STAT253-Homework#1

Sueda Bilen - 150117044

23 03 2020

data1 <- read.csv(file.choose(),header =T, sep =";")  
data1

## GENDER AGE SYSBP DIASBP  
## 1 0 20 145 95  
## 2 0 20 109 70  
## 3 0 17 104 74  
## 4 0 20 130 90  
## 5 0 15 115 70  
## 6 0 16 120 70  
## 7 0 16 122 82  
## 8 0 17 110 70  
## 9 0 16 128 78  
## 10 0 16 118 86  
## 11 0 19 106 72  
## 12 0 16 122 74  
## 13 0 15 90 60  
## 14 0 17 134 84  
## 15 0 16 120 70  
## 16 0 18 144 60  
## 17 0 20 121 70  
## 18 0 17 110 80  
## 19 0 18 104 72  
## 20 0 18 170 110  
## 21 0 15 94 74  
## 22 0 18 140 80  
## 23 0 16 108 80  
## 24 0 17 134 84  
## 25 0 19 118 78  
## 26 0 15 90 60  
## 27 0 16 120 80  
## 28 0 18 112 95  
## 29 0 20 121 70  
## 30 0 15 100 70  
## 31 0 16 130 80  
## 32 0 18 130 88  
## 33 0 20 122 78  
## 34 0 17 120 80  
## 35 0 17 110 70  
## 36 0 18 116 76  
## 37 0 18 130 76  
## 38 0 15 110 72  
## 39 0 18 136 86  
## 40 0 15 114 70  
## 41 0 18 130 80  
## 42 0 16 154 100  
## 43 0 17 116 78  
## 44 0 20 110 70  
## 45 0 15 110 60  
## 46 0 18 112 95  
## 47 0 16 106 68  
## 48 0 15 94 74  
## 49 0 17 106 80  
## 50 0 19 128 92  
## 51 0 17 140 70  
## 52 0 20 105 55  
## 53 0 20 112 76  
## 54 0 19 108 70  
## 55 0 16 120 80  
## 56 0 19 142 70  
## 57 0 16 100 68  
## 58 0 18 110 60  
## 59 0 17 118 86  
## 60 0 18 128 76  
## 61 0 15 124 84  
## 62 0 15 100 60  
## 63 0 20 106 70  
## 64 0 19 132 94  
## 65 0 15 114 72  
## 66 0 16 116 55  
## 67 0 19 96 70  
## 68 0 19 126 74  
## 69 0 16 130 64  
## 70 0 18 120 70  
## 71 0 16 122 82  
## 72 0 18 114 76  
## 73 0 15 118 64  
## 74 0 17 124 78  
## 75 0 16 98 42  
## 76 0 17 100 68  
## 77 0 17 120 70  
## 78 0 15 90 60  
## 79 0 18 120 76  
## 80 0 19 125 74  
## 81 0 15 128 76  
## 82 0 16 170 100  
## 83 0 15 118 80  
## 84 0 16 110 80  
## 85 0 15 122 80  
## 86 0 15 104 74  
## 87 0 15 112 78  
## 88 0 17 136 90  
## 89 0 18 140 100  
## 90 0 18 102 66  
## 91 0 16 120 74  
## 92 0 17 120 70  
## 93 0 18 110 60  
## 94 0 17 120 86  
## 95 0 16 120 70  
## 96 0 19 120 70  
## 97 0 15 103 72  
## 98 0 18 110 60  
## 99 0 15 110 