# Homework 6

Marley Akonnor

8/24/2021

Code line 10 was found on StackOverflow.

```
options(knitr.duplicate.label = "allow")
```

install ggplot

```
library(ggplot2)
```

Step 1: Load the data

```
qualityair <- airquality
```

Step 2: Clean the data and remove NAs

```
sum(is.na(qualityair))
```

```
## [1] 44
```

```
qualityair <- na.omit(qualityair)
```

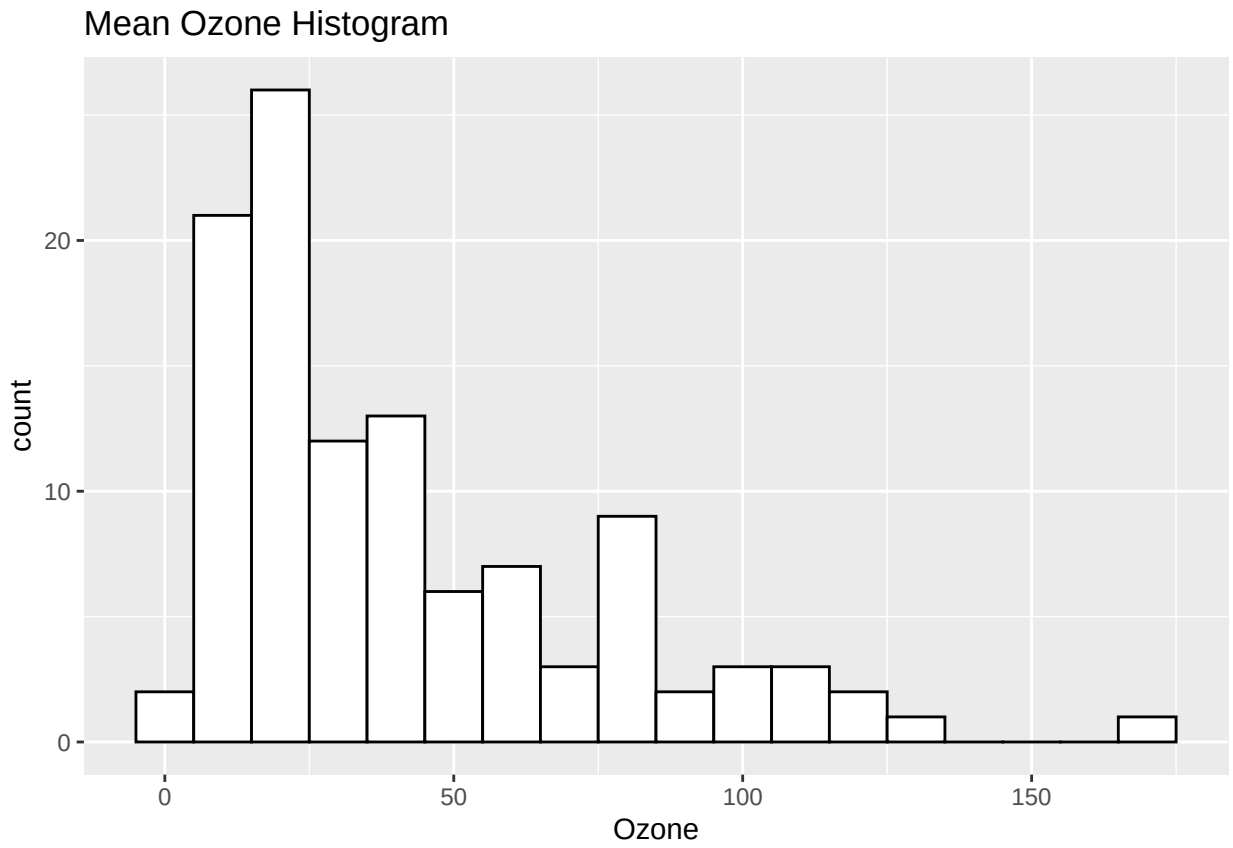
Learn about the dataset

```
?airquality
```

Step 3: Understand the data distribution Create the following visualizations using ggplot:

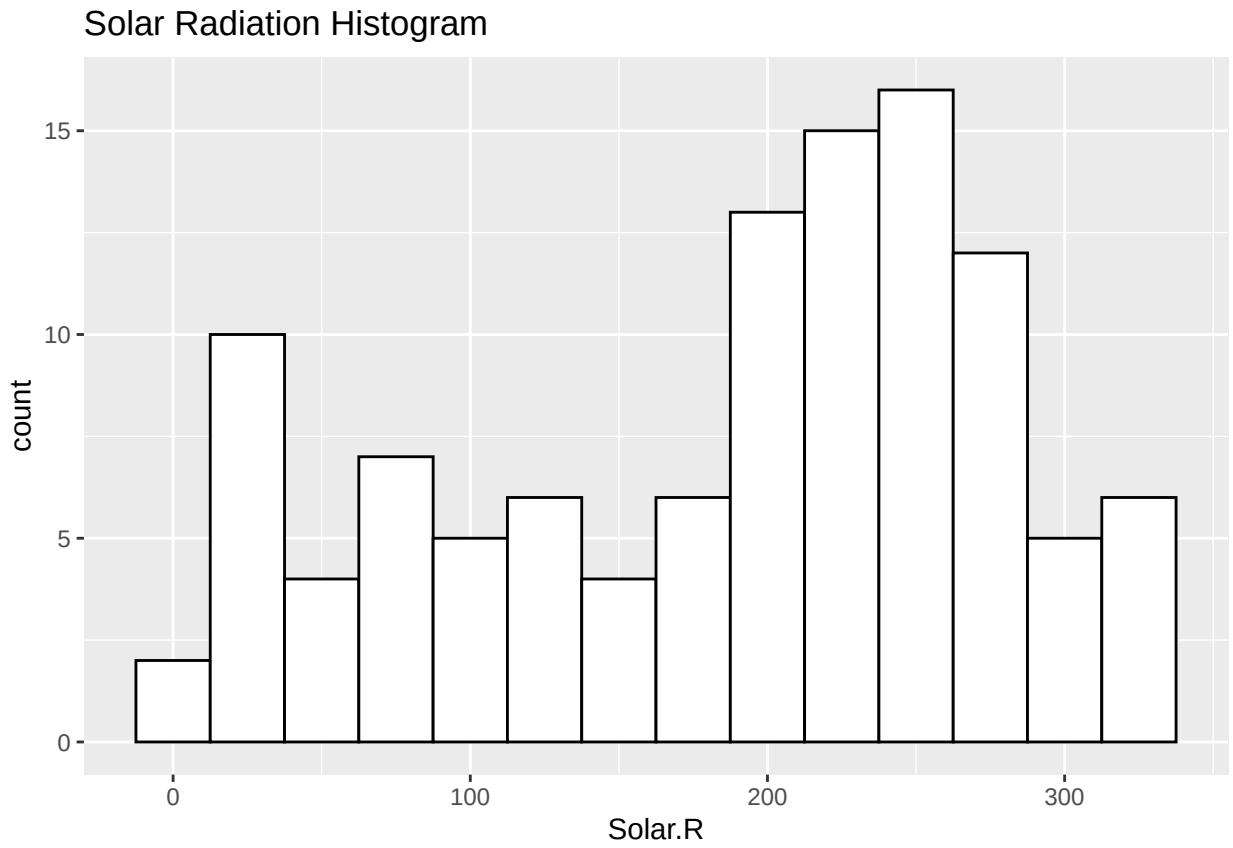
Histograms for each of the variables 3a: Ozone

```
qaHistOz <- ggplot(qualityair, aes(x=Ozone))
qaHistOz <- qaHistOz + geom_histogram(binwidth = 10, color = "black", fill = "white")
qaHistOz <- qaHistOz + ggtitle("Mean Ozone Histogram")
qaHistOz
```



