Critical Thinking - Module 1

David Edwards

1/22/2022

Display the first 6 rows of the built-in Air Quality data set

```
head(airquality)
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1    41     190  7.4   67     5   1
## 2    36     118  8.0   72     5   2
## 3    12     149 12.6   74     5   3
## 4    18     313 11.5   62     5   4
## 5    NA      NA 14.3   56     5   5
## 6    28      NA 14.9   66     5   6
```

Display the summary statistics for all attributes (columns) of the Air Quality data set.

```
summary(airquality)
```

```
##      Ozone           Solar.R           Wind             Temp      
##  Min.   :  1.00   Min.   :  7.0   Min.   : 1.700   Min.   :56.00  
##  1st Qu.: 18.00   1st Qu.:115.8   1st Qu.: 7.400   1st Qu.:72.00  
##  Median : 31.50   Median :205.0   Median : 9.700   Median :79.00  
##  Mean   : 42.13   Mean   :185.9   Mean   : 9.958   Mean   :77.88  
##  3rd Qu.: 63.25   3rd Qu.:258.8   3rd Qu.:11.500   3rd Qu.:85.00  
##  Max.   :168.00   Max.   :334.0   Max.   :20.700   Max.   :97.00  
##  NA's   :37       NA's   :7                                       
##      Month            Day      
##  Min.   :5.000   Min.   : 1.0  
##  1st Qu.:6.000   1st Qu.: 8.0  
##  Median :7.000   Median :16.0  
##  Mean   :6.993   Mean   :15.8  
##  3rd Qu.:8.000   3rd Qu.:23.0  
##  Max.   :9.000   Max.   :31.0  
##                                
```

Select any two attributes of the Air Quality data set. For each of your two selected attributes display:

Ozone

```
mean(airquality$Ozone, na.rm=TRUE)
```

Mean

```
## [1] 42.12931
```

```
median(airquality$Ozone, na.rm=TRUE)
```

Median

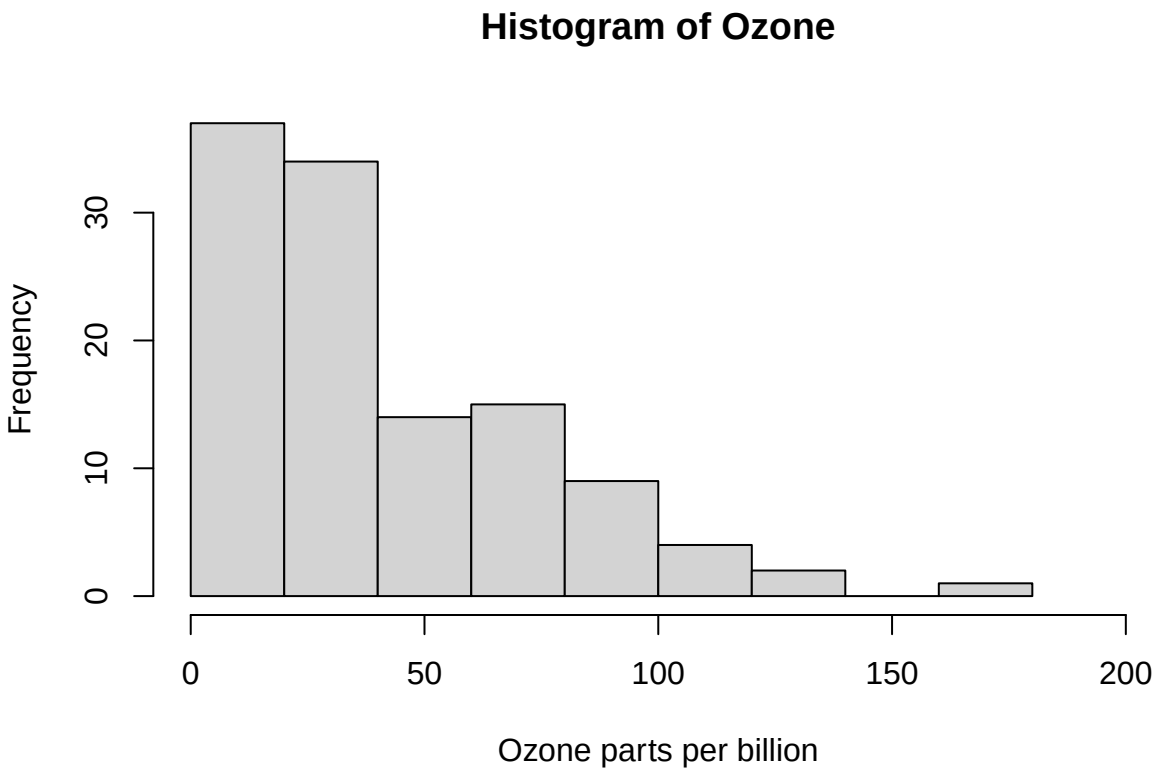
```
## [1] 31.5
```

```
range(airquality$Ozone, na.rm=TRUE)
```

