.20	0	14.25	17.3	21.9	36.8	

This table of summary statistics presents the same means as were in the previous table. It also shows the percentiles for the rainfall and temperature variables. It shows that 75% of rainfall totals were simply 0 millimeters. Only in the top 25% of data are numbers between 0.6 and 132 millimeters. The temperature data shows that the lowest 25% of temperatures were between 0 and 14.2 degrees Celsius, while the highest 25% of temperatures were between 21.9 and 36.8 degrees Celsius. The 50th percentile represents the median and is the same as the one calculated in the above table. The summary statistics also output “top values” for season and month. It shows that fall had the most values at 6716, while spring had the least values at 6529. The variation is due to the number of days within those particular seasons. It also says that May, July, August, and October all have equivalent inputs. This is because each of those months has 31 days so they will have more inputs than months with 30 or 28 days.

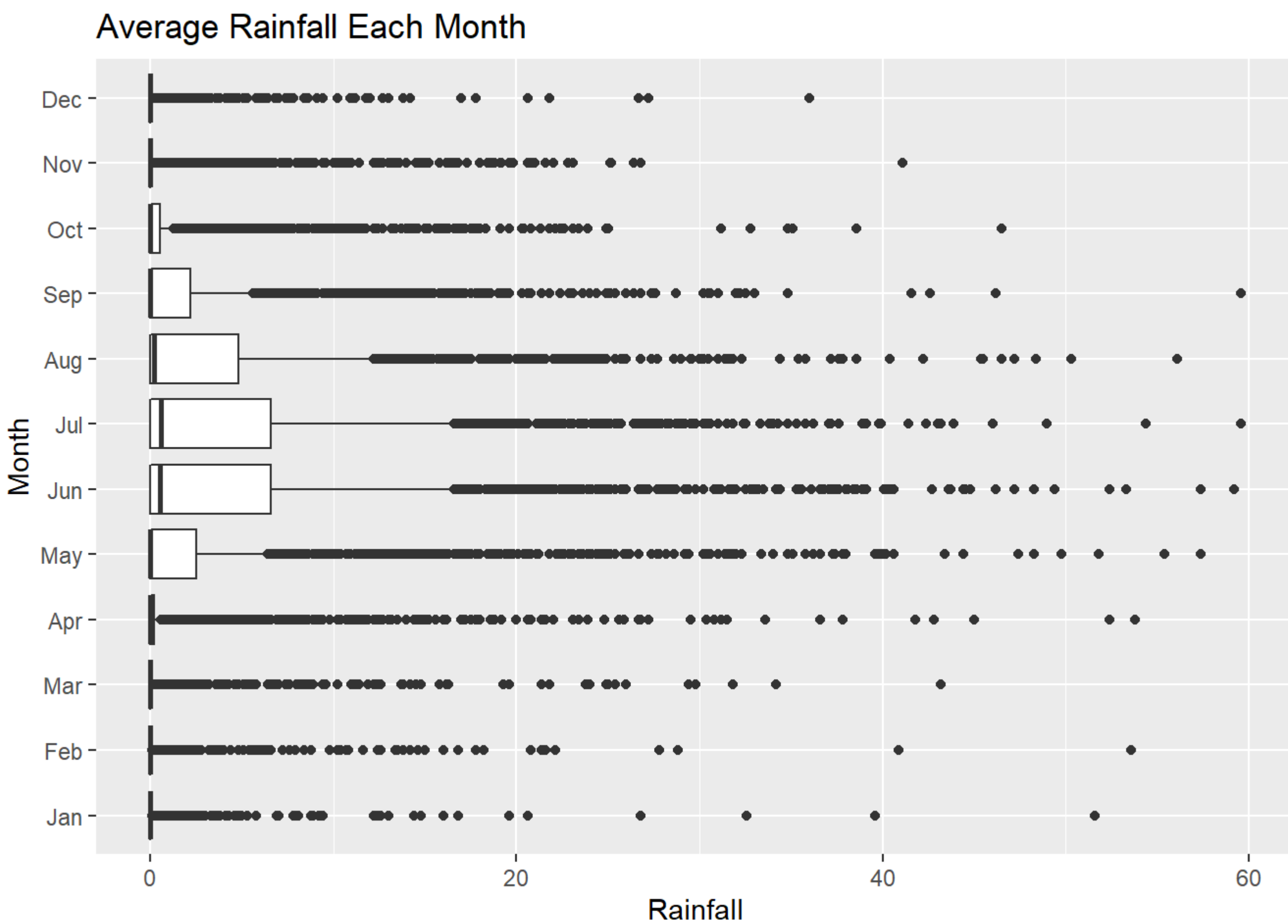
Data Visualizations

In this data set, I will be using the variables of year, month, season, and daily_avg(temp), and rainfall_mm. I will be answering three questions: 1) Has the amount of rainfall increased or decreased over time?
2) Has the average temperature per month each year shifted at all over time?
3) Are the temperature and rainfall correlated in any way?

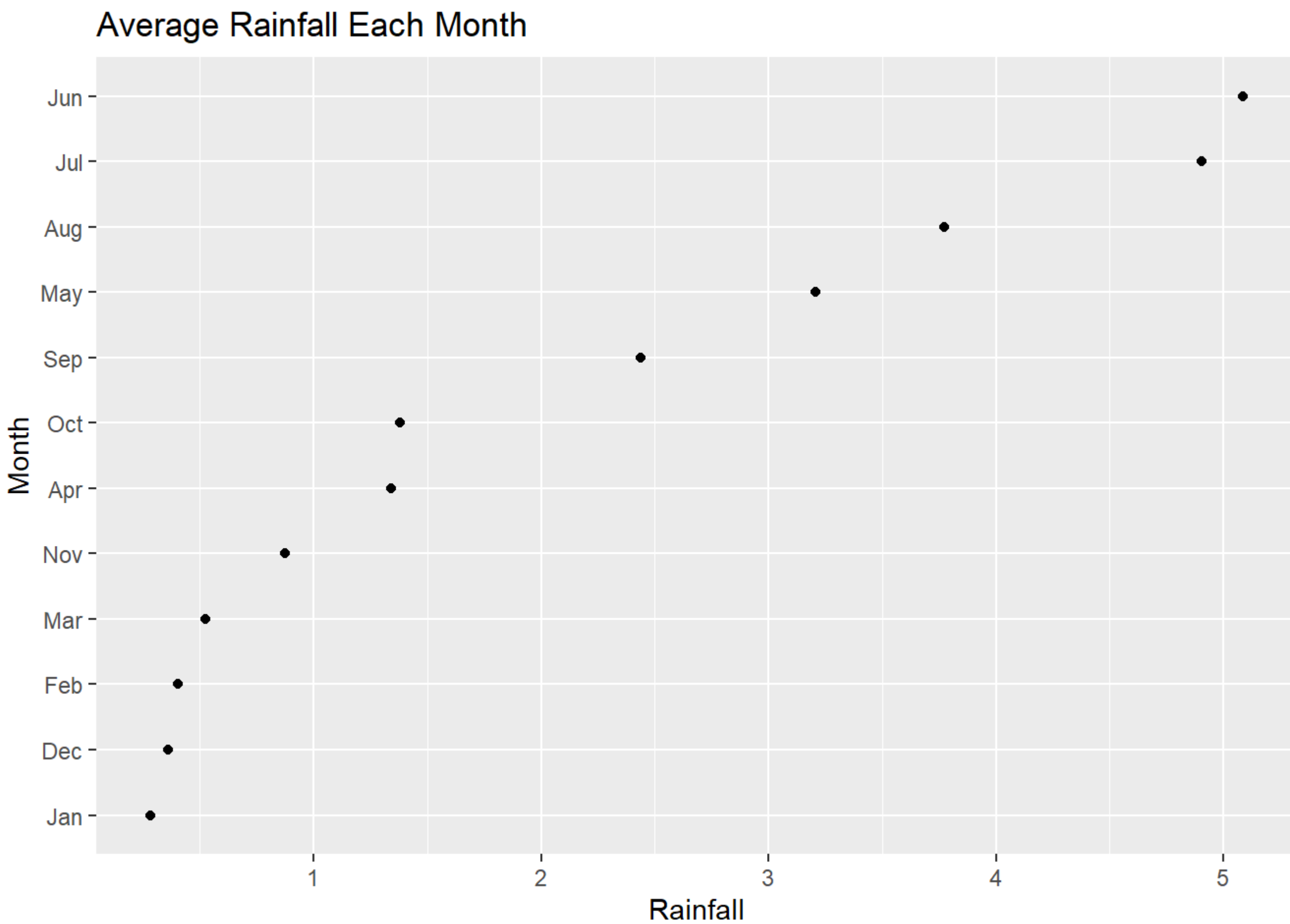
1. Has the amount of rainfall increased or decreased over time?

In order to answer this question, I have to start with a base knowledge of when the rainy season occurs and when the dry season occurs. I then plan on creating graphs of two different time periods in order to compare the rainfalls of those two years. I also plan on creating graphs that compare the average temperatures of each year over time. I had to create some different data sets in order to make the data easier to visualize.

```
# This table was created in order to make month a categorical variable in r and take out the out
liers in rainfall
date <- au_weather %>%
  filter(rainfall_mm < 60) %>%
  mutate(month = cut(Month,
    breaks = c(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12),
    labels = c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov", "De
c")
  )) %>%
  slice(1:26543)

ggplot(data = date, mapping = aes(x = rainfall_mm, y = month, fct_reorder(month, rainfall_mm)))
 + geom_boxplot() + labs(y = "Month", x = "Rainfall", title = "Average Rainfall Each Month")
```




```
# Creating this table allowed me to calculate the average rainfall each month.
avg_rainfall <- date %>%
  select(month, rainfall_mm) %>%
  group_by(month) %>%
  summarize(avg_rain = mean(rainfall_mm))

ggplot(avg_rainfall, aes(avg_rain, fct_reorder(month, avg_rain))) + geom_point()+ labs(y = "Mont
h", x = "Rainfall", title = "Average Rainfall Each Month")
```



I used the box plot in order to look at the general median rainfall totals each month. I then created the scatter plot in order to look at the average rainfall of each month. The graph with averages proved to be easier to interpret because oftentimes, the median value was 0. The averages show that the rainy season is usually the winter months (May-August), and the dry season is usually the summer months (December - March). The graph's purpose was to be able to look at the raw data in a different way and determine when rainfall is the greatest and the least.

```
avg_rainfall_46 <- au_weather %>%
  group_by(Year, Month) %>%
  filter(Year == 1946) %>%
  mutate(rainfall_mm) %>%
  summarize(avg_rain_month = mean(rainfall_mm))
```

