chsi.new.R

jiayuan

Sun Oct 18 14:39:18 2015

```r
setwd("/Users/jiayuan/Documents/MA881/HW6/chsi_dataset")
library(ggplot2)
health <- read.csv("RELATIVEHEALTHIMPORTANCE.csv",header=T)
names(health) <- c('SC', 'CC', 'CN', 'SN', 'SA', 'SIN', 'LBW', 'VLBW', 'Premature',
                   'U18', 'O40', 'Unmarried','LateCare', 'IM', 'IMWNH', 'IMBNH',
                   'IMH', 'IMN', 'IMPN', 'BrstCan', 'ColCan', 'CHD', 'Homicide',
                   'LungCan', 'MVA', 'Stroke', 'Suicide', 'Injury')
attach(health)
Worcester.health <- health[which(CN=="Worcester"),][2,]
Worcester.health
```

```
##      SC CC       CN            SN SA SIN LBW VLBW Premature U18 O40
## 1229 25 27 Worcester Massachusetts MA  2   5   5            5   5   6
##      Unmarried LateCare IM IMWNH IMBNH IMH IMN IMPN BrstCan ColCan CHD
## 1229         7        5  5     7   8   8    7       5       8     8   5
##      Homicide LungCan MVA Stroke Suicide Injury
## 1229        5       7   7      7       5       5
```

```r
num <- 0
for (i in 1:8){
  num[i] <- length(which(Worcester.health==i))
}
num
```

```
## [1]  0  1  0  0 11  1  6  4
```

```r
#Relative health indicator is 5: 11 measures,
#Represent ''Favorable to peers and favorable the U.S. Rate
#Relative health indicator is 6: 1 measures,
#Represent 'Favorable to peers and unfavorable the U.S. Rate
#Relative health indicator is 7: 6 measures,
#Represent 'Unfavorable to peers and favorable the U.S. Rate
#Relative health indicator is 8: 4 measures,
#Represent 'Unfavorable to peers and unfavorable the U.S. Rate

#measures of interest:
#SC, CC
#VLBW, IM, CHD, LungCan
rhi <- cbind(SC, CC, VLBW, IM, CHD, LungCan)
head(rhi)
```

```
##      SC CC VLBW IM CHD LungCan
## [1,]  1  1    8  5   6       8
## [2,]  1  3    8  6   5       8
## [3,]  1  5    8  6   5       8
## [4,]  1  7    8  8   6       8
```

```
## [5,]  1  9    8  8   5       8
## [6,]  1 11    8  8   8       5
```

```
#drop the data of the two situations:  no data available(-2) and no report(-1)
delete=which(VLBW=="-1" | VLBW=="-2" | IM=="-1" | IM=="-2" | CHD=="-1"
             | CHD=="-2" | LungCan=="-1" | LungCan=="-2")
rhi=rhi[-c(delete),]
#new data frame of the six measures after droping the useless data
