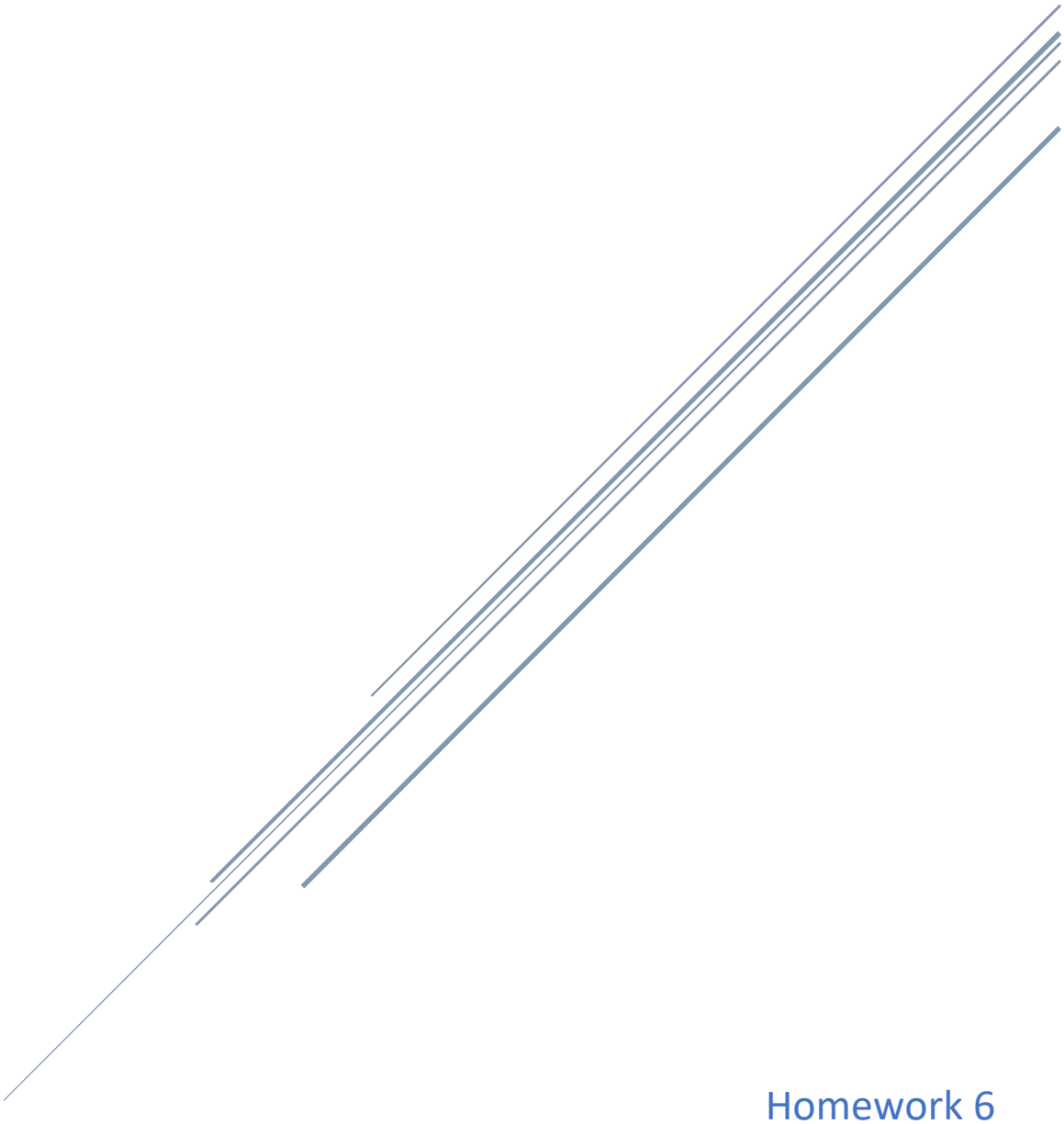
IST 687

Marriah Lewis



Homework 6
Due Week 6 and 11/14/20

#import data

> data("airquality")