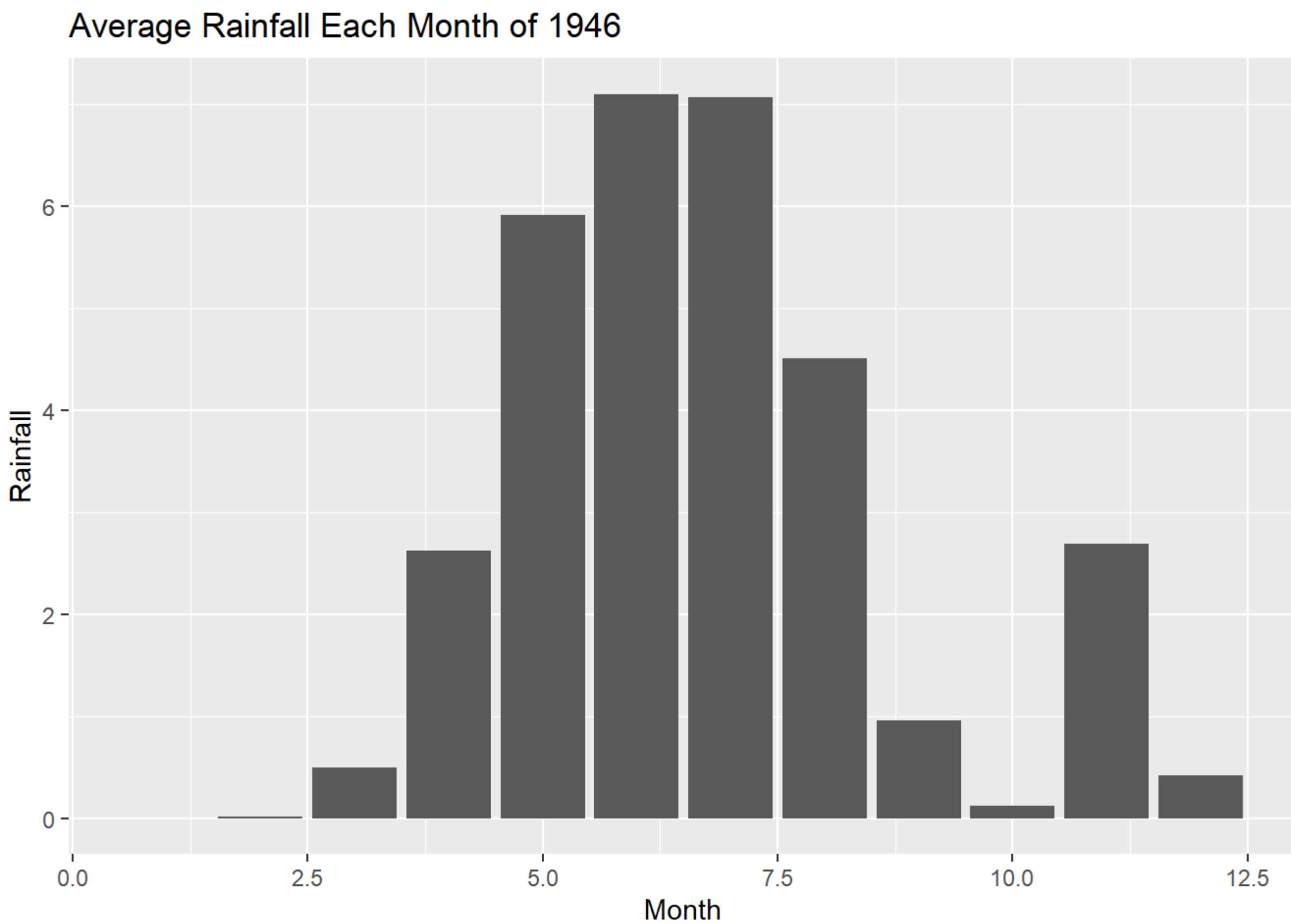
```
## `summarise()` has grouped output by 'Year'. You can override using the `.groups` argument.
```

```
avg_rainfall_16 <- au_weather %>%
  group_by(Year, Month) %>%
  filter(Year == 2016) %>%
  mutate(rainfall_mm) %>%
  summarize(avg_rain_month = mean(rainfall_mm))
```

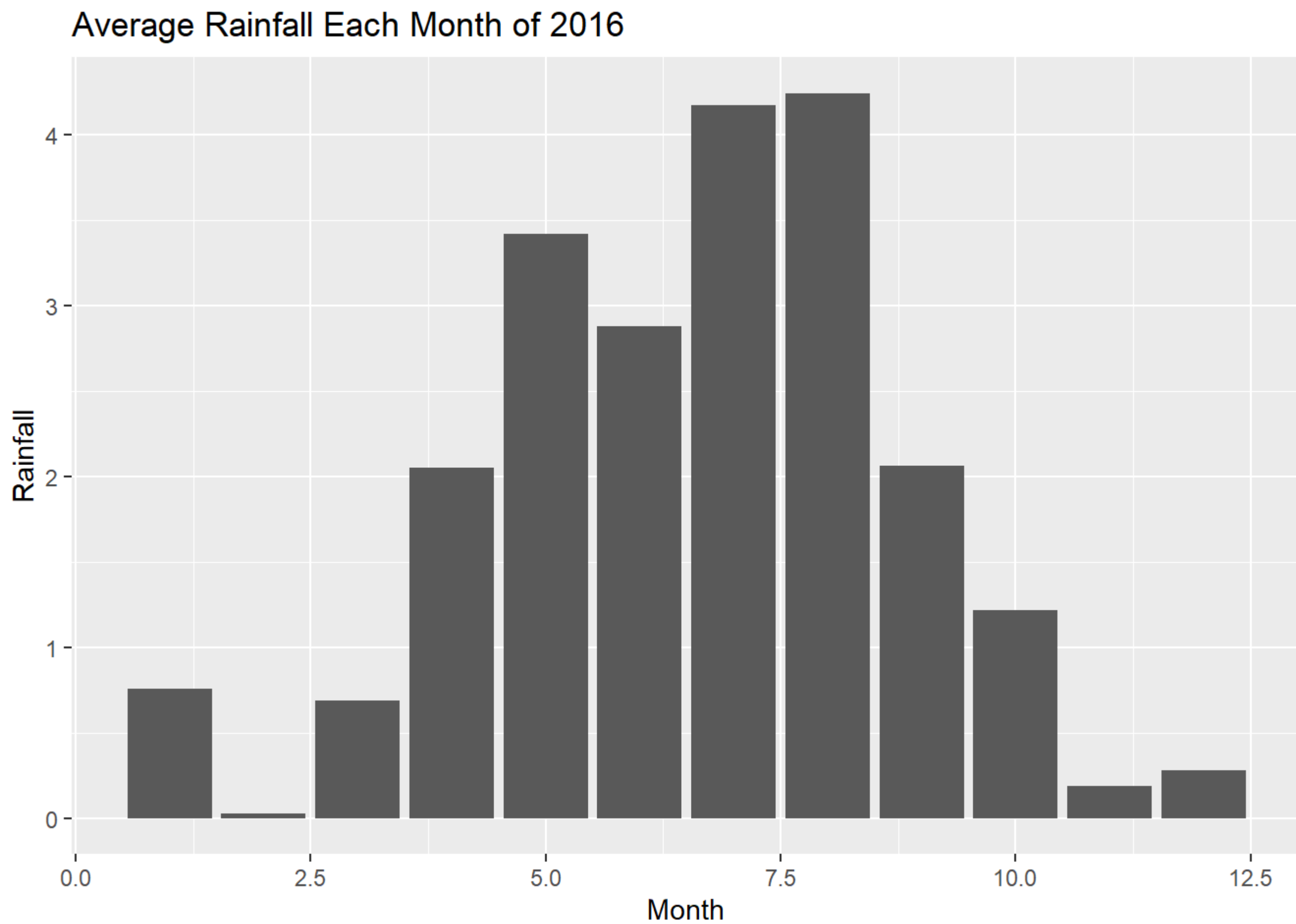
```
## `summarise()` has grouped output by 'Year'. You can override using the `.groups` argument.
```

This code filters out the data set using the years 2016 and 1946. It also creates a variable that is the average rain each month of that particular year.

```
ggplot(data = avg_rainfall_46, mapping = aes(x = Month, y = avg_rain_month)) + geom_col() + labs
(x = "Month", y = "Rainfall", title = "Average Rainfall Each Month of 1946")
```

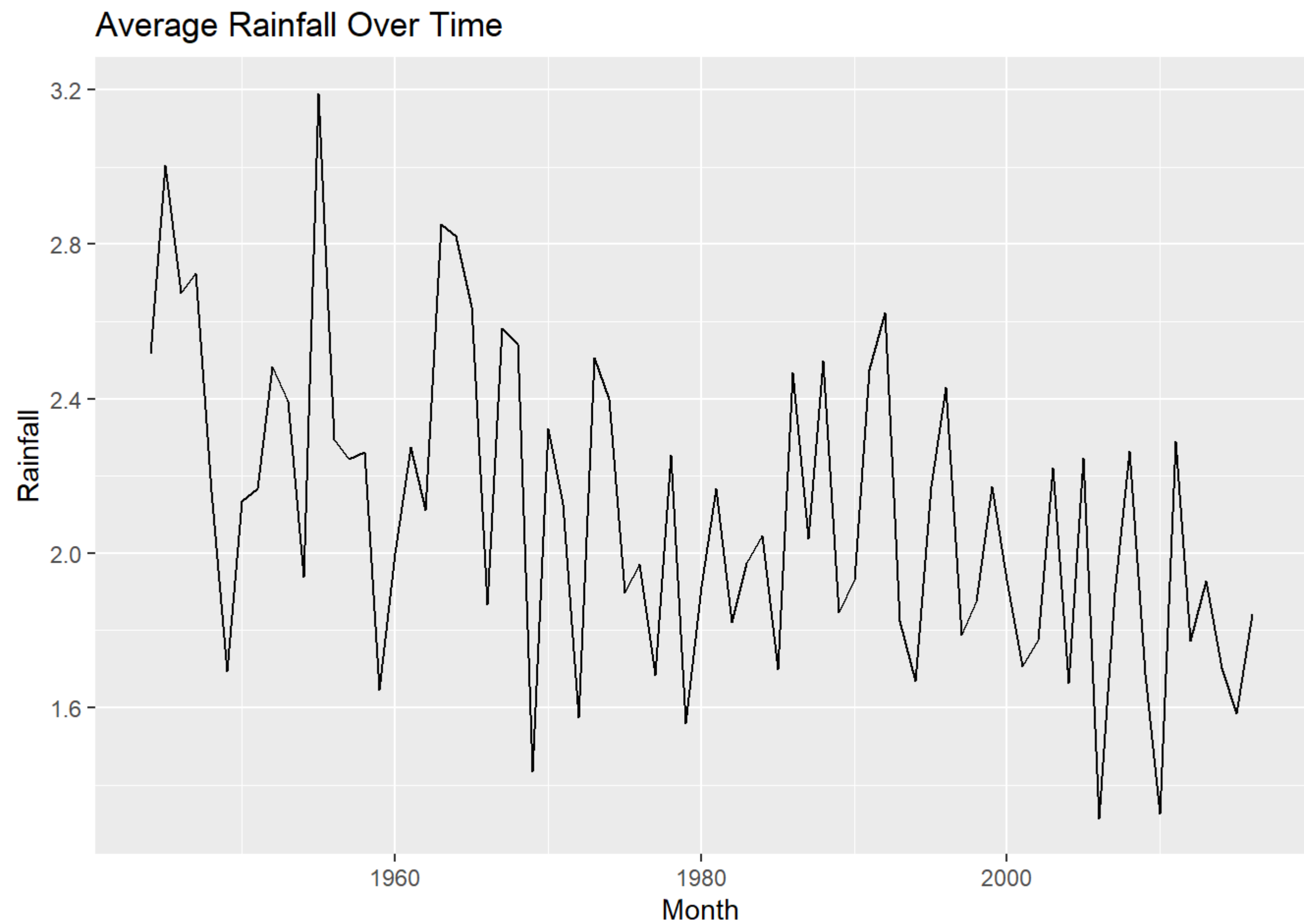



```
ggplot(data = avg_rainfall_16, mapping = aes(x = Month, y = avg_rain_month)) + geom_col()+ labs
(x = "Month", y = "Rainfall", title = "Average Rainfall Each Month of 2016")
```



