Homework 3

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

##
## Attaching package: 'magrittr'

## The following object is masked from 'package:tidyr':
##
##     extract

## Loading required package: sm

## Package 'sm', version 2.2-5.4: type help(sm) for summary information

##
## Please cite as:

##  Hlavac, Marek (2015). stargazer: Well-Formatted Regression and Summary Statistics Tables.

##  R package version 5.2. http://CRAN.R-project.org/package=stargazer

##
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':
##
##     between, first, last

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:data.table':
##
##     hour, isoweek, mday, minute, month, quarter, second, wday,
##     week, yday, year

## The following object is masked from 'package:base':
##
##     date
```

Problem 4

Programming Style Guides

My Takeaway: Every person has their own style of programming their codes. The key is consistency in that style. Ultimately, mainting a consistent style of programming helps facilitate reproducible research, if the style improves the way the code can be interpreted.

I will try to add comments to my codes henceforth to improve my coding style. I will strive to be consistent with the way I define objects and functions.

Problem 5

```
#Cannot run this code chunk, way too many markers
lint(filename = "C:/Users/Shreya/Desktop/STAT 5014/STAT_5014/HW2_Chandrasekharan_Shreya.Rmd")
```

Some things I need to change in my code 1. Spacing seems to be one of the most prominent markers. I will need to start adding spaces after commas and around infix operators. 2. I will need to use only double-quotes as much as possible. 3.I will try not to cross 80 characters per line.

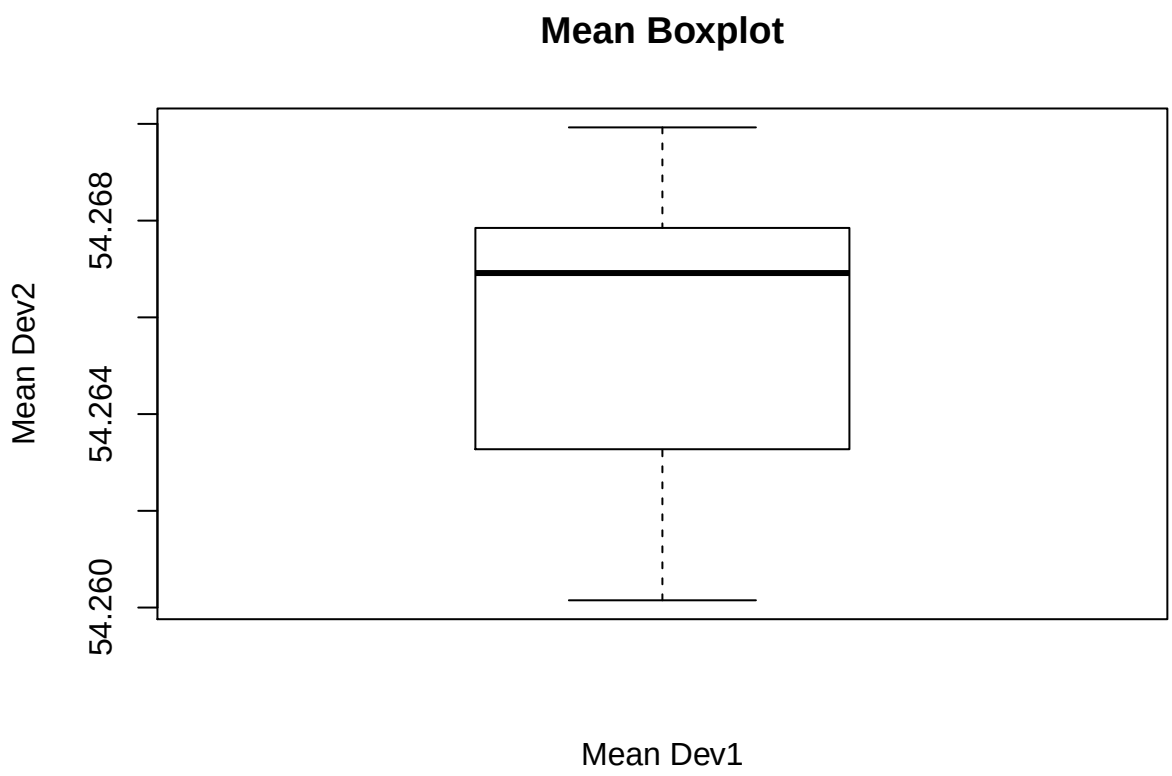
Problem 6