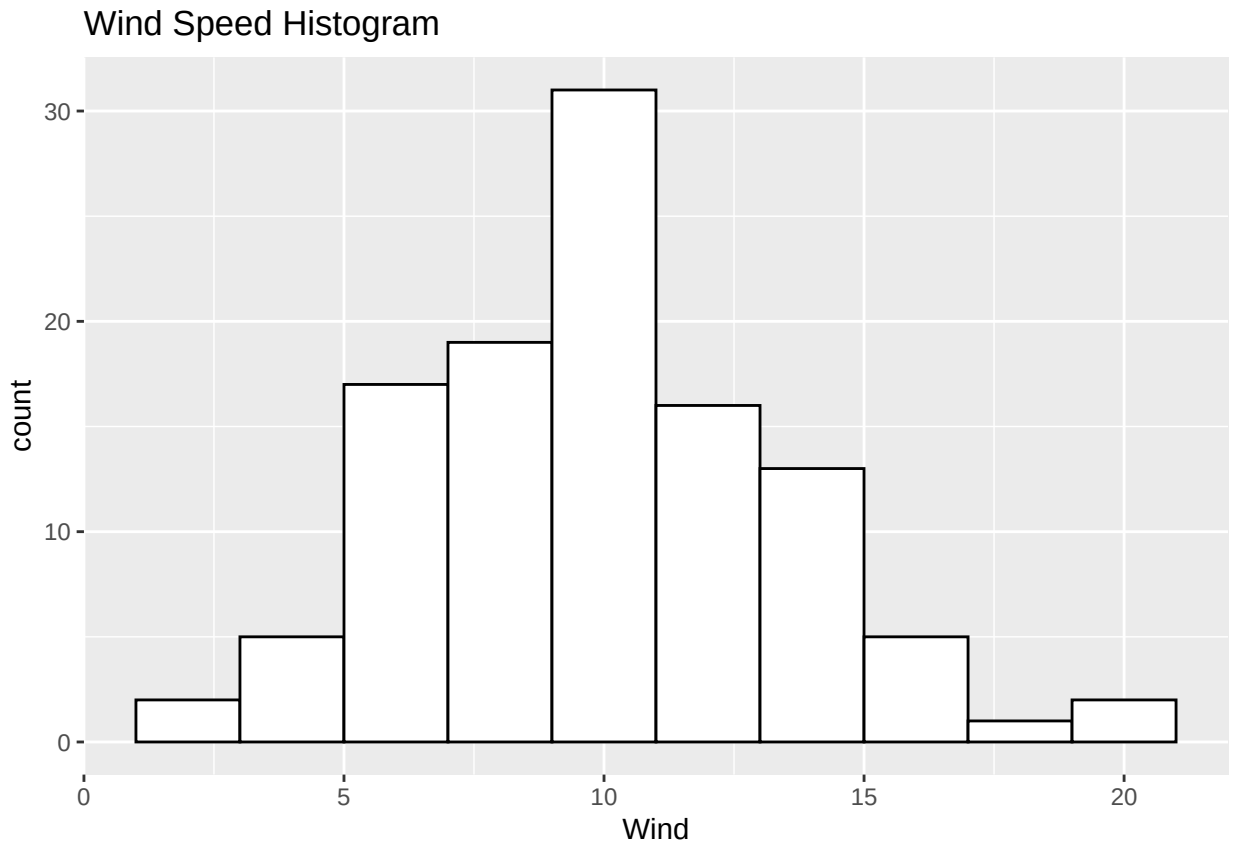

3c: Solar

```
qaHistSol <- ggplot(qualityair, aes(x=Solar.R))
qaHistSol <- qaHistSol + geom_histogram(binwidth = 25, color = "black", fill = "white")
qaHistSol <- qaHistSol + ggtitle("Solar Radiation Histogram")
qaHistSol
```



