Project

Joey Simone

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

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

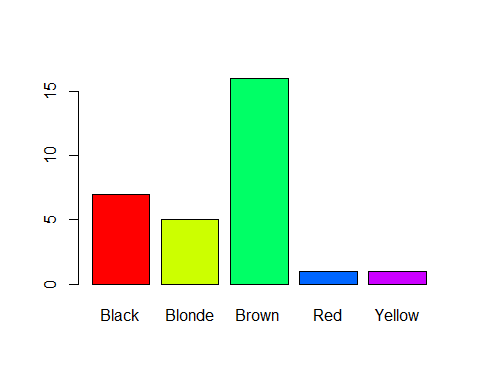
2.Provide summary of data.

summary(data)

## Height\_M Size\_Shoes\_M Color\_Hair\_M Height\_W Size\_Shoes\_W   
## Min. :48.00 Min. : 6.500 Black : 7 Min. :59.0 Min. : 5.00   
## 1st Qu.:66.25 1st Qu.: 9.125 Blonde: 5 1st Qu.:62.0 1st Qu.: 7.25   
## Median :69.50 Median :11.000 Brown :16 Median :65.0 Median : 8.50   
## Mean :68.68 Mean :10.550 Red : 1 Mean :64.8 Mean : 8.80   
## 3rd Qu.:71.75 3rd Qu.:12.000 Yellow: 1 3rd Qu.:68.0 3rd Qu.:10.00   
## Max. :80.00 Max. :14.000 Max. :73.0 Max. :14.00   
## Color\_Hair\_W  
## Black : 5   
## Blonde: 2   
## Brown :22   
## Red : 1   
##   
##

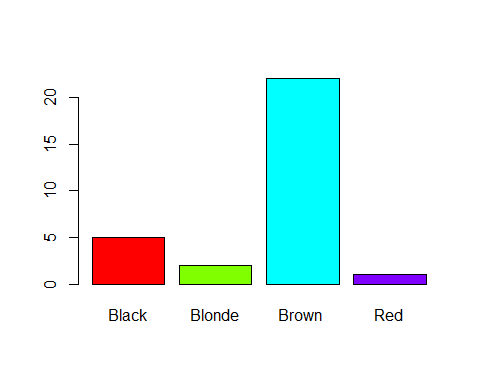
3.(2points) Create a bar plot for hair color of men.

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



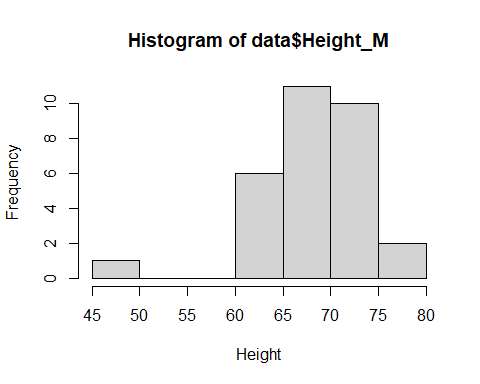
4.(2points) Create a pie chart for hair color of women.

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



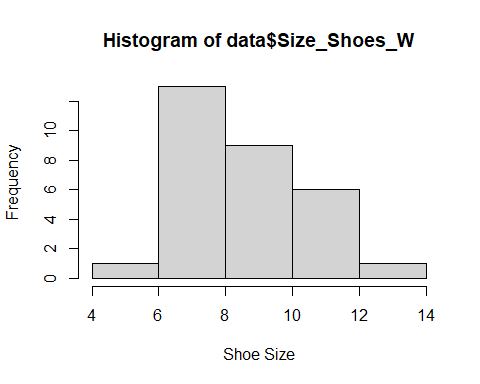
5.(2points) Construct histogram for height of men. (Number of bars can be choosen by default in R-studio)

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



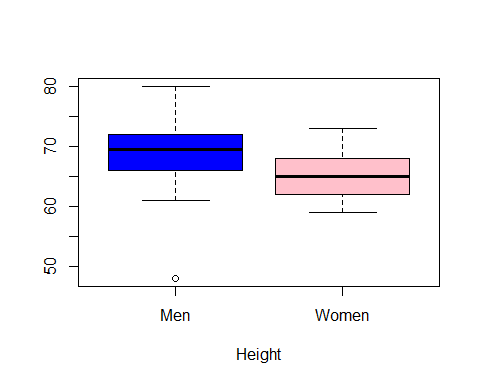
6.(2points) Construct histogram for shoe size for women. (Number of bars can be choosen by default in R-studio)