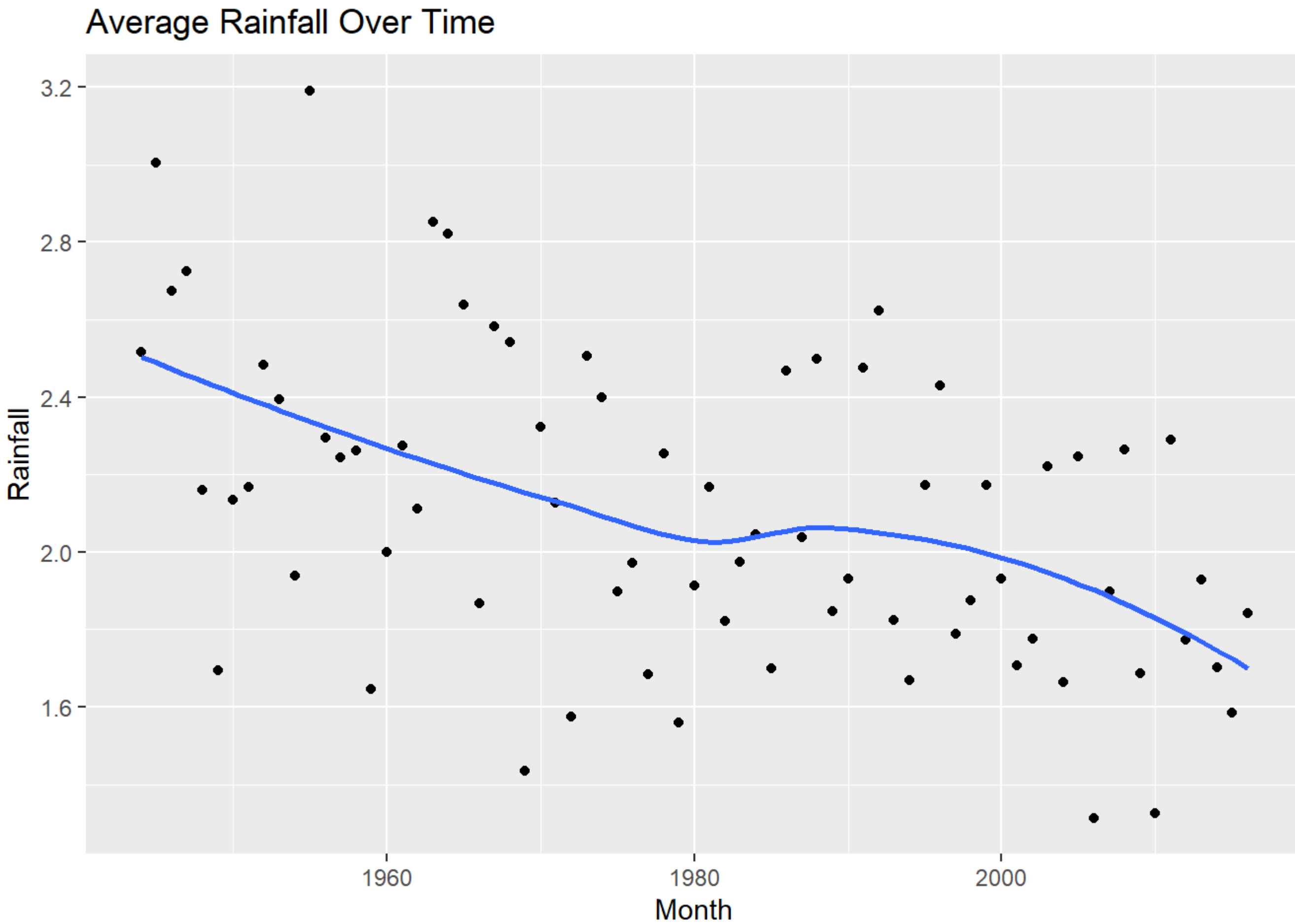
The main difference between these two years in time is the peak months. I had already determined that the winter months averaged the most rainfall so the peak months are between May and August. In 1946, the peak months averaged around 2 millimeters more of rain, however, the rest of the months show little difference between these two years. The graphs look similar, however, the scaling is different as the 2016 graph tops with a value of 4, while the 1946 graph tops with a value of 6. The 2016 graph also has more rain in the dry season. The graph may not necessarily show that rainfall has decreased over the past 70 years because it is a snapshot of two different years, however, it does pose the question of whether rainfall has decreased.

```
# This code creates a different data table with the average rainfall each year from 1946-2016.
avg_rain_year <- au_weather %>%
  select(Year, rainfall_mm) %>%
  group_by(Year) %>%
  mutate(rainfall_mm) %>%
  summarize(avg_rain = mean(rainfall_mm))

ggplot(data = avg_rain_year, mapping = aes(x = Year, y = avg_rain)) + geom_line() + labs(x = "Mo
nth", y = "Rainfall", title = "Average Rainfall Over Time")
```




```
ggplot(data = avg_rain_year, mapping = aes(x = Year, y = avg_rain)) + geom_point() + geom_smooth
(se = FALSE) + labs(x = "Month", y = "Rainfall", title = "Average Rainfall Over Time")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



I created a new set of data that I grouped by year and represented the average rainfall each year. I then visualized the data in two different ways that are simple to understand. The line graph looked as though it was trending downwards, but it was hard to tell with all of the peaks and pits in the graph. In the scatter plot, I inserted a trend line that aided in the visualization of the data. It is evident that rainfall has decreased over time in Western Australia. This trend does not look as though it is going to change in the future which is slightly concerning for Western Australia.

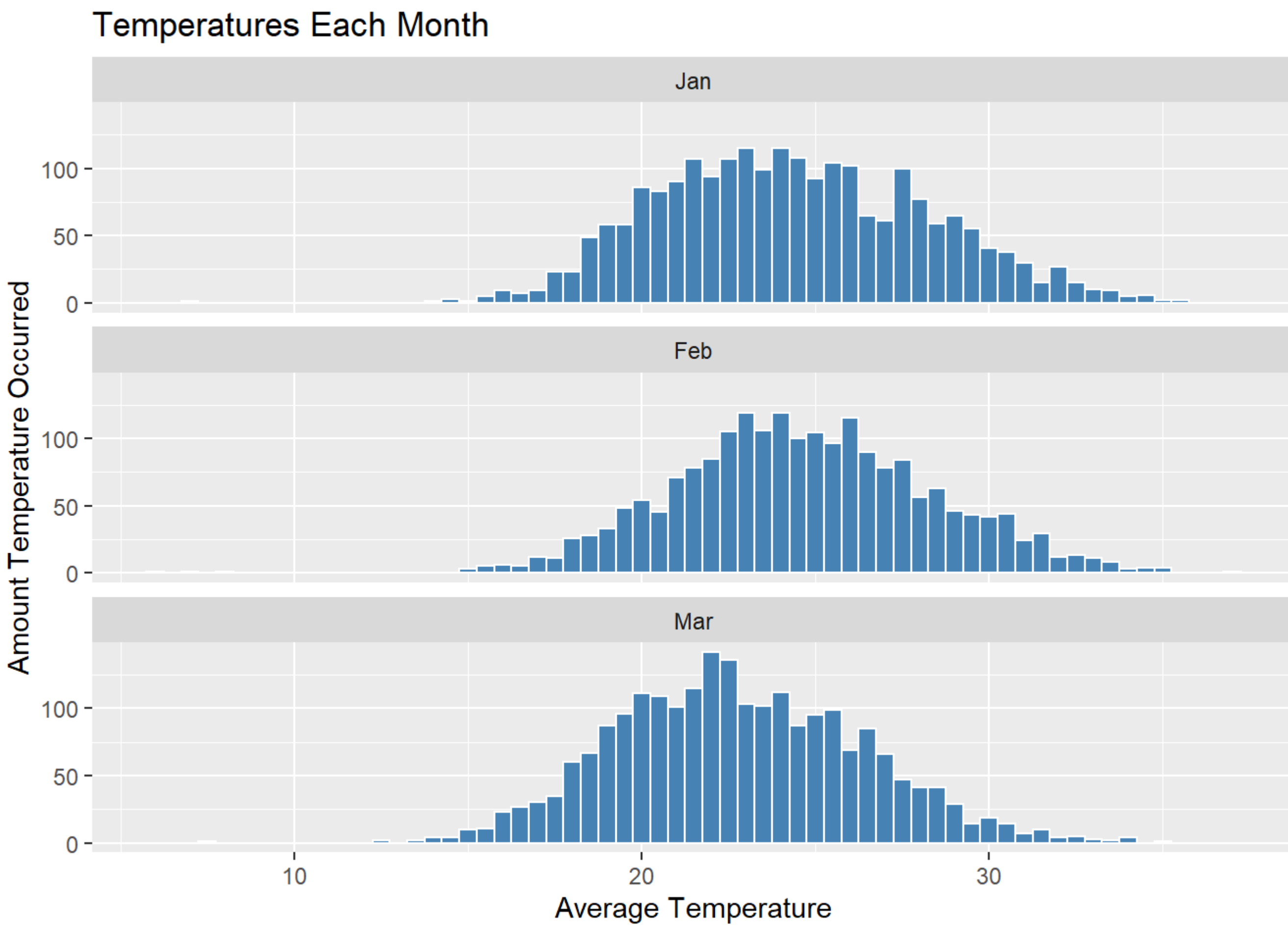
2. Has the average temperature per month each year shifted at all over time?

In order to answer this question, I want to create a base knowledge of when the temperatures are generally the highest and the lowest. I then want to do a similar graph to what I did with rainfall and compare the temperatures in two different years. Finally, I want to see the general trend based on the average temperature each year over time. I created some new data sets in order to make the data easier to visualize.

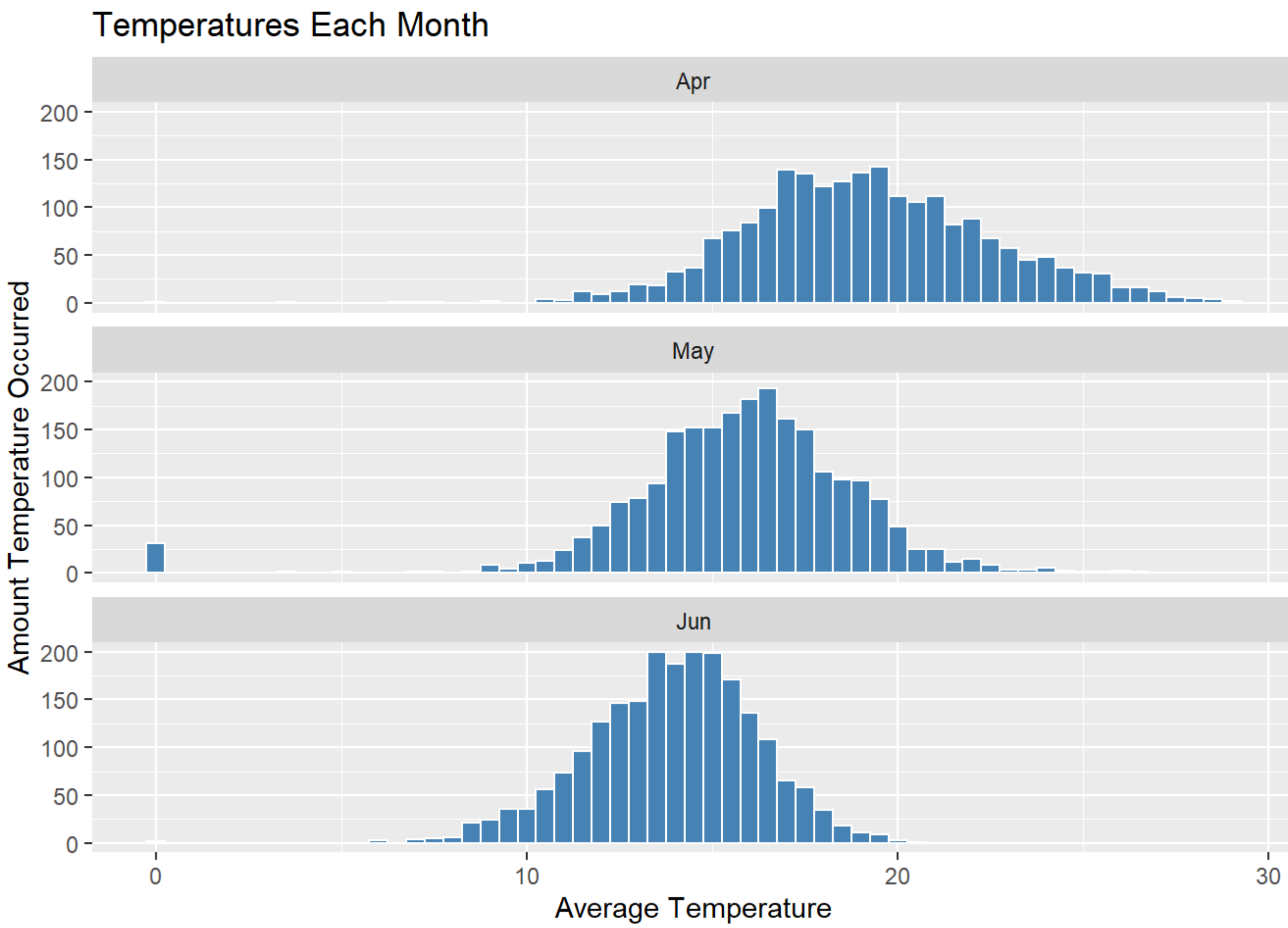
```
# Turn season into cat variable
weather_season <- au_weather %>%
  mutate(daily_avg) %>%
  filter(daily_avg <= 30) %>%
  mutate(rainfall_mm) %>%
  filter(rainfall_mm <= 75) %>%
  mutate(month = cut(Month,
    breaks = c(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12),
    labels = c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov", "De
c")
  )) %>%
  mutate(season = cut(Season,
    breaks = c(0, 1, 2, 3, 4),
    labels = c("Spring", "Summer", "Autumn", "Winter")
  )) %>%
  slice(1:26543)
```

I made this data set so that I could use both month and season as categorical variables. I specifically made it so that I could use the season as the color and cut out any outliers so the scatterplot was easier to see.

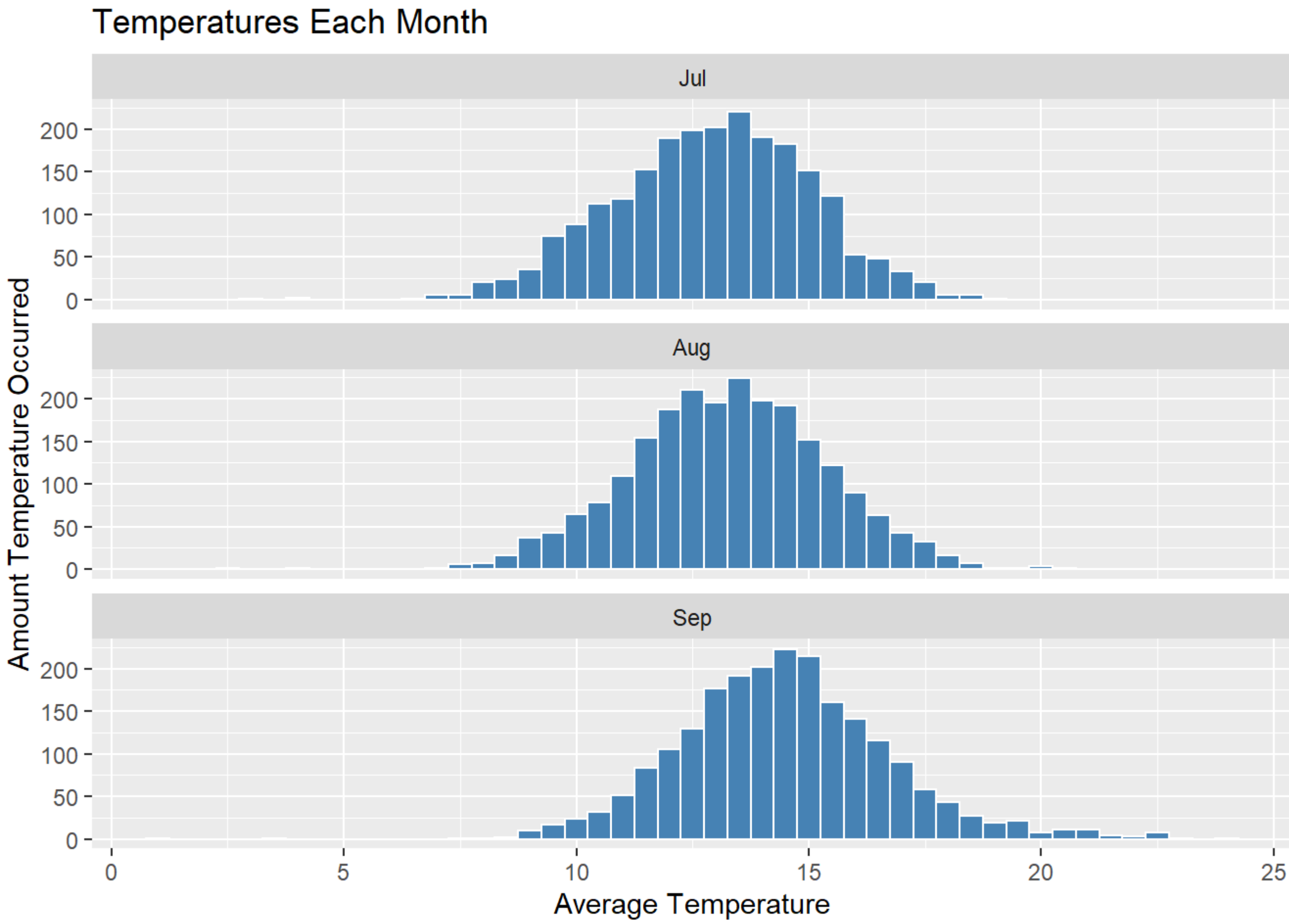
```
Jan_Mar <- date %>%
  select(month, daily_avg) %>%
  filter(month == "Jan" | month == "Feb" | month == "Mar")

