STAT253 – Homework#1

Çağatay USLU - 150113030

21.03.2020

### Initializing dataset from .csv

library(DescTools)  
library(ggplot2)  
  
man <- 0  
woman <- 1  
datasetFile <- "HW1\_Data.csv"  
dataSet <- read.csv(datasetFile, header = TRUE, ";")  
man <- 0  
woman <- 1  
datasetFile <- "HW1\_Data.csv"  
dataSet <- read.csv(datasetFile, header = TRUE, ";")  
men.sysbp <- dataSet[which(dataSet$GENDER == man),]['SYSBP'][, 1]  
men.diasbp <- dataSet[which(dataSet$GENDER == man),]['DIASBP'][, 1]  
women.sysbp <- dataSet[which(dataSet$GENDER == woman),]['SYSBP'][, 1]  
women.diasbp <- dataSet[which(dataSet$GENDER == woman),]['DIASBP'][, 1]

### Answers

*1 -* Four variables exist for each person individually. Except gender, all of others are quantitative variable. Gender is qualitative variable. It is characteristic variable on each person (experimental unit). Other variables; age, sysbp, diasbp are discrete quantitative variable. Because these are countable. Data is multivariate. More than two variables are measured on each person (experimental unit)

*2 -* PASS

*3 -* MEN SYSTOLIC BLOOD PRESSURE

men.sysbp <- dataSet[which(dataSet$GENDER == man),]['SYSBP'][, 1]  
men.sysbp.sorted <- sort(men.sysbp)  
men.sysbp.mean <- mean(men.sysbp)  
men.sysbp.most.frequent.value <- Mode(men.sysbp)  
men.sysbp.most.frequent.value.count <- length(dataSet[which(dataSet$SYSBP == men.sysbp.most.frequent.value),]['SYSBP'][, 1])  
men.sysbp.median <- median(men.sysbp)  
men.sysbp.max <- max(men.sysbp)  
men.sysbp.min <- min(men.sysbp)  
men.sysbp.sum <- sum(men.sysbp)  
men.sysbp.std <- sd(men.sysbp, na.rm = FALSE)  
men.sysbp.fiveNumbersSumy <- quantile(men.sysbp)  
men.sysbp.lowerQuartile <- men.sysbp.fiveNumbersSumy[2]  
men.sysbp.upperQuartile <- men.sysbp.fiveNumbersSumy[4]  
men.sysbp.interQuartile <- men.sysbp.upperQuartile - men.sysbp.lowerQuartile  
men.sysbp.variance <- men.sysbp.std^2  
men.sysbp.range <- men.sysbp.max - men.sysbp.min  
men.sysbp.range.over.std <- men.sysbp.range / men.sysbp.std  
men.sysbp.largest.zscore <- (men.sysbp.max - men.sysbp.mean) / men.sysbp.std  
men.sysbp.smallest.zscore <- (men.sysbp.min - men.sysbp.mean) / men.sysbp.std

Sorted: 80, 90, 98, 98, 100, 100, 100, 100, 100, 100, 100, 104, 104, 104, 106, 106, 106, 106, 108, 108, 109, 109, 110, 110, 110, 110, 110, 110, 110, 110, 110, 110, 110, 110, 112, 112, 112, 114, 114, 114, 115, 115, 116, 116, 117, 118, 118, 118, 120, 120, 120, 120, 120, 120, 120, 120, 120, 120, 120, 120, 120, 120, 121, 121, 122, 122, 124, 124, 124, 124, 125, 125, 126, 126, 128, 128, 128, 128, 128, 128, 130, 130, 130, 130, 130, 130, 130, 130, 134, 134, 136, 136, 136, 140, 140, 140, 146, 150, 156, 168

Mean, 𝑥̅: 118.55

Variance, s^2: 188.4722222

Standard deviation, s: 13.7285186

Lower Quartile: Lower Quartile: 110

Upper Quartile: Upper Quartile: 128

Min value: Min value: 80

Max value: Max value: 168

Range: 88

The range is approximately 6.4100143 standard deviations

Median: 120

IQR: 18

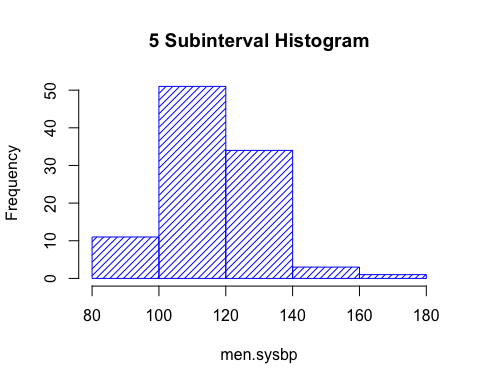
Five-number summaries: 80, 110, 120, 128, 168