```
#Read the given data

data_rds <-
    readRDS("C:/Users/Shreya/Desktop/STAT 5014/STAT_5014/03_good_programming_R_functions/HW3_data.rds")
mean.d1 <- data.frame()
mean.d2 <- data.frame()
sd.d1 <- data.frame()
sd.d2 <- data.frame()
cor.d12 <- data.frame()

#Set loop for mean, stamdard deviation, and correlation

for(i in 1:13){
    mean_dev1 <- mean(filter(data_rds, Observer == i)[, "dev1"])
    mean_dev2 <- mean(filter(data_rds, Observer == i)[, "dev2"])
    sd_dev1 <- sd(filter(data_rds, Observer == i)[, "dev1"])
    sd_dev2 <- sd(filter(data_rds, Observer == i)[, "dev2"])
    cor_dev1_dev2 <- cor(filter(data_rds, Observer == i)[, "dev1"],
                         filter(data_rds, Observer == i)[, "dev2"])

    #Make one data table from many

    mean.d1 <- rbind(mean.d1, mean_dev1)
    mean.d2 <- rbind(mean.d2, mean_dev2)
    sd.d1 <- rbind(sd.d1, sd_dev1)
    sd.d2 <- rbind(sd.d2, sd_dev2)
    cor.d12 <- rbind(cor.d12, cor_dev1_dev2)
}