Apr_Jun <- date %>%
  select(month, daily_avg) %>%
  filter(month == "Apr" | month == "May" | month == "Jun")

Jul_Sep <- date %>%
  select(month, daily_avg) %>%
  filter(month == "Jul" | month == "Aug" | month == "Sep" )

Oct_Dec <- date %>%
  select(month, daily_avg) %>%
  filter(month == "Oct" | month == "Nov" | month == "Dec")

ggplot(data = Jan_Mar, aes(x = daily_avg)) + geom_histogram(binwidth = 0.5, color = "white", fil
l = "steelblue") + facet_wrap(~ month , ncol = 1) + labs(x = "Average Temperature", y = "Amount
 Temperature Occurred", title = "Temperatures Each Month")
```



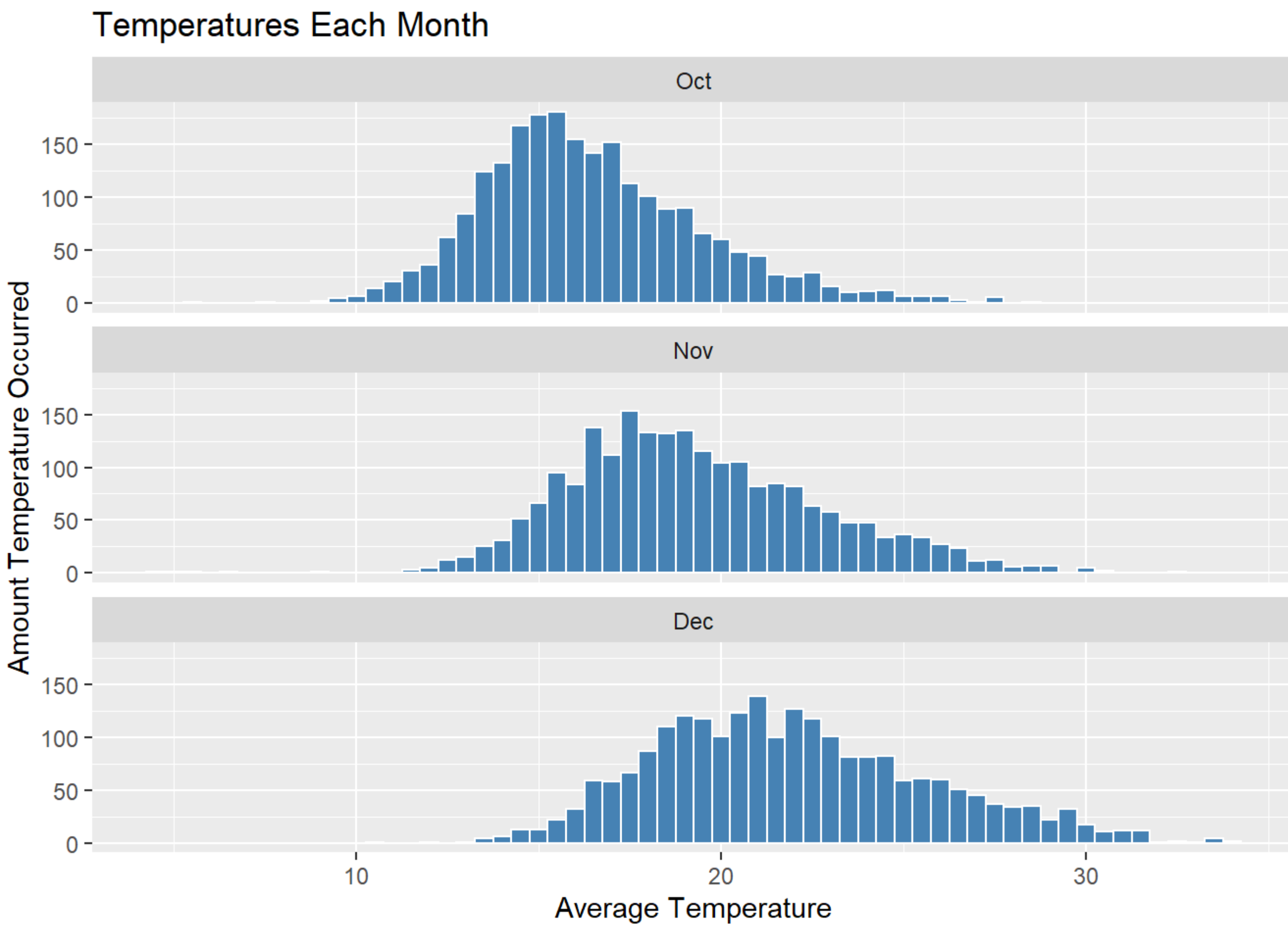

```
ggplot(data = Apr_Jun, aes(x = daily_avg)) + geom_histogram(binwidth = 0.5, color = "white", fil
l = "steelblue") + facet_wrap(~ month , ncol = 1) + labs(x = "Average Temperature", y = "Amount
 Temperature Occurred", title = "Temperatures Each Month")
```



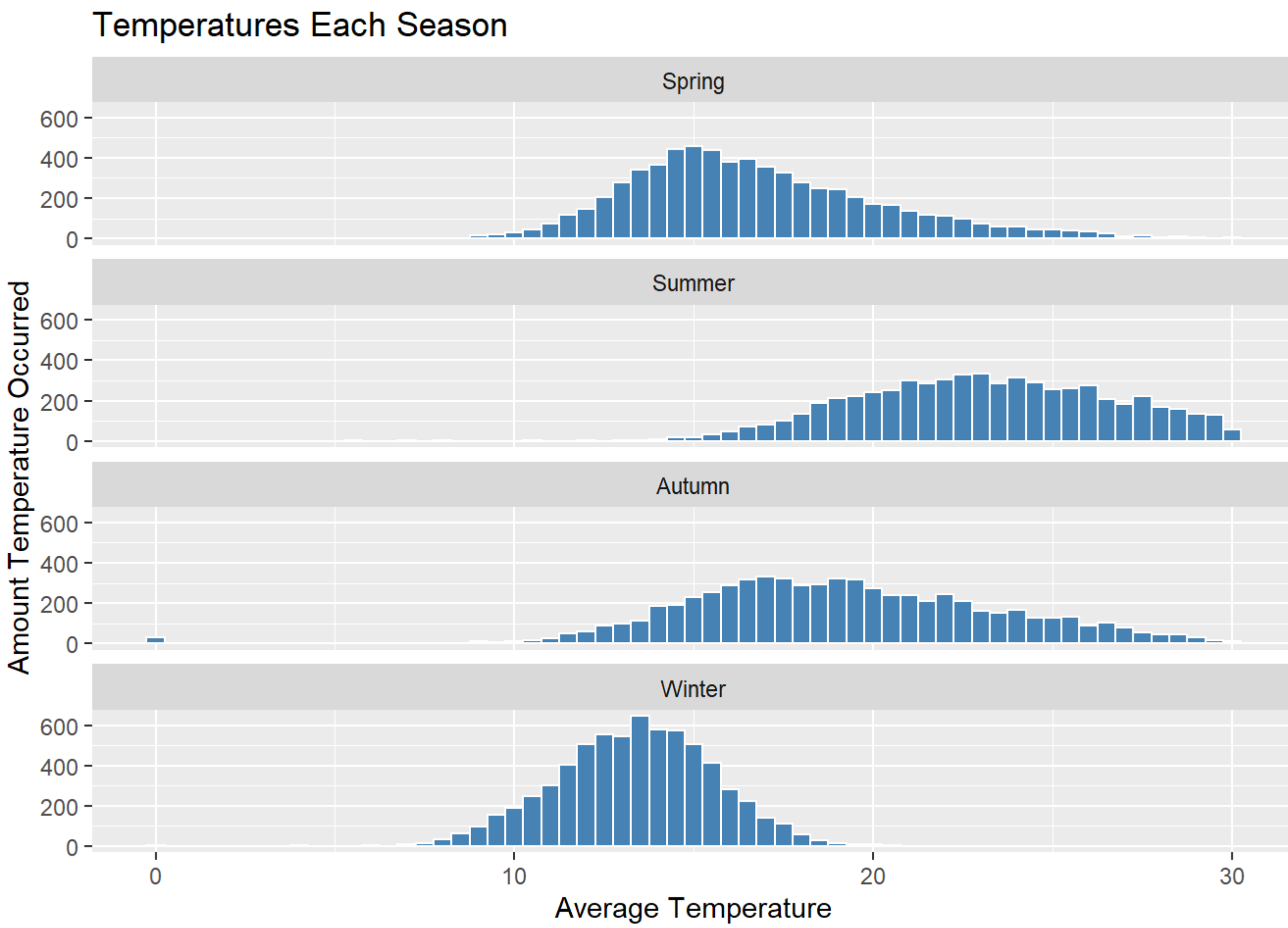

```
ggplot(data = Jul_Sep, aes(x = daily_avg)) + geom_histogram(binwidth = 0.5, color = "white", fil
l = "steelblue") + facet_wrap(~ month , ncol = 1) + labs(x = "Average Temperature", y = "Amount
 Temperature Occurred", title = "Temperatures Each Month")
```




```
ggplot(data = Oct_Dec, aes(x = daily_avg)) + geom_histogram(binwidth = 0.5, color = "white", fill = "steelblue") + facet_wrap(~ month , ncol = 1) + labs(x = "Average Temperature", y = "Amount Temperature Occurred", title = "Temperatures Each Month")
```