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)

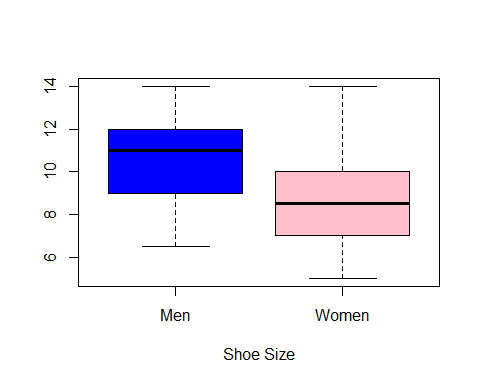


8.(1points) Compare these two box plots and describe the differences between these two box plots. Does any boxplot have outliers? If so, how many outliers?

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?

No outliers, men shoe size is larger, women’s shoes have more range, neither plot has outliers

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 = 63.363, df = 29, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 90 percent confidence interval:  
## 66.84153 70.52514  
## sample estimates:  
## mean of x   
## 68.68333

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

We estimate with 90% confidence that mean height of population of men is between 66.8 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 = 95.471, df = 29, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 63.41182 66.18818  
## sample estimates:  
## mean of x   
## 64.8

**Write a statement that correctly interprets the confidence interval.** We estimate from sample, with 95% confidence, that mean population height of women is between 63.4 inches and 66.2 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 = 25.917, df = 29, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 94 percent confidence interval:  
## 8.135421 9.464579  
## sample estimates:  
## mean of x   
## 8.8

**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 9.5

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**

t.test(data$Height\_W, mu=72, alternative = "less")

##   
## One Sample t-test  
##   
## data: data$Height\_W  
## t = -10.608, df = 29, p-value = 8.508e-12  
## alternative hypothesis: true mean is less than 72  
## 95 percent confidence interval:  
## -Inf 65.95327  
## sample estimates:  
## mean of x   
## 64.8

**5) Write conclusion about your claim.**  P value .000000000008508<.05 alpha, so we can reject null hypothesis and accept alternative hypothesis. We have enough evidence to support claim that average woman in population is under 6 feet tall.

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**

t.test(data$Size\_Shoes\_M, mu=9, alternative = "greater")

##   
## One Sample t-test  
##   
## data: data$Size\_Shoes\_M  
## t = 4.0945, df = 29, p-value = 0.0001547  
## alternative hypothesis: true mean is greater than 9  
## 95 percent confidence interval:  
## 9.906785 Inf  
## sample estimates:  
## mean of x   
## 10.55

**5) Write conclusion about your claim.**  P value .00015 < alpha .05, reject null, accept alternative hypothesis, mens shoe population mu greater than 9.

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. Null Hypothesis is opposite statement, men women. Alternative Hypothesis Men > Women
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:**

Men Mu Height > Women Mu Height **2) Opposite statement:**

Men Mu Height Women Mu Height **3) H\_0: ; H\_1:**

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

t.test(data$Height\_M, data$Height\_W, alternative = "greater")

##   
## Welch Two Sample t-test  
##   
## data: data$Height\_M and data$Height\_W  
## t = 3.0364, df = 48.711, p-value = 0.001919  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## 1.738885 Inf  
## sample estimates:  
## mean of x mean of y   
## 68.68333 64.80000

**5) Write conclusion about your claim.**  p-values .0019 < alpha .05, reject , support , support claim, we have 95% confidence that men are taller than women

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: Men’s Shoe not equal Women’s Shoe**

**2) Opposite statement:Men’s Shoe equals Women’s shoe**

**3) : ; :**

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

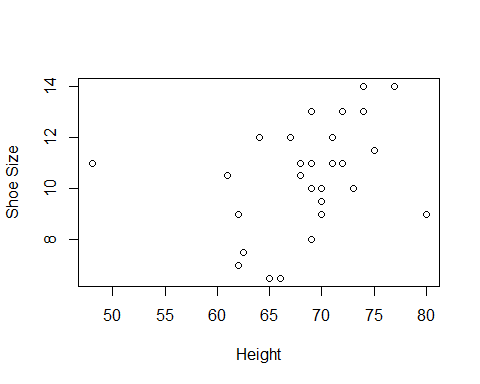
t.test(data$Size\_Shoes\_M, data$Size\_Shoes\_W, alternative = "two.sided")

##   
## Welch Two Sample t-test  
##   
## data: data$Size\_Shoes\_M and data$Size\_Shoes\_W  
## t = 3.4414, df = 57.327, p-value = 0.001087  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 0.7318334 2.7681666  
## sample estimates:  
## mean of x mean of y   
## 10.55 8.80

**5) Write conclusion about your claim.**  p value .001 is less than .05 alpha, reject null hypothesis, support alternative hypothesis, we have enough evidence to prove that men’s average shoe is not equal to women’s average shoe.

