

# Stat 123 Homework 15

Jonathan Wilson

February 18, 2019

```
knitr::opts_knit$set(root.dir =  
"C:\\Users\\jon\\Documents\\School\\R\\HW\\HW15")  
#Make sure you are calling the right lib in the RIGHT ORDER!  
  
library(plyr); library(magrittr); library(dplyr); library(data.table)  
  
##  
## Attaching package: 'dplyr'  
  
## The following objects are masked from 'package:plyr':  
##  
##   arrange, count, desc, failwith, id, mutate, rename, summarise,  
##   summarize  
  
## The following objects are masked from 'package:stats':  
##  
##   filter, lag  
  
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union  
  
##  
## Attaching package: 'data.table'  
  
## The following objects are masked from 'package:dplyr':  
##  
##   between, first, last
```

Data on the daily concentration of fine particles pollutant (PM2.5) for all monitor sites in Utah County is available in .csv files attached to the assignment, with each file having data from a year. These files have the name format "ad\_viz\_plotval\_data-XXXX.csv" (where XXXX is place holder for the year).

Write a function named 'air' that takes a 'year' argument and returns a data.frame containing that data for that year, suppressing the automatic conversion to factors.

```
air <- function(year){  
  file <- sprintf("ad_viz_plotval_data-%s.csv", year)  
  data <- read.table(file, sep=";", header=TRUE, stringsAsFactors=FALSE)  
  return(data)  
}
```

Using lapply, make a list containing a data.frame for each year.

```
data.list <- lapply(2009:2016, air)
```

Combine all the data.frames in the list to one data.frame (without losing any data).

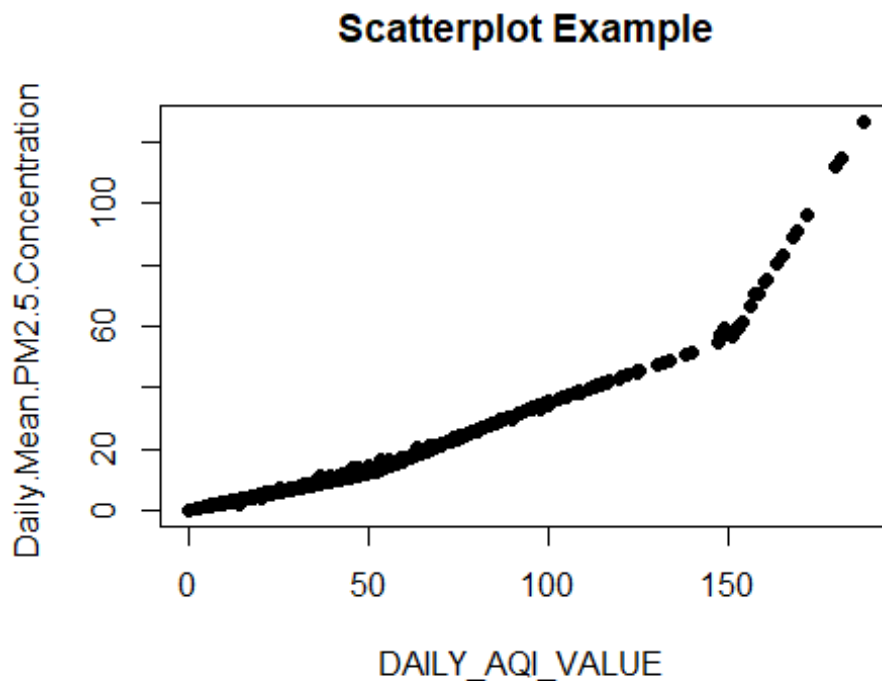
```
data <- rbindlist(data.list)
```

Summarize this big data.frame to yield a data.frame containing only the average “DAILY\_AQI\_VALUE” and the average “Daily.Mean.PM2.5.Concentration” for each value of “Date”.

```
avg.data <- data %>% group_by(Date) %>% summarize(DAILY_AQI_VALUE =  
mean(DAILY_AQI_VALUE), Daily.Mean.PM2.5.Concentration =  
mean(Daily.Mean.PM2.5.Concentration))
```

Produce a scatter plot of the average “Daily.Mean.PM2.5.Concentration” (on the y-axis) and average “DAILY\_AQI\_VALUE” (on the x-axis).

```
attach(avg.data)  
my.plot <- plot(DAILY_AQI_VALUE, Daily.Mean.PM2.5.Concentration,  
main="Scatterplot Example",  
xlab="DAILY_AQI_VALUE ", ylab="Daily.Mean.PM2.5.Concentration ", pch=19)
```



Add vertical lines at 50, 100, and 150 on the x-axis. Hint: See the help for the “abline” function.

```
plot(DAILY_AQI_VALUE, Daily.Mean.PM2.5.Concentration, main="Scatterplot  
Example",  
      xlab="DAILY_AQI_VALUE ", ylab="Daily.Mean.PM2.5.Concentration ", pch=19)  
abline(v=c(50, 100, 150), col=c("green", "blue", "red"))
```

