Project

Joey Simone

1. (6 points for collecting data) Load Your Data from excel file into Rstudio.

data=read.csv("data.csv", stringsAsFactors = T)

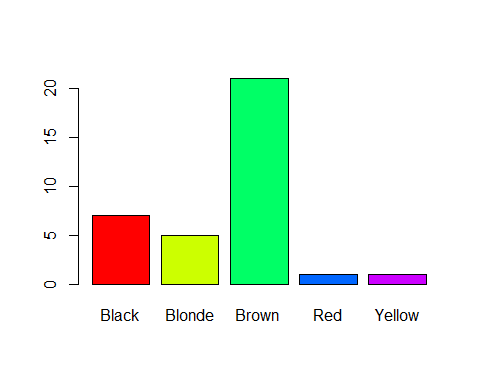
2.Provide summary of data.

summary(data)

## Height\_M Size\_Shoes\_M Color\_Hair\_M Height\_W Size\_Shoes\_W   
## Min. :48.00 Min. : 6.50 Black : 7 Min. :59.00 Min. : 5.000   
## 1st Qu.:67.00 1st Qu.:10.00 Blonde: 5 1st Qu.:62.00 1st Qu.: 7.000   
## Median :69.00 Median :11.00 Brown :21 Median :65.00 Median : 8.500   
## Mean :68.79 Mean :10.84 Red : 1 Mean :64.66 Mean : 8.724   
## 3rd Qu.:71.50 3rd Qu.:12.00 Yellow: 1 3rd Qu.:68.00 3rd Qu.:10.000   
## Max. :80.00 Max. :14.00 Max. :73.00 Max. :14.000   
## NA's :6 NA's :6   
## Color\_Hair\_W Height\_N Size\_Shoes\_N Color\_Hair\_N  
## : 6 Min. :62.00 Min. : 8.50 :33   
## Black : 5 1st Qu.:62.75 1st Qu.:10.12 Black : 1   
## Blonde: 2 Median :63.50 Median :11.75 Purple: 1   
## Brown :21 Mean :63.50 Mean :11.75   
## Red : 1 3rd Qu.:64.25 3rd Qu.:13.38   
## Max. :65.00 Max. :15.00   
## NA's :33 NA's :33

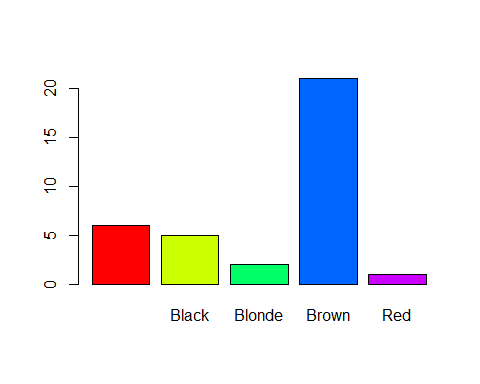
3.(2points) Create a bar plot for hair color of men.

mencol=table(data$Color\_Hair\_M)  
barplot(mencol,  
 col=rainbow(length(mencol)))



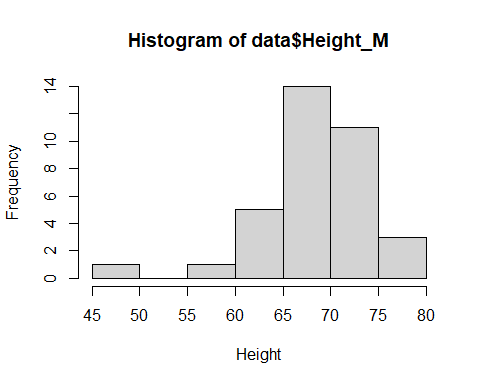
4.(2points) Create a pie chart for hair color of women.

womencol=table(data$Color\_Hair\_W)  
barplot(womencol,  
 col=rainbow(length(womencol)))



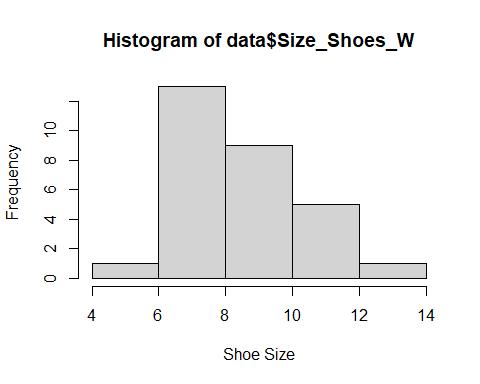
5.(2points) Construct histogram for height of men. (Number of bars can be choosen by default in R-studio)

hist(data$Height\_M,  
 xlab = "Height",  
 freq = T)