Range

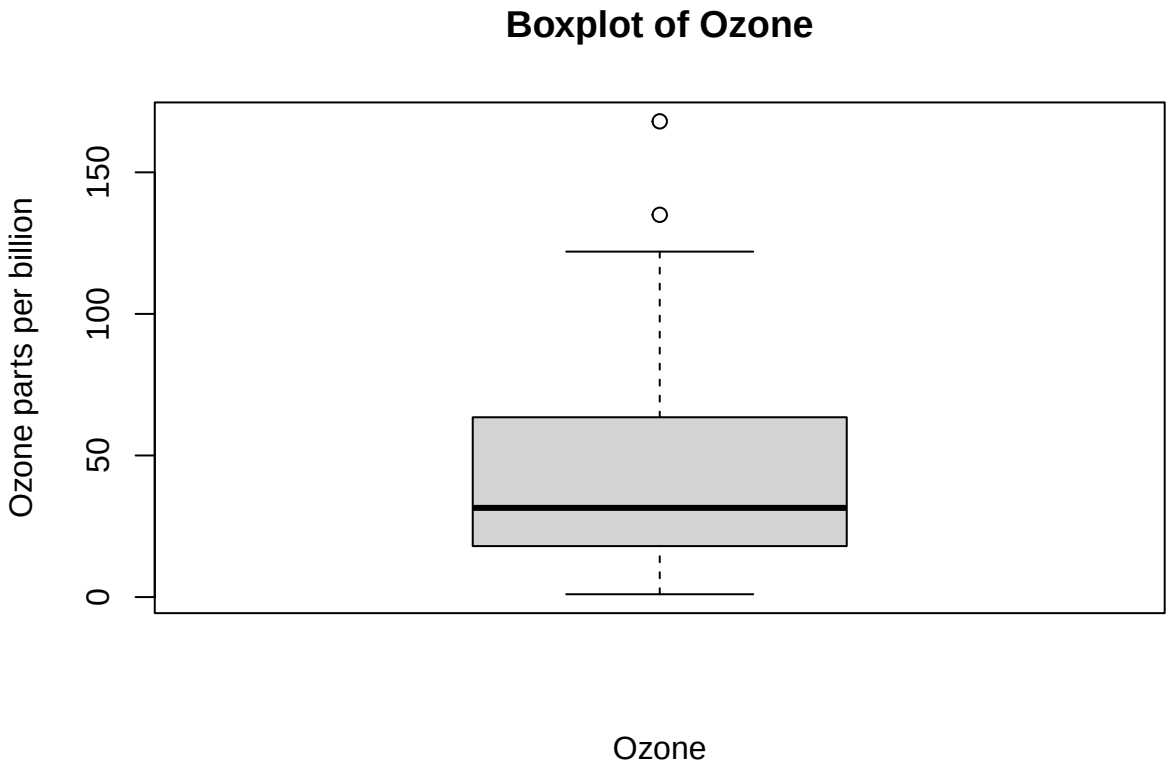
```
## [1]   1 168
```

```
Ozone <- airquality$Ozone[!is.na(airquality$Ozone)]
hist(Ozone, xlim=c(0,200), xlab="Ozone parts per billion")
```