rhi=as.data.frame(rhi)
attach(rhi)
```

```
## The following objects are masked from health:
##
##     CC, CHD, IM, LungCan, SC, VLBW
```

```
#compare the four measures VLBW, IM, CHD, LungCan for differnt counties
#within MA state
MA <- rhi[which(SC=="25"),] #MA has the state identifier of 25 in US
MA
```

```
##        SC CC VLBW IM CHD LungCan
## 1137 25   1    5  5   5       5
## 1138 25   3    8  5   5       8
## 1139 25   5    8  7   8       7
## 1140 25   7    8  5   5       5
## 1141 25   9    5  5   5       8
## 1142 25  11    5  5   5       5
## 1143 25  13    6  5   5       6
## 1144 25  15    5  5   5       5
## 1145 25  17    5  5   5       7
## 1146 25  19    8  8   8       8
## 1147 25  21    5  5   5       7
## 1148 25  23    7  5   5       8
## 1149 25  25    8  5   5       8
## 1150 25  27    5  5   5       7
```

```
nrow(MA) #14 counties in MA
```

```
## [1] 14
```

```
Worcester <- MA[which(MA$CC=="27"),] #Worcester has the county identifier of 27 in MA
Worcester
```

```
##        SC CC VLBW IM CHD LungCan
## 1150 25  27    5  5   5       7
```

```
#    SC CC VLBW IM CHD LungCan
#1150 25 27    5  5   5       7
#VLBW: 5, representing 'Favorable to peers and favorable the U.S. Rate'
#IM: 5, representing 'Favorable to peers and favorable the U.S. Rate'
#CHD: 5, representing 'Favorable to peers and favorable the U.S. Rate'
```

```
#LungCan: 7, representing 'Unfavorable to peers and favorable the U.S. Rate'

MAdata = MA[,3:6]
MA.VLBW <- MAdata$VLBW
MA.IM <- MAdata$IM
MA.CHD <- MAdata$CHD
MA.LungCan <- MAdata$LungCan

mysummary <- function (x) {
  options (digits=4)
  matrix (c( mean(x), sd(x), length(x), min(x),
             quantile(x,.25), median(x),
             quantile(x,.75), max(x), sum(is.na(x))), ncol=9,
        dimnames = list (NULL, c( "Mean", "SD", "n", "Min",
                                  "1st Qu", "Median", "3rd Qu", "Max", "NAs")))}

#summaries of MAdata data frame
summary (MAdata)
```

```
##      VLBW             IM              CHD            LungCan
## Min.   :5.000   Min.   :5.000   Min.   :5.000   Min.   :5.000
## 1st Qu.:5.000   1st Qu.:5.000   1st Qu.:5.000   1st Qu.:5.250
## Median :5.500   Median :5.000   Median :5.000   Median :7.000
## Mean   :6.286   Mean   :5.357   Mean   :5.429   Mean   :6.714
## 3rd Qu.:8.000   3rd Qu.:5.000   3rd Qu.:5.000   3rd Qu.:8.000
## Max.   :8.000   Max.   :8.000   Max.   :8.000   Max.   :8.000
```

