Quiz 6 (R-Studio)

Joey Simone

Given data set with the ages of the best actors and best actresses.

1. Load Data from excel file into Rstudio.

data=read.csv("income.csv")

1. Provide summary of data.

summary(data)

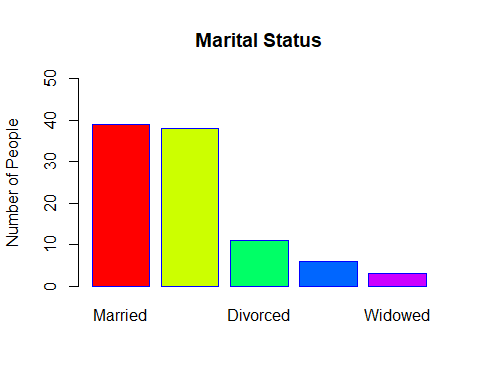
## Age Marital\_Status Degree Gender   
## Min. :23.00 Length:97 Length:97 Length:97   
## 1st Qu.:32.00 Class :character Class :character Class :character   
## Median :40.00 Mode :character Mode :character Mode :character   
## Mean :40.74   
## 3rd Qu.:47.00   
## Max. :79.00   
## Years\_of\_School Income   
## Min. :11.00 Min. : 4999   
## 1st Qu.:12.00 1st Qu.: 19999   
## Median :16.00 Median : 44999   
## Mean :15.49 Mean : 45568   
## 3rd Qu.:17.00 3rd Qu.: 69999   
## Max. :24.00 Max. :108000

1. Create bar plot for Marital Status variable.

marital\_freq <- table(data$Marital\_Status)   
marital\_freq

##   
## Divorced Married Never Married Separated Widowed   
## 11 39 38 6 3

barplot(marital\_freq[order(marital\_freq, decreasing = T)],  
   
 col = rainbow(5),   
 border = "blue",   
 ylim=c(0,50),  
 main = "Marital Status",  
 ylab = "Number of People",   
   
   
   
   
   
 )

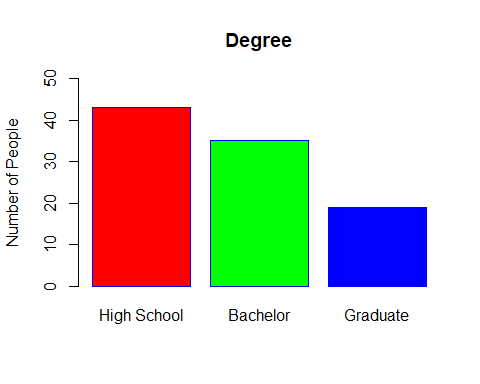


1. (2 point) Create bar plot for Degree variable.

degree\_freq <- table(data$Degree)   
degree\_freq

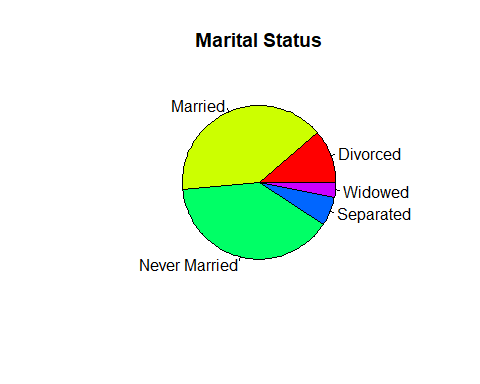
##   
## Bachelor Graduate High School   
## 35 19 43

barplot(degree\_freq[order(degree\_freq, decreasing = T)],  
   
 col = rainbow(3),   
 border = "blue",   
 ylim=c(0,50),  
 main = "Degree",  
 ylab = "Number of People",   
)