Histogram

```
boxplot(Ozone, main="Boxplot of Ozone", xlab="Ozone", ylab="Ozone parts per billion")
```



Temp

```
Temperature <- airquality$Temp[!is.na(airquality$Temp)]
mean(Temperature)
```

Mean

```
## [1] 77.88235
```

```
median(Temperature)
```

Median

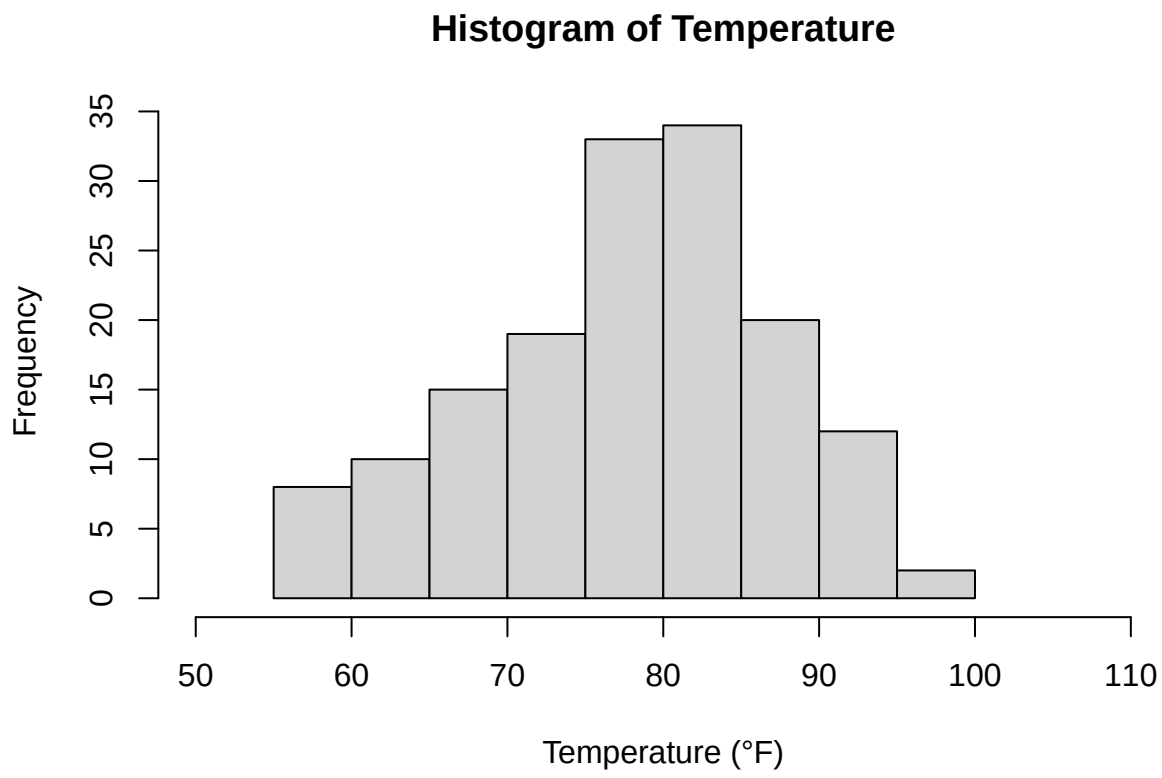
```
## [1] 79
```

```
range(Temperature)
```