60  
## 100 0 16 120 80  
## 101 1 19 112 66  
## 102 1 17 126 70  
## 103 1 19 116 86  
## 104 1 20 90 60  
## 105 1 20 132 102  
## 106 1 19 112 70  
## 107 1 17 102 70  
## 108 1 18 100 60  
## 109 1 16 104 70  
## 110 1 17 100 58  
## 111 1 20 112 70  
## 112 1 15 90 70  
## 113 1 19 90 60  
## 114 1 20 90 60  
## 115 1 20 120 70  
## 116 1 15 88 70  
## 117 1 15 80 50  
## 118 1 19 128 76  
## 119 1 20 110 70  
## 120 1 20 120 90  
## 121 1 20 95 55  
## 122 1 18 90 70  
## 123 1 16 100 60  
## 124 1 15 130 70  
## 125 1 17 112 86  
## 126 1 20 110 56  
## 127 1 19 120 78  
## 128 1 16 96 70  
## 129 1 18 118 70  
## 130 1 18 100 80  
## 131 1 19 102 60  
## 132 1 15 104 70  
## 133 1 20 110 60  
## 134 1 16 104 64  
## 135 1 16 78 50  
## 136 1 15 90 68  
## 137 1 16 110 78  
## 138 1 18 100 66  
## 139 1 16 126 86  
## 140 1 20 118 60  
## 141 1 15 122 64  
## 142 1 19 144 80  
## 143 1 16 100 60  
## 144 1 16 104 56  
## 145 1 15 108 84  
## 146 1 20 118 60  
## 147 1 19 120 70  
## 148 1 19 114 80  
## 149 1 16 116 82  
## 150 1 20 120 70  
## 151 1 16 110 78  
## 152 1 19 100 72  
## 153 1 20 120 60  
## 154 1 18 100 60  
## 155 1 18 112 78  
## 156 1 15 136 76  
## 157 1 16 110 62  
## 158 1 17 110 60  
## 159 1 20 110 82  
## 160 1 19 104 60  
## 161 1 17 120 70  
## 162 1 15 110 70  
## 163 1 17 110 70  
## 164 1 17 114 64  
## 165 1 16 114 80  
## 166 1 18 100 66  
## 167 1 19 130 80  
## 168 1 16 100 70  
## 169 1 15 122 56  
## 170 1 17 100 80  
## 171 1 17 115 80  
## 172 1 18 100 70  
## 173 1 16 130 60  
## 174 1 19 122 84  
## 175 1 16 104 70  
## 176 1 18 120 82  
## 177 1 18 124 70  
## 178 1 20 104 66  
## 179 1 18 120 78  
## 180 1 16 120 70  
## 181 1 16 122 76  
## 182 1 19 120 70  
## 183 1 17 112 80  
## 184 1 19 120 70  
## 185 1 17 100 70  
## 186 1 15 110 60  
## 187 1 20 100 80  
## 188 1 18 100 70  
## 189 1 15 88 68  
## 190 1 16 98 70  
## 191 1 15 110 62  
## 192 1 17 110 80  
## 193 1 20 94 64  
## 194 1 20 150 96  
## 195 1 16 110 80  
## 196 1 15 110 80  
## 197 1 15 104 70  
## 198 1 20 114 70  
## 199 1 16 120 70  
## 200 1 18 92 60