```
mysummary(MA.VLBW)
```

```
##       Mean    SD  n Min 1st Qu Median 3rd Qu Max NAs
## [1,] 6.286 1.437 14   5      5    5.5      8   8   0
```

```
mysummary(MA.IM)
```

```
##       Mean     SD  n Min 1st Qu Median 3rd Qu Max NAs
## [1,] 5.357 0.9288 14   5      5      5      5   8   0
```

```
mysummary(MA.CHD)
```

```
##       Mean    SD  n Min 1st Qu Median 3rd Qu Max NAs
## [1,] 5.429 1.089 14   5      5      5      5   8   0
```

```
mysummary(MA.LungCan)
```

```
##       Mean    SD  n Min 1st Qu Median 3rd Qu Max NAs
## [1,] 6.714 1.267 14   5   5.25      7      8   8   0
```

```
#stem-plot of MAdata data frame
stem(MA.VLBW)
```

```
## 
##   The decimal point is at the |
## 
##   5 | 0000000
##   6 | 0
##   7 | 0
##   8 | 00000
```

```r
stem(MA.IM)
```

```
## 
##   The decimal point is at the |
## 
##   5 | 000000000000
##   6 |
##   7 | 0
##   8 | 0
```
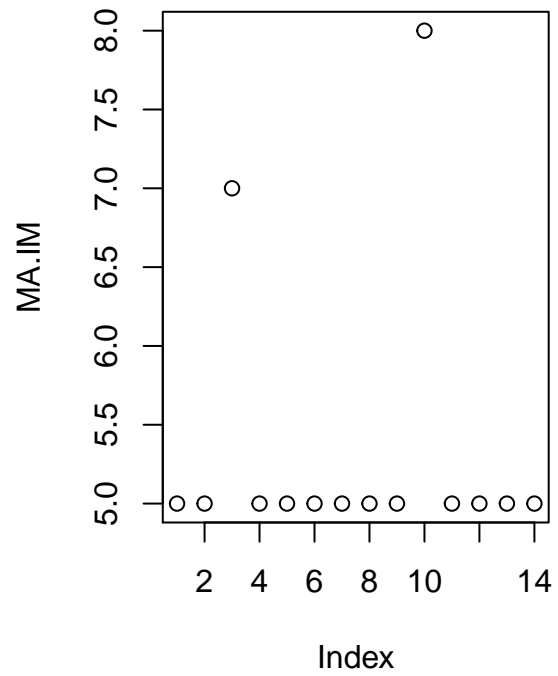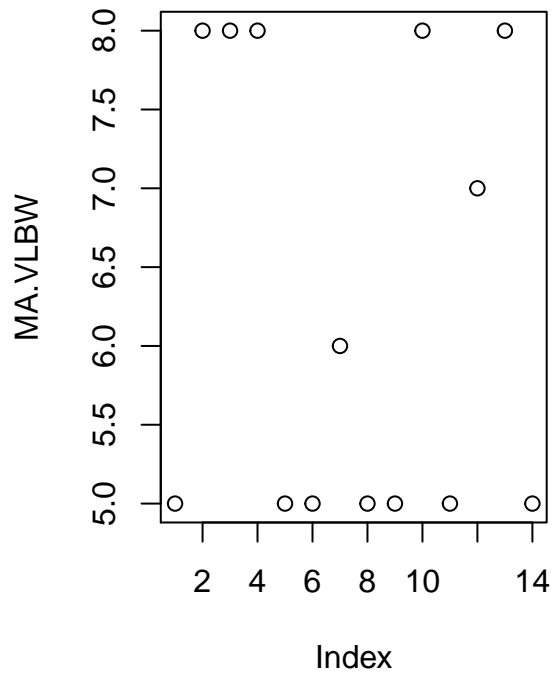
```r
stem(MA.CHD)
```

```
## 
##   The decimal point is at the |
## 
##   5 | 000000000000
##   6 |
##   7 |
##   8 | 00
```
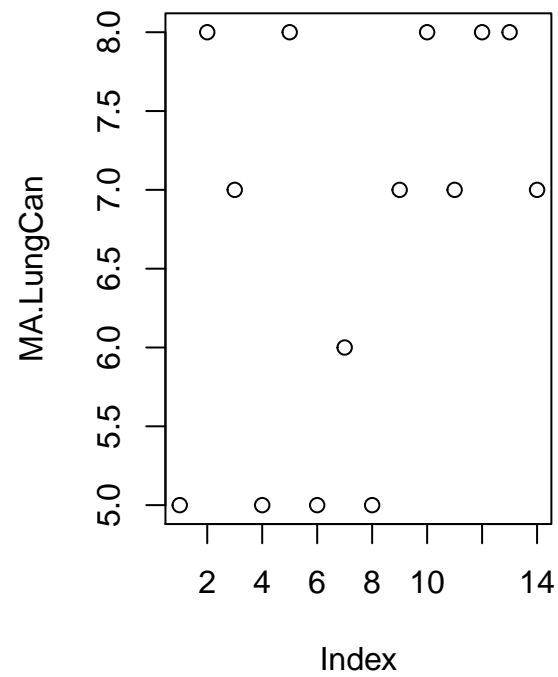
```r
stem(MA.LungCan)
```

```
## 
##   The decimal point is at the |
## 
##   5 | 0000
##   6 | 0
##   7 | 0000
##   8 | 00000
```

```r
#plot of MAdata data frame
par (mfrow=c(1,2))
plot(MA.VLBW)
plot(MA.IM)
```

```
plot(MA.CHD)
plot(MA.LungCan)
```
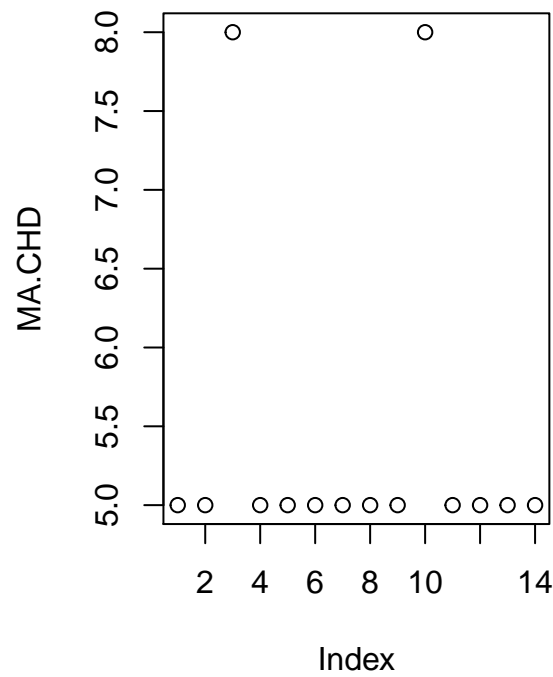


```
#histgram of MAdata data frame
hist(MA.VLBW)
hist(MA.IM)
```
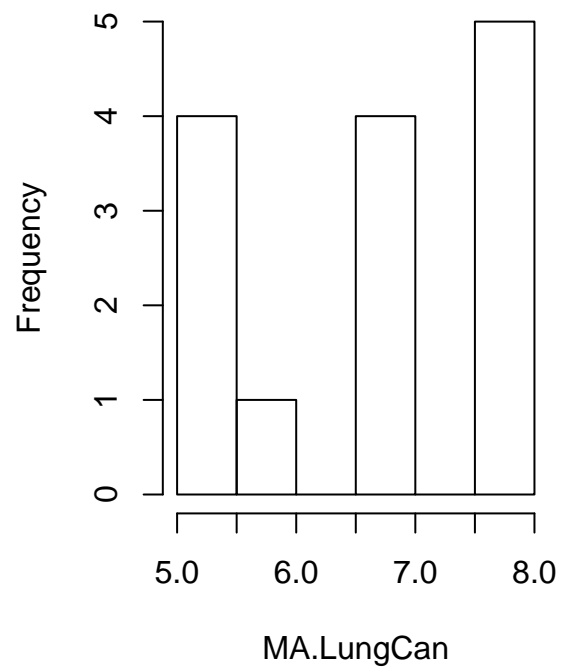
Histogram of MA.VLBW
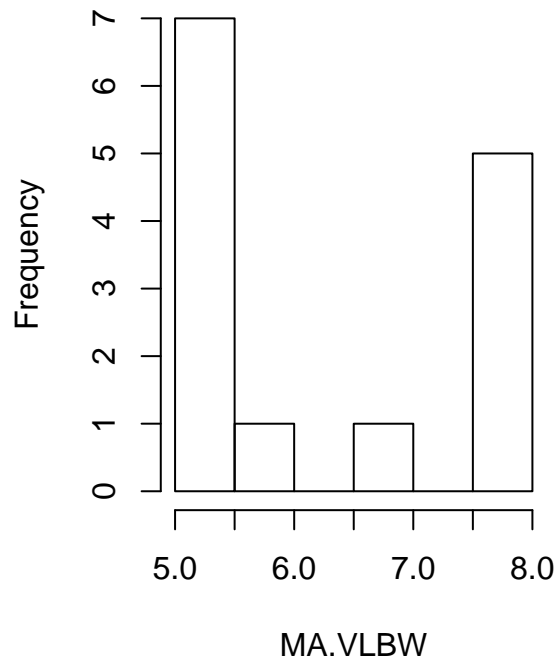
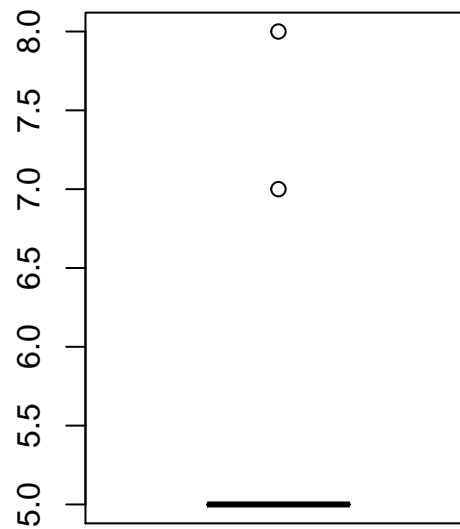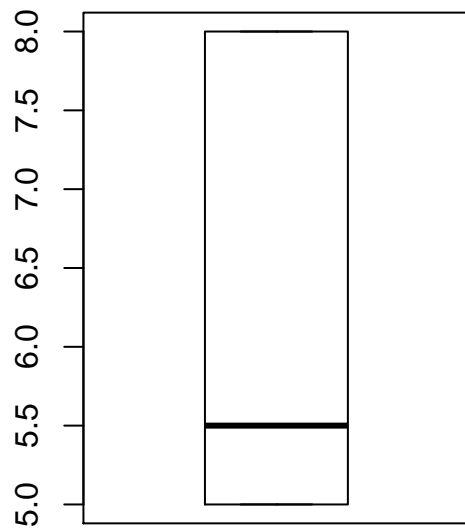## Histogram of MA.IM
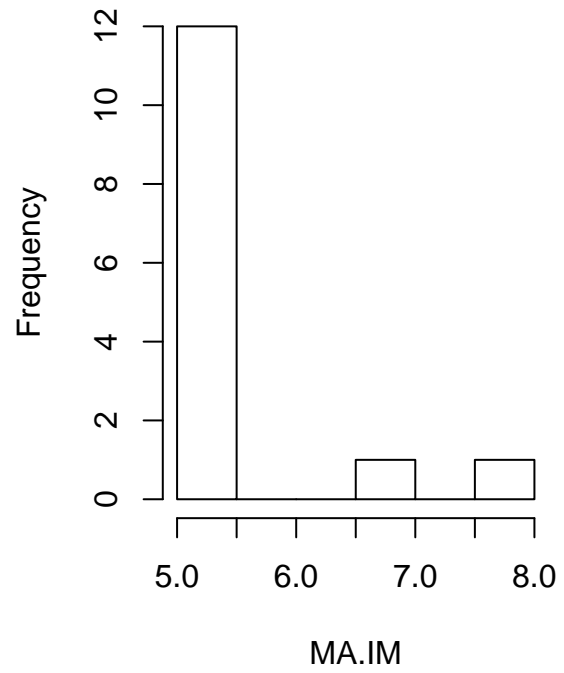

```
hist(MA.CHD)
hist(MA.LungCan)
```

Histogram of MA.CHD
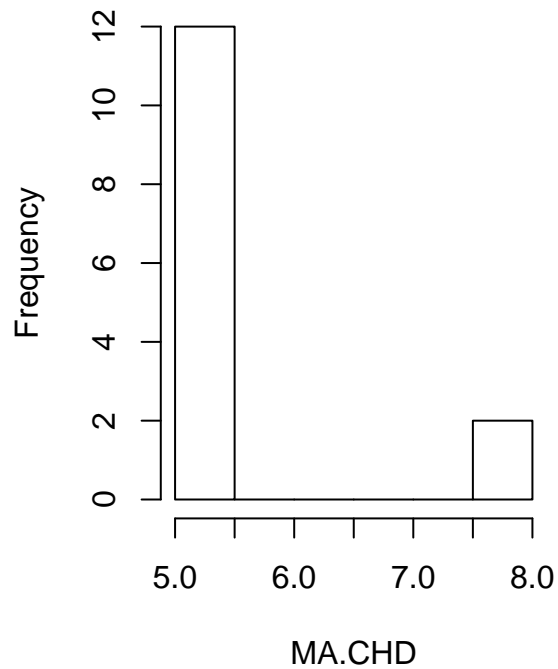
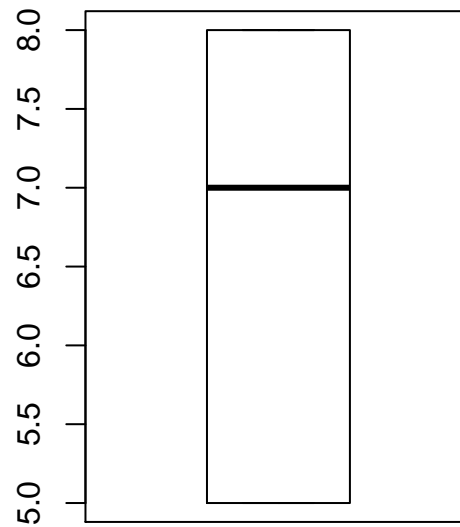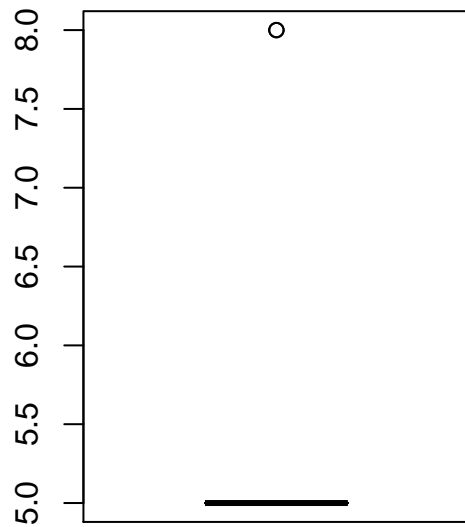## Histogram of MA.LungCan

```
#boxplot of MAdata data frame
boxplot(MA.VLBW)
boxplot(MA.IM)
```
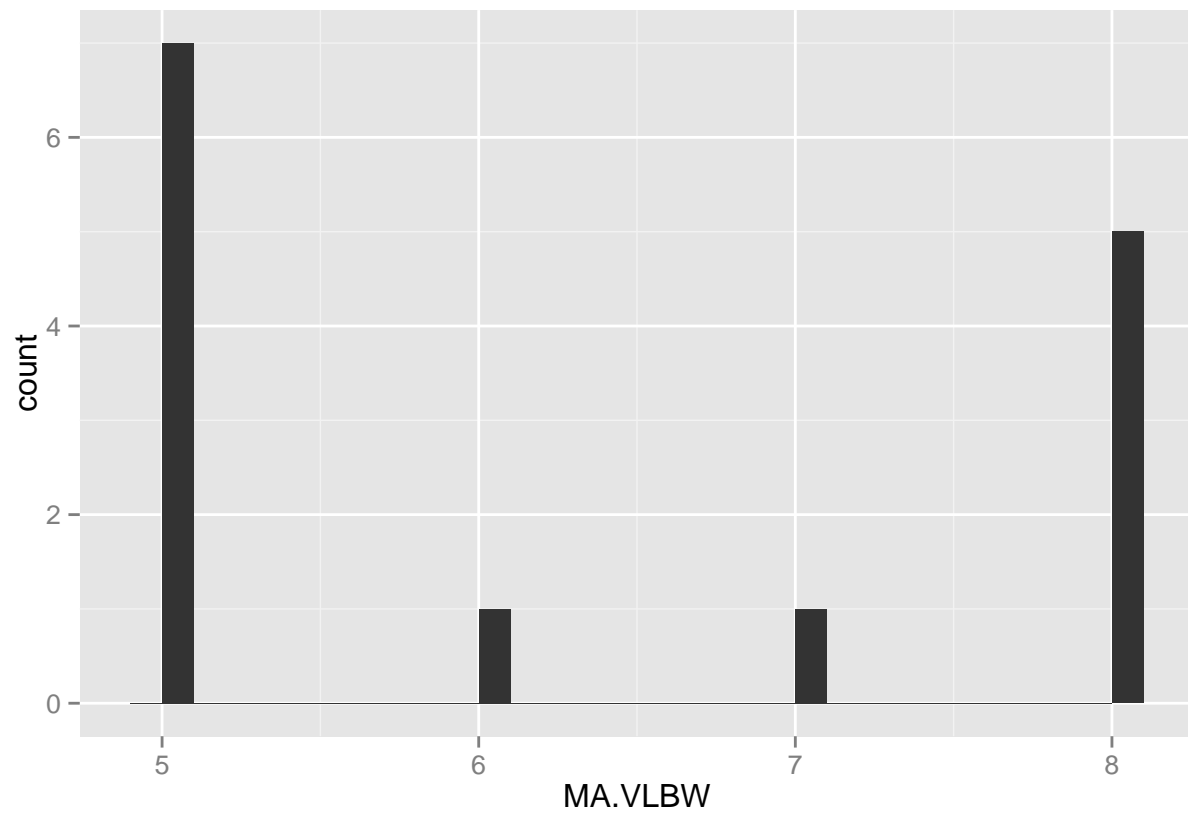