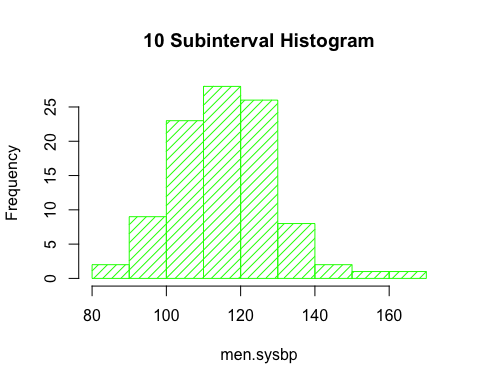
### Box Plot

### Stem Plot

##   
## The decimal point is 1 digit(s) to the right of the |  
##   
## 8 | 0  
## 9 | 088  
## 10 | 000000044466668899  
## 11 | 00000000000022244455667888  
## 12 | 00000000000000112244445566888888  
## 13 | 0000000044666  
## 14 | 0006  
## 15 | 06  
## 16 | 8

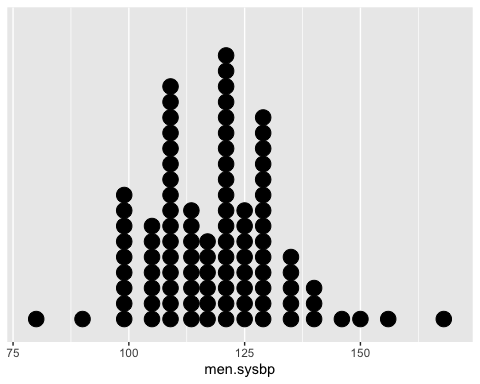
### Histograms





Description

### Dotplot



The data is not mound shaped.

We can not use the Tchebysheff’s Theorem to describe this data set. Because the set of measurement is skewed. The shape is skewed-right.

We can not use the Empirical Rule to describe this data set. Because it’s not mound-shaped.

q.Construct

r.Construct

## z-scores of largest and smallest observations

Smallest observation z-score: 3.601991

Largest observation z-score: 3.601991

*3 -* MEN DIASTOLIC BLOOD PRESSURE

men.diasbp.sorted <- sort(men.diasbp)  
men.diasbp.mean <- mean(men.diasbp)  
men.diasbp.most.frequent.value <- Mode(men.diasbp)  
men.diasbp.most.frequent.value.count <- length(dataSet[which(dataSet$DIASBP == men.diasbp.most.frequent.value),]['DIASBP'][, 1])  
men.diasbp.median <- median(men.diasbp)  
men.diasbp.max <- max(men.diasbp)  
men.diasbp.min <- min(men.diasbp)  
men.diasbp.sum <- sum(men.diasbp)  
men.diasbp.std <- sd(men.diasbp, na.rm = FALSE)  
men.diasbp.fiveNumbersSumy <- quantile(men.diasbp)  
men.diasbp.lowerQuartile <- men.diasbp.fiveNumbersSumy[2]  
men.diasbp.upperQuartile <- men.diasbp.fiveNumbersSumy[4]  
men.diasbp.interQuartile <- men.diasbp.upperQuartile - men.diasbp.lowerQuartile  
men.diasbp.variance <- men.diasbp.std^2  
men.diasbp.range <- men.diasbp.max - men.diasbp.min  
men.diasbp.range.over.std <- men.diasbp.range / men.diasbp.std  
men.diasbp.largest.zscore <- (men.diasbp.max - men.diasbp.mean) / men.diasbp.std  
men.diasbp.smallest.zscore <- (men.diasbp.min - men.diasbp.mean) / men.diasbp.std

Sorted: 40, 46, 50, 50, 52, 54, 54, 56, 56, 60, 60, 60, 60, 60, 60, 60, 60, 60, 60, 61, 62, 62, 62, 62, 62, 64, 64, 65, 65, 65, 66, 66, 68, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 72, 72, 72, 74, 74, 74, 74, 76, 76, 76, 76, 78, 78, 78, 78, 78, 80, 80, 80, 80, 80, 80, 80, 80, 80, 80, 80, 82, 82, 82, 84, 84, 84, 84, 84, 86, 86, 86, 88, 88, 88, 88, 88, 90, 90, 94, 96, 97, 106