#Answer 1: gender is a qualitative variable #Age,systolic blood pressure and diastolic b.p. is quantitative and discrete variables. #This data is multivariate

#Answer 2:For age,systolic bp and diastolic bp variables,we can use pie chart,bar chart,line chart,stem and leaf plots,dotplot and histogram,by this graphical methods we can see the frequency,shape,scale,outliers and the spreading of our data. #subsetting my data

sysbpWomen <- subset(data1, GENDER == 1, select=AGE:SYSBP)  
sysbpMen <- subset(data1, GENDER == 0, select=AGE:SYSBP)  
diasbpWomen <- subset(data1, GENDER == 1, select=c(AGE,DIASBP))  
diasbpMen <- subset(data1, GENDER == 0, select=c(AGE,DIASBP))

#Answer 3:A:

mean(sysbpWomen$SYSBP)

## [1] 109.76

mean(sysbpMen$SYSBP)

## [1] 118.16

mean(diasbpWomen$DIASBP)

## [1] 70.31

mean(diasbpMen$DIASBP)

## [1] 75.21

#Answer 3:B:

var(sysbpWomen$SYSBP)

## [1] 169.2145

var(sysbpMen$SYSBP)

## [1] 220.0954

var(diasbpWomen$DIASBP)

## [1] 92.07465

var(diasbpMen$DIASBP)

## [1] 120.8342

#Answer 3:C

sd(sysbpWomen$SYSBP)

## [1] 13.00825

sd(sysbpMen$SYSBP)

## [1] 14.83561

sd(diasbpWomen$DIASBP)

## [1] 9.595553

sd(diasbpMen$DIASBP)

## [1] 10.99246

#Answer 3:D;Upperand Lower Quartiles

quantile(sysbpWomen$SYSBP,c(0.25,0.75))

## 25% 75%   
## 100 120

quantile(sysbpMen$SYSBP,c(0.25,0.75))

## 25% 75%   
## 110.00 125.25

quantile(diasbpWomen$DIASBP,c(0.25,0.75))

## 25% 75%   
## 62 78

quantile(diasbpMen$DIASBP,c(0.25,0.75))

## 25% 75%   
## 70 80

#Answer 3:E:max and min values

min(sysbpWomen$SYSBP)

## [1] 78

max(sysbpWomen$SYSBP)

## [1] 150

min(sysbpMen$SYSBP)

## [1] 90

max(sysbpMen$SYSBP)

## [1] 170

min(diasbpWomen$DIASBP)

## [1] 50

max(diasbpWomen$DIASBP)

## [1] 102

min(diasbpMen$DIASBP)

## [1] 42

max(diasbpMen$DIASBP)

## [1] 110

#Answer 3:F:Range

rangeSysbpWomen <- (max(sysbpWomen$SYSBP)) - (min(sysbpWomen$SYSBP))  
rangeSysbpWomen

## [1] 72

rangeSysbpMen <- max(sysbpMen$SYSBP) - min(sysbpMen$SYSBP)  
rangeSysbpMen

## [1] 80

rangeDiasbpWomen <- max(diasbpWomen$DIASBP) - min(diasbpWomen$DIASBP)  
rangeDiasbpWomen

## [1] 52

rangeDiasbpMen <- max(diasbpMen$DIASBP) - min(diasbpMen$DIASBP)  
rangeDiasbpMen

## [1] 68

#Answer 3:G

compareSysbpWomen <- rangeSysbpWomen / sd(sysbpWomen$SYSBP)  
compareSysbpWomen

## [1] 5.534949

compareSysbpMen <- rangeSysbpMen / sd(sysbpMen$SYSBP)  
compareSysbpMen

## [1] 5.39243

compareDiasbpWomen <-rangeDiasbpWomen / sd(diasbpWomen$DIASBP)  
compareDiasbpWomen

## [1] 5.419177

compareDiasbpMen <- rangeDiasbpMen / sd(diasbpMen$DIASBP)  
compareDiasbpMen

## [1] 6.186057

#Answer 3:H

median(sysbpWomen$SYSBP)

## [1] 110

median(sysbpMen$SYSBP)

## [1] 118

median(diasbpWomen$DIASBP)

## [1] 70

median(diasbpMen$DIASBP)

## [1] 74

#Answer 3:I

IQR(sysbpWomen$SYSBP)

## [1] 20

IQR(sysbpMen$SYSBP)

## [1] 15.25

IQR(diasbpWomen$DIASBP)

## [1] 16

IQR(diasbpMen$DIASBP)

## [1] 10

#Answer 3:J

fivenum(sysbpWomen$SYSBP)

## [1] 78 100 110 120 150

fivenum(sysbpMen$SYSBP)

## [1] 90.0 110.0 118.0 125.5 170.0

fivenum(diasbpWomen$DIASBP)

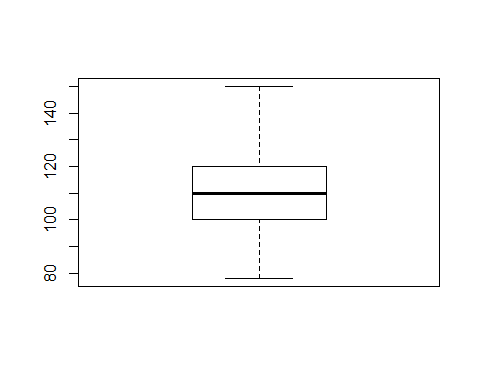
## [1] 50 62 70 78 102

fivenum(diasbpMen$DIASBP)