```
boxplot(MA.CHD)
boxplot(MA.LungCan)
```



```
#qplot of MAdata data frame
qplot(MA.VLBW,data = MAdata, geom = "histogram")
```
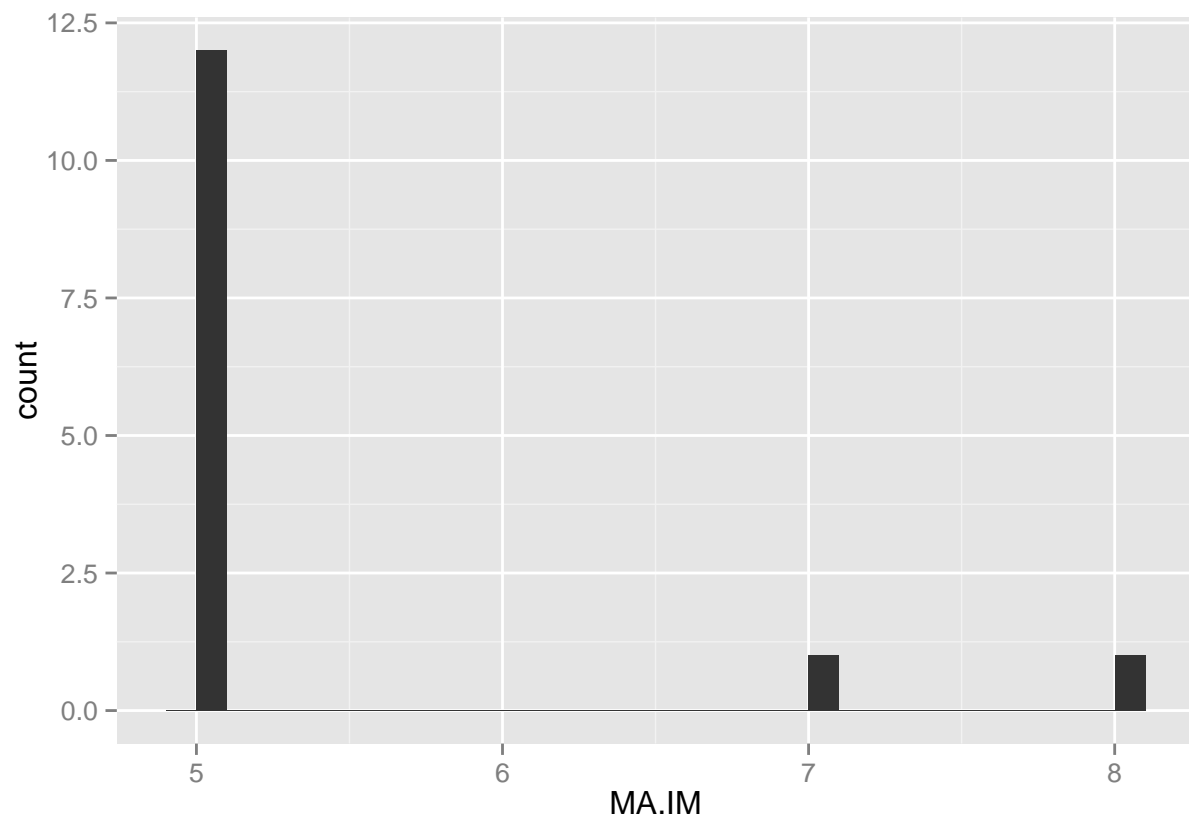
```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```

```
qplot(MA.IM,data = MAdata, geom = "histogram")
```
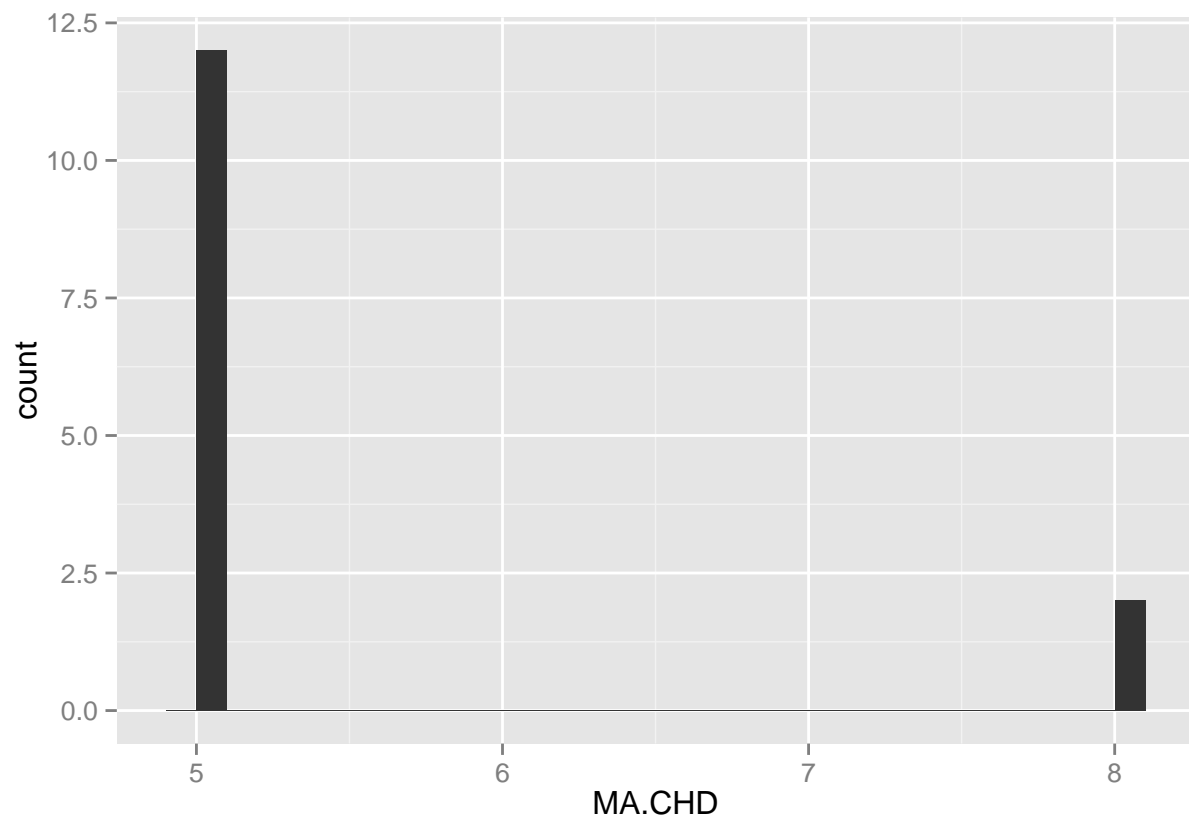
```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```
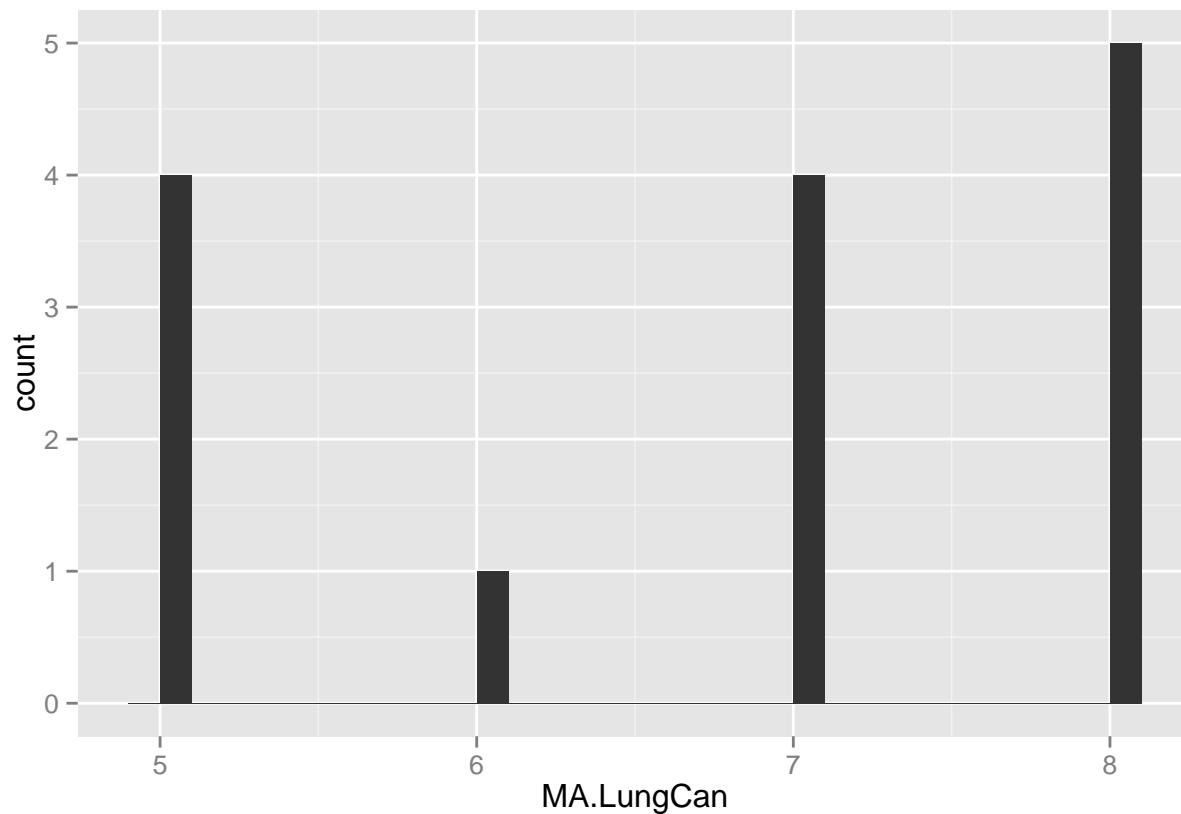
```r
qplot(MA.CHD,data = MAdata, geom = "histogram")
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```

```r
qplot(MA.LungCan,data = MAdata, geom = "histogram")
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```

```r
#use The Elbow Method to look at the percentage of variance
#explained as a function of the number of clusters
MAwss <- (nrow(MAdata)-1)*sum(apply(MAdata,2,var))
for (i in 2:8) {
  MAwss[i] <- sum(kmeans(MAdata,centers=i)$withinss)
}
#plot to deterine the number of clusters
plot(1:8, MAwss,type="b", xlab="Number of Clusters",
     ylab="Within groups sum of squares")
#The location of the elbow in the resulting plot suggests
#a suitable number of clusters for the kmeans is 3
MAK3 <- kmeans(x = MAdata, centers = 3)
MAK3
```

```
## K-means clustering with 3 clusters of sizes 2, 8, 4
##
## Cluster means:
##     VLBW   IM CHD LungCan
## 1 8.000 7.5   8    7.50
## 2 5.125 5.0   5    6.25
## 3 7.750 5.0   5    7.25
##
## Clustering vector:
## 1137 1138 1139 1140 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150
##    2    3    1    3    2    2    2    2    2    1    2    3    3    2
##
## Within cluster sum of squares by cluster:
## [1]   1.00 10.38  7.50
```

```
##  (between_SS / total_SS =  74.6 %)
##
## Available components:
##
## [1] "cluster"      "centers"       "totss"          "withinss"
## [5] "tot.withinss" "betweenss"     "size"           "iter"
## [9] "ifault"
```
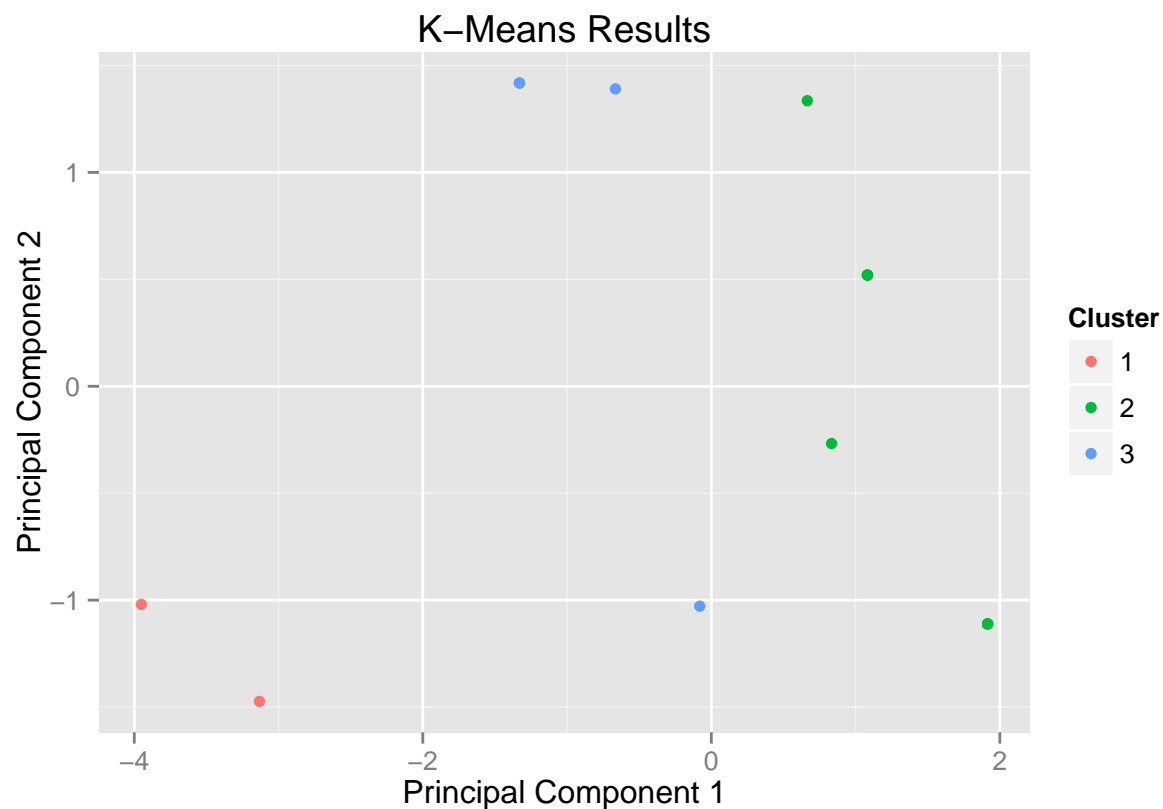
```r
#plot the 3 groups of MAdata data frame
require(useful)
```

```
## Loading required package: useful
```

```r
par (mfrow=c(1,1))
```
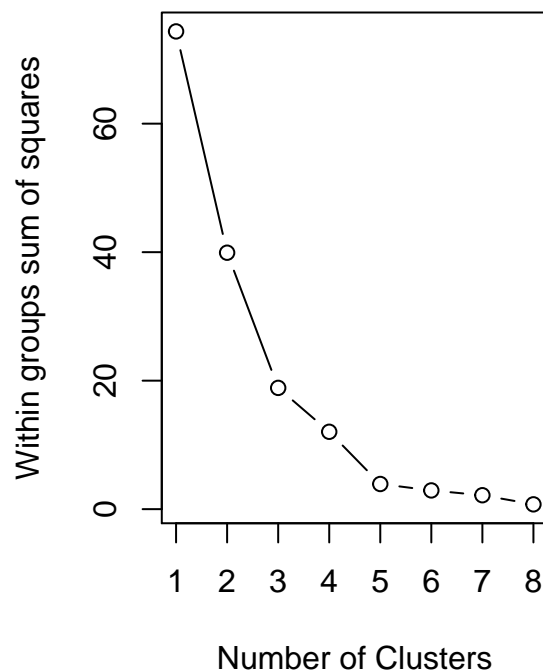


```r
plot(MAK3, data = MAdata)
```

K–Means Results


```r
#compare the four measures VLBW, IM, CHD, LungCan for differnt counties
#within the U.S.