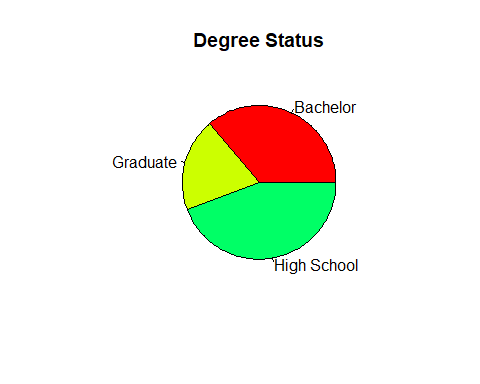
1. Create pie chart for Marital Status variable.

pie(marital\_freq,  
 col = rainbow(5),   
   
 main = "Marital Status")



1. (2 point) Create pie chart for Degree variable.

pie(degree\_freq,  
 col = rainbow(5),   
   
 main = "Degree Status")

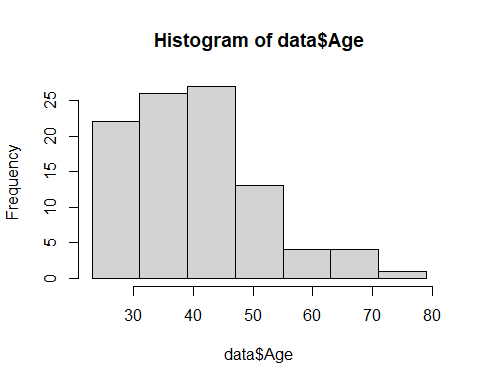


1. (2 point) Construct histogram with seven bars for people’s ages.

age\_freq <- table(data$Age)   
age\_freq

##   
## 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48   
## 3 3 2 4 1 1 5 2 1 4 4 2 2 2 8 3 1 2 3 4 4 2 4 2 6 1   
## 49 50 51 53 55 59 61 62 66 68 79   
## 1 2 3 5 1 1 2 1 1 3 1

hist(data$Age,  
 breaks =seq(min(data$Age),  
 max(data$Age), length.out=8))



1. (1 point) Find the mode for people age’s.

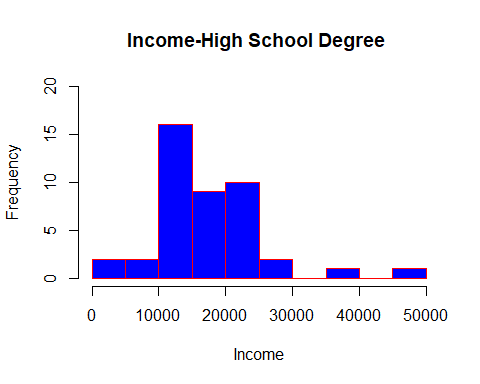
mode=table(as.vector(data$Age))  
names(mode)[mode==max(mode)]

## [1] "37"

**The mode is**  37

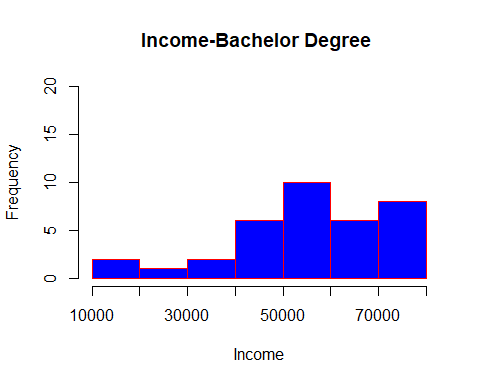
1. Construct histogram for people’s income when people have high school degree. (number of bars can be chosen by default in R-studio) .

hist(data$Income[data$Degree=="High School"],   
   
 main="Income-High School Degree ",  
 ylab="Frequency",  
 xlab="Income",  
 ylim=c(0,20),  
 col="Blue",   
 border="red")



1. (2 point) Construct histogram for people’s income when people have Bachelor Degree (number of bars can be chosen by default in R-studio) .

hist(data$Income[data$Degree=="Bachelor"],   
   
 main="Income-Bachelor Degree ",  
 ylab="Frequency",  
 xlab="Income",  
 ylim=c(0,20),  
 col="Blue",   
 border="red")



1. (2 point) Construct boxplot for people’s income when people have high school degree.