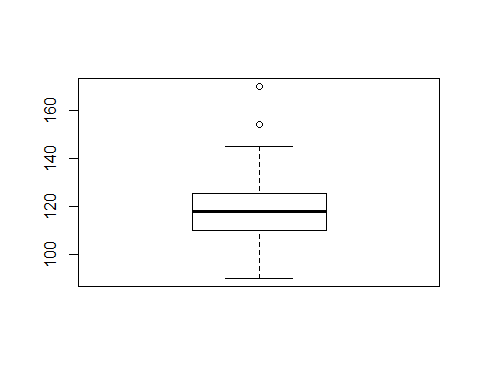
## [1] 42 70 74 80 110

#Answer 3:K

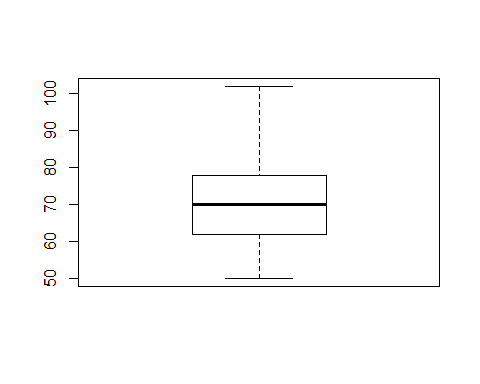
boxplot(sysbpWomen$SYSBP)

 #symmetric distribution and no

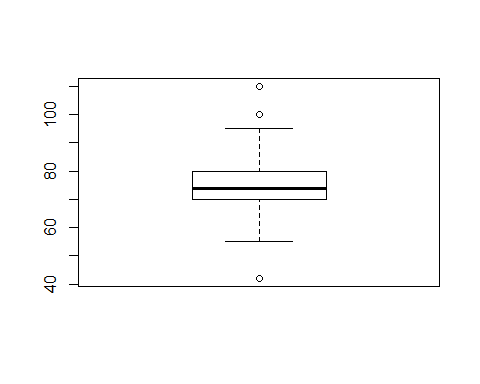
boxplot(sysbpMen$SYSBP)

 #symmetric distribution and 2 outliners

boxplot(diasbpWomen$DIASBP)

 #symmetric distribution and no outliners

boxplot(diasbpMen$DIASBP)

 #skewed right and 3 outliners

#Answer 3:L

stem(sysbpWomen$SYSBP)

##   
## The decimal point is 1 digit(s) to the right of the |  
##   
## 7 | 8  
## 8 | 088  
## 9 | 00000024568  
## 10 | 00000000000000022444444448  
## 11 | 0000000000000002222224444566888  
## 12 | 000000000000022224668  
## 13 | 00026  
## 14 | 4  
## 15 | 0

stem(sysbpMen$SYSBP)

##   
## The decimal point is 1 digit(s) to the right of the |  
##   
## 9 | 0004468  
## 10 | 00002344456666889  
## 11 | 000000000002222444566688888  
## 12 | 00000000000000112222244568888  
## 13 | 00000024466  
## 14 | 000245  
## 15 | 4  
## 16 |   
## 17 | 00

stem(diasbpWomen$DIASBP)

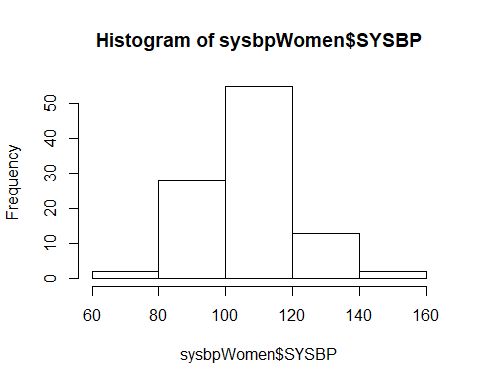
##   
## The decimal point is 1 digit(s) to the right of the |  
##   
## 5 | 0056668  
## 6 | 00000000000000000224444666688  
## 7 | 00000000000000000000000000000000266688888  
## 8 | 00000000000022244666  
## 9 | 06  
## 10 | 2

stem(diasbpMen$DIASBP)