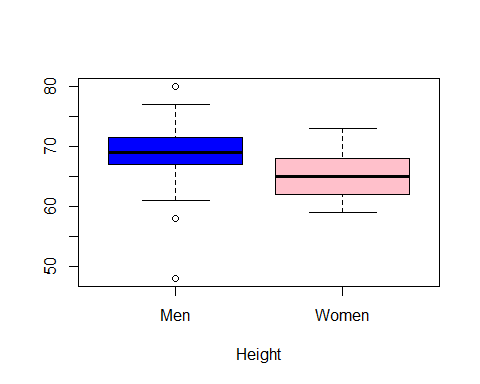
6.(2points) Construct histogram for shoe size for women. (Number of bars can be choosen by default in R-studio)

hist(data$Size\_Shoes\_W,  
 xlab = "Shoe Size",  
 freq = T)



7.(2points) Construct two Box Plots (one for height of men and another for height of women) on the same plot.

boxplot(data$Height\_M,  
 data$Height\_W,  
   
 names = c("Men", "Women"),  
 xlab = "Height",  
 col = c("Blue", "Pink"),  
 horizontal = F)

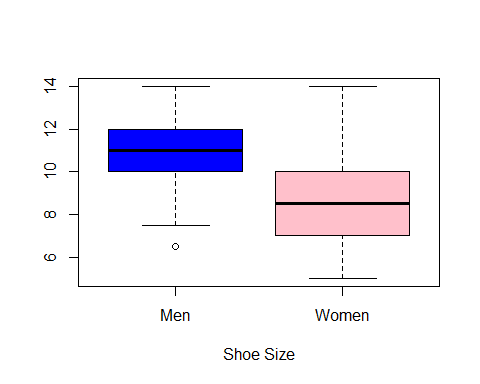


8.(1points) Compare these two box plots and describe the differences between these two box plots. Does any boxplot have outliers? If so, how many outliers?

Women are shorter in all quartiles, only men have outliers, and have a wider range overall. Min height woman shorter than min height man, q1 woman shorter than q1 man, etc. Men have 3 outliers, women have no outliers.

9.(2points) Construct two Box Plots (one for size of shoes for men and another for size of shoes for women) on the same plot.

boxplot(data$Size\_Shoes\_M,  
 data$Size\_Shoes\_W,  
   
 names = c("Men", "Women"),  
 xlab = "Shoe Size",  
 col = c("Blue", "Pink"),  
 horizontal = F)



10.(1points) Compare these two box plots and describe the differences between these two box plots. Does any boxplot have outliers? If so, how many outliers?

Men have one outlier in small, women have wider range overall, and smaller at all critical points

11.(2points) Construct a 90% confidence interval for the estimate of mean of men’s height.

t.test(data$Height\_M, conf.level=0.90)

##   
## One Sample t-test  
##   
## data: data$Height\_M  
## t = 69.322, df = 34, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 90 percent confidence interval:  
## 67.10788 70.46355  
## sample estimates:  
## mean of x   
## 68.78571

**Write a statement that correctly interprets the confidence interval.**

We estimate with 90% confidence that mean height of population of men is between 67.1 and 70.5 inches

12.(2points) Construct a 95% confidence interval for the estimate of mean of women’s height.

t.test(data$Height\_W, conf.level=0.95)

##   
## One Sample t-test  
##   
## data: data$Height\_W  
## t = 94.197, df = 28, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 63.24918 66.06117  
## sample estimates:  
## mean of x   
## 64.65517

**Write a statement that correctly interprets the confidence interval.**

We estimate from sample, with 95% confidence, that mean population height of women is between 64.2 inches and 66.1 inches

13.(2points) Construct a 94% confidence interval for the estimate of mean of shoe size for women.

t.test(data$Size\_Shoes\_W, conf.level=0.94)

##   
## One Sample t-test  
##   
## data: data$Size\_Shoes\_W  
## t = 25.467, df = 28, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 94 percent confidence interval:  
## 8.052649 9.395627  
## sample estimates:  
## mean of x   
## 8.724138