```
ggplot(data = weather_season, aes(x = daily_avg)) + geom_histogram(binwidth = 0.5, color = "white", fill = "steelblue") + facet_wrap(~ season , ncol = 1) + labs(x = "Average Temperature", y = "Amount Temperature Occurred", title = "Temperatures Each Season")
```



I created these data sets in order to break apart the average temperatures for each month. I also created one for the seasons. Each histogram shows the counts for the temperatures each month, and each one shows the counts of temperatures for each season. It is easy to tell the distributions of temperatures in order to determine the approximate averages of temperatures. December to March have the highest temperatures overall. May to August have the lowest temperatures overall. This is because the seasons are opposite of North America due to Australia being in the southern hemisphere.

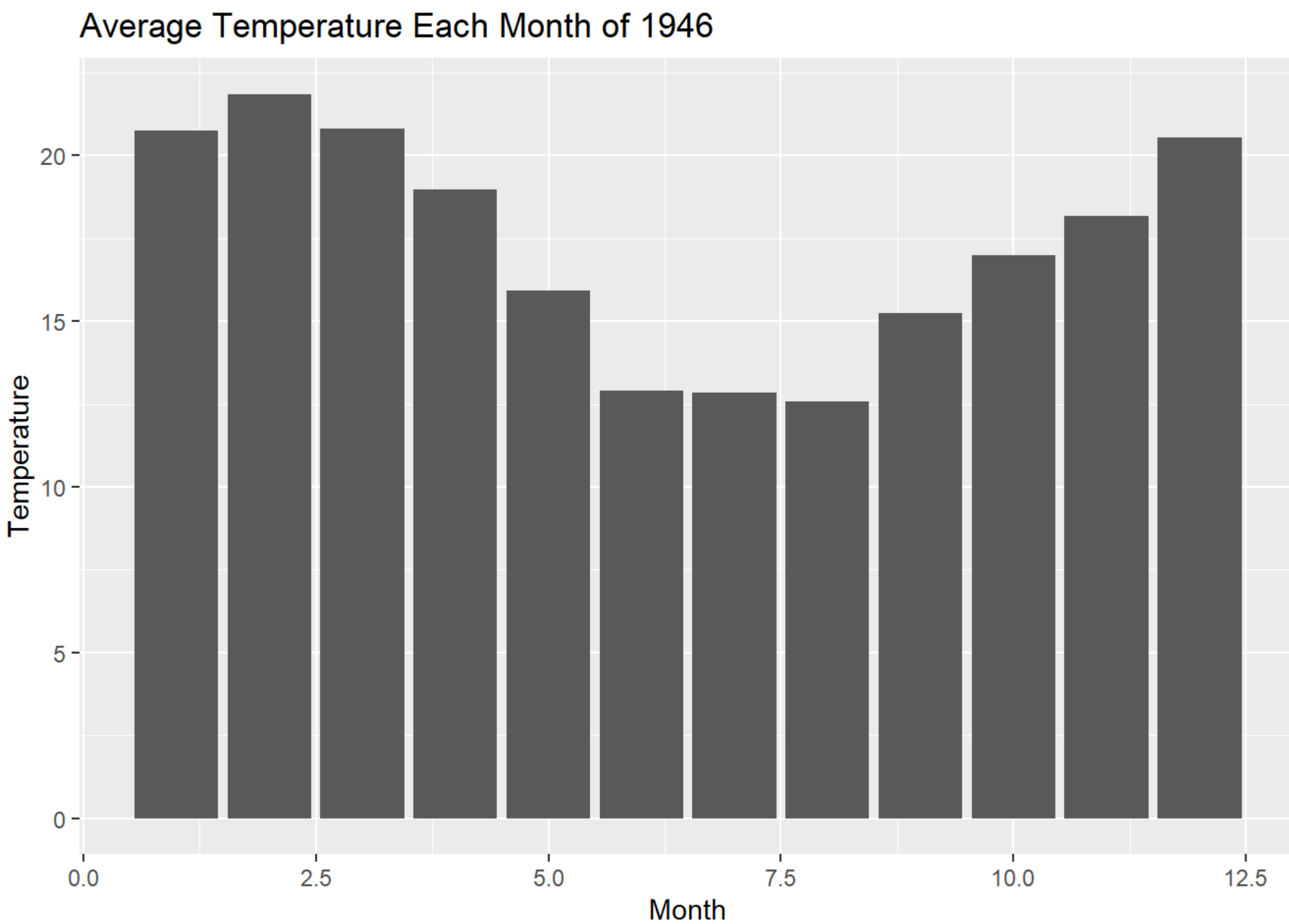
```
# These data sets were created in order to average the temperature each month of the years 1946
 and 2016
avg_temp_46 <- au_weather %>%
  group_by(Year, Month) %>%
  filter(Year == 1946) %>%
  mutate(daily_avg) %>%
  summarize(avg_temp = mean(daily_avg)) %>%
  mutate(date = make_datetime(Year, Month))
```

```
## `summarise()` has grouped output by 'Year'. You can override using the `.groups` argument.
```

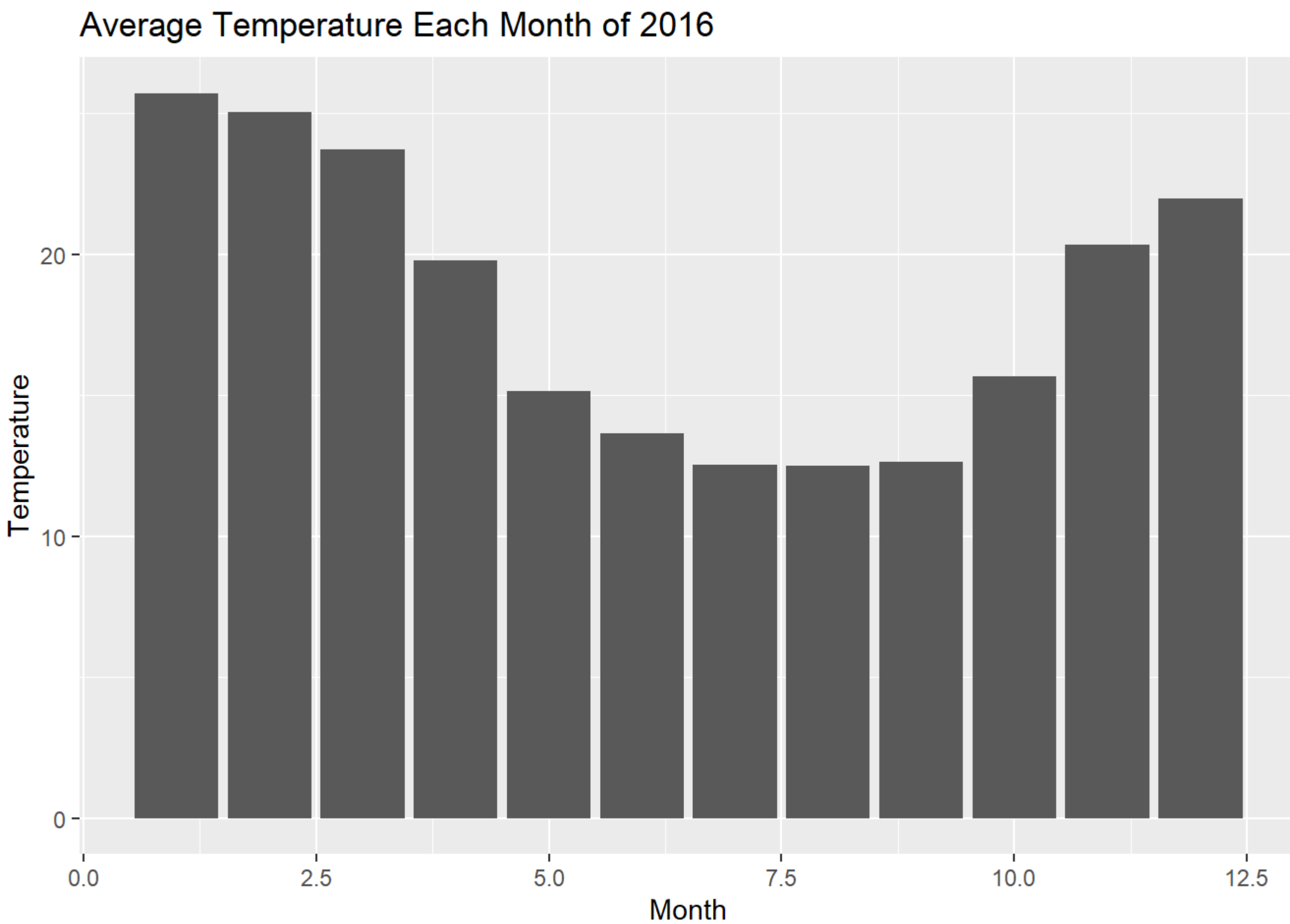
```
avg_temp_16 <- au_weather %>%
  group_by(Year, Month) %>%
  filter(Year == 2016) %>%
  mutate(daily_avg) %>%
  summarize(avg_temp = mean(daily_avg))
```

```
## `summarise()` has grouped output by 'Year'. You can override using the `.groups` argument.
```

```
ggplot(data = avg_temp_46, mapping = aes(x = Month, y = avg_temp)) + geom_col() + labs(x = "Mont
h", y = "Temperature", title = "Average Temperature Each Month of 1946")
```

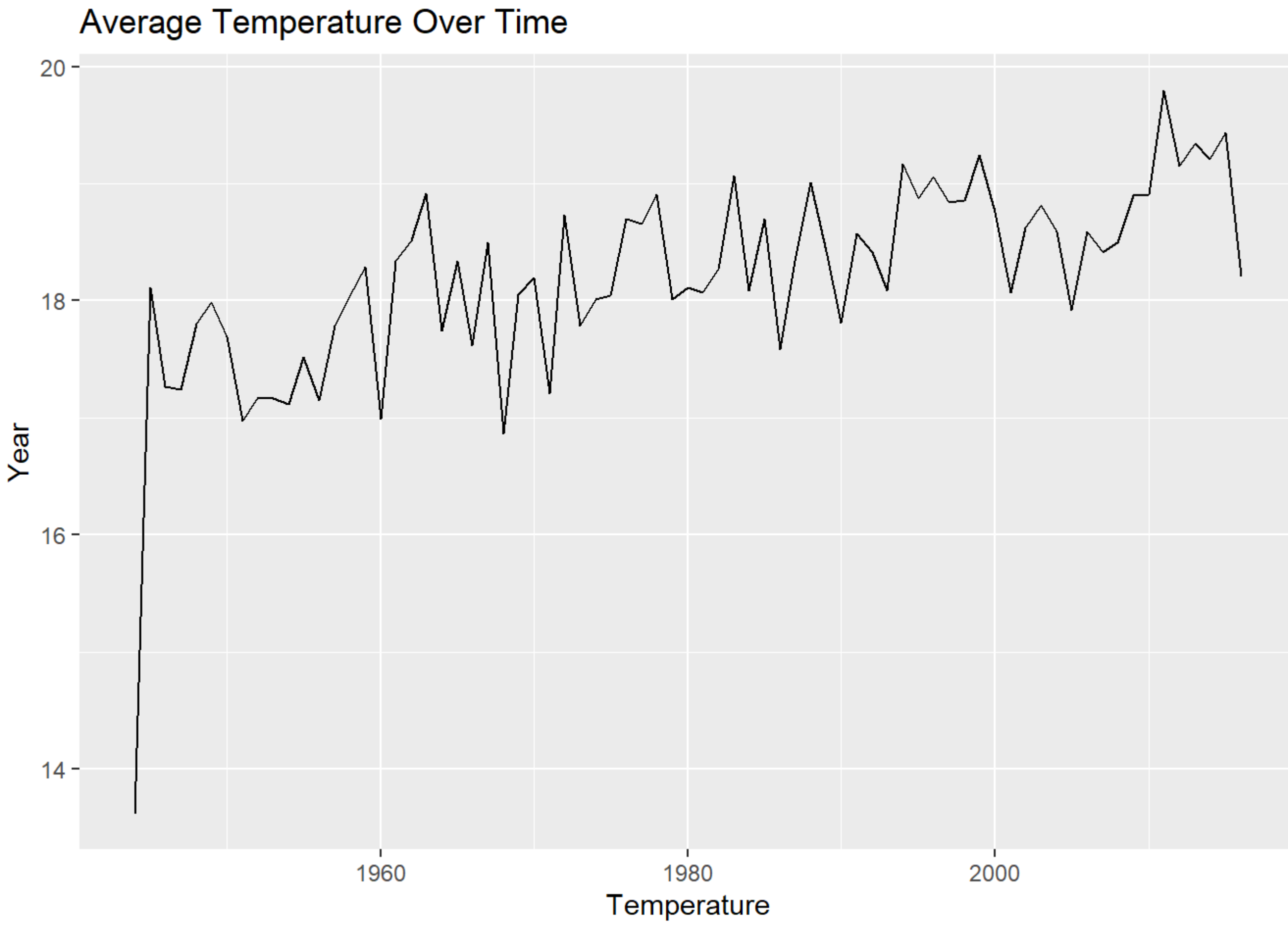