rhidata = rhi[,3:6]
#summaries of rhidata dataset
summary (rhidata)
```

```
##       VLBW            IM             CHD           LungCan
##  Min.   :5.00   Min.   :5.00   Min.   :5.00   Min.   :5.00
##  1st Qu.:5.00   1st Qu.:5.00   1st Qu.:5.00   1st Qu.:5.00
##  Median :6.00   Median :6.00   Median :6.00   Median :6.00
##  Mean   :6.39   Mean   :6.53   Mean   :6.64   Mean   :6.65
##  3rd Qu.:8.00   3rd Qu.:8.00   3rd Qu.:8.00   3rd Qu.:8.00
##  Max.   :8.00   Max.   :8.00   Max.   :8.00   Max.   :8.00
```

```r
mysummary(VLBW)
```

```
##        Mean    SD     n Min 1st Qu Median 3rd Qu Max NAs
## [1,] 6.393 1.356 2860   5      5      6      8   8   0
```

```r
mysummary(IM)
```

```
##        Mean    SD     n Min 1st Qu Median 3rd Qu Max NAs
## [1,] 6.529 1.386 2860   5      5      6      8   8   0
```

```r
mysummary(CHD)
```

```
##        Mean    SD    n Min 1st Qu Median 3rd Qu Max NAs
## [1,] 6.641 1.338 2860   5      5      6      8   8   0
```
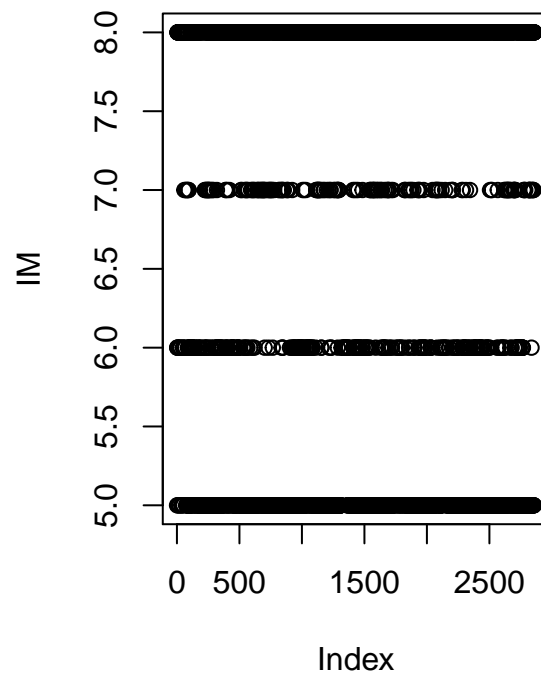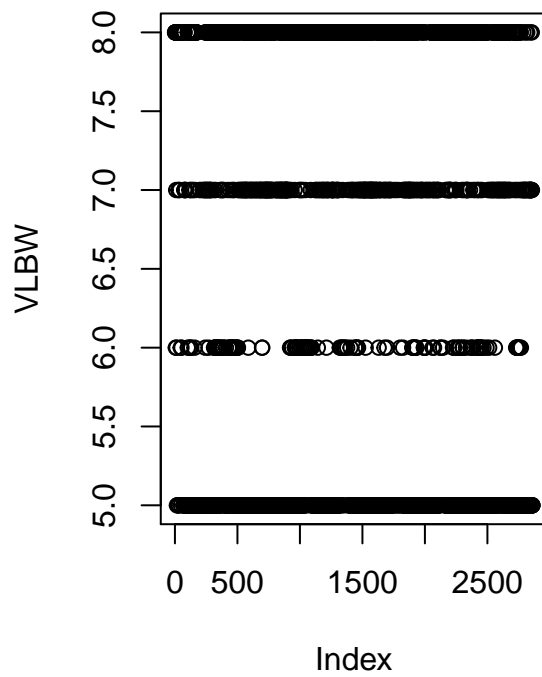
```r
mysummary(LungCan)
```

```
##        Mean    SD    n Min 1st Qu Median 3rd Qu Max NAs
## [1,] 6.647 1.317 2860   5      5      6      8   8   0
```

```r
#plot of rhidata dataset
par (mfrow=c(1,2))
plot(VLBW)
plot(IM)
```
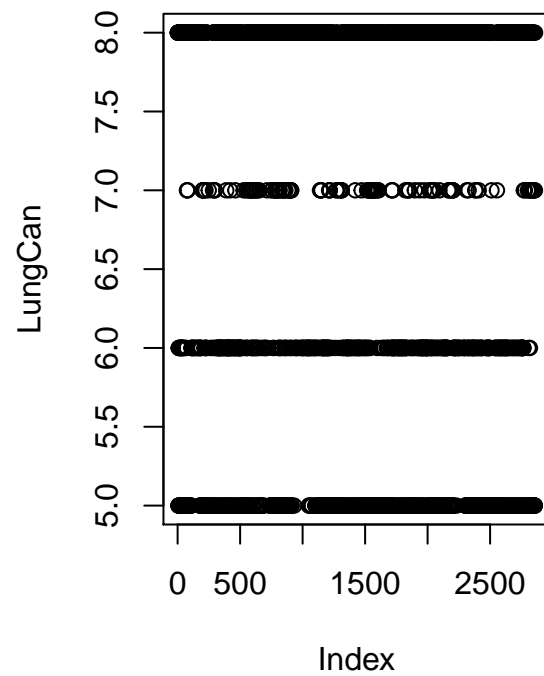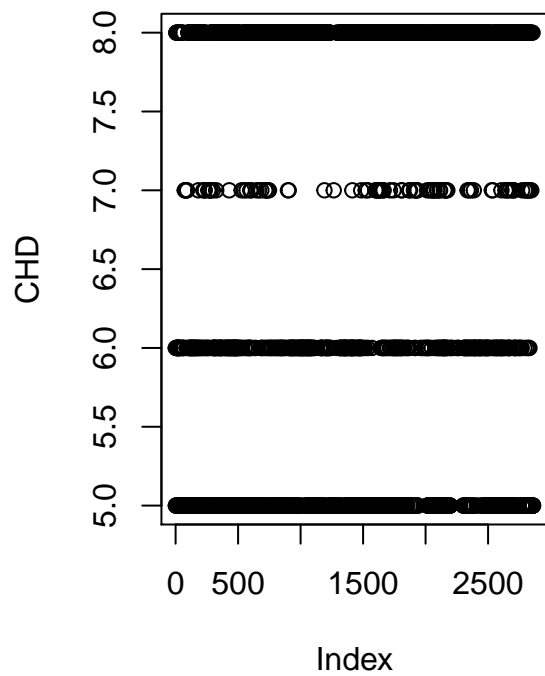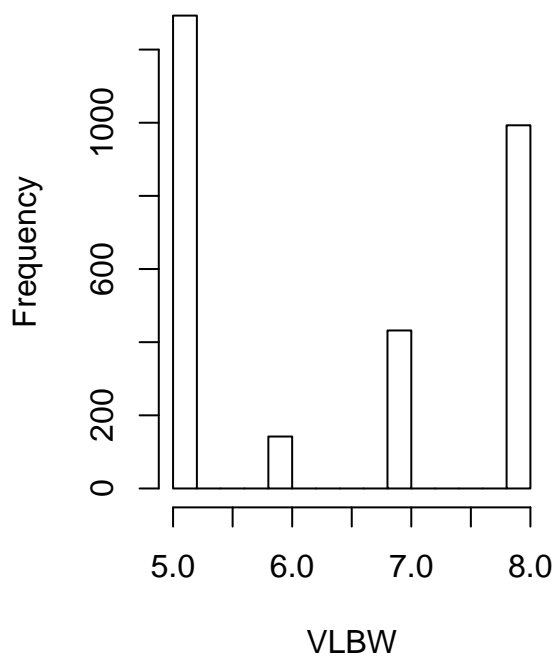


```r
plot(CHD)
plot(LungCan)
```

```
#histgram of rhidata dataset
hist(VLBW)
hist(IM)
```

Histogram of VLBW

Histogram of IM
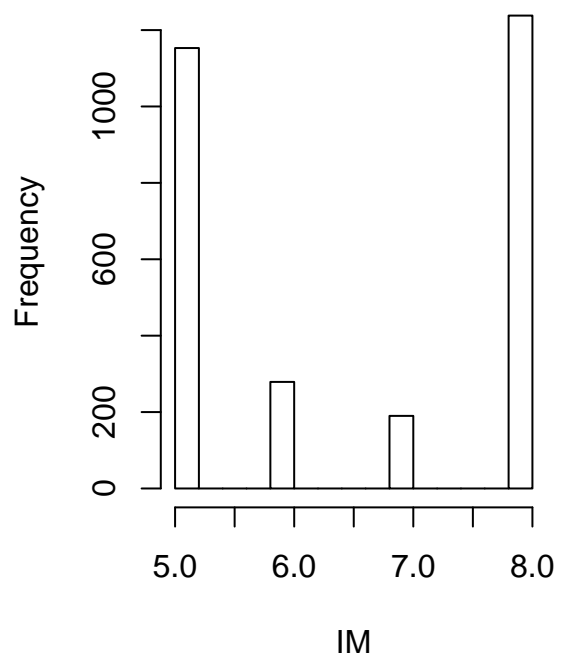

```
hist(CHD)
hist(LungCan)
```

Histogram of CHD
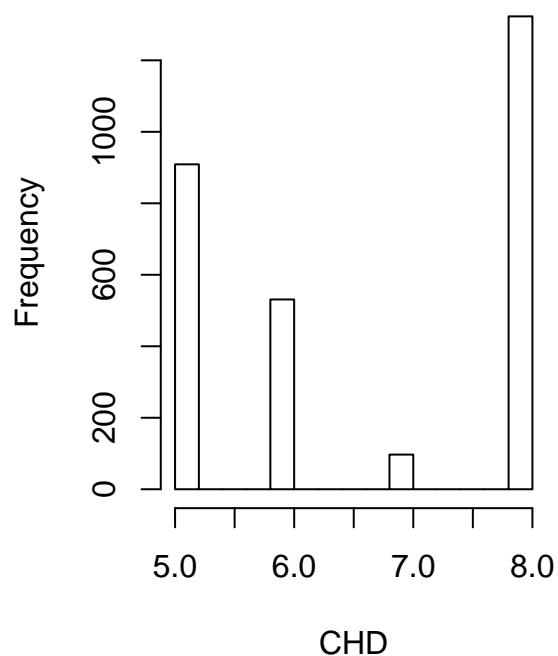