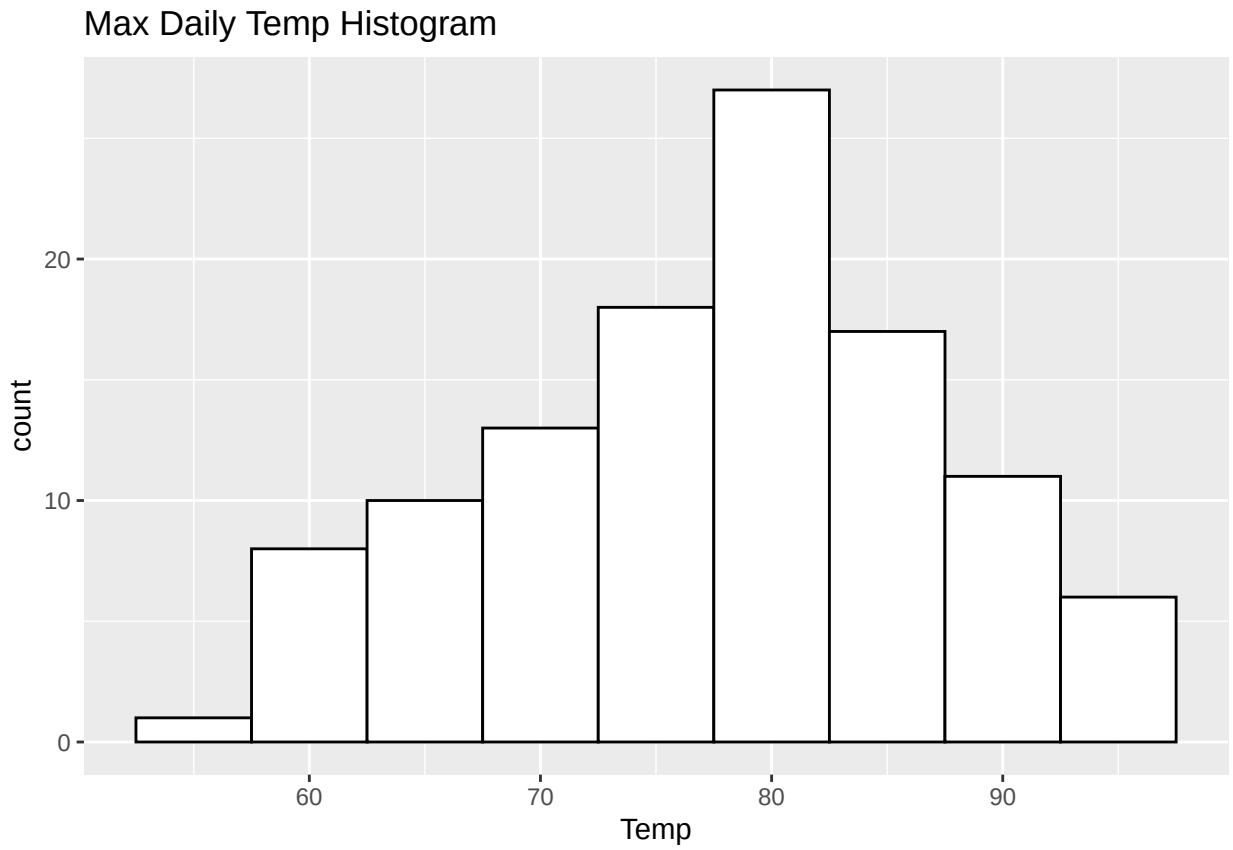

3d: Wind

```
qaHistWind <- ggplot(qualityair, aes(x=Wind))
qaHistWind <- qaHistWind + geom_histogram(binwidth = 2, color = "black", fill = "white")
qaHistWind <- qaHistWind + ggtitle("Wind Speed Histogram")
qaHistWind
```



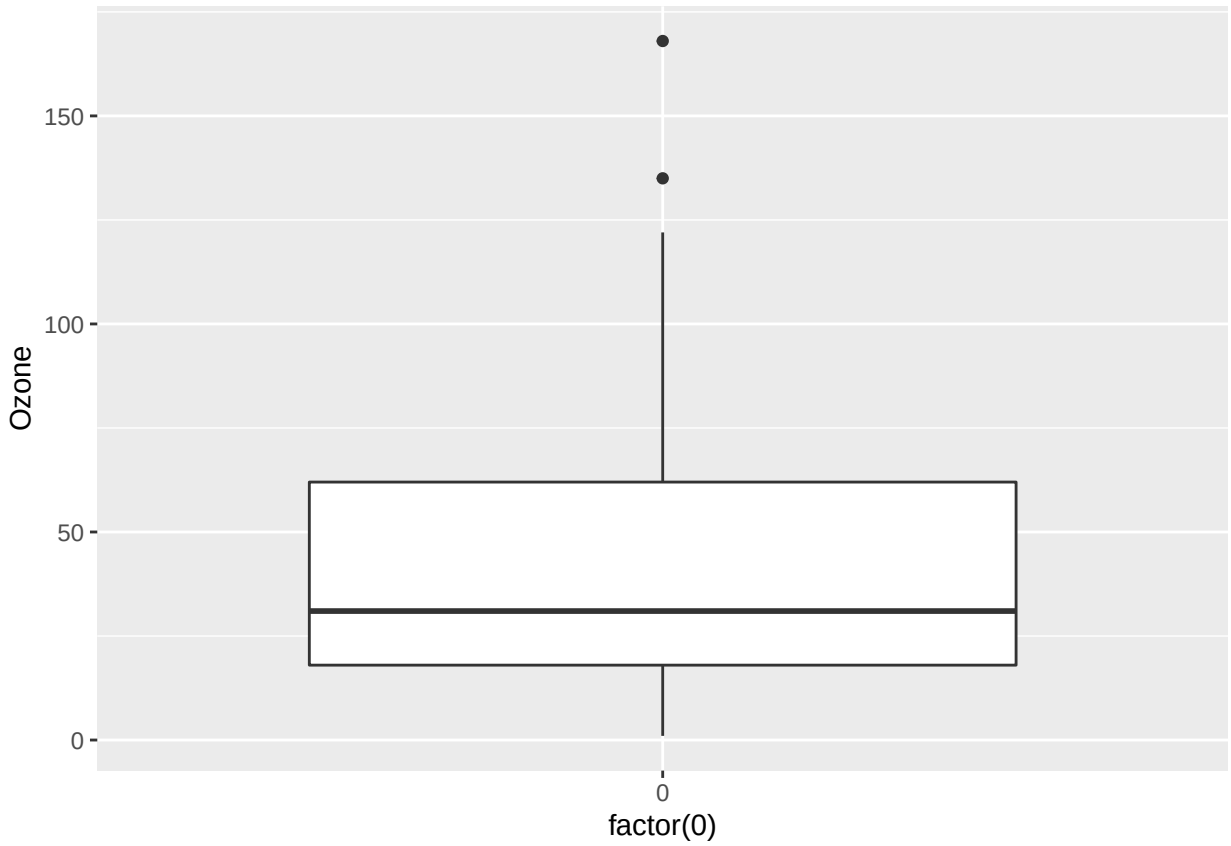

3e: Temp

```
qaHistTemp <- ggplot(qualityair, aes(x=Temp))
qaHistTemp <- qaHistTemp + geom_histogram(binwidth = 5, color = "black", fill = "white")
qaHistTemp <- qaHistTemp + ggtitle("Max Daily Temp Histogram")
qaHistTemp
```