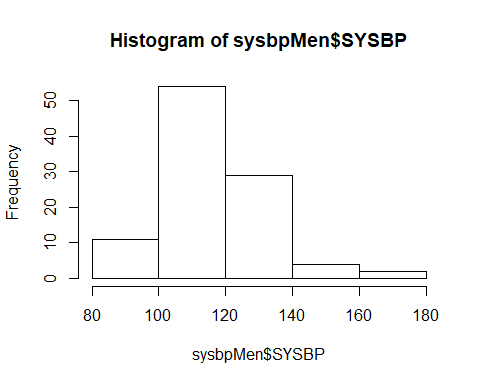
##   
## The decimal point is 1 digit(s) to the right of the |  
##   
## 4 | 2  
## 5 | 55  
## 6 | 0000000000446888  
## 7 | 00000000000000000000022222444444446666666888888  
## 8 | 00000000000002244466668  
## 9 | 0024555  
## 10 | 000  
## 11 | 0

#Answer 3:M;1

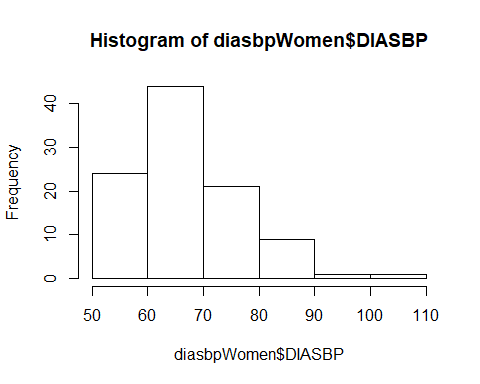
hist(sysbpWomen$SYSBP , breaks = 4)



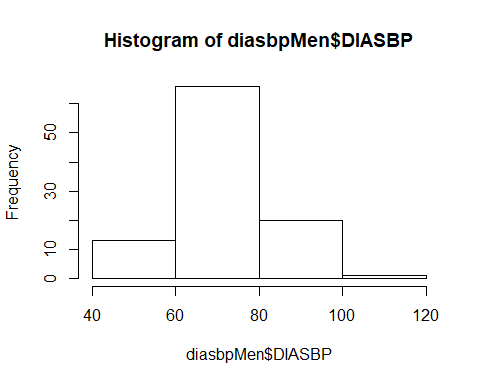
hist(sysbpMen$SYSBP , breaks = 4)



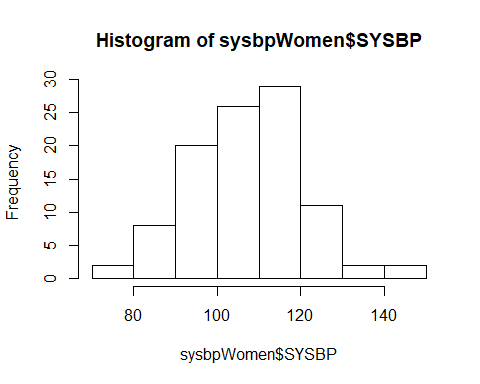
hist(diasbpWomen$DIASBP , breaks = 4)



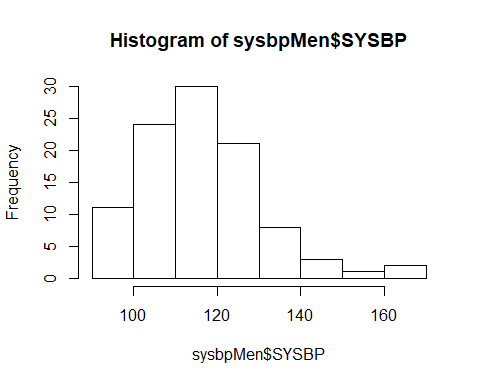
hist(diasbpMen$DIASBP , breaks = 4)

