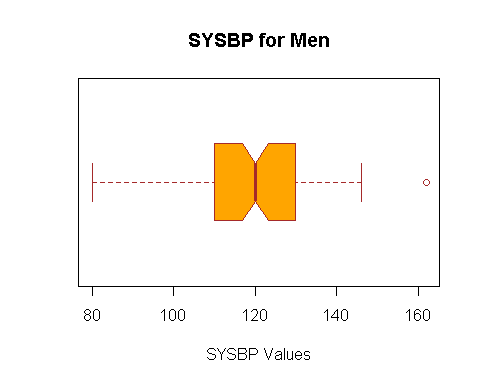
150116034\_enes\_garip.R

Garip

2020-03-21

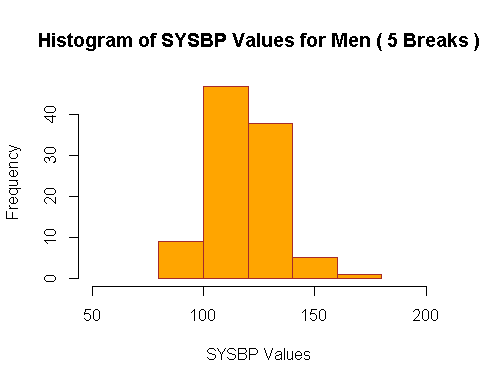
# Stat2053 HW1   
# Enes Garip  
# Variable Structure: "SYSBPorDIASBP"+"operation"+"0forMen 1forWomen"   
# Example:DIASBPmax1 ----> maximum value of DIASBP values of women  
  
HW1\_Dataset<-read.csv2("HW1\_Data.csv",header = TRUE,",")  
rdata<-HW1\_Dataset[c(1:4)] #rdata holds the dataset that do not contain NA.  
  
genderMale<-subset(rdata,GENDER=='0') #For men, genderMale holds the subset.  
genderFemale<-subset(rdata,GENDER=='1') #Similarly, for women, genderFemale holds the subset.  
  
##############SYSBP VALUES FOR MEN###############  
 #  
SYSBPmean0<-mean(genderMale$SYSBP) #   
SYSBPvar0<-var(genderMale$SYSBP) #  
SYSBPsd0<-sd(genderMale$SYSBP) #  
SYSBPquan0<-quantile(genderMale$SYSBP) #  
SYSBPmax0<-max(genderMale$SYSBP) #  
SYSBPmin0<-min(genderMale$SYSBP) #  
SYSBPrange0<-SYSBPmax0 - SYSBPmin0 #  
SYSBPsdc0<-SYSBPrange0/SYSBPsd0 #  
SYSBPmedian0<-median(genderMale$SYSBP) #  
SYSBPiqr0<-IQR(genderMale$SYSBP) #  
SYSBPsummary0<-summary(genderMale$SYSBP) #  
 #  
boxplot(genderMale$SYSBP,   
 main = "SYSBP for Men",  
 xlab = "SYSBP Values",  
 col = "orange",  
 border = "brown",  
 horizontal = TRUE,  
 notch = TRUE  
) # Boxplot of SYSBP values for Men.



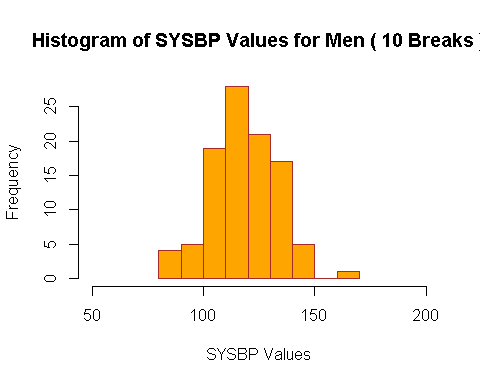
#  
stem(genderMale$SYSBP) # Stem and leaf plot of SYSBP values for Men.