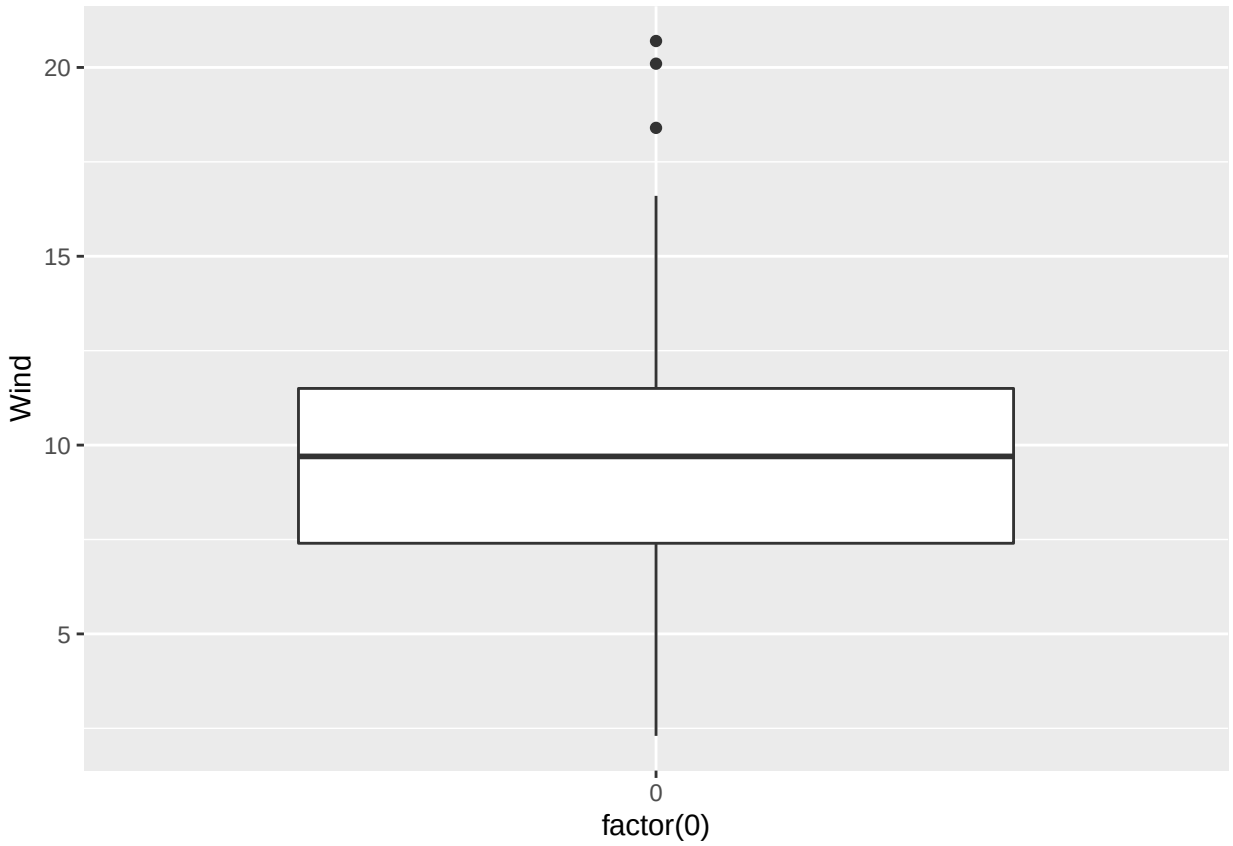

Boxplot for Ozone

```
qaBoxPlotOz <- ggplot(qualityair, aes(x=factor(0), Ozone)) + geom_boxplot()
qaBoxPlotOz
```




Boxplot for wind values (round the wind to get a good number of "buckets")

```
qaBoxPlotWind <- ggplot(qualityair, aes(x = factor(0), Wind)) + geom_boxplot()
qaBoxPlotWind
```