#get rid of the na

> airdata <-na.omit(airquality)

> colSums(is.na(airdata))

Ozone	Solar.R	Wind	Temp	Month	Day
0	0	0	0	0	0

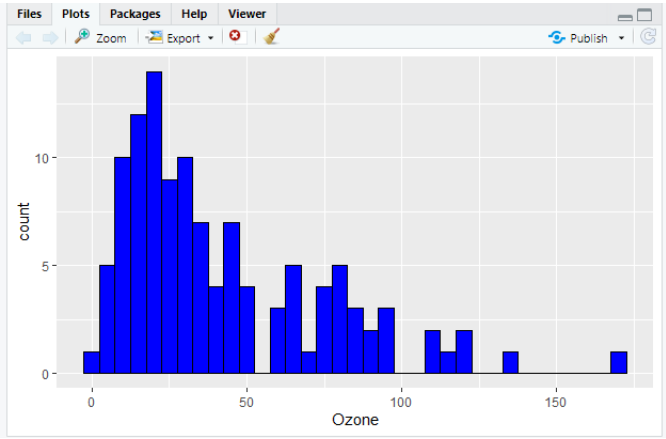
#install ggplot

>install.packages("ggplot")

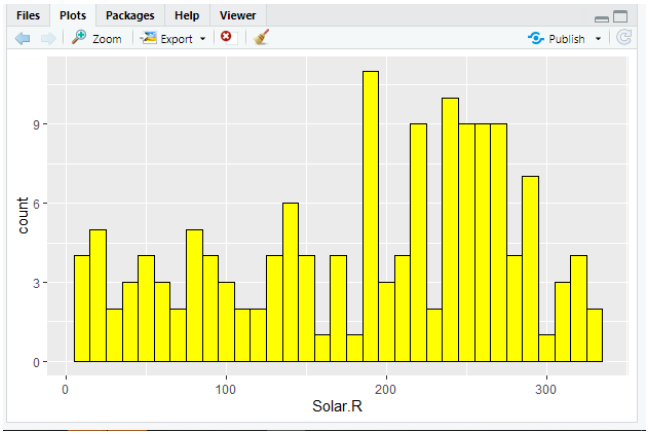
>library(ggplot2)

#Histograms for each of the variables and change the width of the histogram bin

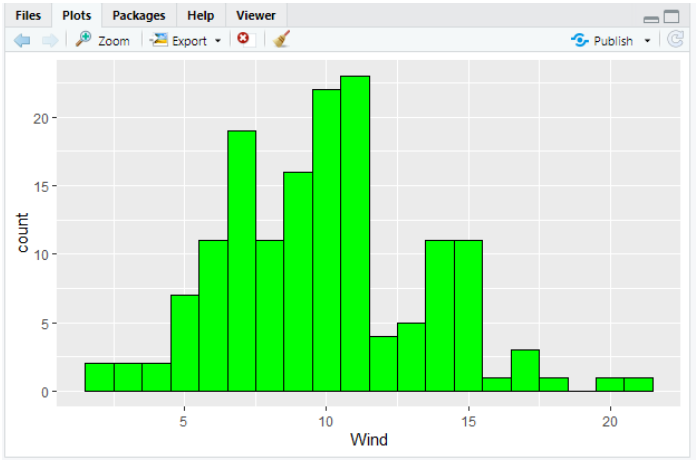
> ggplot(airdata, aes(x = Ozone)) + geom_histogram(binwidth = 5, fill="blue", color="black")