Histogram of LungCan
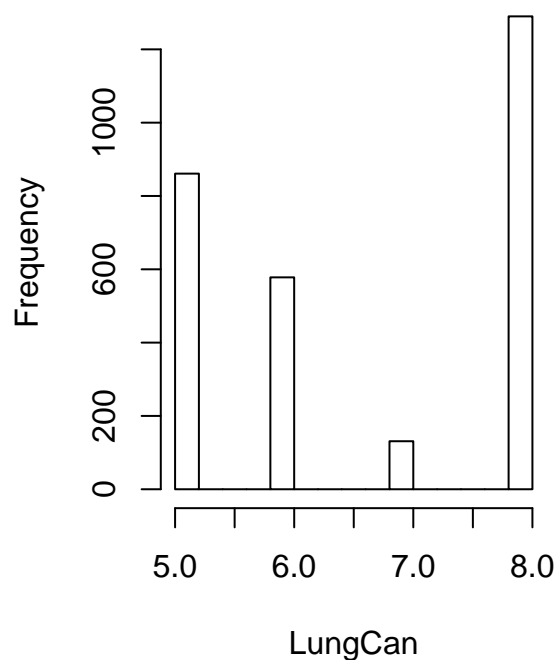

```
#boxplot of rhidata dataset
boxplot(VLBW)
boxplot(IM)
boxplot(CHD)
boxplot(LungCan)
```
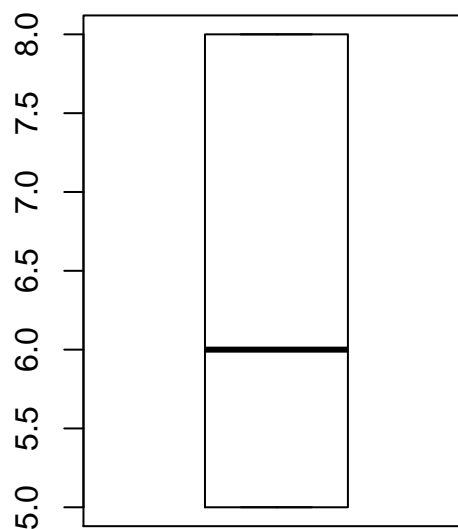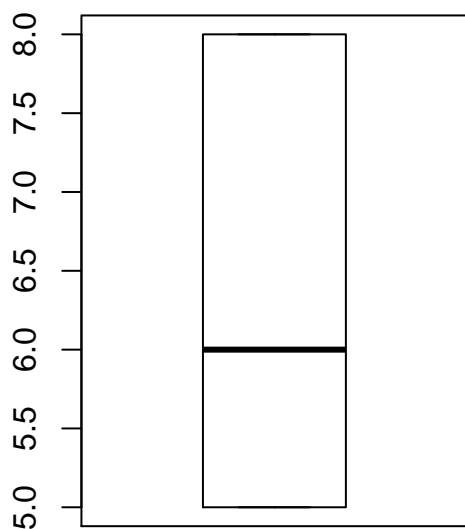


```
#qplot of rhidata dataset
qplot(VLBW,data = rhidata, geom = "histogram")
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```
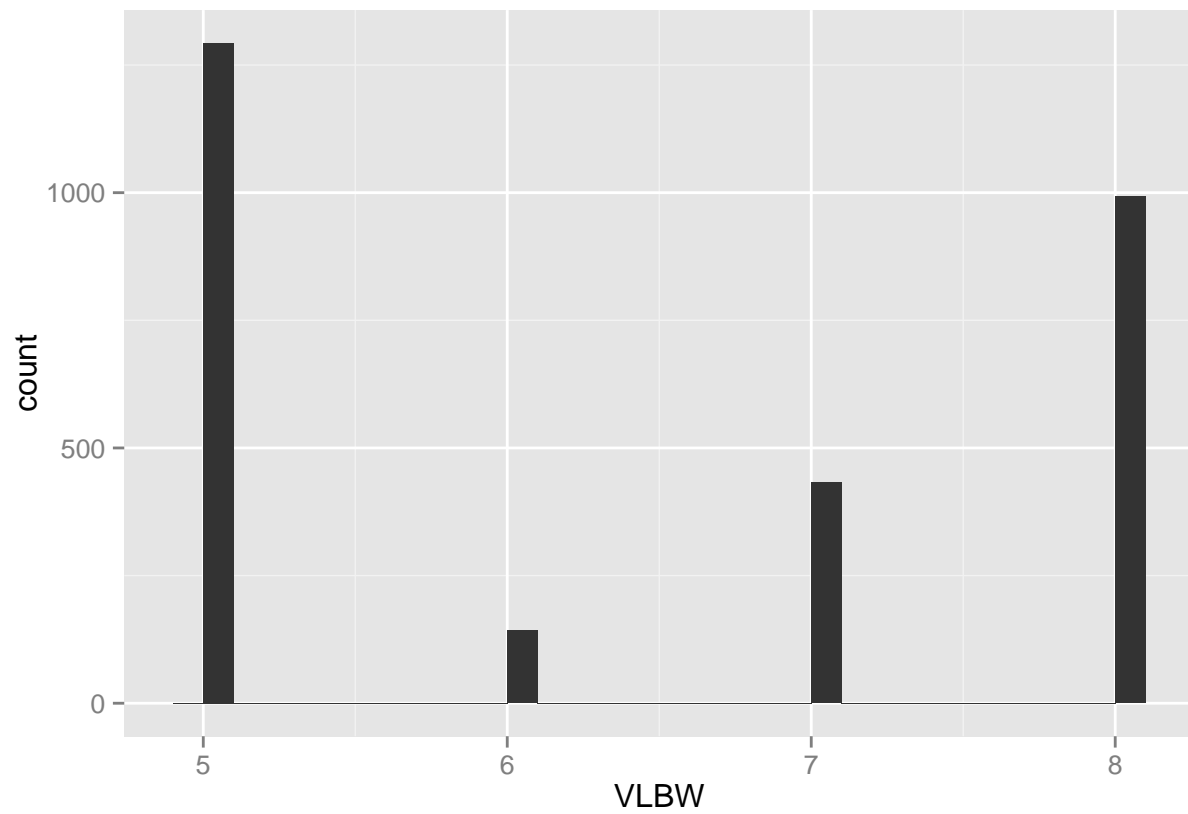
```
qplot(IM,data = rhidata, geom = "histogram")
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```
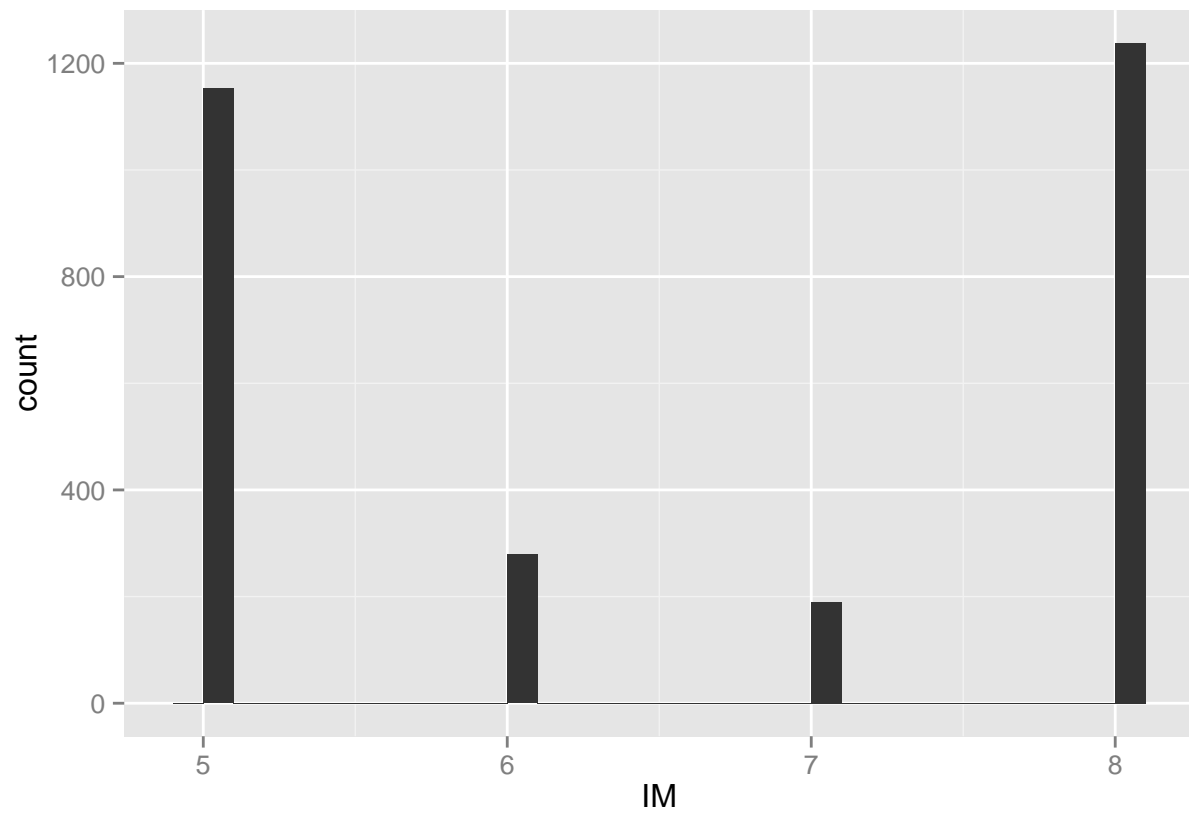
```
qplot(CHD,data = rhidata, geom = "histogram")
```

stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
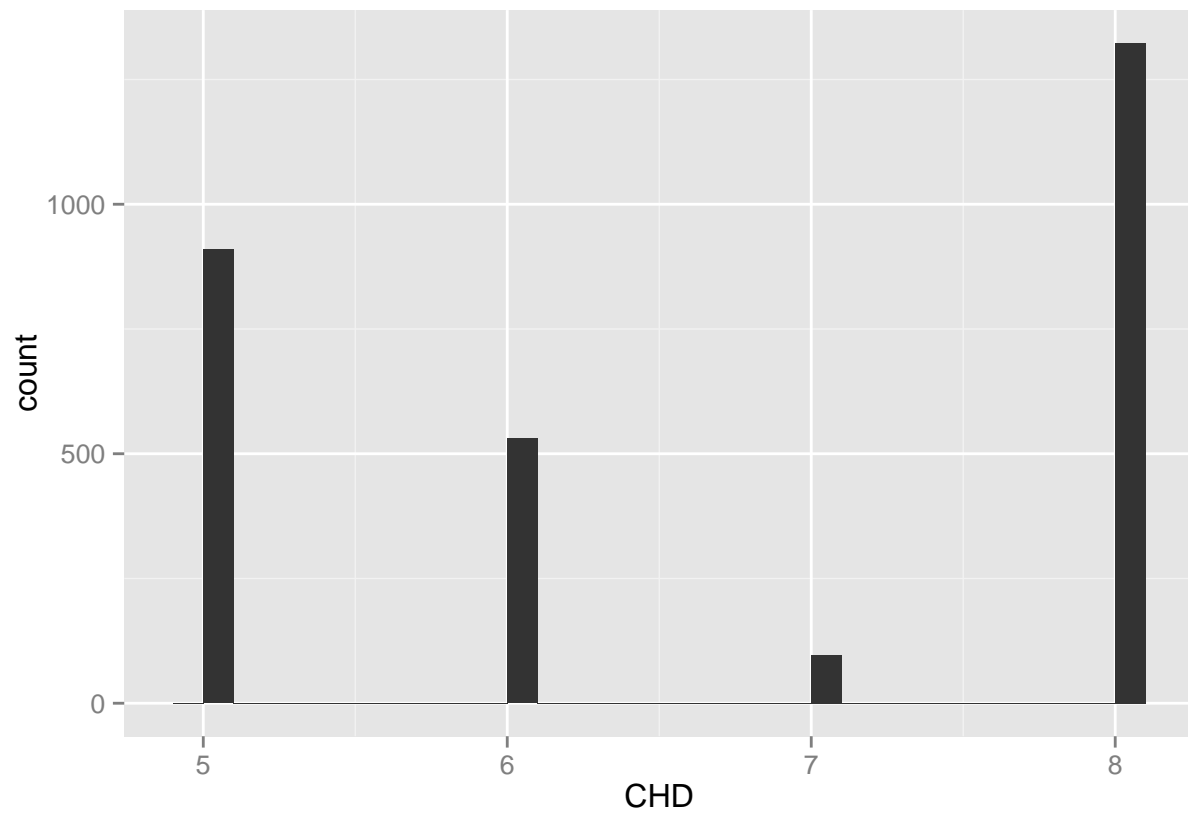
```
qplot(LungCan,data = rhidata, geom = "histogram")
```

stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
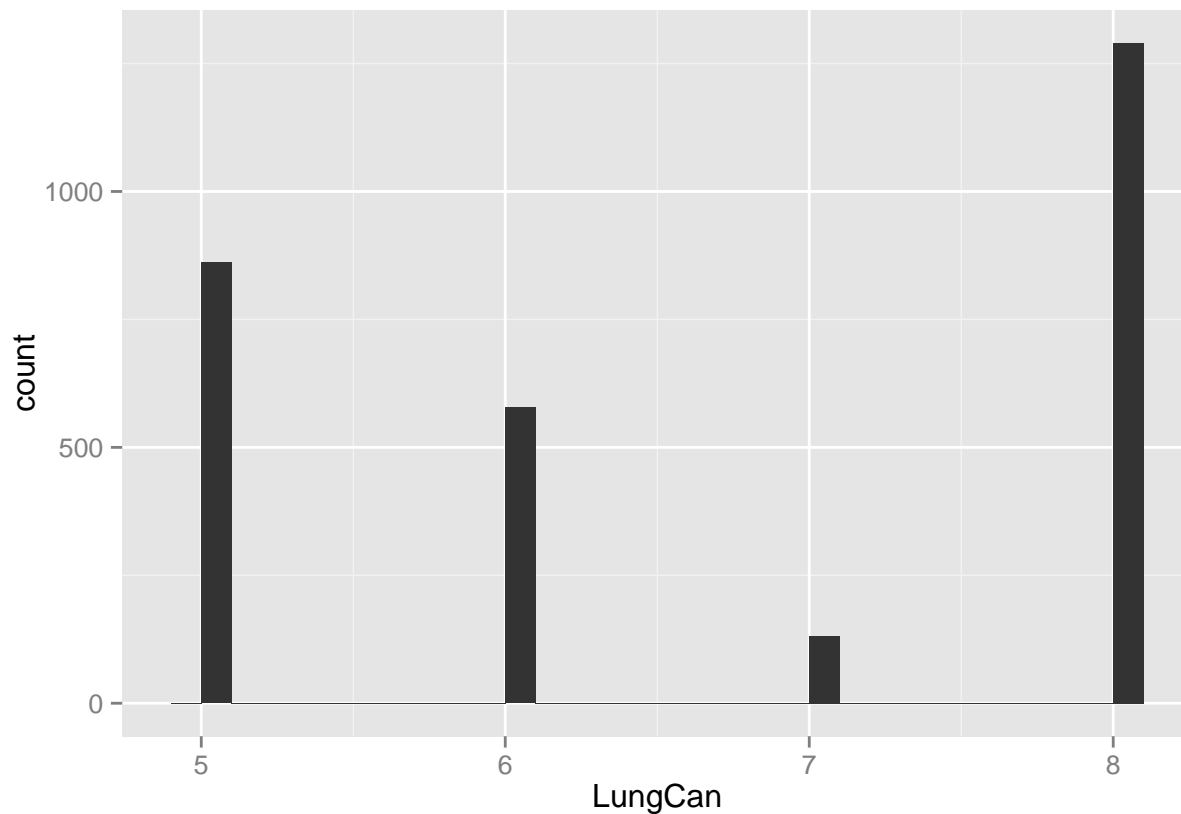
```r
#use The Elbow Method to look at the percentage of variance
#explained as a function of the number of clusters
wss <- (nrow(rhidata)-1)*sum(apply(rhidata,2,var))
for (i in 2:15) {
  wss[i] <- sum(kmeans(rhidata,centers=i)$withinss)
}
#plot to deterine the number of clusters
plot(1:15, wss,type="b", xlab="Number of Clusters",
     ylab="Within groups sum of squares")
#The location of the elbow in the resulting plot suggests
#a suitable number of clusters for the kmeans is 4
rhidataK4 <- kmeans(x = rhidata, centers = 4)
rhidataK4
```

```
## K-means clustering with 4 clusters of sizes 373, 740, 802, 945
##
## Cluster means:
##     VLBW    IM   CHD LungCan
## 1 5.598 5.828 7.901   5.416
## 2 5.661 5.816 5.245   5.364
## 3 5.810 5.948 7.141   7.941
## 4 7.776 7.857 6.814   7.039
##