#Table

table_rds <- cbind(1:13, mean.d1, mean.d2, sd.d1, sd.d2, cor.d12)
colnames(table_rds) <- c("Observer", "Mean Dev1", "Mean Dev2", "SD Dev1",
                         "SD Dev2", "Correlation Dev1 Dev2" )
knitr::kable(table_rds, caption = "Observed Data")
```

Table 1: Observed Data

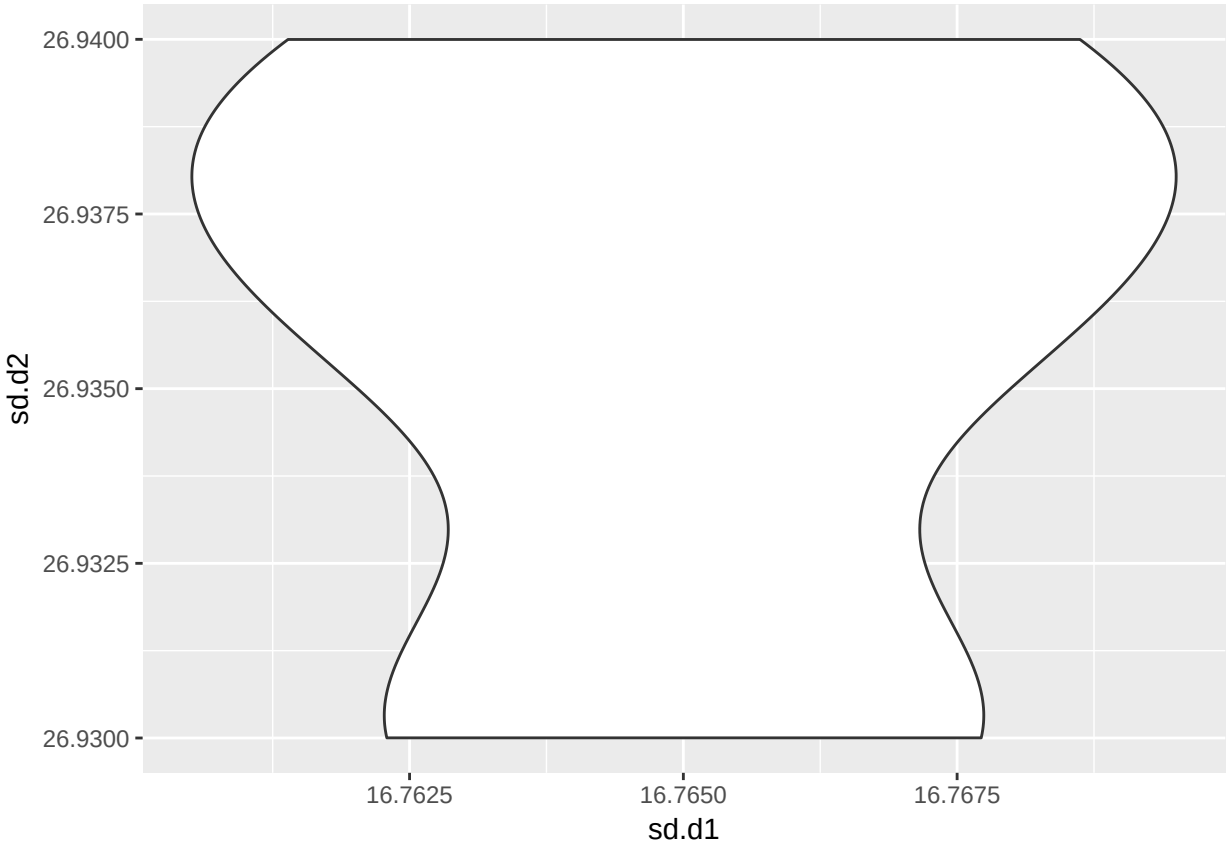
Observer	Mean Dev1	Mean Dev2	SD Dev1	SD Dev2	Correlation Dev1 Dev2
1	54.26610	47.83472	16.76983	26.93974	-0.0641284
2	54.26873	47.83082	16.76924	26.93573	-0.0685864
3	54.26732	47.83772	16.76001	26.93004	-0.0683434
4	54.26327	47.83225	16.76514	26.93540	-0.0644719
5	54.26030	47.83983	16.76774	26.93019	-0.0603414
6	54.26144	47.83025	16.76590	26.93988	-0.0617148
7	54.26881	47.83545	16.76670	26.94000	-0.0685042
8	54.26785	47.83590	16.76676	26.93610	-0.0689797
9	54.26588	47.83150	16.76885	26.93861	-0.0686092
10	54.26734	47.83955	16.76896	26.93027	-0.0629611
11	54.26993	47.83699	16.76996	26.93768	-0.0694456
12	54.26692	47.83160	16.77000	26.93790	-0.0665752
13	54.26015	47.83972	16.76996	26.93000	-0.0655833

```
#Boxplot

boxplot(mean.d1, mean.d2, xlab = "Mean Dev1", ylab = "Mean Dev2", main =  "Mean Boxplot")
```




```
#Violinplot

sd_plot_violin <- ggplot(table_rds, aes(x = sd.d1, y = sd.d2))
sd_plot_violin + geom_violin()
```

```
## Don't know how to automatically pick scale for object of type data.frame. Defaulting to continuous.
## Don't know how to automatically pick scale for object of type data.frame. Defaulting to continuous.
```



Problem 7

```
#Get link

BP_link <- "http://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/BloodPressure.dat"

#CHeck what's needed in data

BP_raw <- read.table(BP_link, header = FALSE, skip = 2, fill = TRUE,
                     stringsAsFactors = FALSE)

BP_tidy <- BP_raw[, -5]

#Column 5 is unneeded

#Change column names for clarity

colnames(BP_tidy) <- c("Day", "Device 1", "Device 2", "Device 3", "Doctor 1",
                       "Doctor 2", "Doctor 3")