 #Answer 3:M;2

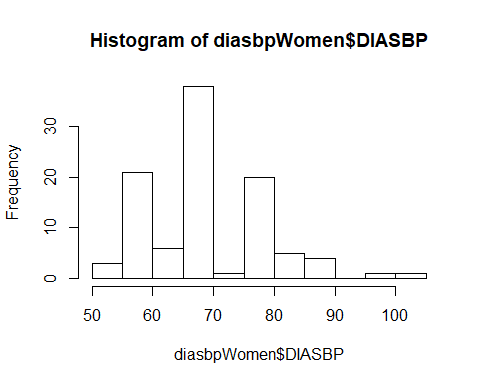
hist(sysbpWomen$SYSBP)



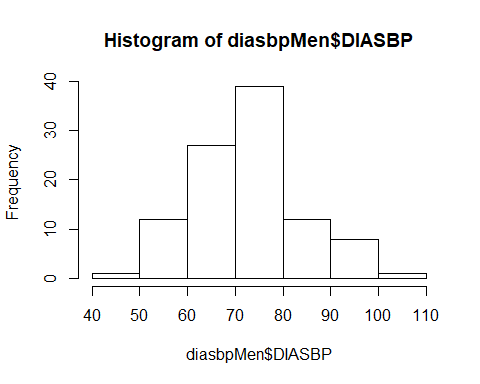
hist(sysbpMen$SYSBP)



hist(diasbpWomen$DIASBP)

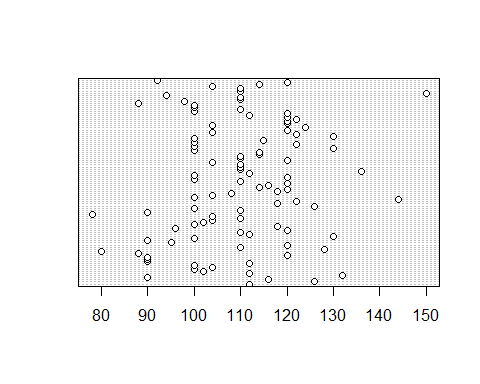


hist(diasbpMen$DIASBP)

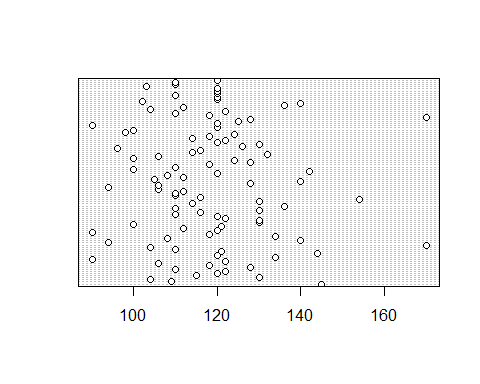
 #Answer 3:M:3 second version is better than first,because we can see clearly our variable in second one.But first one we can miss them.

#Answer 3:N

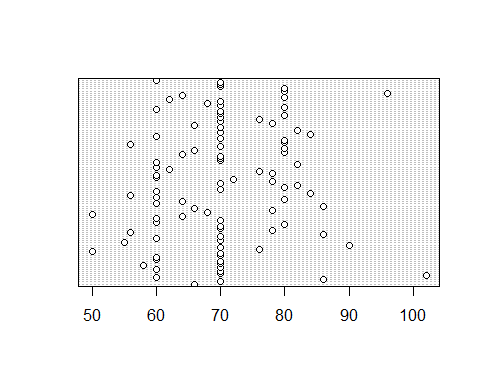
dotchart(sysbpWomen$SYSBP)



