Stat 123 Homework 6

Jonathan Wilson

January 25, 2019

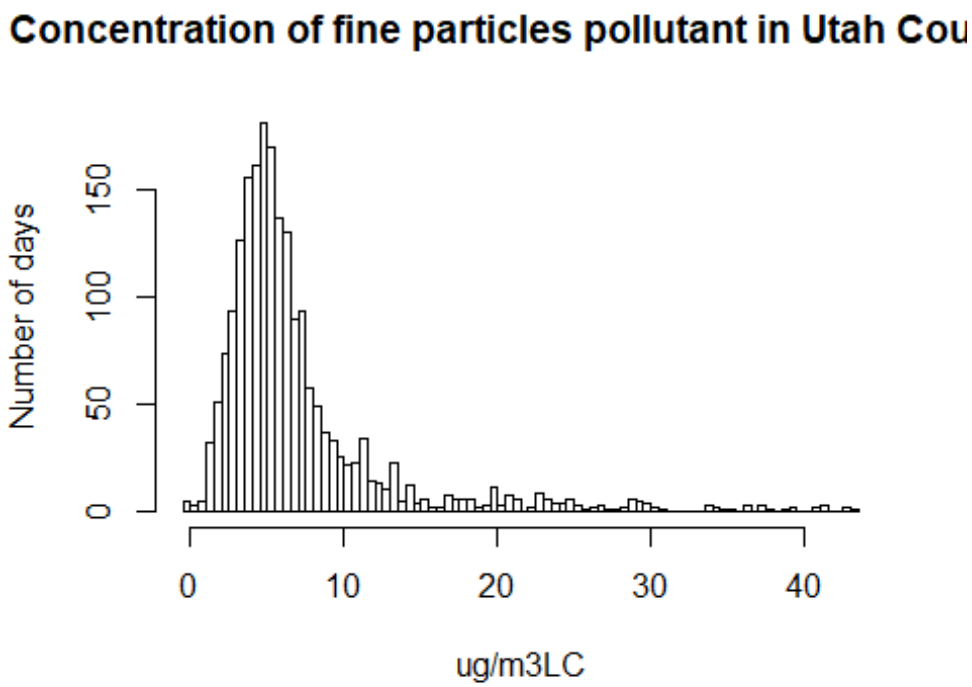
```
#Make working directory global
knitr::opts_knit$set(root.dir =
"C:\\Users\\jon\\Documents\\School\\R\\HW\\HW6")
```

Read in the data on the daily concentration of fine particles pollutant (PM2.5) in Utah County for all of 2014 into a data.frame in R.

```
air <- read.table("daily_airquility.csv", sep=",", header=TRUE,
                  stringsAsFactors=FALSE)
pol <- data.frame(air$Daily.Mean.PM2.5.Concentration)
names(pol) <- c("pollution")
#pol$Date <- c(air$Date)
```

Using R's basic graphics, make a histogram of the "Daily.Mean.PM2.5.Concentration" variable with axis labels and a title.