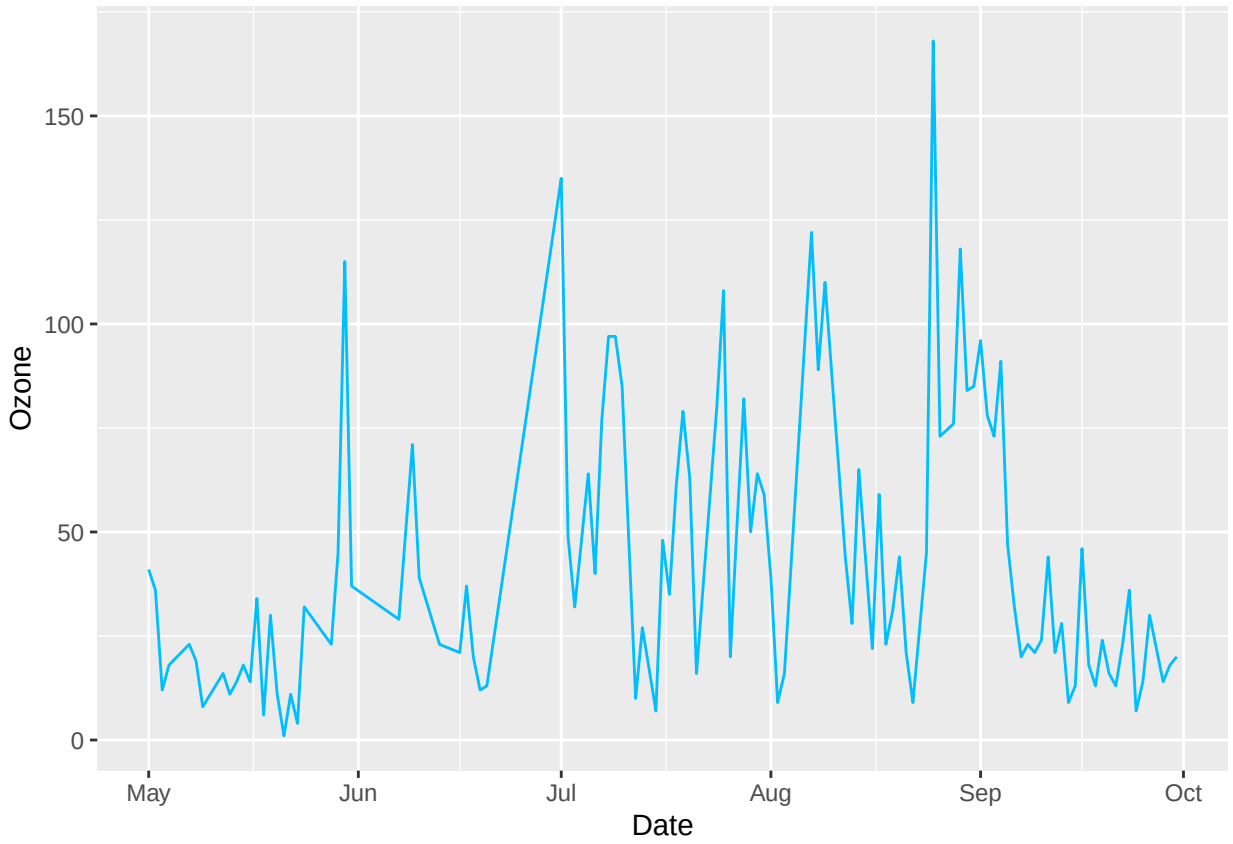

Step 4:Explore how the data changes over time Then create line charts for ozone, temp, wind and solar.R (one line chart for each, and then one chart with 4 lines, each having a different color)

Create a new "Date" column in the Quality Air dataframe

```
qualityair$Date <- as.Date(paste("1973", qualityair$Month, qualityair$Day, sep="-"))
```

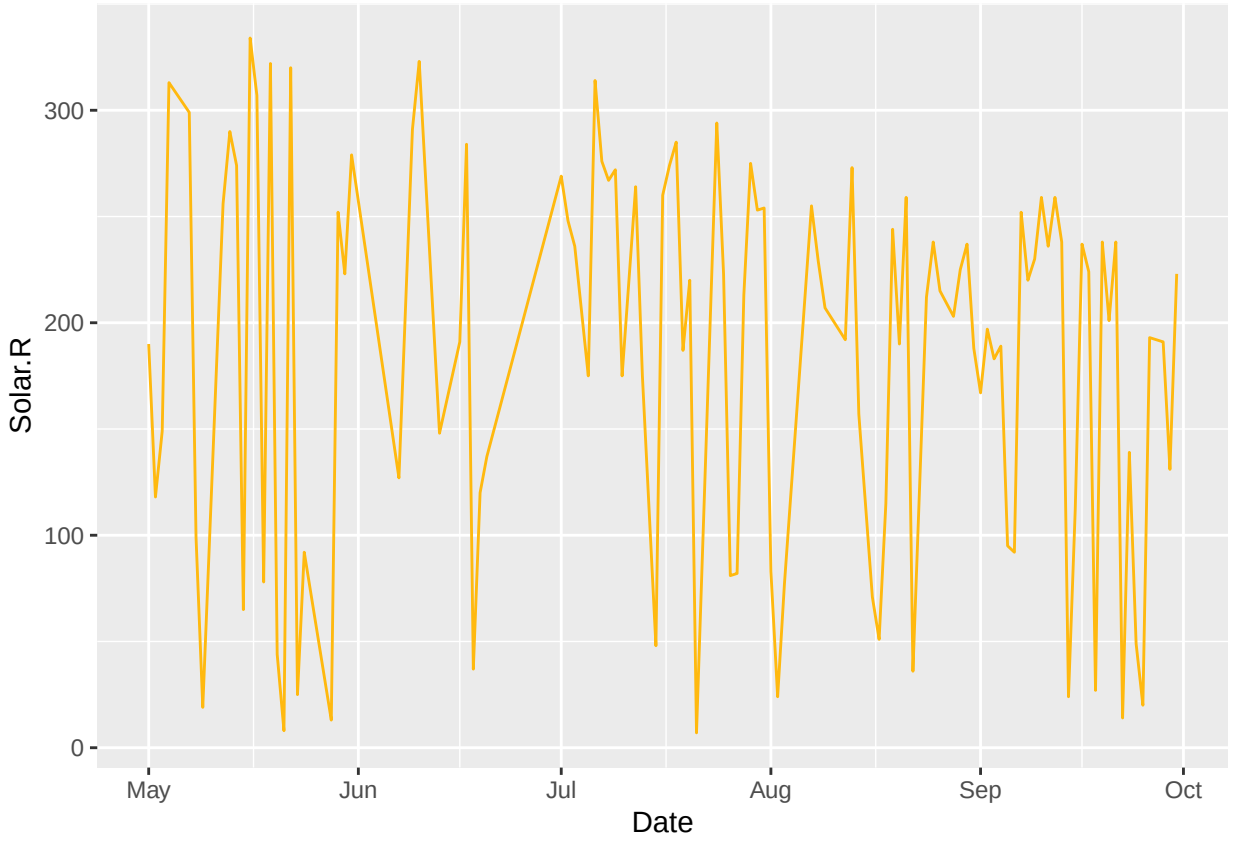
Ozone

```
qaLineOz <- ggplot(qualityair, aes(x = Date, y = Ozone)) + geom_line(color = "deepskyblue")
qaLineOz
```