Mean, 𝑥̅: 72.35

Variance, s^2: 141.9469697

Standard deviation, s: 11.91415

Lower Quartile: Lower Quartile: 63.500000

Upper Quartile: Upper Quartile: 80.000000

Min value: Min value: 40.000000

Max value: Max value: 106.000000

Range: 66

The range is approximately 5.5396315 standard deviations

Median: 70

IQR: 16.5

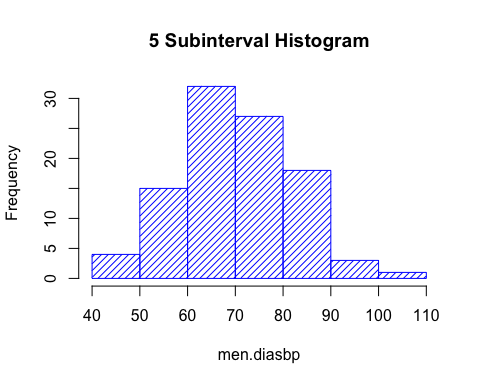
Five-number summaries: 40, 63.5, 70, 80, 106

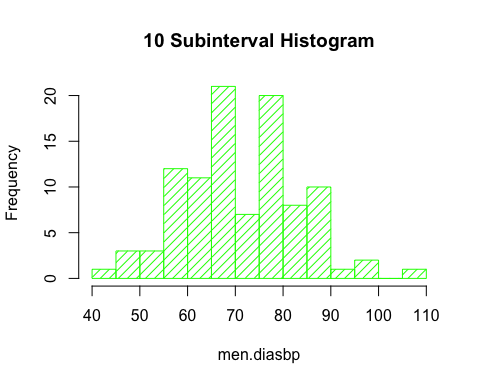
### Box Plot

### Stem Plot

##   
## The decimal point is 1 digit(s) to the right of the |  
##   
## 4 | 06  
## 5 | 0024466  
## 6 | 000000000012222244555668  
## 7 | 0000000000000000002224444666688888  
## 8 | 000000000002224444466688888  
## 9 | 00467  
## 10 | 6

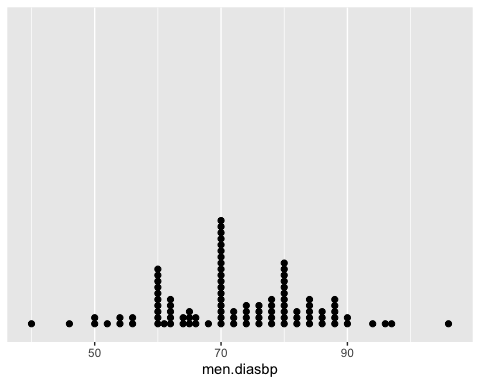
### Histograms





Description

### Dotplot



The data is not mound shaped.

We can not use the Tchebysheff’s Theorem to describe this data set. Because the set of measurement is skewed. The shape is skewed-right.

We can not use the Empirical Rule to describe this data set. Because it’s not mound-shaped.

q.Construct

r.Construct

## z-scores of largest and smallest observations

Smallest observation z-score: 3.601991

Largest observation z-score: 3.601991

*3 -* WOMEN SYSTOLIC BLOOD PRESSURE

women.sysbp.sorted <- sort(women.sysbp)  
women.sysbp.mean <- mean(women.sysbp)  
women.sysbp.most.frequent.value <- Mode(women.sysbp)  
women.sysbp.most.frequent.value.count <- length(dataSet[which(dataSet$SYSBP == women.sysbp.most.frequent.value),]['SYSBP'][, 1])  
women.sysbp.median <- median(women.sysbp)  
women.sysbp.max <- max(women.sysbp)  
women.sysbp.min <- min(women.sysbp)  
women.sysbp.sum <- sum(women.sysbp)  
women.sysbp.std <- sd(women.sysbp, na.rm = FALSE)  
women.sysbp.fiveNumbersSumy <- quantile(women.sysbp)  
women.sysbp.lowerQuartile <- women.sysbp.fiveNumbersSumy[2]  
women.sysbp.upperQuartile <- women.sysbp.fiveNumbersSumy[4]  
women.sysbp.interQuartile <- women.sysbp.upperQuartile - women.sysbp.lowerQuartile  
women.sysbp.variance <- women.sysbp.std^2  
women.sysbp.range <- women.sysbp.max - women.sysbp.min  
women.sysbp.range.over.std <- women.sysbp.range / women.sysbp.std  
women.sysbp.largest.zscore <- (women.sysbp.max - women.sysbp.mean) / women.sysbp.std  
women.sysbp.smallest.zscore <- (women.sysbp.min - women.sysbp.mean) / women.sysbp.std