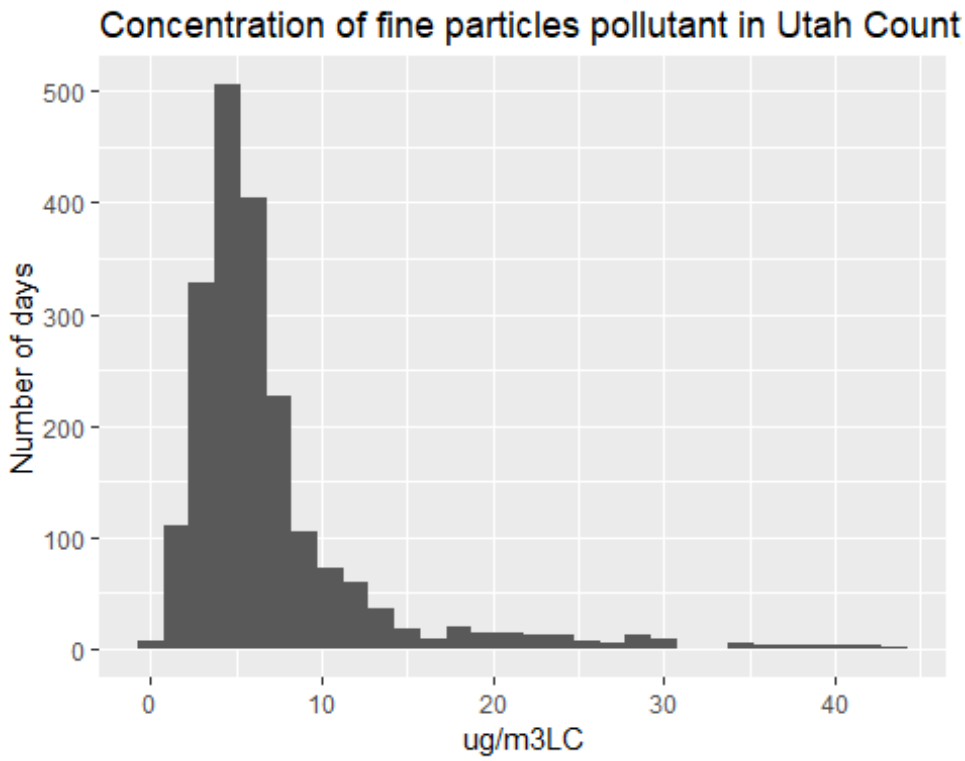
```
hist(pol$pollution, main="Concentration of fine particles pollutant in Utah
County", breaks=100, xlab="ug/m3LC",
     ylab = "Number of days")
```



Using the ggplot2 package, make a histogram of the “Daily.Mean.PM2.5.Concentration” variable with axis labels and a title.