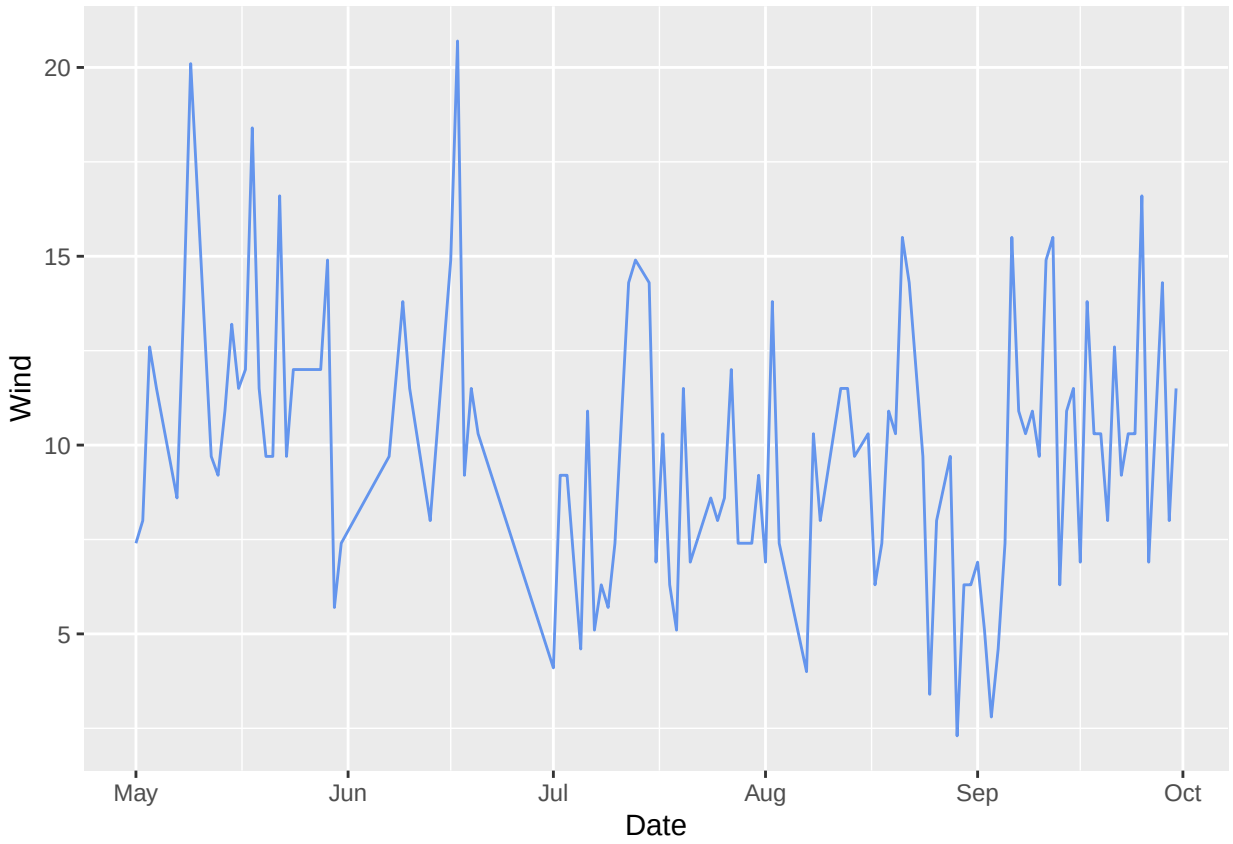

Solar

```
qaLineSolar <- ggplot(qualityair, aes(x = Date, y = Solar.R)) + geom_line(color = "darkgoldenrod1")
qaLineSolar
```



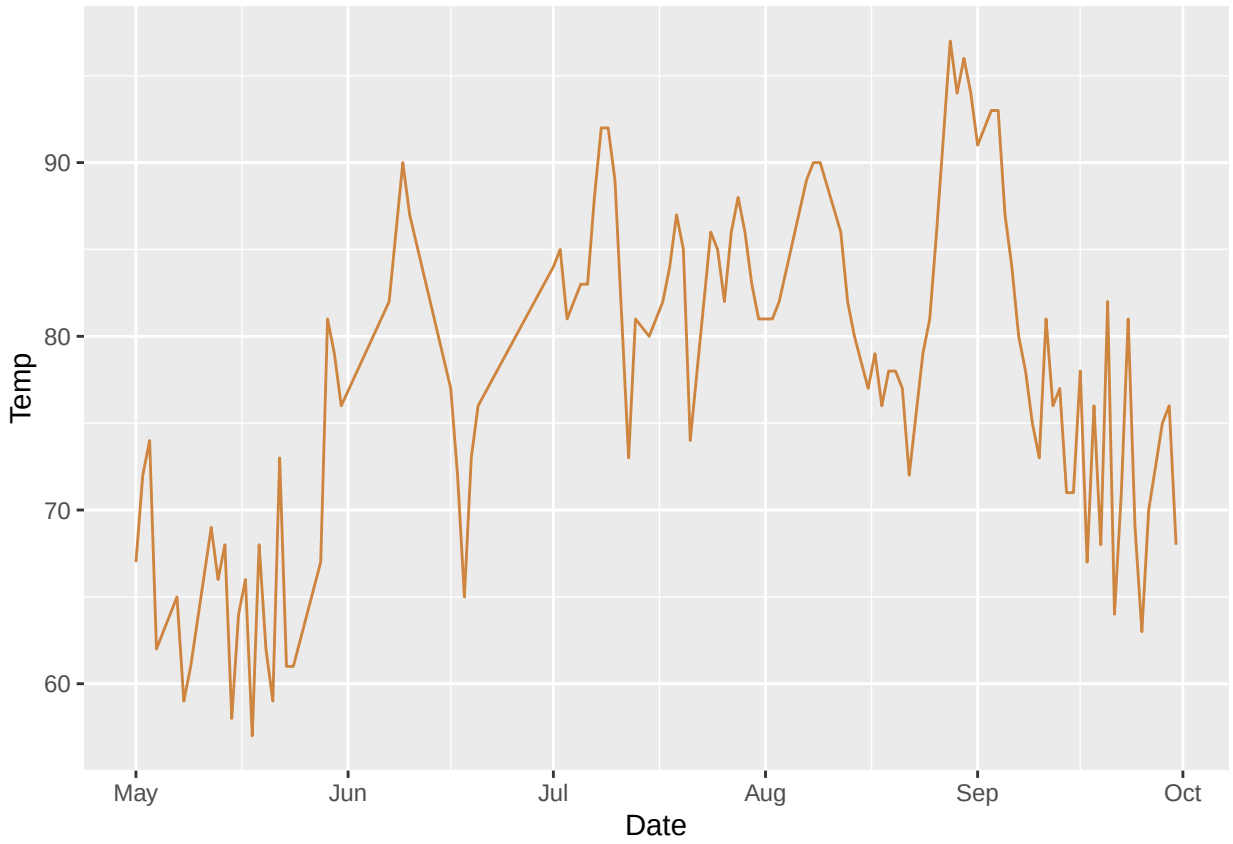

Wind

```
qaLineWind <- ggplot(qualityair, aes(x = Date, y = Wind)) + geom_line(color = "cornflowerblue")
qaLineWind
```