##   
## The decimal point is 1 digit(s) to the right of the |  
##   
## 8 | 00  
## 9 | 00  
## 10 | 000004666888888  
## 11 | 000000000224456667888888  
## 12 | 000000000000022444444666668  
## 13 | 00000002444467788  
## 14 | 000000025566  
## 15 |   
## 16 | 2

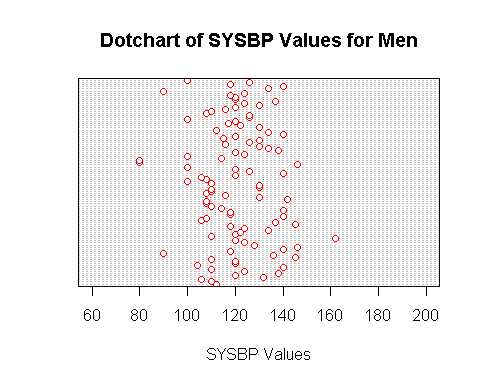
#  
 #  
hist(genderMale$SYSBP,  
 main="Histogram of SYSBP Values for Men ( 5 Breaks )",  
 xlab="SYSBP Values",  
 xlim=c(50,200),  
 border = "brown",  
 col="orange",  
 breaks = 5  
) # Histogram with 5 breaks...



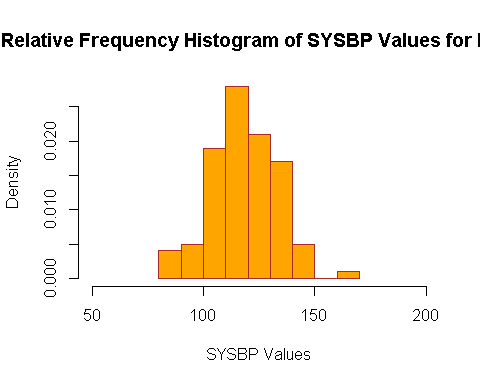
hist(genderMale$SYSBP,  
 main="Histogram of SYSBP Values for Men ( 10 Breaks )",  
 xlab="SYSBP Values",  
 xlim=c(50,200),  
 border = "brown",  
 col="orange",  
 breaks = 10  
) # Histogram with 10 breaks...



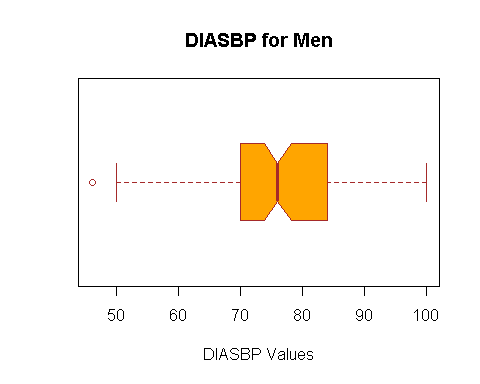
#  
dotchart(genderMale$SYSBP,  
 main="Dotchart of SYSBP Values for Men",  
 xlab = "SYSBP Values",  
 col = "red",  
 xlim=c(60,200),  
   
) # Dotchart of Men's SYSBP



#  
hist(genderMale$SYSBP,  
 main="Relative Frequency Histogram of SYSBP Values for Men",  
 xlab="SYSBP Values",  
 xlim=c(50,200),  
 border = "brown",  
 col="orange",  
 breaks = 10,  
 prob="TRUE"  
) # Relative Frequency Histogram...



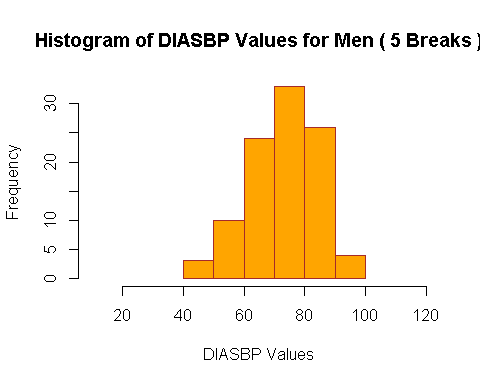
#  
SYSBPzmax0<-(SYSBPmax0-SYSBPmean0)/SYSBPsd0 # Z-score for maximum value  
SYSBPzmin0<-(SYSBPmin0-SYSBPmean0)/SYSBPsd0 # Z-score for minimum value  
 #  
#################################################  
  