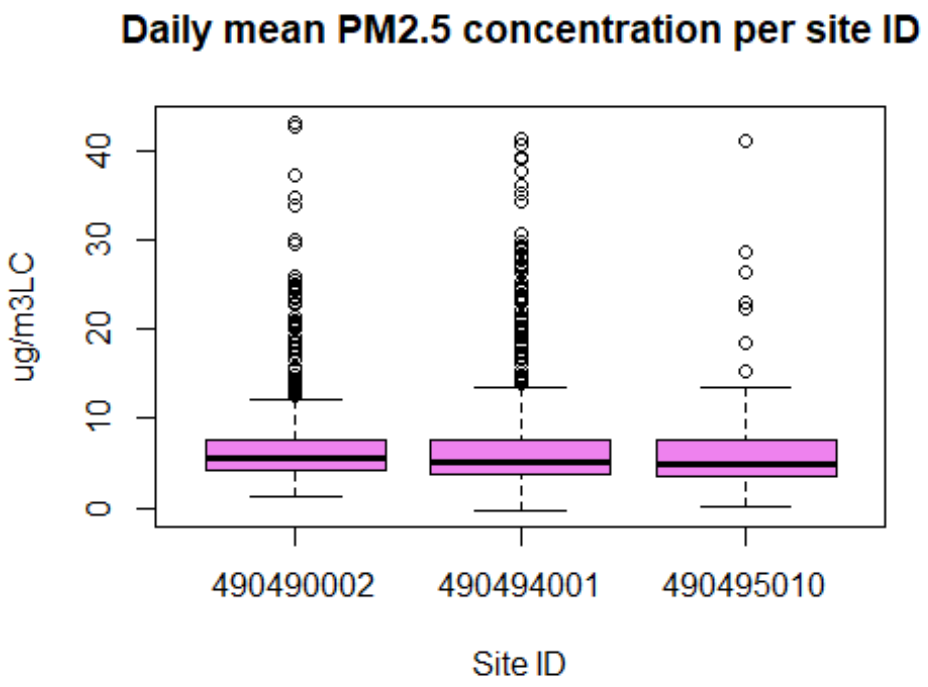
```
#install.packages("ggplot2")
library(ggplot2)
g <- ggplot(data=pol, aes(x=pollution)) + geom_histogram() +
ggtitle("Concentration of fine particles pollutant in Utah County")
g + xlab("ug/m3LC") + ylab("Number of days")

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



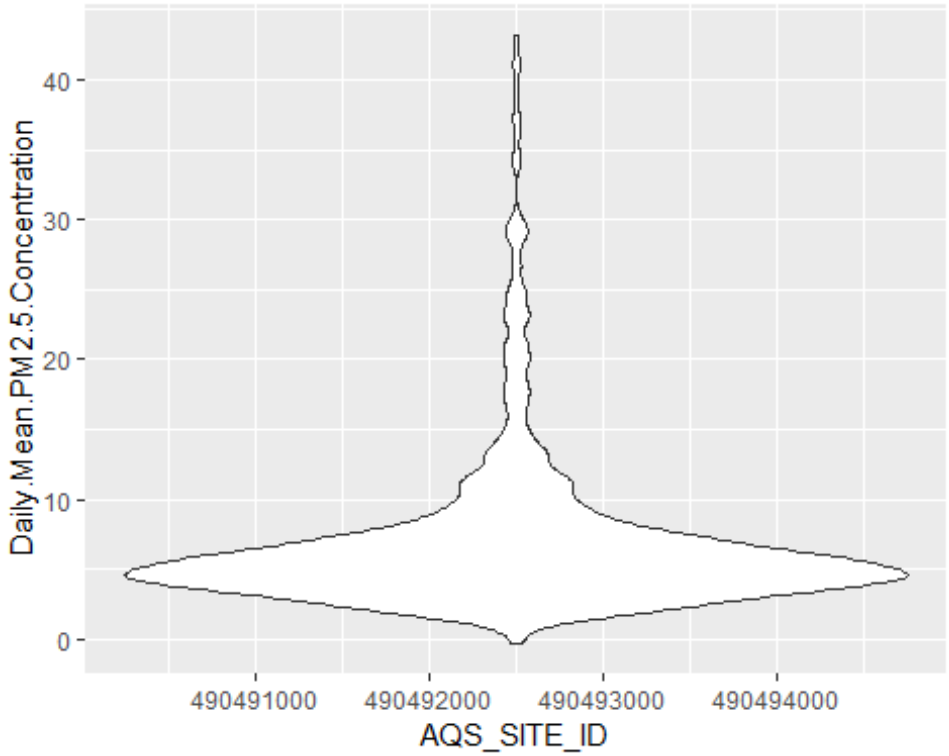
Using either R’s basic graphics or ggplot2, produce side-by-side boxplots for the “Daily.Mean.PM2.5.Concentration” variable for each value of the “AQS_SITE_ID” variable.

```
boxplot(pol$pollution ~ air$AQS_SITE_ID, col="violet", main="Daily mean PM2.5
concentration per site ID", ylab="ug/m3LC", xlab="Site ID", horizontal=FALSE)
```



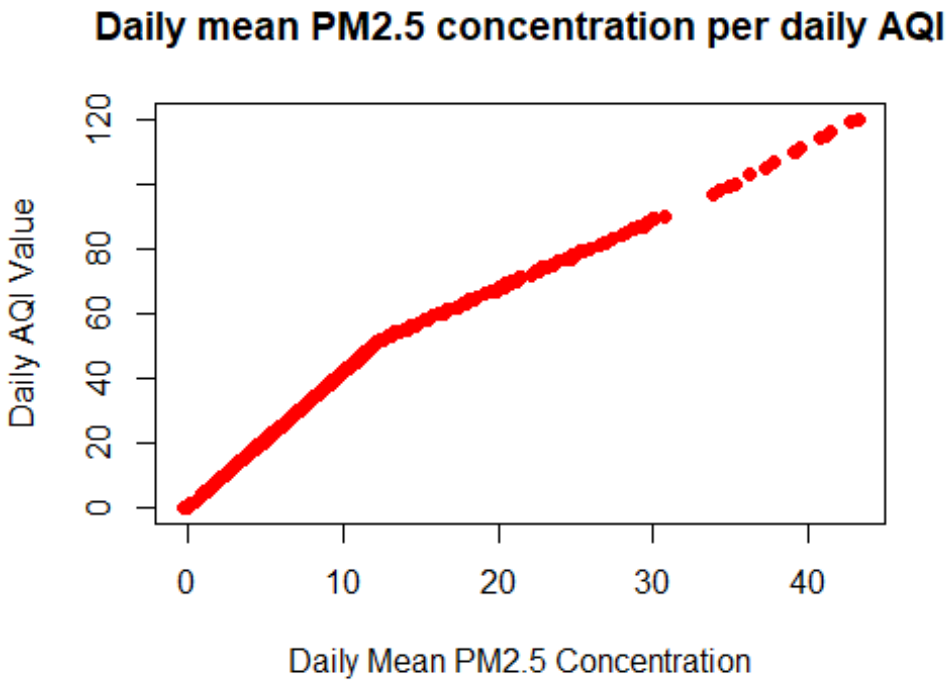
Using ggplot2, produce side-by-side violin plots for the "Daily.Mean.PM2.5.Concentration" variable for each value of the "AQS_SITE_ID" variable.