Sorted: 80, 82, 88, 88, 88, 90, 90, 92, 94, 96, 96, 96, 98, 98, 98, 98, 98, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 102, 104, 104, 104, 105, 106, 106, 106, 108, 108, 108, 108, 110, 110, 110, 110, 110, 110, 110, 110, 110, 110, 110, 110, 110, 110, 110, 110, 110, 110, 110, 112, 112, 112, 114, 114, 114, 114, 114, 115, 115, 116, 118, 118, 118, 118, 120, 120, 120, 120, 120, 120, 120, 120, 120, 122, 124, 126, 126, 126, 126, 128, 130, 130, 130, 140, 160

Mean, 𝑥̅: 109.01

Variance, s^2: 153.6261616

Standard deviation, s: 12.3946021

Lower Quartile: Lower Quartile: 100.000000

Upper Quartile: Upper Quartile: 116.500000

Min value: Min value: 80.000000

Max value: Max value: 160.000000

Range: 80

The range is approximately 6.4544226 standard deviations

Median: 110

IQR: 16.5

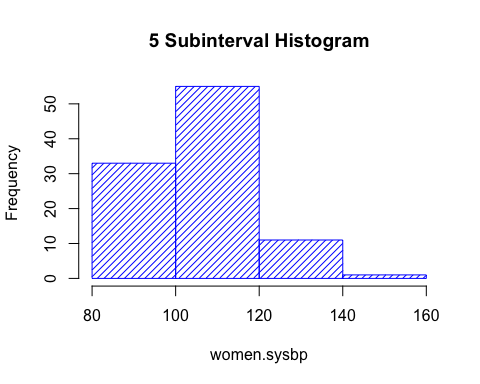
Five-number summaries: 80, 100, 110, 116.5, 160

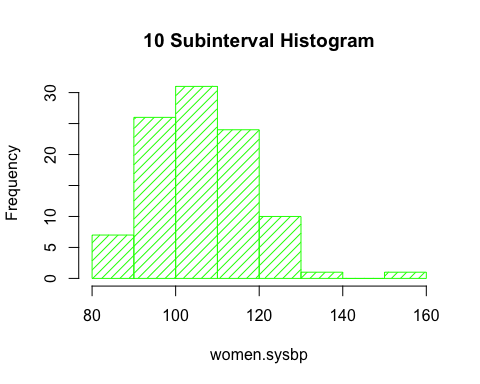
### Box Plot

### Stem Plot

##   
## The decimal point is 1 digit(s) to the right of the |  
##   
## 8 | 02888  
## 9 | 002466688888  
## 10 | 0000000000000000244456668888  
## 11 | 0000000000000000000222444445568888  
## 12 | 0000000002466668  
## 13 | 000  
## 14 | 0  
## 15 |   
## 16 | 0

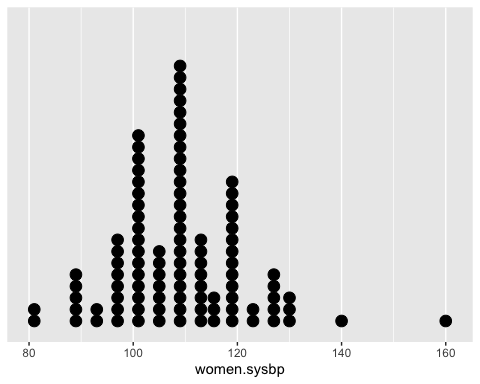
### Histograms





Description

### Dotplot



The data is not mound shaped.

We can not use the Tchebysheff’s Theorem to describe this data set. Because the set of measurement is skewed. The shape is skewed-right.

We can not use the Empirical Rule to describe this data set. Because it’s not mound-shaped.