##############DIASBP VALUES FOR MEN##############  
DIASBPmean0<-mean(genderMale$DIASBP) #  
DIASBPvar0<-var(genderMale$DIASBP) #  
DIASBPsd0<-sd(genderMale$DIASBP) #  
DIASBPquan0<-quantile(genderMale$DIASBP) #  
DIASBPmax0<-max(genderMale$DIASBP) #  
DIASBPmin0<-min(genderMale$DIASBP) #  
DIASBPrange0<-DIASBPmax0 - DIASBPmin0 #  
DIASBPsdc0<-DIASBPrange0/DIASBPsd0 #  
DIASBPmedian0<-median(genderMale$DIASBP) #  
DIASBPiqr0<-IQR(genderMale$DIASBP) #  
DIASBPsummary0<-summary(genderMale$DIASBP) #  
 #  
boxplot(genderMale$DIASBP,   
 main = "DIASBP for Men",  
 xlab = "DIASBP Values",  
 col = "orange",  
 border = "brown",  
 horizontal = TRUE,  
 notch = TRUE  
) # Boxplot of DIASBP values for Men.



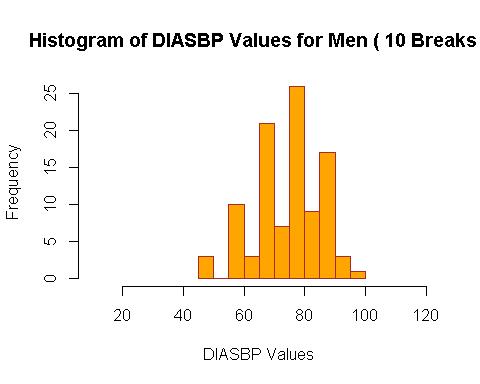
#  
stem(genderMale$DIASBP) # Stem and leaf plot of DIASBP values for Men.

##   
## The decimal point is 1 digit(s) to the right of the |  
##   
## 4 | 6  
## 5 | 0088  
## 6 | 000000004456668  
## 7 | 000000000000000002224444666666688888  
## 8 | 000000000000002244444446666688  
## 9 | 0000000000455  
## 10 | 0

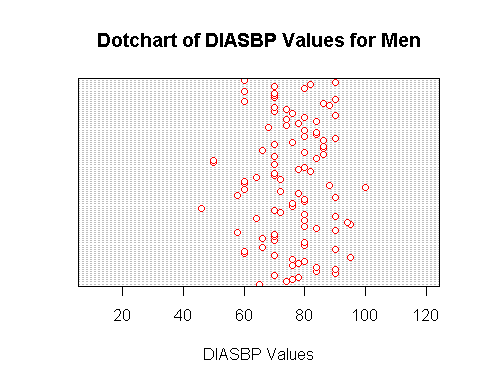
#  
 #  
hist(genderMale$DIASBP,  
 main="Histogram of DIASBP Values for Men ( 5 Breaks )",  
 xlab="DIASBP Values",  
 xlim=c(10,120),  
 border = "brown",  
 col="orange",  
 breaks = 5  
) # Histogram with 5 breaks...



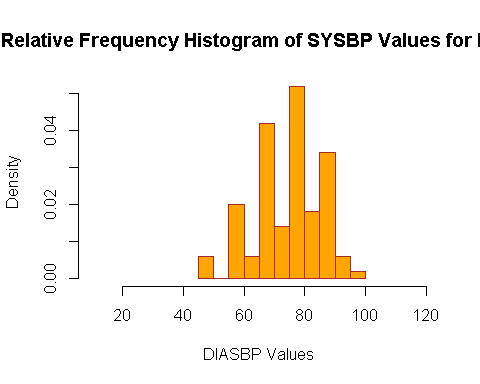
#  
hist(genderMale$DIASBP,  
 main="Histogram of DIASBP Values for Men ( 10 Breaks )",  
 xlab="DIASBP Values",  
 xlim=c(10,120),  
 border = "brown",  
 col="orange",  
 breaks = 10  
) # Histogram with 10 breaks...



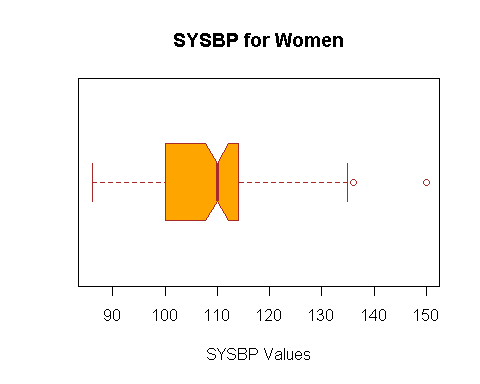
#  
dotchart(genderMale$DIASBP,  
 main="Dotchart of DIASBP Values for Men",  
 xlab = "DIASBP Values",  
 col = "red",  
 xlim=c(10,120),  
   
) # Dotchart of Men's DIASBP



#  
hist(genderMale$DIASBP,  
 main="Relative Frequency Histogram of SYSBP Values for Men",  
 xlab="DIASBP Values",  
 xlim=c(10,120),  
 border = "brown",  
 col="orange",  
 breaks = 10,  
 prob="TRUE"  
) # Relative Frequency Histogram...