# Find IQR  
IQR(data$Income [data$Degree=="High School"])

## [1] 9999.5

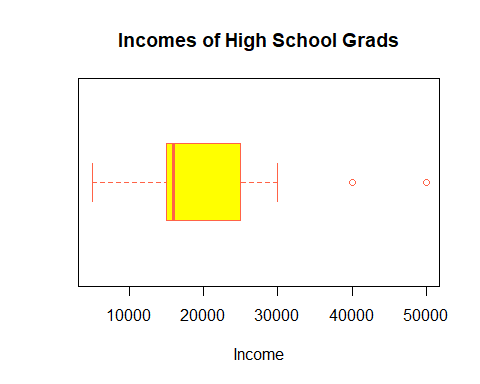
#Find outliers:  
# 1) outliers<Q\_1-1.5\*IQR  
outlier\_min\_value=quantile(data$Income [data$Degree=="High School"], 0.25)-1.5\*IQR(data$Income [data$Degree=="High School"])  
outlier\_min\_value

## 25%   
## 0.75

data[data<outlier\_min\_value]

## character(0)

boxplot(data$Income [data$Degree=="High School"],  
 main="Incomes of High School Grads",  
 xlab="Income",  
 # ylim=c(50,110),  
   
 col="yellow",  
 border= "tomato1" ,  
 horizontal = T)



# 2)outliers<Q\_3+1.5\*IQR  
outlier\_max\_value=quantile(data$Income [data$Degree=="High School"], 0.75)+1.5\*IQR(data$Income [data$Degree=="High School"])  
outlier\_max\_value

## 75%   
## 39998.75

data[data>outlier\_max\_value]