#Better presentation
```

```
knitr::kable(BP_tidy, caption = "Blood Pressure Measurement")
```

Table 2: Blood Pressure Measurement

Day	Device 1	Device 2	Device 3	Doctor 1	Doctor 2	Doctor 3
1	133.34	133.36	133.45	126.54	127.36	131.88
2	110.94	110.85	110.92	124.69	128.86	132.39
3	118.54	118.56	118.67	125.46	129.43	134.43
4	137.94	137.80	137.77	125.95	130.72	134.28
5	139.52	139.62	139.59	125.90	130.13	134.44
6	139.23	139.11	139.36	127.85	132.03	137.37
7	117.96	117.81	117.85	125.55	132.05	132.17
8	119.59	119.42	119.48	125.80	129.87	134.97
9	116.12	116.00	115.93	125.11	128.09	133.97
10	128.38	128.48	128.41	125.75	131.94	132.68
11	125.17	125.25	125.34	128.77	130.05	134.75
12	134.62	134.41	134.55	125.26	131.13	134.29
13	136.14	136.07	136.22	126.26	130.91	133.38
14	131.21	131.03	130.96	125.68	128.83	135.67
15	132.51	132.86	132.65	124.47	129.46	134.39

Problem 8

$$f(x) = 3^x - sin(x) + cos(5x) \tag{1}$$

```
#Define function

newt_func <- function(fn, tol = 1E-18, x0 = 1, n = 20){
    h <- 0.001
    i <- 1
    x1 <- x0
    k <- n

#Try to reproduce what I saw online, pray it works

    while(i<=n){
        dx_func <- (fn(x0+h) - fn(x0))/h
    x1 <- (x0 - (fn(x0)/dx_func))
    k[i] < x1
    i <- i + 1
    if (abs(x1-x0) < tol) break
    x0 <- x1
    }
    return(k[1:(i + 1)])
}

fn <- function(x){3^x - sin(x) + cos(5*x)}
a <- newt_func(fn, x0 = 0, n = 5)
a
```

```
## [1]  5 NA NA NA NA NA NA
```

```
#It doesn't work :(

#The code looks like what it's "supposed" to look like (courtesy: A LOT of Google-ing).
#But I am not sure of what output I am getting.
```

Problem 9

```
#Oh no! Not this again!

    Car_Gebreken_raw <- read.csv("Open_Data_RDW__Gebreken.csv",stringsAsFactors = F,
                                 nrows=200, header=T,quote = '"')
    Car_Geconstat_raw <- read.csv("Open_Data_RDW__Geconstateerde_Gebreken.csv",
                                  stringsAsFactors = F, nrows=200, header=T)
    Car_Person_raw <- read.csv("Personenauto_basisdata.csv",stringsAsFactors = F,
                               nrows=200, header=T)

    Car_Gebreken_raw.colclass <- sapply(Car_Gebreken_raw,class)
    Car_Geconstat_raw.colclass <- sapply(Car_Geconstat_raw,class)
    Car_Person_raw.colclass <- sapply(Car_Person_raw,class)

    print("Gebreken")
    print(Car_Gebreken_raw.colclass)
    print("Geconstat")
    print(Car_Geconstat_raw.colclass)
    print("Personen")
    print(Car_Person_raw.colclass)

    #this had the defect code and description
     Car_Gebreken_select <- fread(input = "Open_Data_RDW__Gebreken.csv",
                                  header = T, select=c(1,6), showProgress=F)
    #this has the license plate, inspection date and defect code
    Car_Geconstat_select <- fread(input = "Open_Data_RDW__Geconstateerde_Gebreken.csv",

                                  header=T, select=c(1,3,5),showProgress=F)
    #this has the license plate, make and model of vehicle
    Car_Person_select <- fread(input = "Personenauto_basisdata.csv",
                               header=T, showProgress = F, select = c(1,3,4))

    Car_Geconstat_select_2017 <-
      Car_Geconstat_select[grep("2017",
Car_Geconstat_select$"Meld datum door keuringsinstantie")]
```

```
    merge_License_plate <- merge(Car_Gebreken_select, Car_Geconstat_select,
                                 Car_Person_select, by="Kenteken")
    merge_Defect_code <- merge(Car_Gebreken_select,
                               Car_Geconstat_select, Car_Person_select,
                               by="Gebrek identificatie")
```

CONFESSION: I tried and tried, and got nowhere with this particular question. I have simply copied what I had submitted in my previous homework, which itself was incorrect/incomplete. My biggest challenge still remains merging huge datasets.