q.Construct

r.Construct

## z-scores of largest and smallest observations

Smallest observation z-score: 4.1138876

Largest observation z-score: 4.1138876

*3 -* WOMEN DIASTOLIC BLOOD PRESSURE

women.diasbp.sorted <- sort(women.diasbp)  
women.diasbp.mean <- mean(women.diasbp)  
women.diasbp.most.frequent.value <- Mode(women.diasbp)  
women.diasbp.most.frequent.value.count <- length(dataSet[which(dataSet$DIASBP == men.diasbp.most.frequent.value),]['DIASBP'][, 1])  
women.diasbp.median <- median(women.diasbp)  
women.diasbp.max <- max(women.diasbp)  
women.diasbp.min <- min(women.diasbp)  
women.diasbp.sum <- sum(women.diasbp)  
women.diasbp.std <- sd(women.diasbp, na.rm = FALSE)  
women.diasbp.fiveNumbersSumy <- quantile(women.diasbp)  
women.diasbp.lowerQuartile <- women.diasbp.fiveNumbersSumy[2]  
women.diasbp.upperQuartile <- women.diasbp.fiveNumbersSumy[4]  
women.diasbp.interQuartile <- women.diasbp.upperQuartile - women.diasbp.lowerQuartile  
women.diasbp.variance <- women.diasbp.std^2  
women.diasbp.range <- women.diasbp.max - women.diasbp.min  
women.diasbp.range.over.std <- women.diasbp.range / women.diasbp.std  
women.diasbp.largest.zscore <- (women.diasbp.max - women.diasbp.mean) / women.diasbp.std  
women.diasbp.smallest.zscore <- (women.diasbp.min - women.diasbp.mean) / women.diasbp.std

Sorted: 50, 50, 54, 54, 56, 58, 58, 60, 60, 60, 60, 60, 60, 60, 60, 60, 60, 60, 60, 62, 64, 64, 64, 65, 65, 65, 66, 66, 66, 66, 66, 68, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 70, 72, 72, 74, 74, 74, 74, 74, 76, 76, 76, 76, 76, 78, 78, 78, 78, 78, 78, 80, 80, 80, 80, 80, 80, 80, 80, 80, 82, 82, 85, 86, 90, 90, 94, 116

Mean, 𝑥̅: 70.64

Variance, s^2: 91.9094949

Standard deviation, s: 9.586944

Lower Quartile: Lower Quartile: 65

Upper Quartile: Upper Quartile: 76

Min value: Min value: 50

Max value: Max value: 116

Range: 66

The range is approximately 6.8843627 standard deviations

Median: 70

IQR: 11

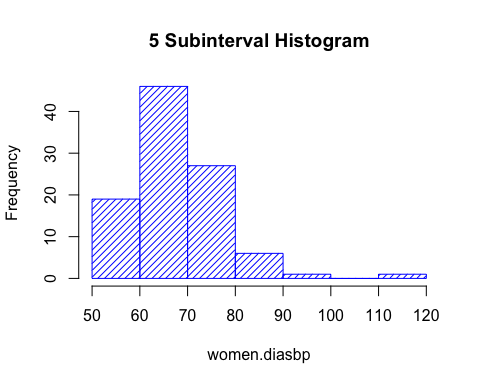
Five-number summaries: 50, 65, 70, 76, 116

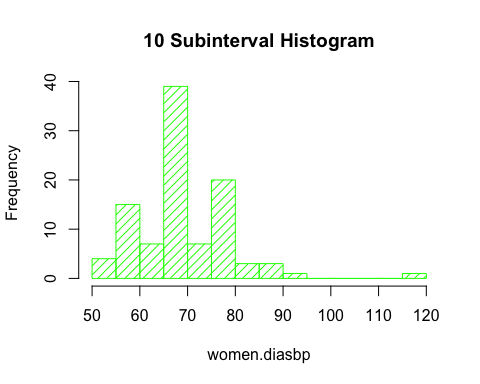
### Box Plot

### Stem Plot

##   
## The decimal point is 1 digit(s) to the right of the |  
##   
## 5 | 0044688  
## 6 | 0000000000002444555666668  
## 7 | 000000000000000000000000000000000224444466666888888  
## 8 | 0000000002256  
## 9 | 004  
## 10 |   
## 11 | 6

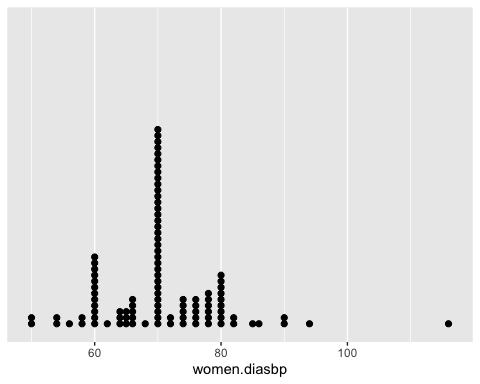
### Histograms





Description

### Dotplot



The data is not mound shaped.

We can not use the Tchebysheff’s Theorem to describe this data set. Because the set of measurement is skewed. The shape is skewed-right.

We can not use the Empirical Rule to describe this data set. Because it’s not mound-shaped.

q.Construct

r.Construct

## z-scores of largest and smallest observations

Smallest observation z-score: 4.1138876

Largest observation z-score: 4.1138876