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

plot(data$Height\_M, data$Size\_Shoes\_M,  
 xlab = "Height",  
 ylab = "Shoe Size")



1. (2points) Find a linear correlation coefficient between height of men and size of men’s shoes .

cor.test(data$Height\_M, data$Size\_Shoes\_M)

##   
## Pearson's product-moment correlation  
##   
## data: data$Height\_M and data$Size\_Shoes\_M  
## t = 1.9416, df = 28, p-value = 0.06231  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.01803869 0.62693476  
## sample estimates:  
## cor   
## 0.3444692

**What is a linear correlation coefficient?** .34, moderate correlation.

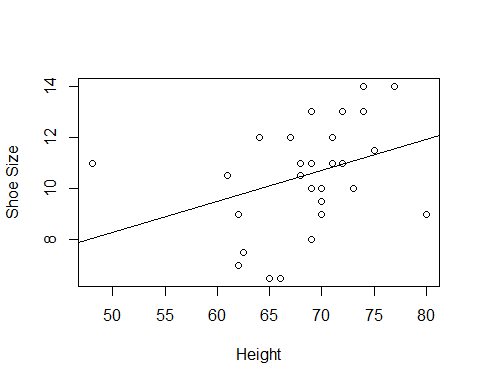
1. (2points) Find regression line for height of men (x-variable) and size of men’s shoes (y-variable)

menshoeline=lm(data$Size\_Shoes\_M~data$Height\_M)  
menshoeline

##   
## Call:  
## lm(formula = data$Size\_Shoes\_M ~ data$Height\_M)  
##   
## Coefficients:  
## (Intercept) data$Height\_M   
## 2.2874 0.1203

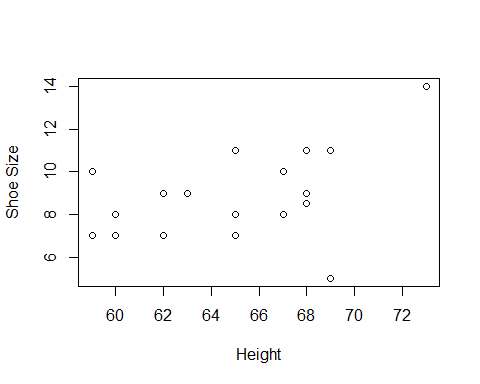
**What is a regression line?** 20) (2points) Construct a scatter plot for height of men (x-axis) and size of men’s shoes (y-axis) and the regression line on the same plot.

plot(data$Height\_M, data$Size\_Shoes\_M,  
 xlab = "Height",  
 ylab = "Shoe Size")  
  
abline(menshoeline)



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

plot(data$Height\_W, data$Size\_Shoes\_W,  
 xlab = "Height",  
 ylab = "Shoe Size")



1. (2points) Find a linear correlation coefficient between height of women and size of women’s shoes .

cor.test(data$Height\_W, data$Size\_Shoes\_W)

##   
## Pearson's product-moment correlation  
##   
## data: data$Height\_W and data$Size\_Shoes\_W  
## t = 3.3742, df = 28, p-value = 0.002183  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.2200011 0.7522196  
## sample estimates:  
## cor   
## 0.5376559

**What is a linear correlation coefficient? What is the strength of relationship?**  .53, Moderate Correlation

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

womenshoeline=lm(data$Size\_Shoes\_W~data$Height\_W)  
womenshoeline

##   
## Call:  
## lm(formula = data$Size\_Shoes\_W ~ data$Height\_W)  
##   
## Coefficients:  
## (Intercept) data$Height\_W   
## -8.629 0.269

**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.

plot(data$Height\_W, data$Size\_Shoes\_W,  
 xlab = "Height",  
 ylab = "Shoe Size")  
  
abline(womenshoeline)