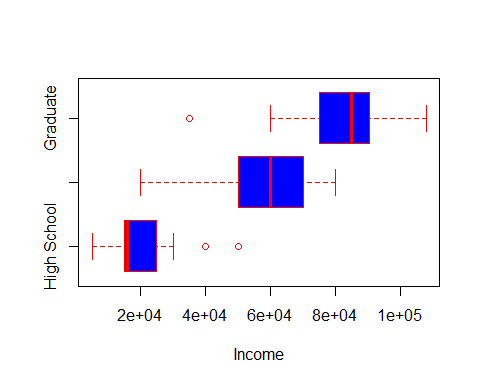
## [1] "Married" "Separated" "Separated" "Never Married"  
## [5] "Separated" "Never Married" "Divorced" "Never Married"  
## [9] "Divorced" "Married" "Never Married" "Married"   
## [13] "Divorced" "Never Married" "Married" "Never Married"  
## [17] "Married" "Never Married" "Never Married" "Never Married"  
## [21] "Never Married" "Never Married" "Divorced" "Married"   
## [25] "Never Married" "Never Married" "Never Married" "Never Married"  
## [29] "Never Married" "Never Married" "Married" "Divorced"   
## [33] "Never Married" "Never Married" "Never Married" "Divorced"   
## [37] "Married" "Divorced" "Never Married" "Never Married"  
## [41] "Married" "Married" "Divorced" "Married"   
## [45] "Married" "Married" "Never Married" "Married"   
## [49] "Separated" "Married" "Married" "Never Married"  
## [53] "Never Married" "Never Married" "Married" "Married"   
## [57] "Married" "Never Married" "Separated" "Never Married"  
## [61] "Never Married" "Married" "Married" "Married"   
## [65] "Widowed" "Separated" "Never Married" "Never Married"  
## [69] "Never Married" "Married" "Never Married" "Married"   
## [73] "Married" "Married" "Married" "Never Married"  
## [77] "Never Married" "Widowed" "Married" "Married"   
## [81] "Married" "Married" "Married" "Married"   
## [85] "Widowed" "Divorced" "Never Married" "Never Married"  
## [89] "Married" "Married" "Never Married" "Married"   
## [93] "Divorced" "Married" "Married" "Married"   
## [97] "Divorced" "High School" "Bachelor" "High School"   
## [101] "High School" "High School" "Graduate" "Graduate"   
## [105] "Bachelor" "Bachelor" "High School" "Graduate"   
## [109] "High School" "High School" "High School" "High School"   
## [113] "High School" "Bachelor" "Bachelor" "Bachelor"   
## [117] "Bachelor" "Bachelor" "High School" "Graduate"   
## [121] "Bachelor" "High School" "Graduate" "Bachelor"   
## [125] "Bachelor" "High School" "High School" "High School"   
## [129] "Bachelor" "Bachelor" "Bachelor" "High School"   
## [133] "Graduate" "High School" "High School" "High School"   
## [137] "Bachelor" "High School" "High School" "Bachelor"   
## [141] "Bachelor" "Graduate" "Bachelor" "Bachelor"   
## [145] "High School" "High School" "High School" "High School"   
## [149] "High School" "High School" "High School" "Bachelor"   
## [153] "High School" "High School" "High School" "Bachelor"   
## [157] "High School" "High School" "High School" "High School"   
## [161] "High School" "Graduate" "Graduate" "Graduate"   
## [165] "High School" "High School" "Graduate" "High School"   
## [169] "Graduate" "Graduate" "Bachelor" "Bachelor"   
## [173] "Graduate" "Bachelor" "High School" "Bachelor"   
## [177] "High School" "Graduate" "Bachelor" "Bachelor"   
## [181] "High School" "Bachelor" "Bachelor" "Bachelor"   
## [185] "Bachelor" "High School" "Graduate" "Bachelor"   
## [189] "Graduate" "Bachelor" "Bachelor" "Graduate"   
## [193] "Graduate" "Bachelor" "Female" "Female"   
## [197] "Male" "Female" "Female" "Female"   
## [201] "Female" "Female" "Female" "Male"   
## [205] "Male" "Male" "Male" "Male"   
## [209] "Female" "Female" "Female" "Male"   
## [213] "Female" "Male" "Female" "Male"   
## [217] "Female" "Female" "Female" "Female"   
## [221] "Male" "Male" "Male" "Male"   
## [225] "Male" "Male" "Male" "Female"   
## [229] "Female" "Female" "Male" "Male"   
## [233] "Female" "Female" "Female" "Male"   
## [237] "Female" "Male" "Male" "Male"   
## [241] "Female" "Male" "Male" "Male"   
## [245] "Male" "Male" "Female" "Female"   
## [249] "Female" "Male" "Female" "Female"   
## [253] "Male" "Female" "Male" "Female"   
## [257] "Male" "Female" "Male" "Female"   
## [261] "Female" "Male" "Male" "Female"   
## [265] "Male" "Male" "Female" "Male"   
## [269] "Female" "Male" "Male" "Male"   
## [273] "Female" "Male" "Female" "Male"   
## [277] "Male" "Female" "Female" "Female"   
## [281] "Male" "Male" "Female" "Male"   
## [285] "Male" "Female" "Female" "Female"   
## [289] "Male" "Female" "Female" " 40000"   
## [293] " 59999" " 74999" " 59999" " 50000"   
## [297] " 49999" " 59999" " 39999" " 64999"   
## [301] " 74999" " 59999" " 89999" " 59999"   
## [305] " 44999" " 69999" " 59999" " 49999"   
## [309] " 64999" " 59999" " 75000" " 74999"   
## [313] " 90999" " 54999" " 44999" " 44999"   
## [317] " 39999" " 99999" " 84999" " 74999"   
## [321] " 99999" " 88000" " 79000" " 54999"   
## [325] " 66000" " 76000" " 66000" " 80000"   
## [329] "108000" " 78000" " 74999" " 76999"   
## [333] " 74999" " 74999" " 59999" "104999"   
## [337] " 69999" " 84999" " 59999" " 49999"   
## [341] " 87999" " 78000" " 68000"

