In Class Exercise 3

Rahul Avadhoot (Last 4 Digits of Student ID: 1787)

October 11, 2017

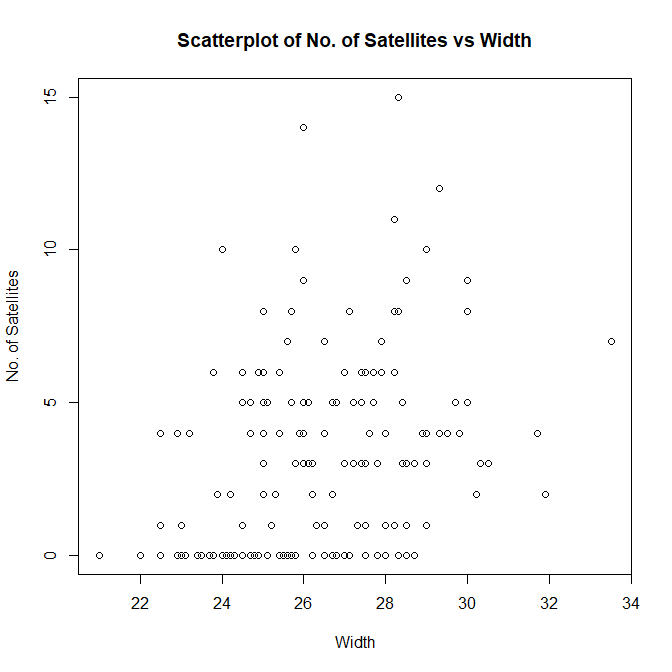
## Question 1

1. The crab.txt dataset will be used for this question. The data are available from canvas or from web: <http://faculty.washington.edu/tathornt/Biost509/DataSets/crab.txt>:

crab <- read.table("http://faculty.washington.edu/tathornt/Biost509/DataSets/crab.txt",header=TRUE)

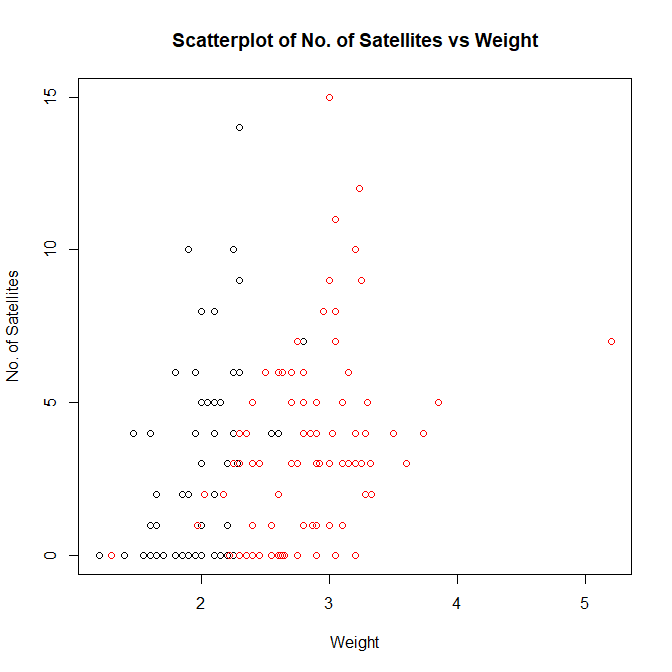
1. Provide a plot of the number of satellites versus width

plot(x=crab$width,y=crab$n.satellites,  
 ylab="No. of Satellites",   
 xlab="Width",  
 main="Scatterplot of No. of Satellites vs Width",  
 pin=1)



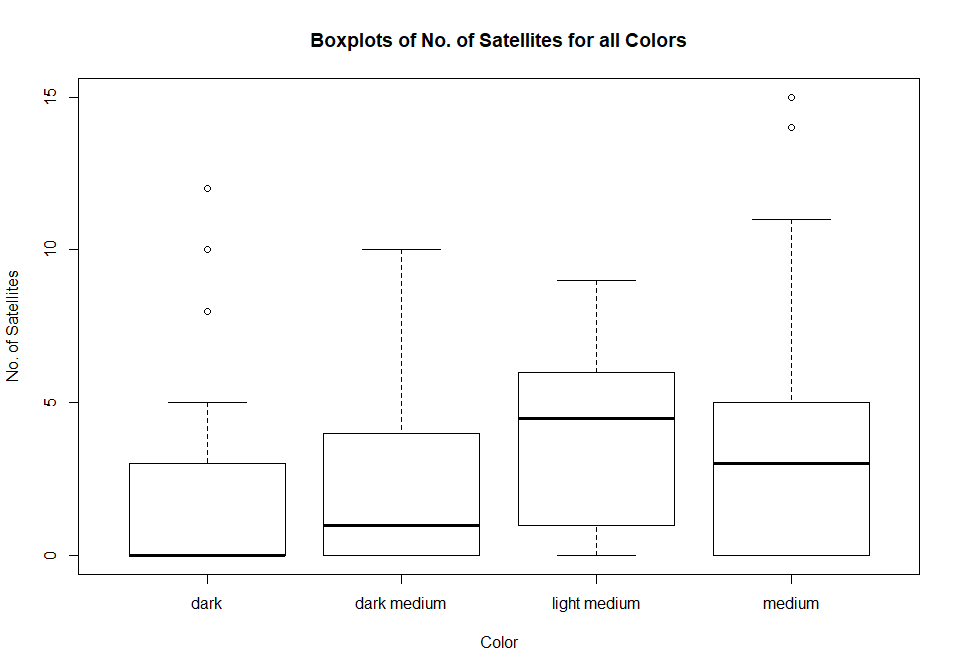
1. Provide a plot of the number of satellites versus weight, and color-code the points differently if their width exceeds 26cm