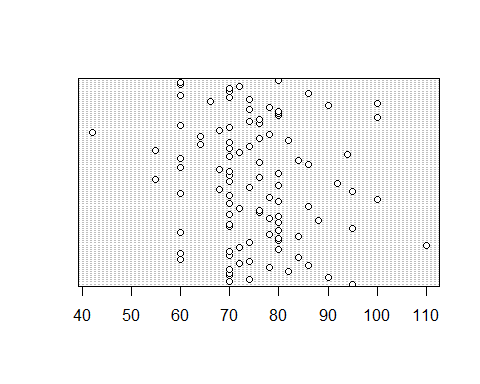
dotchart(sysbpMen$SYSBP)



dotchart(diasbpWomen$DIASBP)



dotchart(diasbpMen$DIASBP)

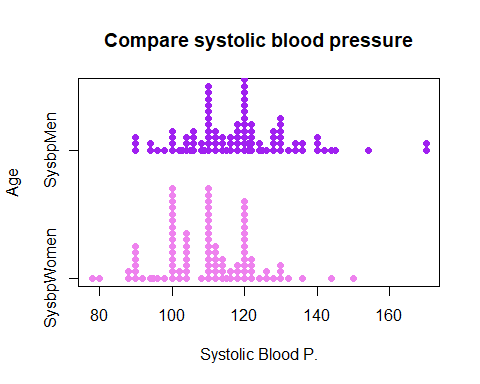
 #only diasbpmen is moundshaped

#Answer 3:O we can use tchebyseff’s theorem for all of our data because in tchebyseff’s theorem says that “Applies that any set of measurements; large or small,mound-shaped or skewed”

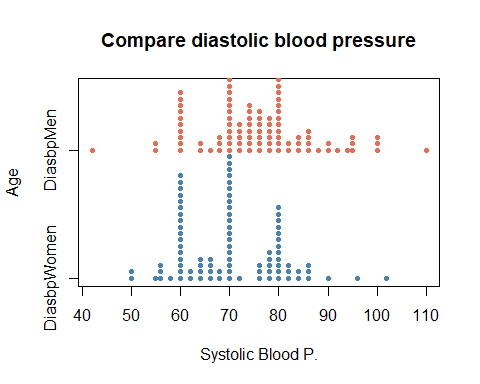
#Answer 3:P we can use empirical law for diasbpMen and sysbpWomen graphs,because they are mound shaped.

#Answer 3:Q

compareDotPlotSys <- list('SysbpWomen' = sysbpWomen$SYSBP,'SysbpMen' = sysbpMen$SYSBP)  
stripchart(compareDotPlotSys,  
 main = 'Compare systolic blood pressure',  
 xlab = 'Systolic Blood P.',  
 ylab = 'Age',  
 col = c('violet', 'purple'),  
 pch = 16,  
 method = 'stack')

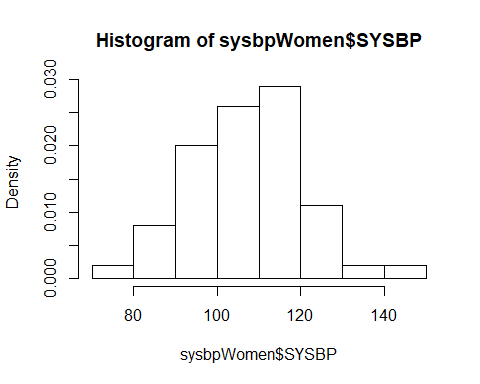


compareDotPlotDias <- list('DiasbpWomen' = diasbpWomen$DIASBP,'DiasbpMen' = diasbpMen$DIASBP)  
stripchart(compareDotPlotDias,  
 main = 'Compare diastolic blood pressure',  
 xlab = 'Systolic Blood P.',  
 ylab = 'Age',  
 col = c('steelblue', 'coral2'),  
 pch = 20,  
 method = 'stack')