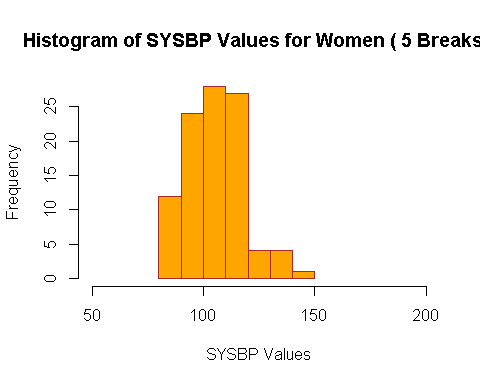
#  
DIASBPzmax0<-(DIASBPmax0-DIASBPmean0)/DIASBPsd0 # Z-score for maximum value  
DIASBPzmin0<-(DIASBPmin0-DIASBPmean0)/DIASBPsd0 # Z-score for minimum value  
#################################################  
  
##############SYSBP VALUES FOR WOMEN#####################  
 #  
 #  
SYSBPmean1<-mean(genderFemale$SYSBP) #   
SYSBPvar1<-var(genderFemale$SYSBP) #  
SYSBPsd1<-sd(genderFemale$SYSBP) #  
SYSBPquan1<-quantile(genderFemale$SYSBP) #  
SYSBPmax1<-max(genderFemale$SYSBP) #  
SYSBPmin1<-min(genderFemale$SYSBP) #  
SYSBPrange1<-SYSBPmax1 - SYSBPmin1 #  
SYSBPsdc1<-SYSBPrange1/SYSBPsd1 #  
SYSBPmedian1<-median(genderFemale$SYSBP) #  
SYSBPiqr1<-IQR(genderFemale$SYSBP) #  
SYSBPsummary1<-summary(genderFemale$SYSBP) #  
 #  
boxplot(genderFemale$SYSBP,  
 main = "SYSBP for Women",  
 xlab = "SYSBP Values",  
 col = "orange",  
 border = "brown",  
 horizontal = TRUE,  
 notch = TRUE  
) #



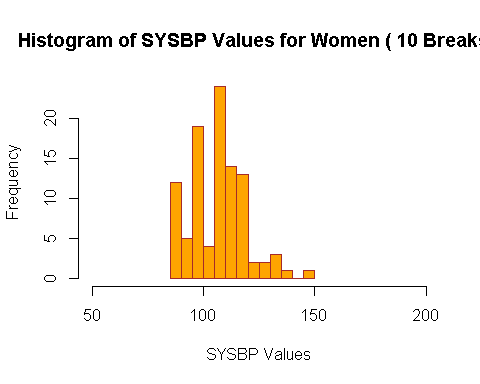
#  
stem(genderFemale$SYSBP) #

##   
## The decimal point is 1 digit(s) to the right of the |  
##   
## 8 | 688  
## 9 | 000000000224446668888  
## 10 | 000000000000244566888  
## 11 | 000000000000000000022224444444445688  
## 12 | 000000000022  
## 13 | 004456  
## 14 |   
## 15 | 0

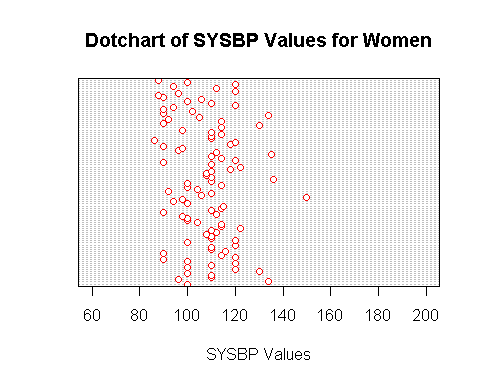
#  
 #  
hist(genderFemale$SYSBP,  
 main="Histogram of SYSBP Values for Women ( 5 Breaks )",  
 xlab="SYSBP Values",  
 xlim=c(50,200),  
 border = "brown",  
 col="orange",  
 breaks = 5  
) #