Range

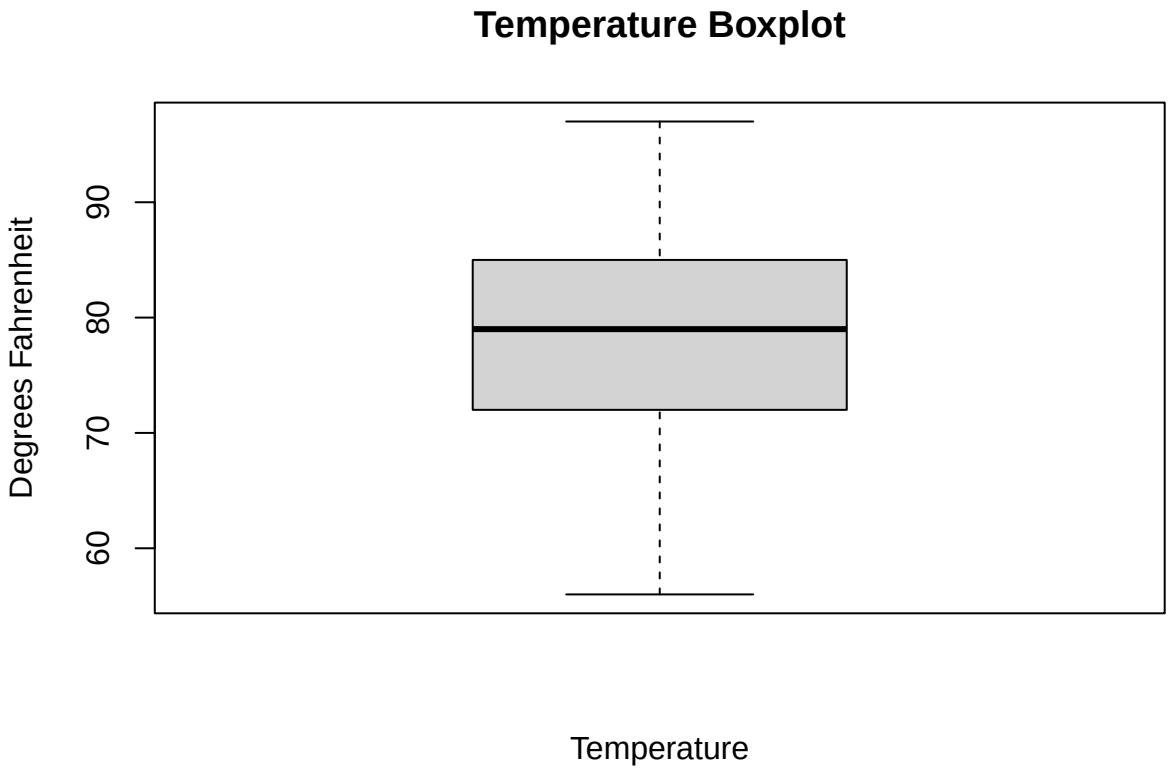
```
## [1] 56 97
```

```
hist(Temperature, xlim=c(50,110), xlab="Temperature (\u00B0F)")
```



Histogram

```
boxplot(Temperature, main="Temperature Boxplot", xlab="Temperature", ylab="Degrees Fahrenheit")
```



For the same two attributes of the Air Quality data set that you selected in step 3, display the correlation between these two attributes

statistically via a correlation number

Creating the corellation resulted in an NA because some Ozone levels were NA. I filtered out the NA Ozone measurements, and the corresponding Temperature elements as well.

```
Ozone <- airquality$Ozone[!is.na(airquality$Ozone)]
Temperature <- airquality$Temp[!is.na(airquality$Ozone)]
cor(Ozone, Temperature, method=c("pearson", "kendall", "spearman"))
```

```
## [1] 0.6983603
```

visually via a scatterplot diagram.

```
plot(Ozone, Temperature, main="Scatterplot Temp and Ozone", xlab="Ozone (ppb) ", ylab="Temperature (\u0
lines(lowess(Ozone, Temperature), col="blue")
```