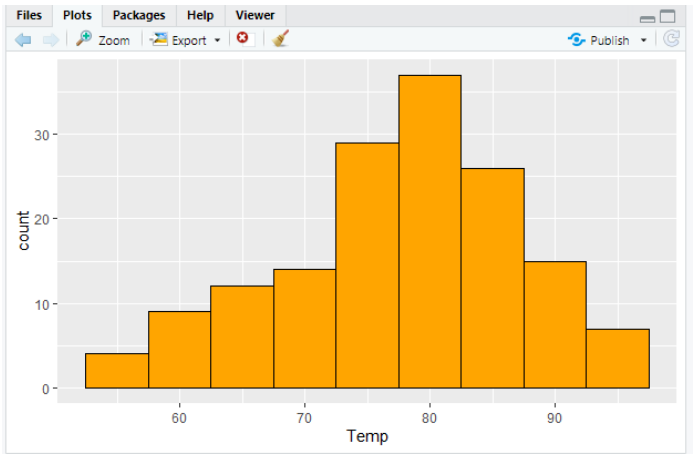
> ggplot(airdata, aes(x = Solar.R)) + geom_histogram(binwidth = 10, fill="yellow", color="black")



> ggplot(airdata, aes(x = Wind)) + geom_histogram(binwidth = 1, fill="green", color="black")>ggplot(airdata,aes(airdata$Temp))+geom_histogram(binwidth = 4)

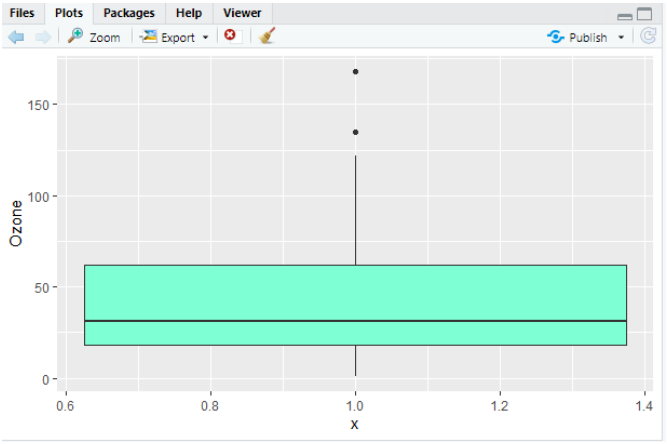


> ggplot(airdata, aes(x = Temp)) + geom_histogram(binwidth = 5, fill="orange", color="black")



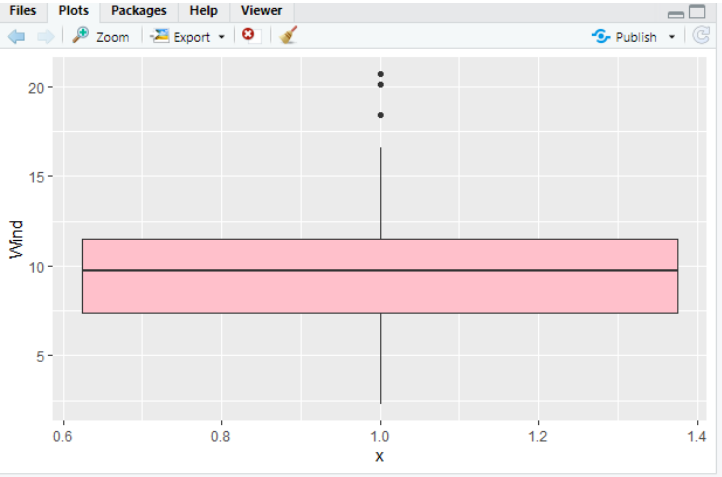
#Boxplot for Ozone

ggplot(airdata, aes(y = Ozone, x = 1)) + geom_boxplot(fill="aquamarine")