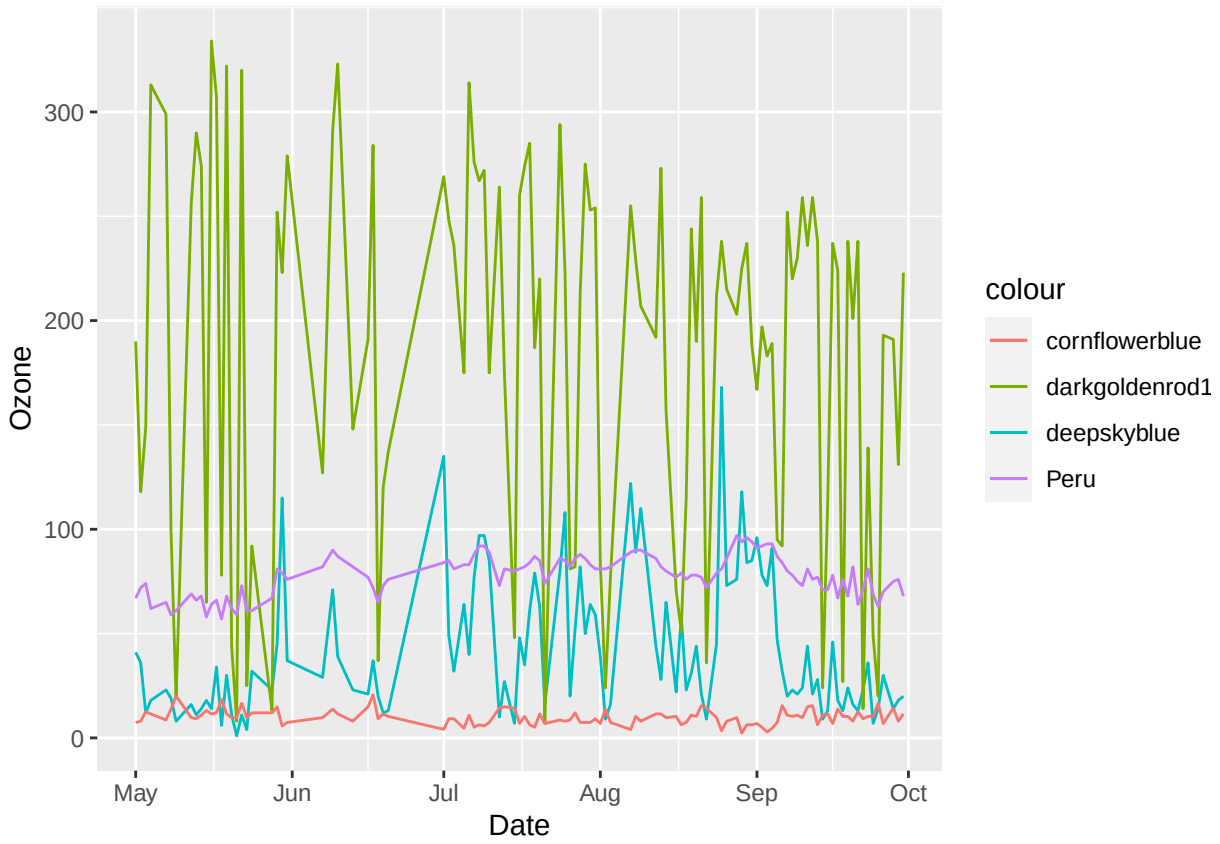

Temp

```
qaLineTemp <- ggplot(qualityair, aes(x = Date, y = Temp)) + geom_line(color = "Peru")
qaLineTemp
```