plot(x=crab$weight,  
 y=crab$n.satellites,  
 ylab="No. of Satellites",   
 xlab="Weight",  
 main="Scatterplot of No. of Satellites vs Weight",  
 col=ifelse(crab$width > 26, "red","black"))



1. Illustrate the distributions of number of satellites, for crabs of different colors. (There are several ways to do this)

boxplot(n.satellites~interaction(color),  
 data=crab,  
 xlab='Color',  
 ylab='No. of Satellites',  
 main='Boxplots of No. of Satellites for all Colors')



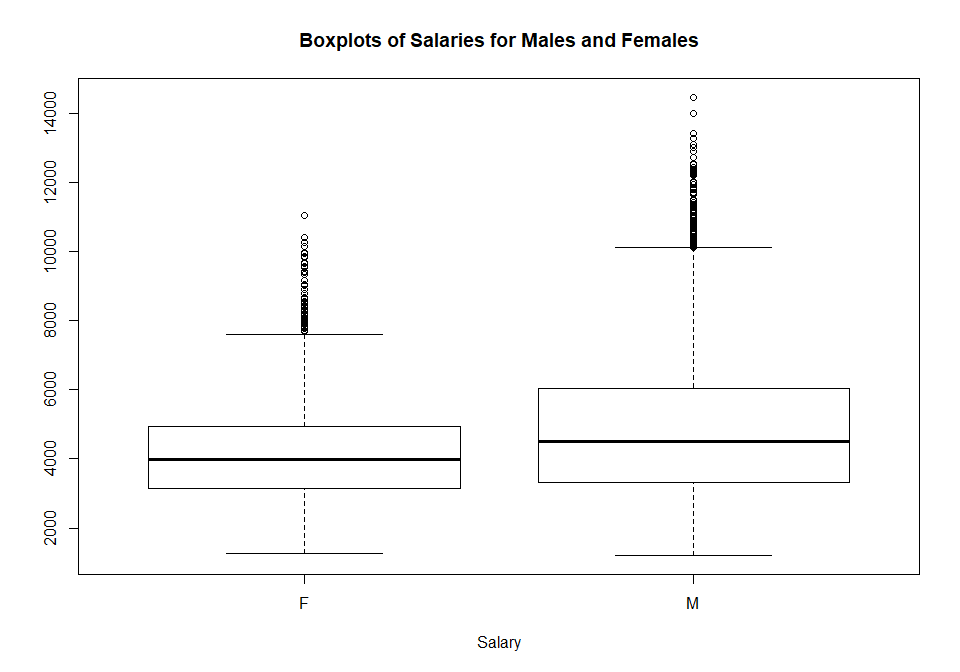
## Question 2

The salary.txt dataset will be used for this question. The data are available from canvas or from the web: <http://faculty.washington.edu/tathornt/Biost509/DataSets/salary.txt>

salary <- read.table("http://faculty.washington.edu/tathornt/Biost509/DataSets/salary.txt",header=TRUE)

1. Provide boxplots of salary by gender. Briefly describe any differences in the average or median distribution of salary by gender.