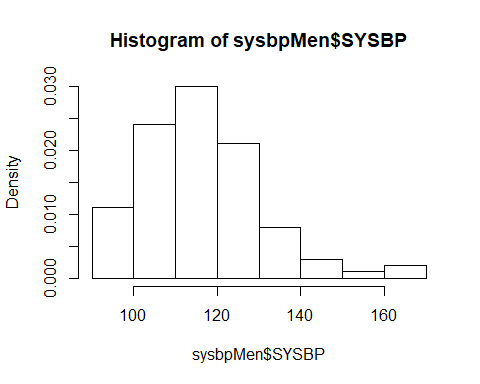


#Answer 3:R

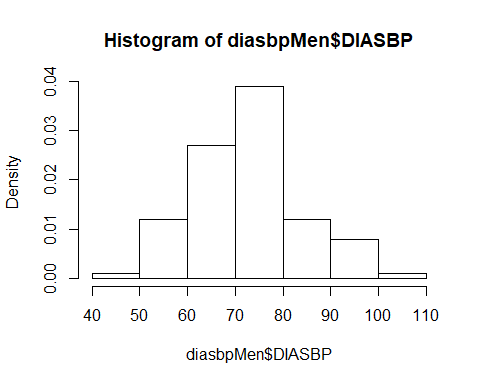
hist(sysbpWomen$SYSBP,prob = TRUE)



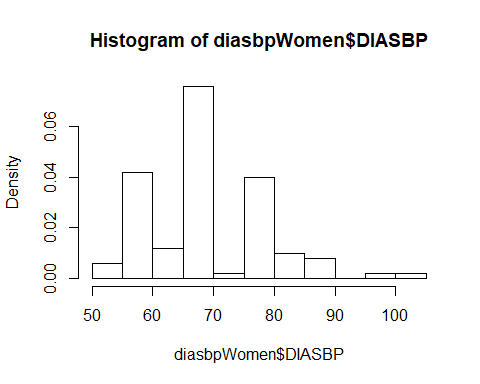
hist(sysbpMen$SYSBP,prob = TRUE)



hist(diasbpMen$DIASBP,prob = TRUE)



hist(diasbpWomen$DIASBP,prob = TRUE)



#Answer 3:S

zSysWomenLargest = (max(sysbpWomen$SYSBP)-mean(sysbpWomen$SYSBP))/sd(sysbpWomen$SYSBP)  
zSysWomenLargest

## [1] 3.093422

zSysWomenSmallest = (min(sysbpWomen$SYSBP)-mean(sysbpWomen$SYSBP))/sd(sysbpWomen$SYSBP)  
zSysWomenSmallest

## [1] -2.441528

zSysMenLargest = (max(sysbpMen$SYSBP)-mean(sysbpMen$SYSBP))/sd(sysbpMen$SYSBP)  
zSysMenLargest

## [1] 3.494295

zSysMenSmallest = (min(sysbpMen$SYSBP)-mean(sysbpMen$SYSBP))/sd(sysbpMen$SYSBP)  
zSysMenSmallest

## [1] -1.898136

zDiasMenLargest = (max(diasbpMen$DIASBP)-mean(diasbpMen$DIASBP))/sd(diasbpMen$DIASBP)  
zDiasMenLargest

## [1] 3.164896

zDiasMenSmallest = (min(diasbpMen$DIASBP)-mean(diasbpMen$DIASBP))/sd(diasbpMen$DIASBP)  
zDiasMenSmallest

## [1] -3.021161

zDiasWomenLargest = (max(diasbpWomen$DIASBP)-mean(diasbpWomen$DIASBP))/sd(diasbpWomen$DIASBP)  
zDiasWomenLargest

## [1] 3.302571

zDiasWomenSmallest = (min(diasbpWomen$DIASBP)-mean(diasbpWomen$DIASBP))/sd(diasbpWomen$DIASBP)  
zDiasWomenSmallest

## [1] -2.116605

#In systolic blood pressure for women data;there can be a outlier because Largest z-score is bigger than 3. #In systolic blood pressure for men data;there can be a outlier because Largest z-score is bigger than 3. #In diastolic blood pressure for men data;there can be a outlier because largest and smallest z-scores are bigger than 3. #In diastolic blood pressure for women data;there can be a outlier because largest z-score is bigger than 3.