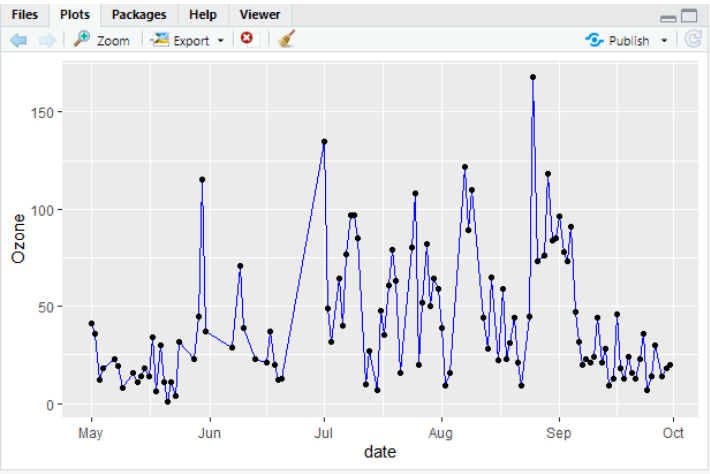
#Boxplot for wind values

```
> ggplot(airdata, aes(y = Wind, x = 1)) + geom_boxplot(fill="pink")
```

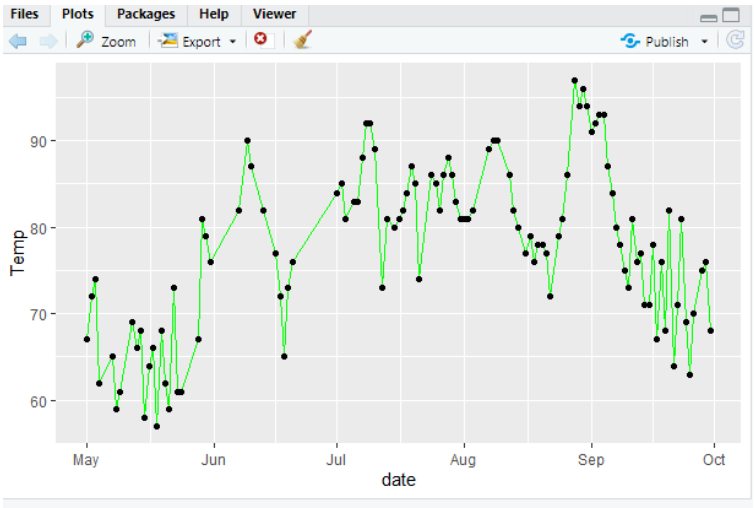


#Line charts for ozone, temp, wind, and solar R. using ggplot

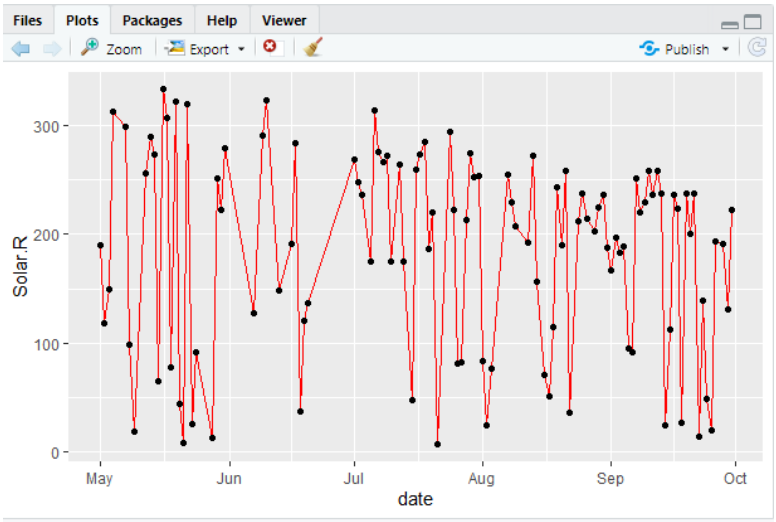
```
> airdata$date <- as.Date(paste("1973", airdata$Month,airdata$Day,
+                               sep="-"))
> ggplot(airdata, aes(x= date, y = Ozone)) +
+     geom_line(color = "blue") + geom_point()
```