```
ggplot(data = avg_temp_16, mapping = aes(x = Month, y = avg_temp)) + geom_col() + labs(x = "Mont
h", y = "Temperature", title = "Average Temperature Each Month of 2016")
```



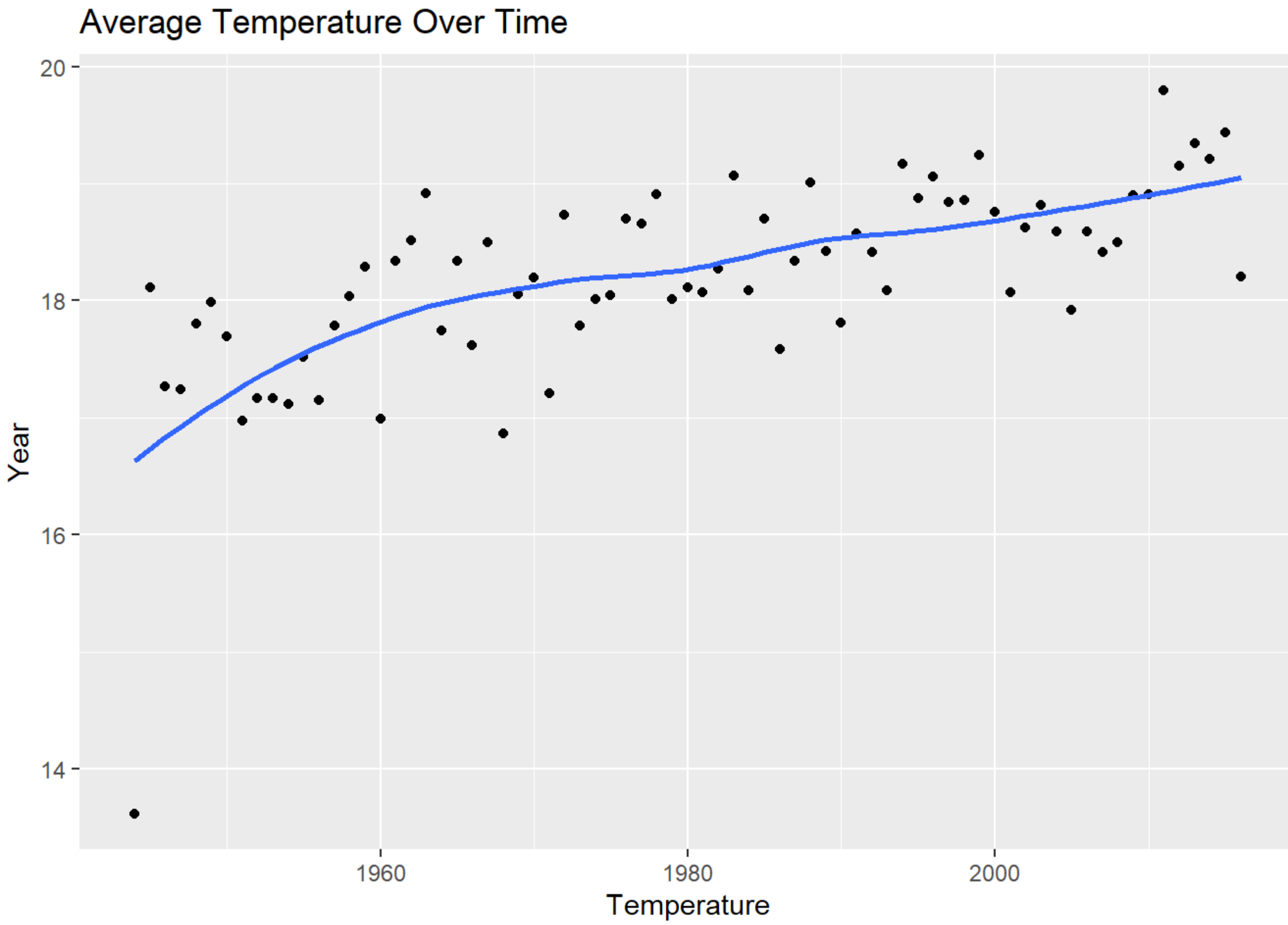
These column graphs represent the average temperature for each month in 1946 and each month in 2016. Upon looking at these graphs, the general trend of increasing and decreasing is the same, however, the scale is slightly different. In 2016, the temperatures are slightly higher, especially in the peak months than in 1946. This could possibly indicate an increase in overall temperature from 1946 to 2016. It is hard to tell if there was actually a shift in temperature over the last 70 years because these two graphs only show a snapshot in time of those two years.

```
# This data set creates a variable called "avg_temp" which is the average temperature each year
 from 1944 to 2016
avg_temp_year <- au_weather %>%
  select(Year, daily_avg) %>%
  group_by(Year) %>%
  mutate(daily_avg) %>%
  summarize(avg_temp = mean(daily_avg))

ggplot(data = avg_temp_year, mapping = aes(x = Year, y = avg_temp)) + geom_line()+ labs(x = "Tem
perature", y = "Year", title = "Average Temperature Over Time")
```




```
ggplot(data = avg_temp_year, mapping = aes(x = Year, y = avg_temp)) + geom_point() + geom_smooth
(se = FALSE) + labs(x = "Temperature", y = "Year", title = "Average Temperature Over Time")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



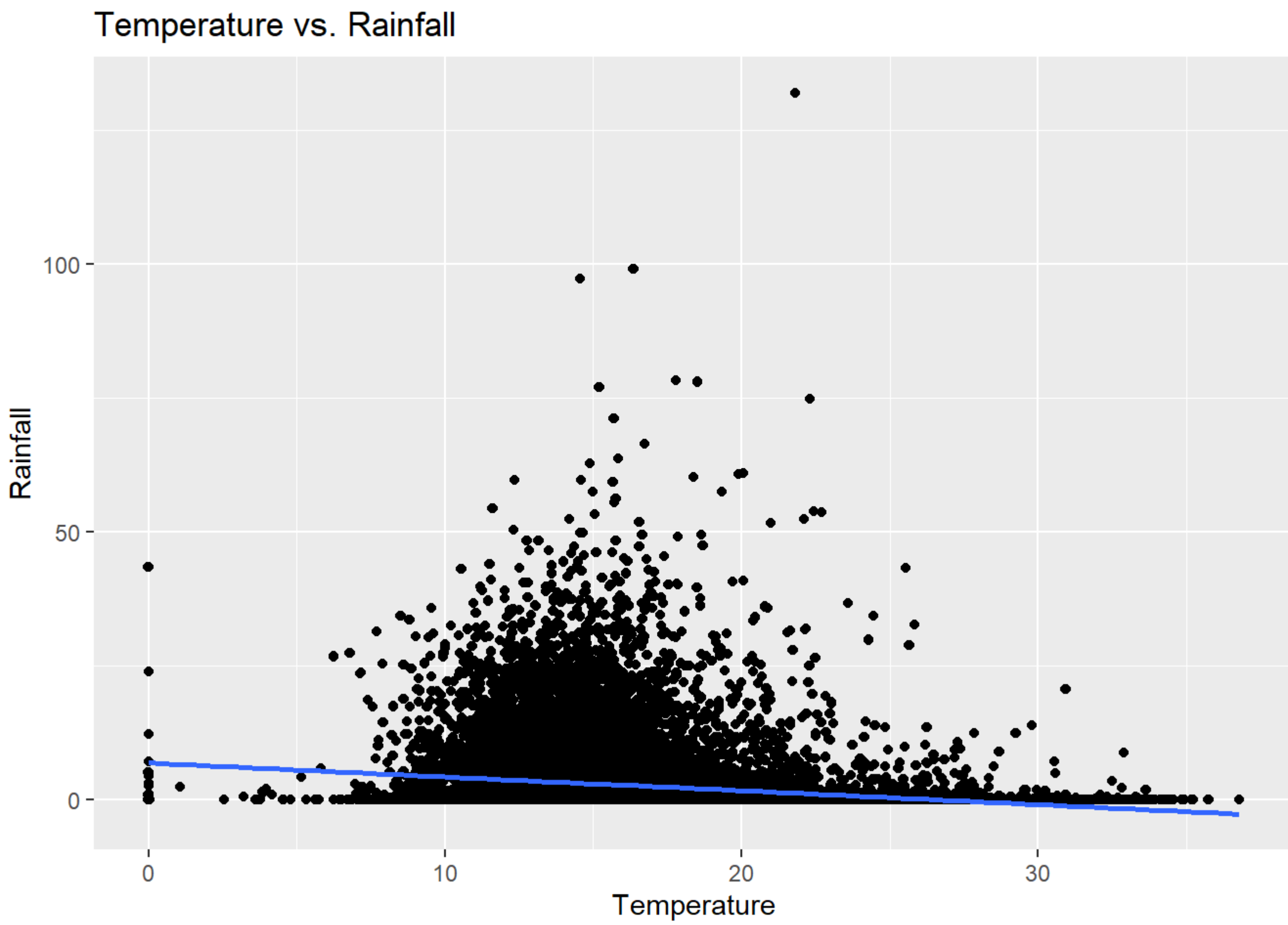
These two graphs indicate a slight increase in temperature from 1944 to 2016. The reason that the first point is so low is because in 1944, the temperature was not recorded for half of the year. It is particularly evident in the scatter plot that there is a steady increase in temperature over time which is generally concerning. The temperature can only increase so much before it becomes an issue. The graph shows an increase of around 2.5 degrees Celsius (about 6 degrees Fahrenheit). Even though one could argue that there are many dips over time, the general trend is still increasing.

3. Are the temperature and rainfall correlated in any way?

In order to answer this final question, I first need to create a scatter plot with the basis of correlation between temperature and rainfall. I then created two different graphs based on the seasons to see whether the seasons have an effect on the temperature and rainfall.

```
ggplot(data = au_weather, mapping = aes(x = daily_avg, y = rainfall_mm)) + geom_point() + geom_jitter() + geom_smooth(method = "lm", se = FALSE)  + labs(x = "Temperature", y = "Rainfall", title = "Temperature vs. Rainfall")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

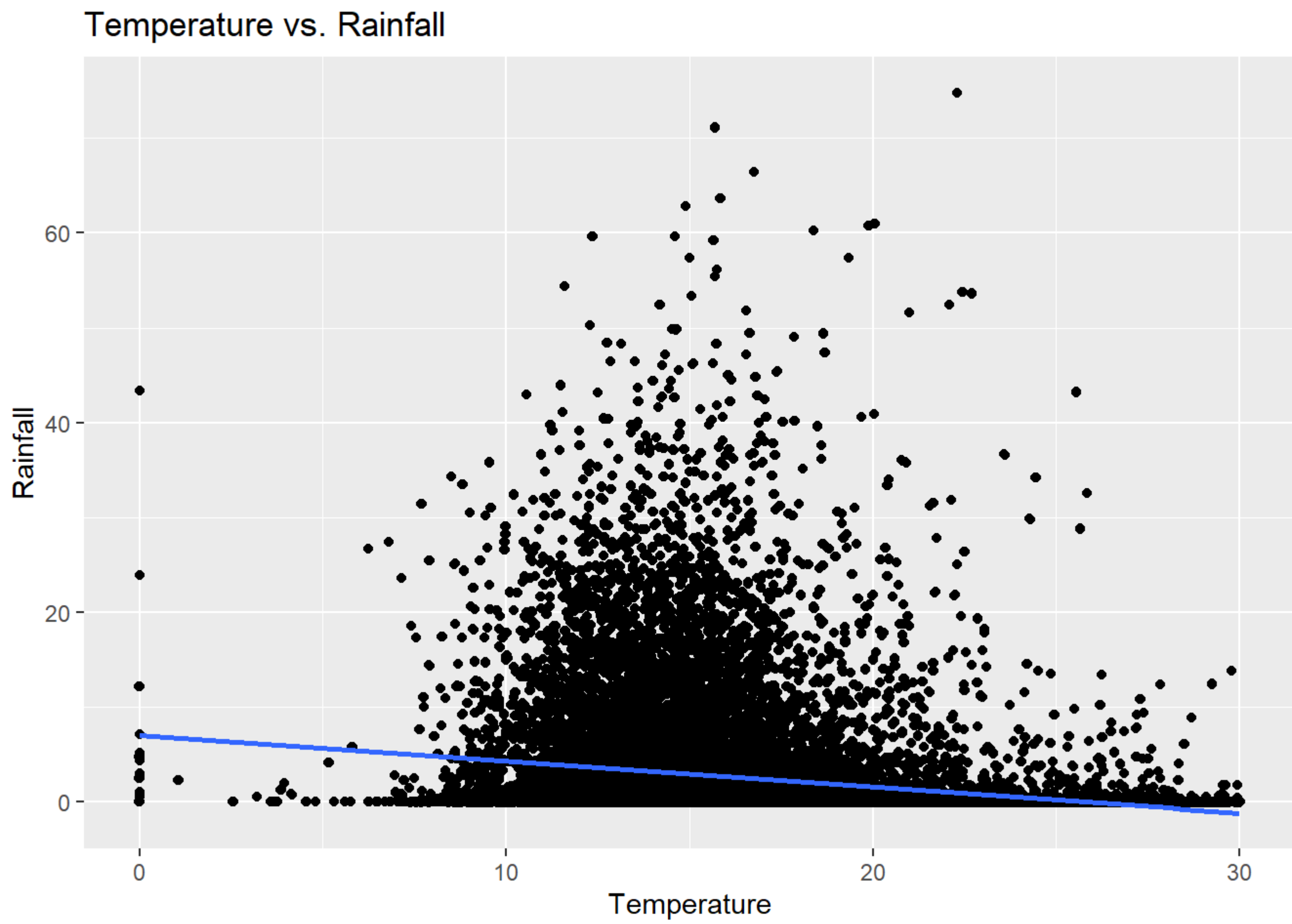