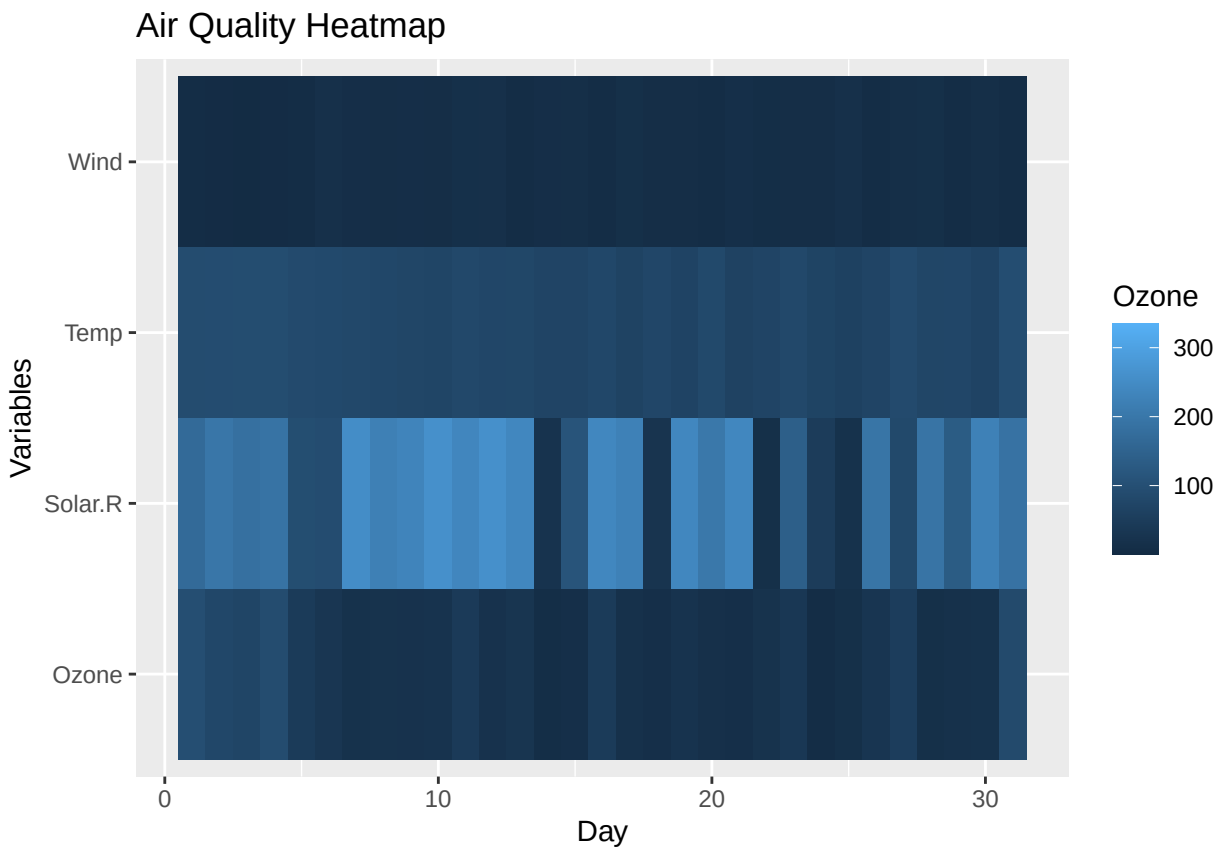

Create one line chart with 4 lines

```
qaMultiLinePlot <- ggplot(qualityair, aes(x = Date)) +
  geom_line(aes(y = Ozone, color = "deepskyblue")) +
  geom_line(aes(y = Solar.R, color = "darkgoldenrod1")) +
  geom_line(aes(y = Wind, color = "cornflowerblue")) +
  geom_line(aes(y = Temp, color = "Peru"))
qaMultiLinePlot
```



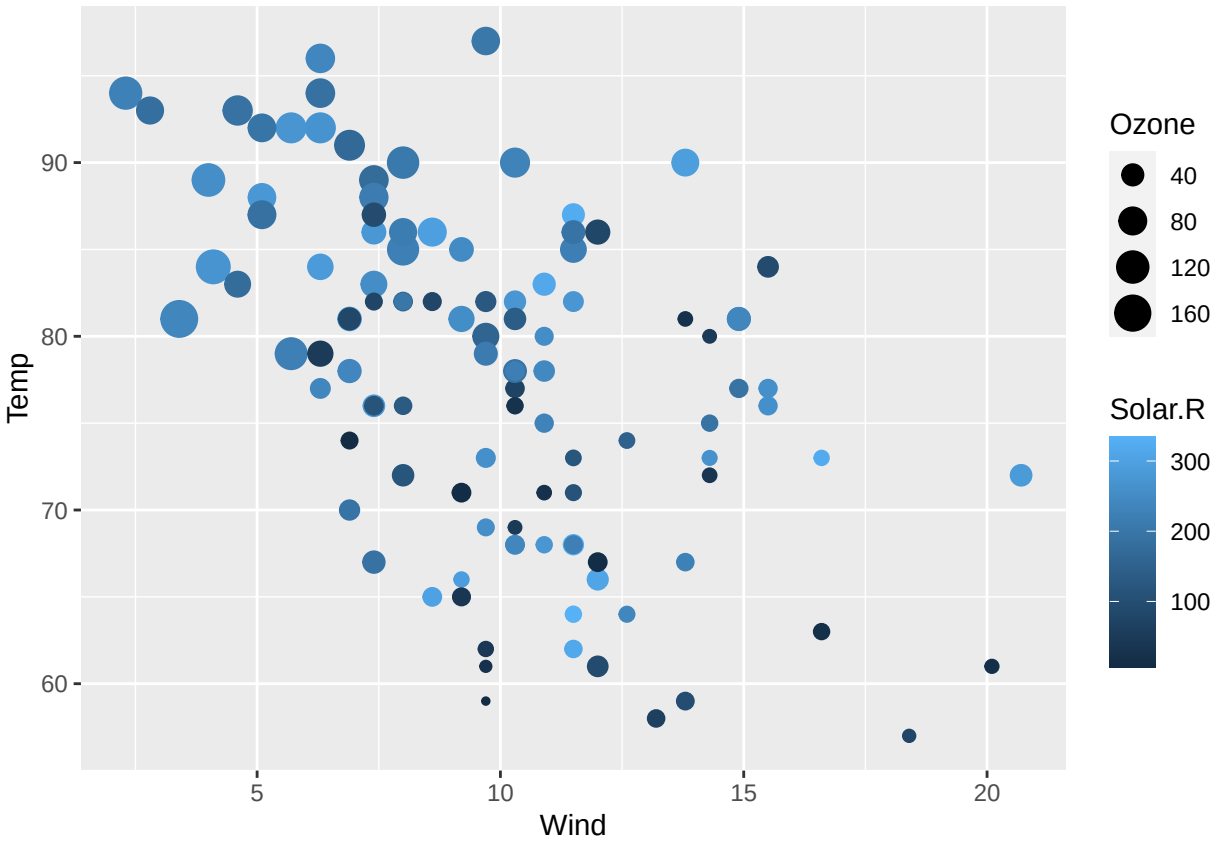

Step 4: Look at all the data via a Heatmap

```
qaHeatMap <- ggplot(qualityair, aes(x = Day)) +
  geom_tile(aes(y = "Ozone", fill = Ozone)) +
  geom_tile(aes(y = "Temp", fill = Temp)) +
  geom_tile(aes(y = "Wind", fill = Wind)) +
  geom_tile(aes(y = "Solar.R", fill = Solar.R))
qaHeatMap <- qaHeatMap + ggtitle("Air Quality Heatmap")
qaHeatMap <- qaHeatMap + xlab("Day")
qaHeatMap <- qaHeatMap +  ylab("Variables")
qaHeatMap
```




Step 5: Look at all the data via a scatter chart

```
ggplot (qualityair, aes(x = Wind, y = Temp)) +
  geom_point(aes(size = Ozone, color = Solar.R))
```




Step 6: Final Analysis There seems to be an relationship with Ozone, Temperature, and Solar Radiation. The higher the temperature the more solar radiaiton and ozone.

Histograms are the easiest to read however the color coded scatterplot showed the most information regarding potential relationships.