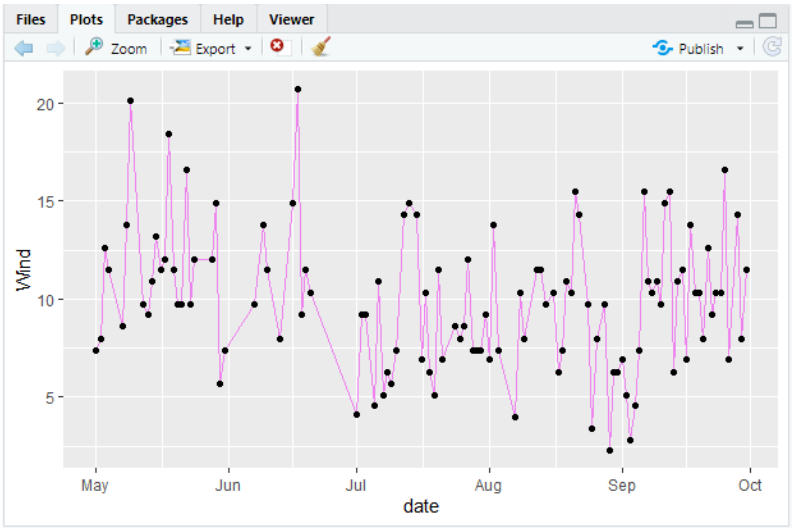

```
> ggplot(airdata, aes(x= date, y = Temp)) +
+     geom_line(color = "green") + geom_point()
```




```
> ggplot(airdata, aes(x= date, y = Solar.R)) +
+     geom_line(color = "red") + geom_point()
```




```
> ggplot(airdata, aes(x= date, y = Wind)) +
+     geom_line(color = "violet") + geom_point()
```

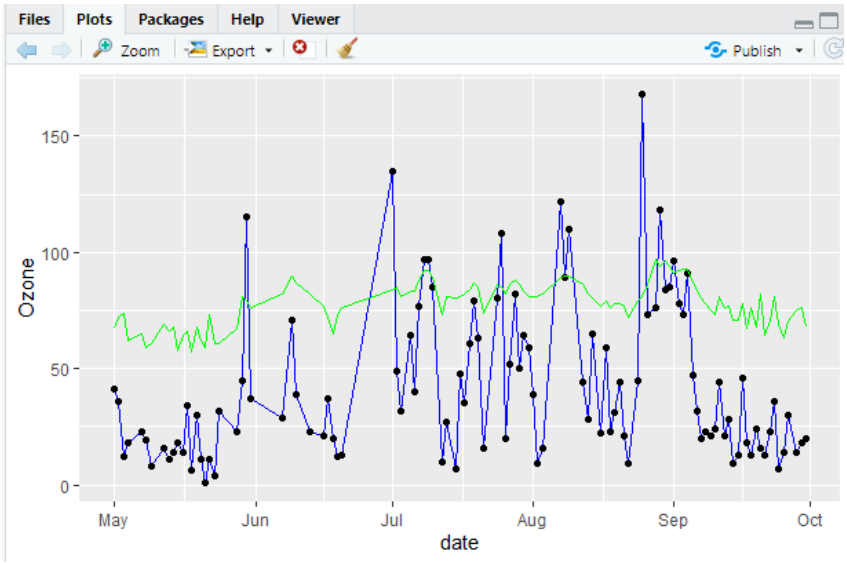


#One chart with 4 lines, each having a different color

> graph <- ggplot(airdata, aes(x= date, y = Ozone)) + geom_line(color = "blue") + geom_point()

