

R Homework 5  
Fisher Ankney  
November 27th, 2018  
Statistics 5193

### Question 1

Consider the social media data on Brightspace. See the previous day's script for importing the data. Construct side by side boxplots of Facebook time for STAT2023 students and STAT 5063 students. First, run `demo("colors")` and `help(boxplot)` and read the help file. For full credit...

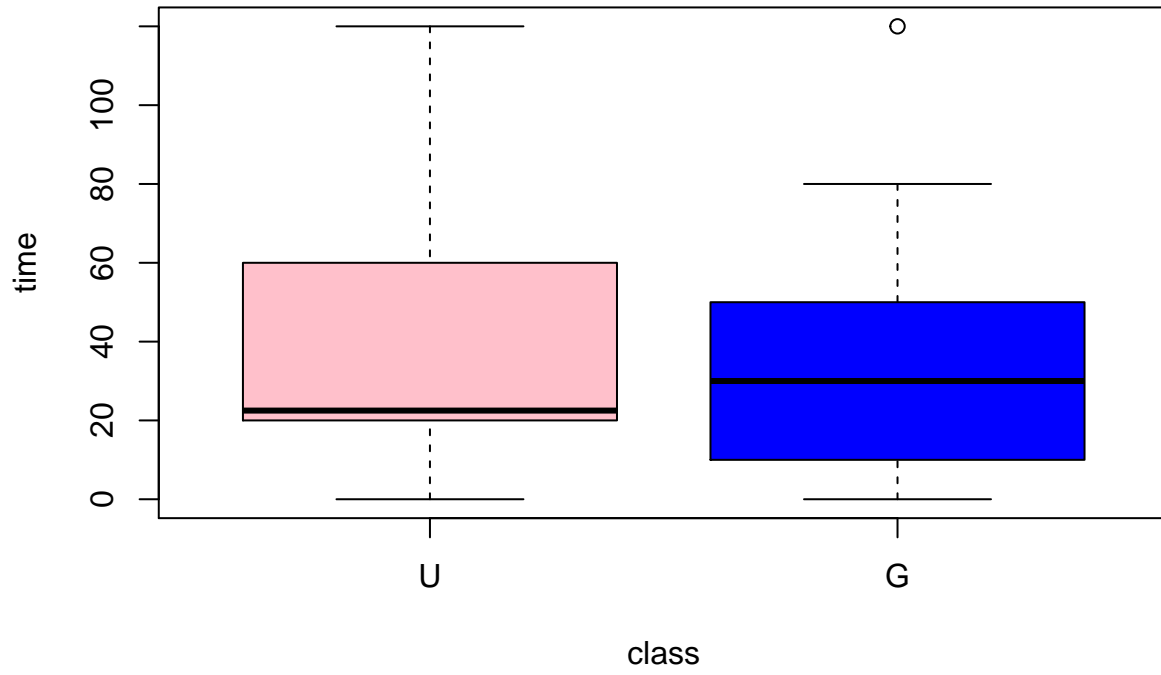
- The boxplots should be labeled “U” and “G” for undergrad and graduate students.
- The undergraduate boxplot should be pink and the graduate boxplot should be blue.
- The y axis must be labeled “time” and the x axis must be labeled “class”.
- The title of the plot should read “Facebook Time vs. Class”

```
library(readxl)
StudentData <- read_excel("/Users/fisher/Documents/data_science/r_stat_5193/data/StudentData.xlsx")

demo("colors")
help(boxplot)

boxplot(StudentData$Fbtime ~ StudentData$Class,
        names = c("U", "G"),
        col = c("pink", 'blue'),
        ylab = "time",
        xlab = 'class',
        main = 'Facebook Time vs. Class')
```

**Facebook Time vs. Class**



## Question 2

Consider the social media data on Brightspace. Construct a scatter plot of Introversion level vs. Facebook time. Hint, before creating the plot, you'll want to create a new data set or vectors with the outlier above removed. See part iv. For full credit...

- The y axis should be labeled “Introversion”, the x axis should be labeled “FB time” and the title should read “FB vs. Introversion vs. Gender”.
- The plotting characters of non-outlier data should be an “M” for Male and “F” for female. Note the pch parameter only accepts vectors (not factors).
- The “M” and “F” points on the plot should be different colors.
- The outlier identified in question 1 should have a different plotting character. Hint: you'll want to use the points() function to add this point to the plot.
- Add 2 least squares regression lines for regressing Introversion level on Facebook time to the plot: one line should be the least squares regression line after the outlier is removed and the other line is the least squares regression line with the outlier still included in the data set. The lines should be different colors and types.

```
StudentDataTrim <- StudentData[-23,]
StudentDataOut <- StudentData[23,]

gender <- as.factor(StudentDataTrim$Gender)

plot(StudentDataTrim$Introvert ~ StudentDataTrim$Fbtime,
     main = 'FB vs. Introversion vs. Gender',
     ylab = 'Introversion',
     xlab = 'FB time',
     pch = StudentDataTrim$Gender,
     col = c("pink", "blue")[gender],
     )

points(StudentDataOut$Introvert ~ StudentDataOut$Fbtime,
       pch = 'O')

with_outlier <- lm(StudentData$Introvert ~ StudentData$Fbtime)
without_outlier <- lm(StudentDataTrim$Introvert ~ StudentDataTrim$Fbtime)

abline(with_outlier, lty = 2, col = "slateblue")
abline(without_outlier, col = "firebrick")
```

FB vs. Introversion vs. Gender