```
no_outliers <- au_weather %>%
  select(daily_avg, rainfall_mm) %>%
  mutate(daily_avg) %>%
  filter(daily_avg <= 30) %>%
  mutate(rainfall_mm) %>%
  filter(rainfall_mm <= 75)
```

After creating the previous scatter plot, I decided that there were a few outliers that were skewing the data making it more difficult to visualize. Using the code above, I created a filtered data set with none of the outliers I perceived as skewing the data.

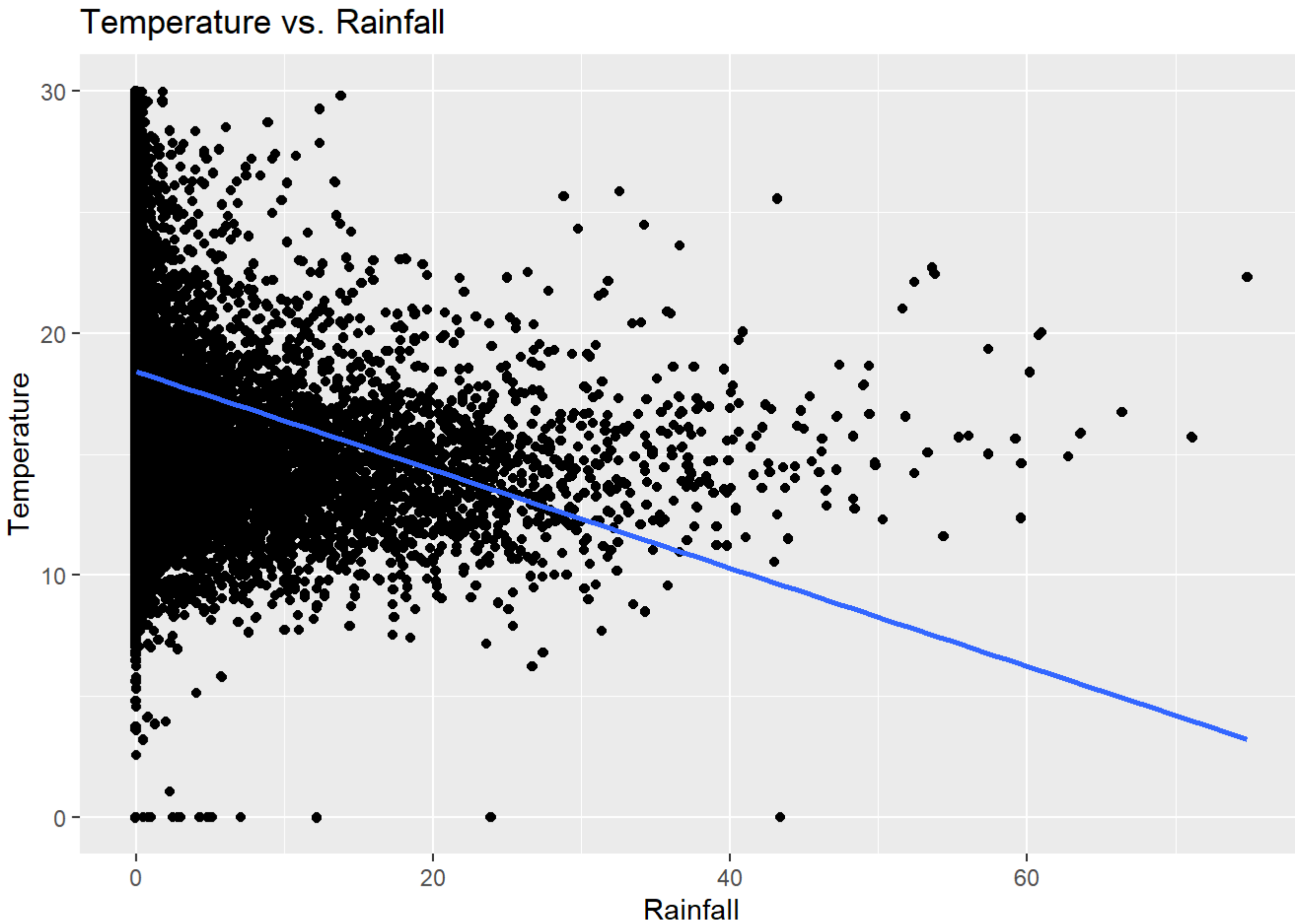
```
ggplot(data = no_outliers, mapping = aes(x = daily_avg, y = rainfall_mm)) + geom_point() + geom_
jitter() + geom_smooth(method = "lm", se = FALSE)  + labs(x = "Temperature", y = "Rainfall", tit
le = "Temperature vs. Rainfall")
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
ggplot(data = no_outliers, mapping = aes(y = daily_avg, x = rainfall_mm)) + geom_point() + geom_
jitter() + geom_smooth(method = "lm", se = FALSE)  + labs(y = "Temperature", x = "Rainfall", tit
le = "Temperature vs. Rainfall")
```

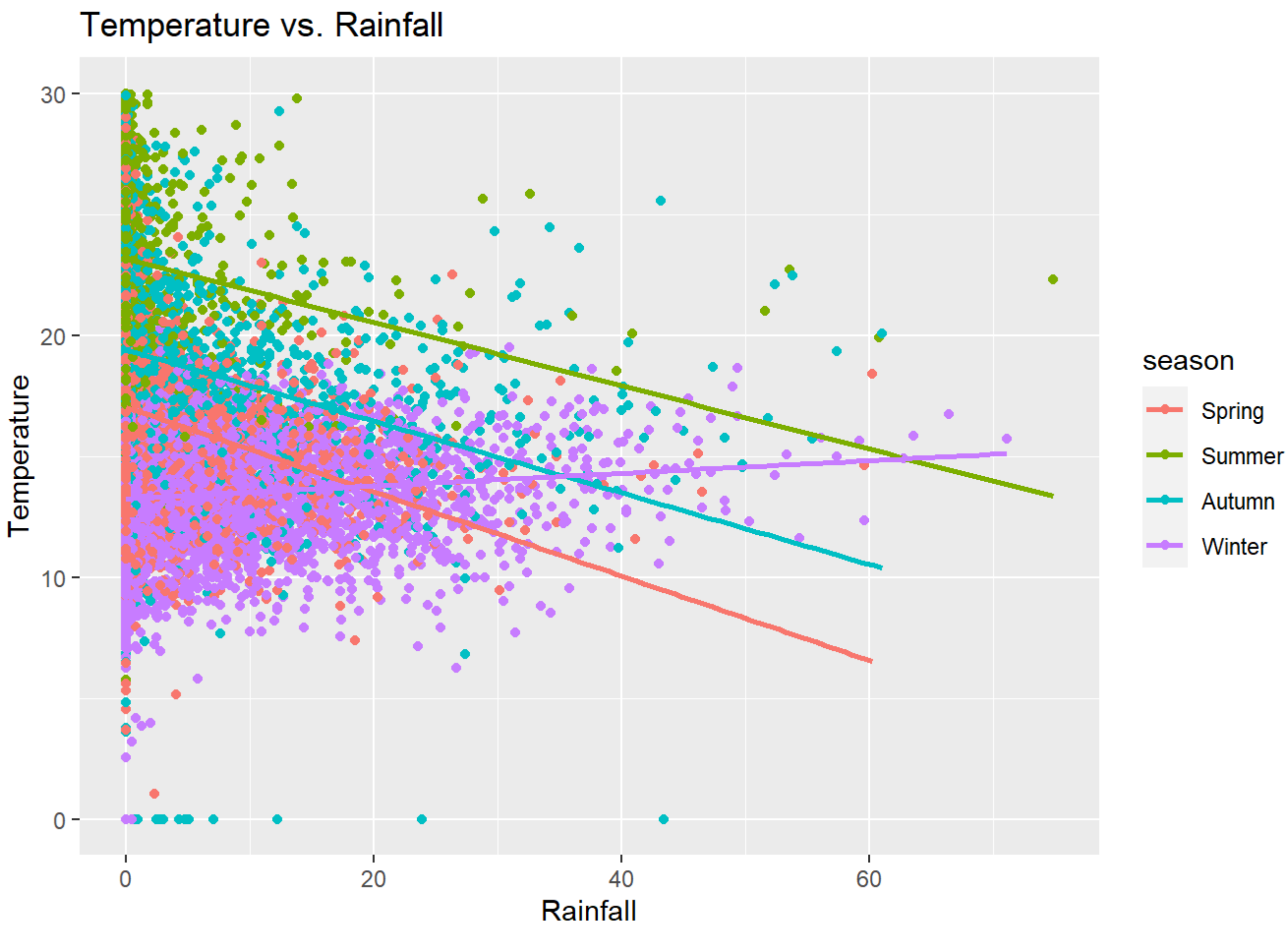
```
## `geom_smooth()` using formula 'y ~ x'
```



This is a graph of the correlation between temperature and rainfall in Western Australia. There seems to be a trend in the graph of temperature increasing and rainfall decreasing. Looking at both of the scatterplots above, it can be seen that at a lower temperature (temp < 15), the amount of precipitation seems to be more than temperatures that are greater (temp > 15). The trend line seems to slope downward after the temperature of 15.

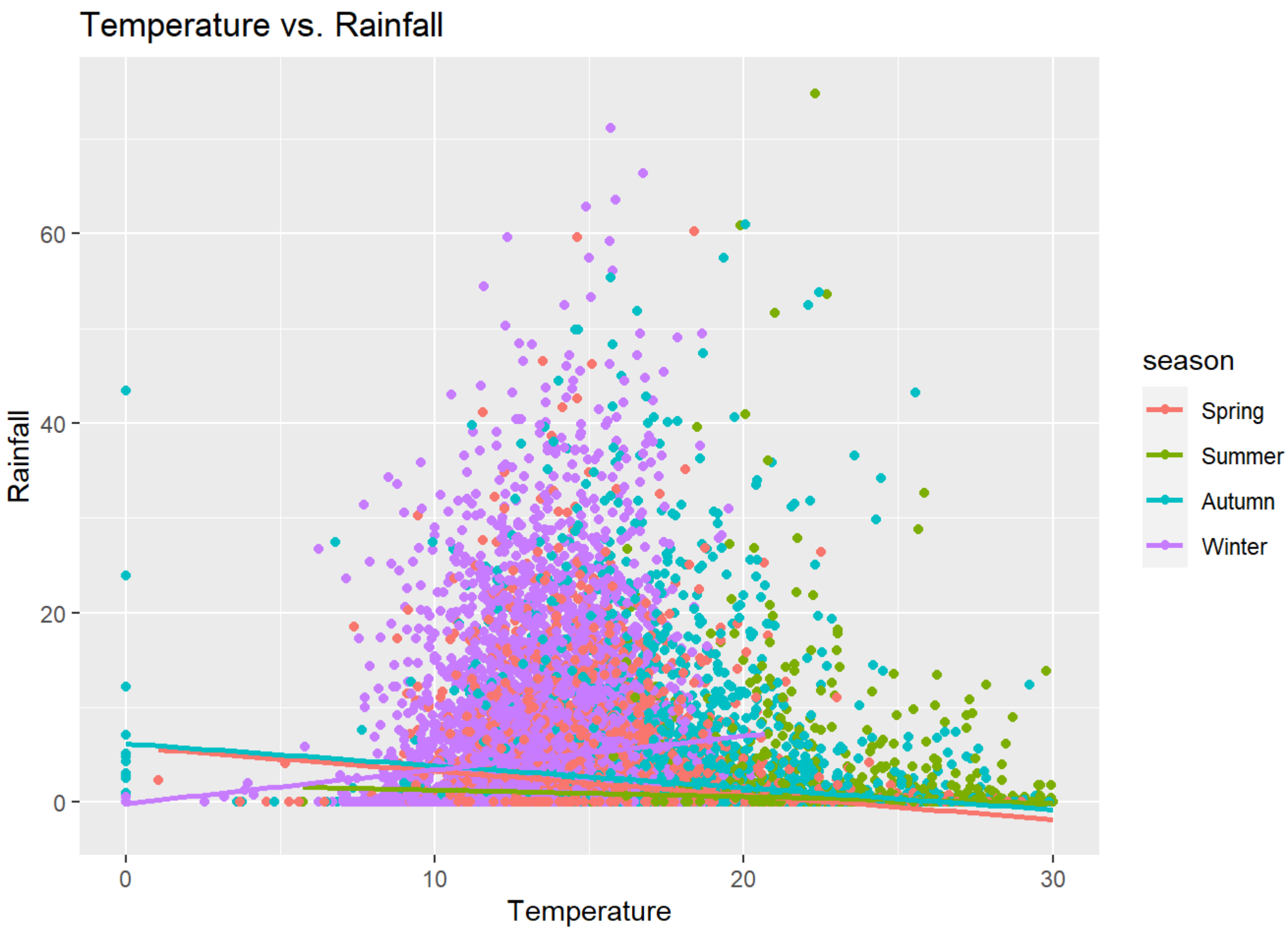
```
ggplot(data = weather_season, mapping = aes(y = daily_avg, x = rainfall_mm, color = season)) + g
eom_point() + geom_smooth(method = "lm", se = FALSE)  + labs(y = "Temperature", x = "Rainfall",
 title = "Temperature vs. Rainfall")
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
ggplot(data = weather_season, mapping = aes(x = daily_avg, y = rainfall_mm, color = season)) + g
eom_point() + geom_smooth(method = "lm", se = FALSE)  + labs(x = "Temperature", y = "Rainfall",
 title = "Temperature vs. Rainfall")
```

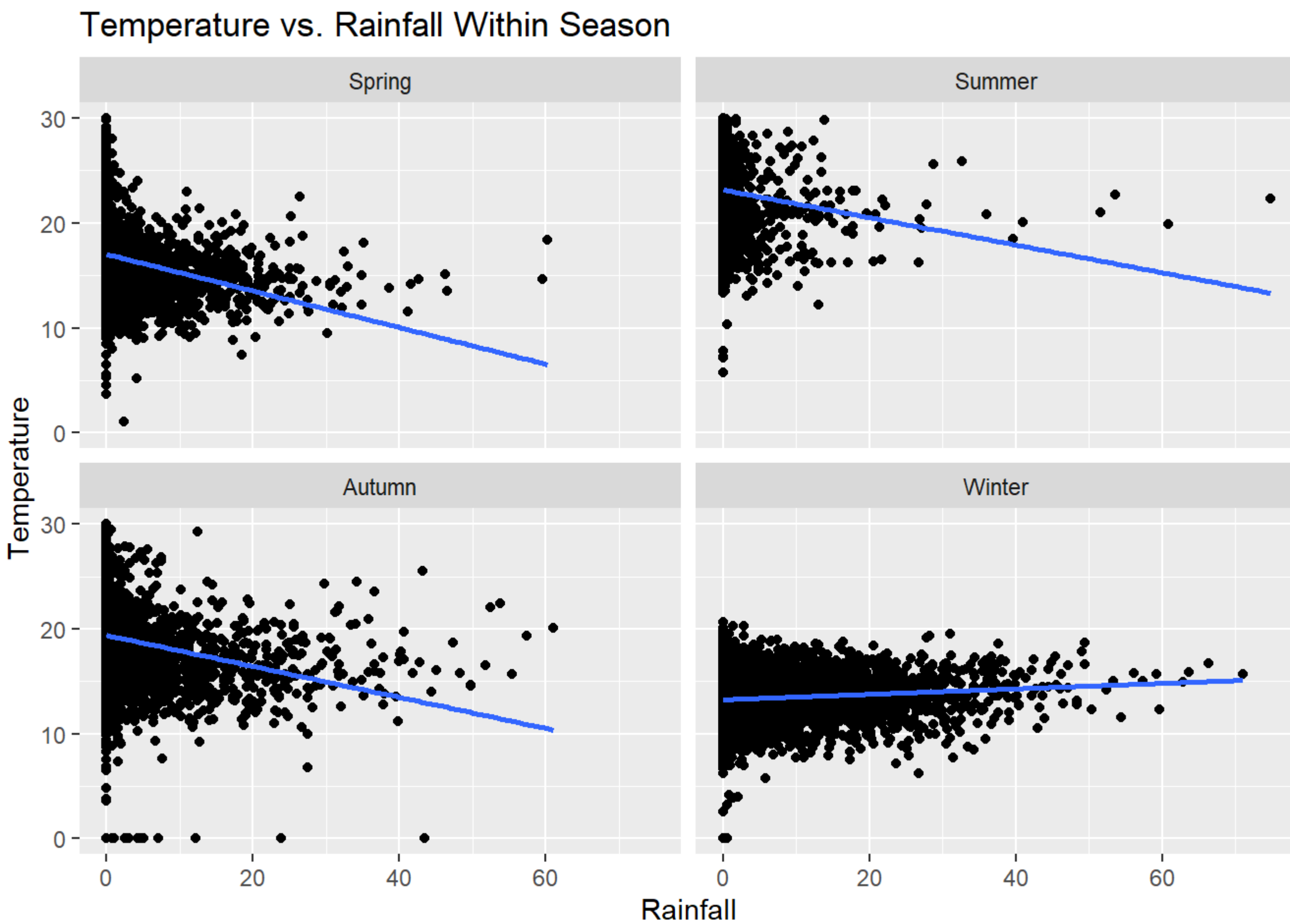
```
## `geom_smooth()` using formula 'y ~ x'
```



The graph above is a scatter plot of rainfall vs temperature with a color variable of season. The scatterplot represents the same trend for most seasons showing, as the temperature increases, rainfall decreases. The one season this does not hold true for is winter where the trend is actually reversed. As the temperature increases, rainfall also increases. This could be due to the fact that winter is the rainy season in this part of Australia. The temperature is warm during the winter in Australia so this direct relationship between rainfall and temperature is justified.