## Clustering vector:
##     [1] 3 4 4 4 4 4 1 4 2 4 4 4 2 4 2 3 4 4 4 3 4 3 4 4 2 4 4 3 2 4 2 4 4 4
##    [35] 4 3 4 4 3 1 4 4 4 4 4 4 4 3 4 4 4 3 4 2 4 4 4 4 2 4 4 4 4 4 4 4 4 3
##    [69] 3 4 3 2 3 2 4 2 4 4 4 1 3 1 3 3 3 3 2 2 2 2 2 1 2 2 3 2 2 2 2 2 2 3
##   [103] 3 3 1 3 3 4 2 3 2 4 2 2 4 4 3 4 4 4 4 1 3 2 3 3 3 4 4 3 3 3 4 3 4 4
```

20

```
##  [137] 3 4 3 4 4 3 3 4 4 3 4 4 4 2 4 3 4 3 4 4 3 3 3 4 4 2 4 4 4 4 1 3 2 2
##  [171] 4 4 3 4 3 3 2 3 2 2 2 1 3 2 1 2 2 2 3 1 2 3 2 1 1 2 3 2 1 4 3 2 2 2
##  [205] 2 2 3 1 3 2 3 2 2 1 2 2 2 2 2 3 2 1 2 1 1 3 4 1 2 1 2 3 1 2 2 2 3 2
##  [239] 2 4 2 2 2 2 2 2 4 2 2 2 4 2 2 4 2 2 2 2 4 2 1 2 2 2 2 2 2 2 2 2 2 1
##  [273] 2 2 2 2 2 2 1 4 2 2 4 4 3 3 4 4 4 4 4 4 4 4 3 4 3 4 3 4 2 4 2 3 4 4
##  [307] 4 4 4 3 3 4 3 2 3 3 4 4 1 4 1 4 1 3 1 4 4 3 1 1 3 2 1 3 4 2 3 4 1 4
##  [341] 3 2 1 4 2 2 3 2 3 4 4 3 3 3 3 1 4 2 3 1 4 1 3 3 4 3 4 4 3 4 3 4 4 4
##  [375] 4 2 2 4 3 4 4 3 4 3 2 4 4 4 4 4 4 4 4 3 4 4 3 2 4 4 2 4 4 4 4 1 2 4
##  [409] 2 4 2 4 3 1 4 4 4 4 4 4 4 2 2 4 3 4 4 4 4 2 3 3 4 3 4 1 2 2 2 4 4 3
##  [443] 4 3 3 4 3 4 2 4 1 4 4 2 4 4 2 1 3 4 3 2 2 3 4 2 3 3 3 1 4 2 4 4 3 4
##  [477] 4 4 4 2 4 4 4 4 4 3 4 2 4 3 2 2 4 4 4 2 2 4 1 2 4 2 4 2 4 3 4 4 4 2
##  [511] 2 4 2 2 2 1 2 4 2 2 3 2 2 2 2 2 2 3 2 2 4 2 2 2 4 2 2 2 2 2 2 2 4 2
##  [545] 1 2 3 2 3 1 4 4 4 2 3 3 4 3 4 3 1 3 3 3 4 3 3 4 4 3 2 4 3 4 1 4 4 3
##  [579] 3 3 3 4 3 3 2 3 2 3 2 3 2 4 4 2 3 3 3 3 1 1 2 2 3 4 2 3 3 2 3 4 3
##  [613] 3 3 2 3 3 4 4 2 3 3 2 4 3 4 4 2 4 3 4 3 4 3 3 1 4 4 4 2 4 4 1 1 4 1 3 4
##  [647] 2 4 2 2 4 3 3 4 4 4 2 3 3 4 4 3 3 4 4 1 4 1 3 4 3 3 4 4 1 4 3 2 3 3
##  [681] 2 4 4 3 3 3 4 3 3 1 4 3 2 4 4 4 3 4 3 1 3 2 1 4 3 4 3 4 3 4 