```
ggplot(air, aes(y=Daily.Mean.PM2.5.Concentration, x=AQS_SITE_ID)) +
geom_violin()
```



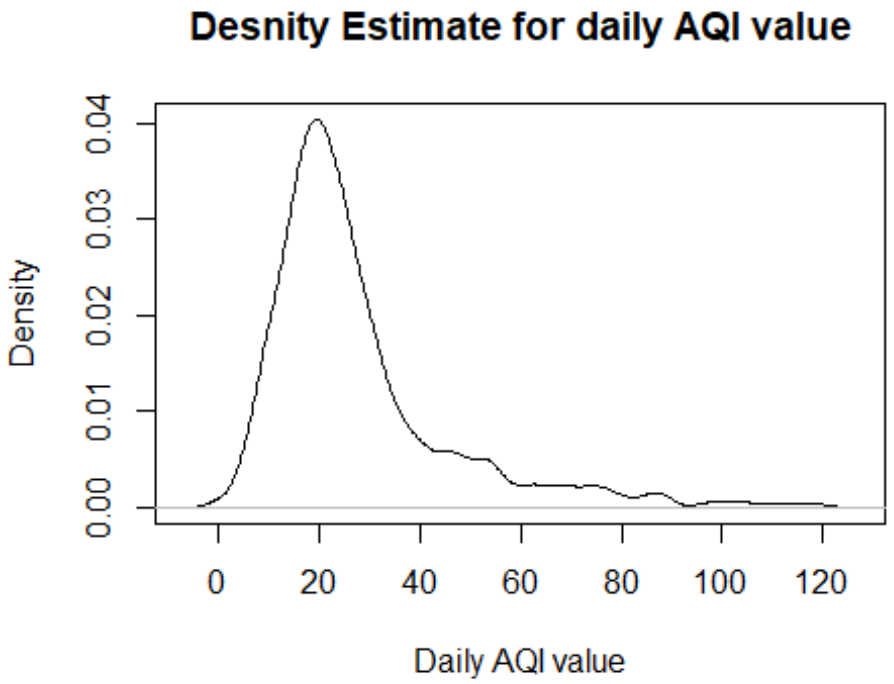
Using either R's basic graphics or ggplot2, produce a scatter plot with "Daily.Mean.PM2.5.Concentration" on the x-axis and "DAILY_AQI_VALUE" on the y-axis, where the color of each point indicates the value of the "POC" variable.

```
plot(pol$pollution, air$DAILY_AQI_VALUE,
       pch = 16, col = 2,
       xlab = "Daily Mean PM2.5 Concentration",
       ylab = "Daily AQI Value",
       main="Daily mean PM2.5 concentration per daily AQI")
```



Using either plotting library, plot a density estimate for the 'DAILY_AQI_VALUE' with axis labels and title.

```
d <- density(air$DAILY_AQI_VALUE)
plot(d, main="Desnity Estimate for daily AQI value", xlab = "Daily AQI
value") # plots the results
```


Desnity Estimate for daily AQI value
Density
0.00
0.01
0.02
0.03
0.04
0
20
40
60
80
100
120
Daily AQI value