#######################  
t.test(data$Size\_Shoes\_M, conf.level=0.94)

##   
## One Sample t-test  
##   
## data: data$Size\_Shoes\_M  
## t = 33.075, df = 34, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 94 percent confidence interval:  
## 10.20503 11.48069  
## sample estimates:  
## mean of x   
## 10.84286

**Write a statement that correctly interprets the confidence interval.**

From 94% confidence Interval Margin of error, Mean Woman’s Shoe is between 8.1 and 9.4, but Mean Man’s Shoe is between 10.2 and 11.5

1. (4 points) Write two claims for any two different variables of your data (Height\_M or Height\_W or Size\_Shoes\_M or Size\_Shoes\_W). Make your claim as alternative hypothesis and show that a Null hypothesis is wrong, so it means that an alternative hypothesis (a claim) is true. Each claim should have different comparison sign: less, greater or not equal.

Please indicate: 1) What is your claim? 2) What is the opposite statement? 3) What is null and alternative hypothesis. 4) What is p-value. 5) Write conclusion about your claim.

Example:

1. Claim: The average (mean) size of shoes for women is less than 9.
2. Opposite statement: The average size of shoes for women is greater or equal 9.
3. H\_0: mean=9 H\_1: mean<9
4. Creat R-code to find p-value

t.test(data$Size\_Shoes\_W, mu=9, alternative=“less”)

1. Write Conclusion: Since p-value = 0.001507<alpha=0.05, we reject H\_0. We have enough evidence to support the claim.

Part a)

**1) Claim:**  Mean Woman under 6 feet tall (72 inches)

**2) Opposite statement:** Mean Woman greater than or equal to 72 inch

**3) H\_0: 72=mean ;H\_1: 72>mean**

**4) Creat R-code to find p-value**

**5) Write conclusion about your claim.**

Part b)

**1) Claim:**  Mean Man’s Shoe Greater than Size 9

**2) Opposite statement:** Mean Man’s Shoe less than or equal to 9

**3) H\_0: 9=mean ; H\_1: 9>mean**

**4) Creat R-code to find p-value**

**5) Write conclusion about your claim.**

15.(2points) Compare mean of height of men and mean of height of women. You can make a claim with any sign that you think is true (less, greater or not equal).

Please indicate:

1. What is your claim?
2. What is your opposite statement?
3. What is null and alternative hypothesis.
4. What is p-value.
5. Write your conclusion about your claim.

Example:

1. Claim: The mean height of men is greater than the mean height of women.
2. Opposite statement: The mean height of men is less or equal to the mean height of women.
3. H\_0: mean of height of men=mean of height of women.

* H\_1: mean of height of men>mean of height of women.

1. You type R-code to find p-value

t.test(dataHeight\_W, alternative=“greater”)

1. Write conclusion: Since p-value = 1.571e-14 < alpha=0.05, we reject H\_0. We have enough evidence to support the claim.

**1) Claim:**

**2) Opposite statement:**

**3) H\_0: mean ; H\_1: mean**

**4) Creat R-code to find p-value**

**5) Write conclusion about your claim.**

16.(2points) Compare mean of shoe size of men and mean of shoe size of women. You can make a claim with any sign that you think is true (less, greater or not equal).

**1) Claim:**

**2) Opposite statement:**

**3) H\_0: mean ; H\_1: mean**

**4) Creat R-code to find p-value**

**5) Write conclusion about your claim.**

1. (2points) Construct a scatter plot for height of men (x-axis) and size of men’s shoes (y-axis)
2. (2points) Find a linear correlation coefficient between height of men and size of men’s shoes .

**What is a linear correlation coefficient?**

1. (2points) Find regression line for height of men (x-variable) and size of men’s shoes (y-variable)

**What is a regression line?**

1. (2points) Construct a scatter plot for height of men (x-axis) and size of men’s shoes (y-axis) and the regression line on the same plot.
2. (2points) Construct a scatter plot for height of women (x-axis) and size of women’s shoes (y-axis)
3. (2points) Find a linear correlation coefficient between height of women and size of women’s shoes .

**What is a linear correlation coefficient? What is the strength of relationship?**

1. (2points) Find regression line for height of women (x-variable) and size of women’s shoes (y-variable)

**What is a regression line?**

1. (2points) Construct a scatter plot for height of women (x-axis) and size of women’s shoes (y-axis) and the regression line on the same plot.