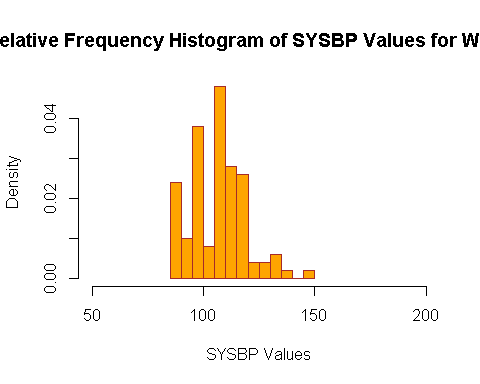
#  
hist(genderFemale$SYSBP,  
 main="Histogram of SYSBP Values for Women ( 10 Breaks )",  
 xlab="SYSBP Values",  
 xlim=c(50,200),  
 border = "brown",  
 col="orange",  
 breaks = 10  
) #



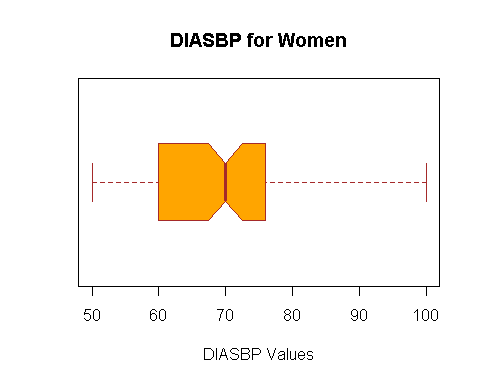
#  
dotchart(genderFemale$SYSBP,  
 main="Dotchart of SYSBP Values for Women",  
 xlab = "SYSBP Values",  
 col = "red",  
 xlim=c(60,200),  
   
) #



hist(genderFemale$SYSBP,  
 main="Relative Frequency Histogram of SYSBP Values for Women",  
 xlab="SYSBP Values",  
 xlim=c(50,200),  
 border = "brown",  
 col="orange",  
 breaks = 10,  
 prob="TRUE"  
) #



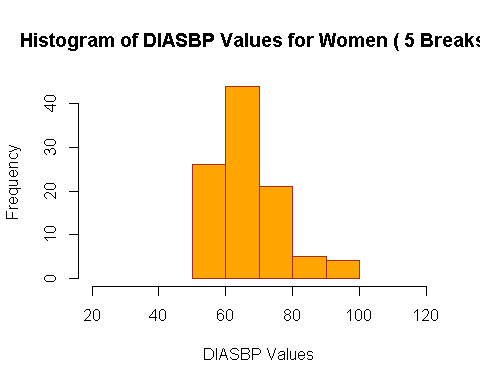
#  
SYSBPzmax1<-(SYSBPmax1-SYSBPmean1)/SYSBPsd1 #  
SYSBPzmin1<-(SYSBPmin1-SYSBPmean1)/SYSBPsd1 #  
 #  
#########################################################   
  
##############DIASBP VALUES FOR WOMEN####################  
 #  
DIASBPmean1<-mean(genderFemale$DIASBP) #  
DIASBPvar1<-var(genderFemale$DIASBP) #  
DIASBPsd1<-sd(genderFemale$DIASBP) #  
DIASBPquan1<-quantile(genderFemale$DIASBP) #  
DIASBPmax1<-max(genderFemale$DIASBP) #  
DIASBPmin1<-min(genderFemale$DIASBP) #  
DIASBPrange1<-DIASBPmax1 - DIASBPmin1 #  
DIASBPsdc1<-DIASBPrange1/DIASBPsd1 #  
DIASBPmedian1<-median(genderFemale$DIASBP) #  
DIASBPiqr1<-IQR(genderFemale$DIASBP) #  
DIASBPsummary1<-summary(genderFemale$DIASBP) #  
 #  
boxplot(genderFemale$DIASBP,   
 main = "DIASBP for Women",  
 xlab = "DIASBP Values",  
 col = "orange",  
 border = "brown",  
 horizontal = TRUE,  
 notch = TRUE  
) # Boxplot of DIASBP values for Women.