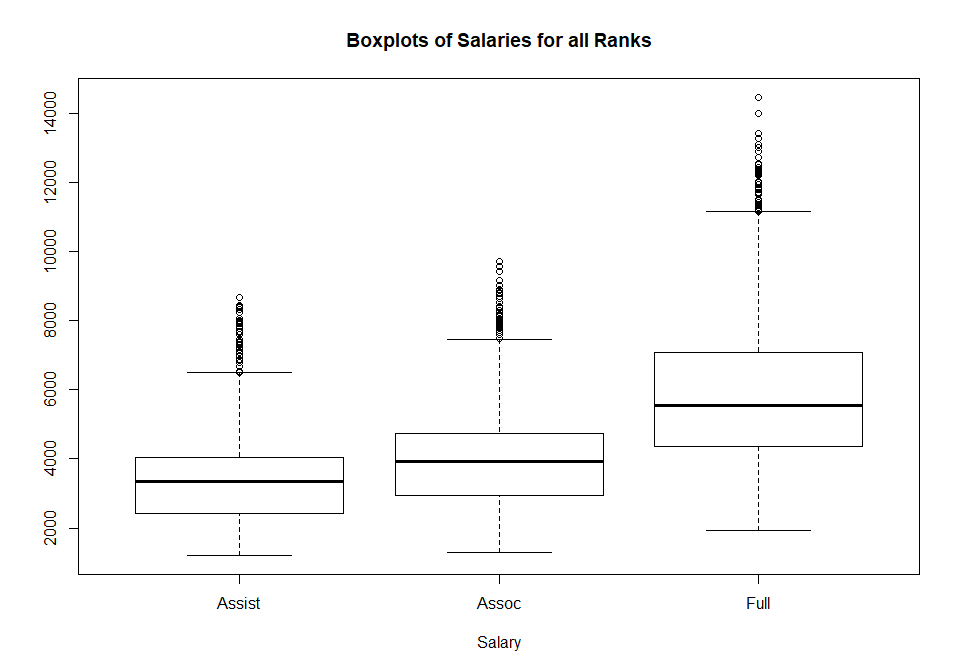
boxplot(salary~interaction(gender),  
 data=salary,  
 xlab="Salary",  
 main="Boxplots of Salaries for Males and Females")



All distributions are left skewed with median salary of Males > Females. The mean and median of male salaries are about $4800 and $4500 respectively. The mean and median of female salaries are about $4200 and $4000 respectively.

1. Provide boxplots of salary by rank. Briefly describe any differences in the average or median distribution of salaries by rank.

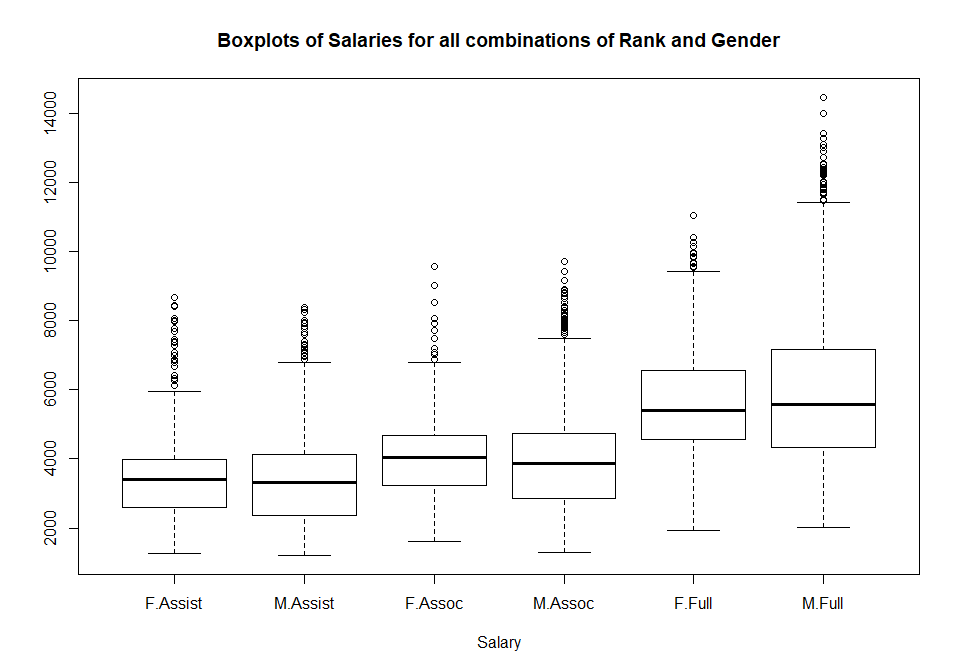
boxplot(salary~interaction(rank),  
 data=salary,  
 xlab="Salary",  
 main="Boxplots of Salaries for all Ranks")



All distributions are left skewed with median salary of Full > Assoc > Assist. Mean and Median of Assistant Professors salaries are about $3400 and $3350 respectively. Mean and Median of Associate Professors salaries are about $3950 and $3900 respectively. Mean and Median of Full Professors salaries are about $5800 and $5550 respectively.

1. Now provide boxplots of salary for each combination of rank and gender using the boxplot() function. [Hint: Use the boxplot() function with the interaction() function for rank and gender: boxplot(salary~interaction(gender,rank),data=salary) ]. Does it appear as if the average or median salary distribution is significantly different between males and females within each rank? Briefly explain.

boxplot(salary~interaction(gender,rank),  
 data=salary,  
 xlab="Salary",  
 main="Boxplots of Salaries for all combinations of Rank and Gender")



The median salary distribution is significantly different between males and females within each rank. It is observed that the median salary of females in the Assistant and Associate Ranks is higher than that of males. But among Full Professors, males have a higher median salary as compared to females.

1. Redo question, but use the varwidth option with the boxplot() function. [Hint: use the help function for boxplot to learn with the varwidth option does.]

boxplot(salary~interaction(gender,rank),  
 data=salary,  
 varwidth=TRUE,  
 xlab="Salary",  
 main="Boxplots for all combinations of Rank and Gender")