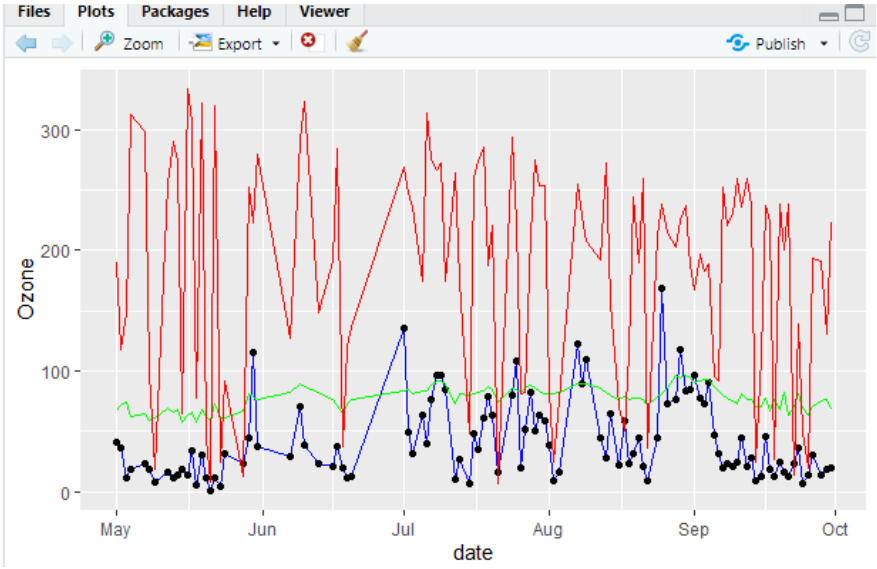
> graph <- graph + geom_line(aes(x = date, y = Temp), color = "green")

>graph



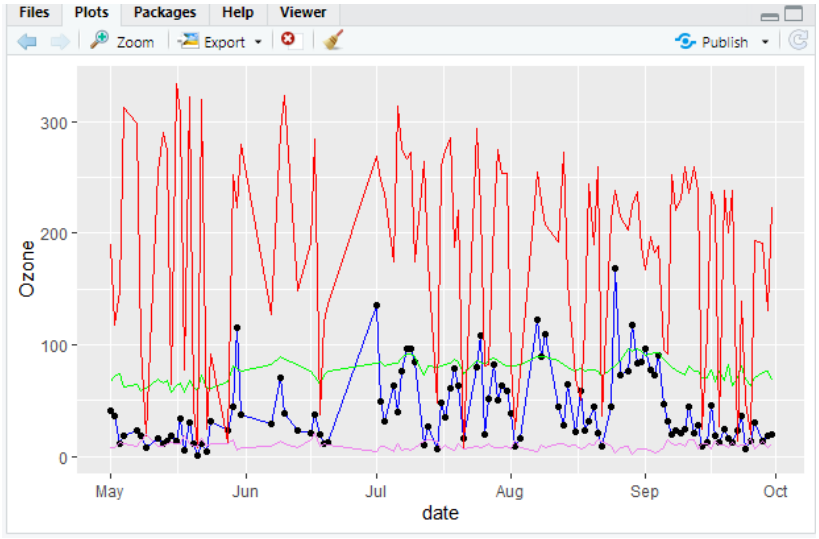
> graph <- graph + geom_line(aes(x = date, y = Solar.R), color = "red")

>graph



> graph <- graph + geom_line(aes(x = date, y = Wind), color = "violet")

> graph



#Creating a heatmap

>install.packages("tidyr")

>install.packages("reshape2")

> library(reshape2)

> library(tidyr)

Attaching package: ‘tidyr’

The following object is masked from ‘package:reshape2’:

Smiths

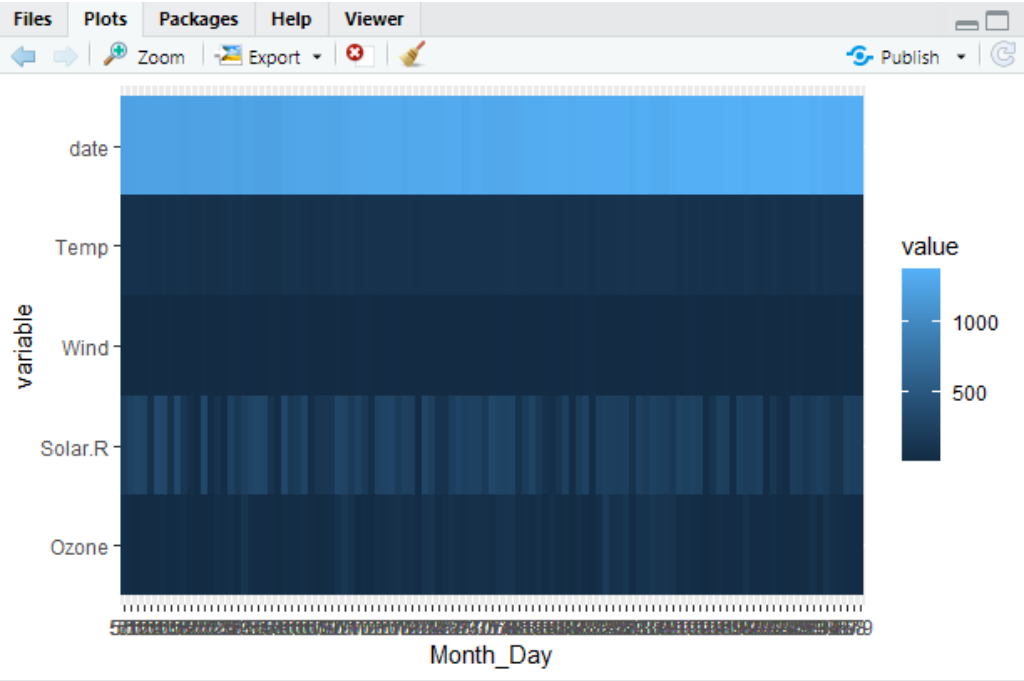
> AirMonth_Day <- unite(airdata,Month_Day,Month,Day, sep = "/")

> AirMonth_DayMelt <- melt(AirMonth_Day, id=c("Month_Day"))

Warning message:

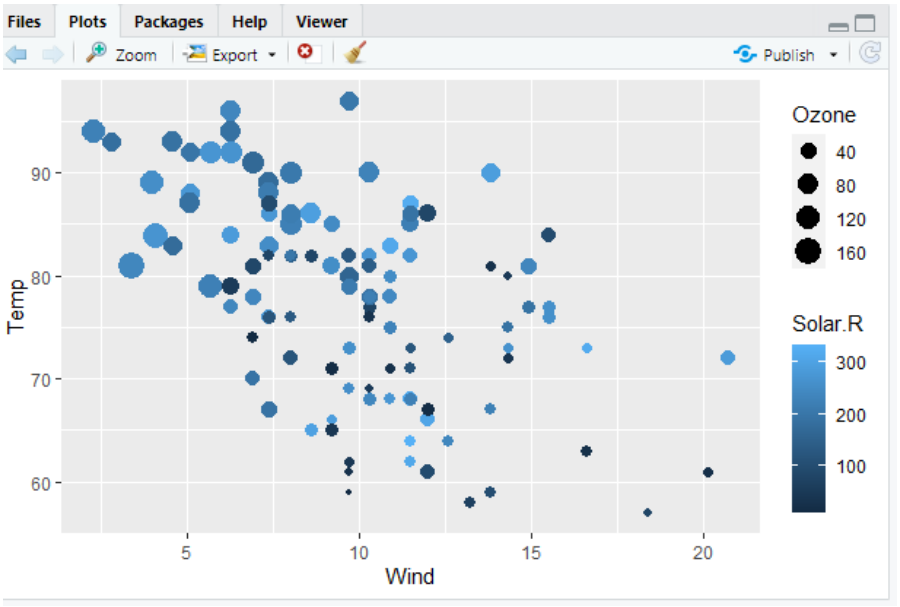
attributes are not identical across measure variables; they will be dropped

> ggplot(AirMonth_DayMelt, aes(x =Month_Day, y = variable, fill = value)) + geom_tile()



#Create a scatter plot

```
> ggplot (airdata, aes (Wind, Temp)) +
+     geom_point ( aes ( size = Ozone, color = Solar.R))
```



Step 6: Final Analysis

- There does not seem to be a pattern with the data, but a larger sample size may reveal more over a longer period. The most useful visualizations to me is the histograms. Histograms are easier to organize and structure.