**Does boxplot have outliers? If so, how many outliers?**

Yes, two high school grads, at 40k and 50k

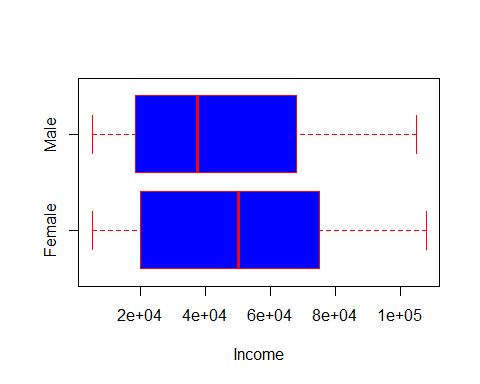
1. Construct three Box Plots for people’s income (one with people who have High School Degree, the other who have Bachelor degree and last one who have Graduate degree) on the same plot.

boxplot(data$Income[data$Degree=="High School"],   
 data$Income[data$Degree=="Bachelor"],  
 data$Income[data$Degree=="Graduate"],   
 names=c("High School", "Bachelor","Graduate" ),  
   
 xlab="Income",  
 col="Blue",   
 border="red",  
 horizontal = T)



13 (2 point) Construct two Box Plots for people’s income (one for Female and the other for Male) on the same plot.

boxplot(data$Income[data$Gender=="Female"],   
 data$Income[data$Gender=="Male"],  
 names=c("Female", "Male"),  
   
 xlab="Income",  
 col="Blue",   
 border="red",  
 horizontal = T)

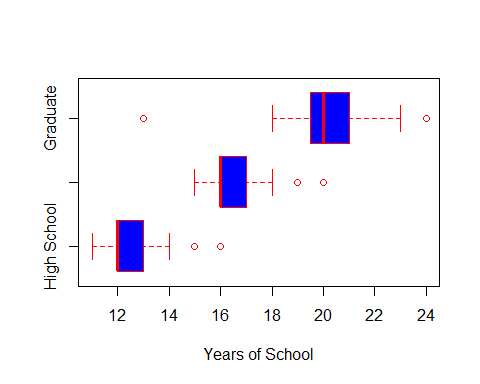


1. (1 point) Compare these two box plots and describe the differences between these two box plots. Does anyone boxplot has outliers? If so, how many outliers?

No Outliers, women have higher max, higher median, larger range.

1. (2 point) Construct three Box Plots for people’s number years in school (one with people who have High School Degree, the other who have Bachelor degree and last one who have Graduate degree) on the same plot.

boxplot(data$Years\_of\_School[data$Degree=="High School"],   
 data$Years\_of\_School[data$Degree=="Bachelor"],  
 data$Years\_of\_School[data$Degree=="Graduate"],   
 names=c("High School", "Bachelor","Graduate" ),  
   
 xlab="Years of School",  
 col="Blue",   
 border="red",  
 horizontal = T)



1. (1 point) Compare these three box plots and describe the differences between these two box plots. Does anyone boxplot has outliers? If so, how many outliers?

All plots have two outliers, there is a dependent relationship between degree acheived and number of years in school, high school least, graduate most. Interestingly, for Bachelor, 1st quartile and mean are touching, that’s strange.

1. Find the mean of people’s income when people have high school degree.

mean(data$Income[data$Degree=="High School"])

## [1] 19234.88

1. (1 point) Find the mean of people’s income for Male.

mean(data$Income[data$Gender=="Male"])

## [1] 43728.62

1. (1 point) Find the median of people’s income when people have Graduate Degree.

median(data$Income[data$Degree=="Graduate"])

## [1] 84999

1. (1 point) Find three standard deviations for people’s income (one with people who have High School Degree, the other who have Bachelor degree and last one who have Graduate degree)

sd(data$Income[data$Degree=="High School"])

## [1] 8355.942

sd(data$Income[data$Degree=="Bachelor"])

## [1] 15930.41

sd(data$Income[data$Degree=="Graduate"])

## [1] 17123.79

**Compare these standard deviations:**

Standard Deviation for high school is least, deviation for Bachelor is almost twice as much, deviation for Graduate is a little more.