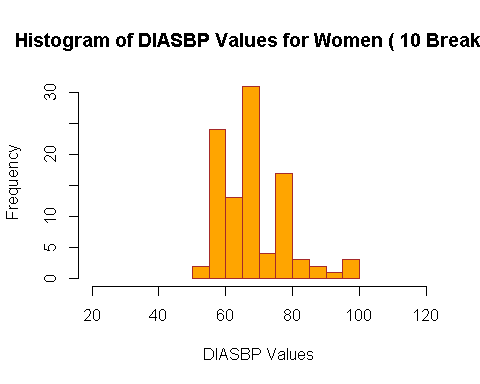
#  
stem(genderFemale$DIASBP) # Stem and leaf plot of DIASBP values for Women.

##   
## The decimal point is 1 digit(s) to the right of the |  
##   
## 5 | 046888  
## 6 | 0000000000000000000022444444445556668888888  
## 7 | 000000000000000000000244466666888  
## 8 | 00000000022488  
## 9 | 28  
## 10 | 00

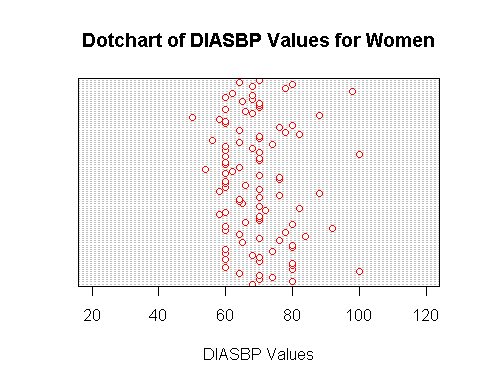
#  
 #  
hist(genderFemale$DIASBP,  
 main="Histogram of DIASBP Values for Women ( 5 Breaks )",  
 xlab="DIASBP Values",  
 xlim=c(20,120),  
 border = "brown",  
 col="orange",  
 breaks = 5  
) # Histogram with 5 breaks...



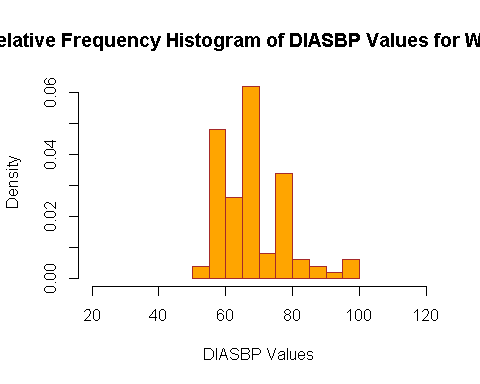
#  
hist(genderFemale$DIASBP,  
 main="Histogram of DIASBP Values for Women ( 10 Breaks )",  
 xlab="DIASBP Values",  
 xlim=c(20,120),  
 border = "brown",  
 col="orange",  
 breaks = 10  
) # Histogram with 10 breaks...



#  
dotchart(genderFemale$DIASBP,  
 main="Dotchart of DIASBP Values for Women",  
 xlab = "DIASBP Values",  
 col = "red",  
 xlim=c(20,120),  
   
) # Dotchart of Women's DIASBP



#  
hist(genderFemale$DIASBP,  
 main="Relative Frequency Histogram of DIASBP Values for Women",  
 xlab="DIASBP Values",  
 xlim=c(20,120),  
 border = "brown",  
 col="orange",  
 breaks = 10,  
 prob="TRUE"  
) # Relative Frequency Histogram...



#  
DIASBPzmax1<-(DIASBPmax1-DIASBPmean1)/DIASBPsd1 # Z-score for maximum value  
DIASBPzmin1<-(DIASBPmin1-DIASBPmean1)/DIASBPsd1 # Z-score for minimum value  
#########################################################

**ANSWER OF THE OTHER QUESTIONS**

**1)** The variables are quantitative and also continuous. The data is bivariate.

**2)** Bar charts, box-plot, histograms, stem and leaf plot can describe this dataset. We can find frequency, range, outliers, and relative frequency using graphical methods.

**g)** The range is approximately 5 times of standard deviation.

**k)** There are values some of them are outliers.

**m)** More intervals can describe the data with more detail so using more intervals is better.

**n)** The data is mound-shaped.

**o)** Yes, because Tchebysheff’s Theorem can be used for almost any dataset.

**p)** Yes, because the dataset is mound-shaped.

**s)** Some of z-scores is greater than 3 so these values can be an outlier.