```
ggplot(data = weather_season, mapping = aes(y = daily_avg, x = rainfall_mm)) + geom_point() + geom_smooth(method = "lm", se = FALSE) + facet_wrap(~ season)  + labs(x = "Rainfall", y = "Temperature", title = "Temperature vs. Rainfall Within Season")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



These graphs are the same as the above graphs just separated into the different seasons. The seasons of autumn, spring, and summer all have the same correlation of rainfall increasing as the temperature decreases. Winter is the opposite showing, as the rainfall increases, temperature increases.

Data Modeling

1-variable

```
rainfall_model <- lm(rainfall_mm ~ daily_avg, data = au_weather)
get_regression_table(rainfall_model)
```

```
## # A tibble: 2 x 7
##   term      estimate std_error statistic p_value lower_ci upper_ci
##   <chr>        <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
## 1 intercept    6.81      0.127      53.5       0    6.56     7.06
## 2 daily_avg   -0.259     0.007     -38.5       0   -0.272   -0.245
```

Equation: $\widehat{rainfall}$ = 6.814 + -0.259(daily_avg)

Based on this equation, it is seen that the y-intercept of the graph would be 6.814. This is consistent with the graph of temperature vs. rainfall. It is also shown that as the variable “daily_avg” (temperature) increases by one degree Celsius, the average rainfall per day decreases by 0.259 millimeters. This is consistent with the same graph of temperature vs. rainfall. The correlation of -0.259 is also consistent with the correlational value calculated in the summary statistics between rainfall and temperature. Overall, this model can attest to the fact that as the temperature increases, the amount of rainfall per day does decrease.

2-variable

```
temp_model_interaction <- lm(daily_avg ~ rainfall_mm * season, data = weather_season)
get_regression_table(temp_model_interaction)
```

```
## # A tibble: 8 x 7
##   term                        estimate std_error statistic p_value lower_ci upper_ci
##   <chr>                          <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
## 1 intercept                     17.0       0.045    383.     0       16.9     17.1
## 2 rainfall_mm                   -0.175     0.01     -17.8    0       -0.194   -0.155
## 3 season: Summer                 6.14      0.063     98.1    0        6.01     6.26
## 4 season: Autumn                 2.38      0.063     38.0    0        2.26     2.50
## 5 season: Winter                -3.78      0.065    -57.8    0       -3.91    -3.65
## 6 rainfall_mm:seasonSumm~        0.043     0.02       2.19   0.029    0.005    0.082
## 7 rainfall_mm:seasonAutu~        0.027     0.013      2.12   0.034    0.002    0.051
## 8 rainfall_mm:seasonWint~        0.201     0.011     18.2    0        0.179    0.222
```

One equation: $\widehat{temp}$ = 17.031 - 0.175(rainfall) + 6.136 * $1_{summer}(x)$ + 2.378 * $1_{autumn}(x)$ - 3.778 * $1_{winter}(x)$ + 0.043(rainfall) * $1_{summer}(x)$ + 0.027(rainfall) * $1_{autumn}(x)$ + 0.201(rainfall) * $1_{winter}(x)$

Based upon the table of values above, we can create four separate equations for each season listed below.

$\widehat{temp}_{summer}$ = 23.167 – 0.132 * (rainfall)

$\widehat{temp}_{autumn}$ = 19.049 – 0.148 * (rainfall)

$\widehat{temp}_{winter}$ = 13.253 + 0.026 * (rainfall)

$\widehat{temp}_{spring}$ = 17.031 – 0.175 * (rainfall)

These four equations represent how rainfall, temperature, and season are all related to one another. The first equation represents a prediction of temperature in the summertime (December - February). This equation shows a y-intercept of 23.167 and that seems to be consistent with the graph of rainfall and temperature in relation to the season. The slope seems to be a negative 0.132 meaning that in the summer, as the temperature increases by one degree Celsius, rainfall decreases by 0.132 millimeters. This also seems to be consistent with the idea that as the temperature increases, rainfall decreases.

The second equation represents a prediction of the temperature in the autumn (March-May) based on the rainfall. This equation has a y-intercept of 19.049 and this seems to be accurate in relation to the graph of rainfall and temperature distributed by season. The slope of the line is a negative 0.148 showing a negative correlation between rainfall and temperature in the autumn. This would mean that as the temperature increases by 1 degree Celsius, rainfall decreases by 0.148 millimeters. This is consistent with the idea that as the temperature increases, rainfall decreases.

The third equation represents a prediction of the temperature in the winter (June - August) based on the amount of rainfall. This equation gives a y-intercept of 13.253 and that is consistent with the graph above of temperature vs rainfall distributed by season. The y-intercept of this equation is slightly smaller than that of the other equations. This change is also seen in the slope which is a positive 0.026. This means that as the temperature increases by one degree Celsius in the winter, the rainfall also increases by 0.026 millimeters. This shows a positive correlation between rainfall and temperature, so as the temperature increases, rainfall increases. Although the slop is also small, this relationship is still the opposite of what is seen among the other three seasons. I believe this change in the relationship between rainfall and temperature may be due to the fact that the winter is the rainy season in Australia. The temperature is increasing throughout the winter, but the rainfall is also increasing throughout the winter. This data is not consistent with the graph of temperature and rainfall showing there can be a different correlation.

The fourth and final equation is a prediction of the temperature in the spring (September - November) based on the amount of rain. This equation is the baseline for the cumulative equation meaning all other equations are based on this one. The equation gives a y-intercept of 17.031 and this is consistent with the graph of rainfall vs temperature distributed by season. The equation also gives a slope of a negative 0.175. This would mean that as the temperature increases in the spring by one degree Celsius, the rainfall decreases by one millimeter. This represents a negative correlation between temperature and rainfall, or as the temperature increases, rainfall decreases. This idea is consistent with both the seasons of winter and autumn. It is also consistent with the idea that as the temperature increases rainfall decreases.