4 4 3 4
##  [715] 4 3 4 3 3 4 3 4 3 3 3 4 3 4 4 3 4 3 3 4 3 3 3 4 1 3 3 1 2 1 2 1 1 1
##  [749] 1 1 2 2 3 2 2 2 3 1 3 1 2 1 1 4 3 1 1 4 3 1 1 2 2 1 1 4 4 2 2 3 3 1
##  [783] 3 4 1 1 4 2 3 3 1 2 4 3 2 4 3 3 3 1 4 3 2 3 3 1 3 1 4 3 4 3 2 2 1 3
##  [817] 3 3 2 1 4 2 4 2 2 4 2 1 1 3 1 1 4 4 2 2 2 4 4 3 4 3 2 2 4 4 3 4 3 4
##  [851] 4 3 4 1 1 2 2 2 4 4 4 4 4 2 2 4 3 2 1 3 4 2 1 2 2 4 3 2 2 2 2 4 1 2
##  [885] 4 2 1 2 2 4 1 3 2 4 4 2 2 2 4 3 2 2 4 2 3 3 4 2 3 2 3 2 1 4 4 3 3 3
##  [919] 3 4 3 3 3 3 3 4 3 3 3 3 3 3 2 3 3 1 4 3 3 3 4 1 4 1 3 3 3 3 3 2 3 4 3
##  [953] 3 3 3 3 1 2 3 3 4 3 3 3 4 3 3 3 3 3 3 4 3 3 4 3 3 3 3 3 3 3 3 3 4 3
##  [987] 3 4 2 3 2 3 3 3 3 2 3 3 1 3 2 3 3 2 3 1 3 3 4 3 3 3 3 3 3 3 3 3 4 3
## [1021] 2 3 3 2 4 3 2 3 3 3 3 3 4 2 4 4 4 3 2 4 4 4 3 4 4 4 4 4 4 3 2 4 3 3
## [1055] 4 1 3 4 3 4 4 3 2 4 4 4 4 4 2 4 4 2 4 4 4 3 4 4 4 4 4 3 4 3 4 4
## [1089] 4 3 4 4 4 2 3 4 3 1 3 2 2 3 2 3 3 3 3 2 1 2 3 3 4 4 4 3 4 3 4 4 4 1
## [1123] 4 3 2 4 2 4 2 4 4 2 3 4 4 4 2 3 4 2 3 2 2 2 2 4 2 3 3 2 4 4 4 1 3 1
## [1157] 3 2 3 1 4 1 4 1 3 2 2 4 2 4 4 2 2 3 4 3 4 2 1 4 1 1 2 4 2 3 4 3 4 1
## [1191] 4 4 1 2 1 1 4 3 3 2 1 3 2 1 4 4 4 1 3 4 3 3 2 1 3 3 3 3 2 4 4 4 4 4
## [1225] 3 3 1 4 2 2 4 2 3 3 2 2 1 2 2 2 2 2 2 2 4 4 2 2 2 2 2 2 2 2 2 2 2 4
## [1259] 2 2 4 2 2 2 4 2 4 2 2 2 2 2 4 2 2 2 4 3 2 2 2 2 2 2 3 2 2 1 2 4 2
## [1293] 2 2 2 1 2 2 2 2 2 2 2 2 2 4 1 2 2 1 3 2 2 4 4 4 4 4 4 4 4 4 2 4 4
## [1327] 4 4 4 1 4 3 2 4 3 4 3 4 4 4 4 4 4 3 4 4 3 4 4 4 4 4 4 4 4 4 4 4 3 4
## [1361] 4 4 2 4 4 2 4 3 3 4 3 4 3 4 4 3 4 3 4 4 4 4 4 3 4 4 4 4 4 4 1 3 4 3
## [1395] 4 2 3 1 4 3 3 3 3 3 2 3 4 4 4 2 1 2 3 3 2 1 2 2 3 3 2 2 3 2 3 1 3 3
## [1429] 2 4 3 4 2 3 3 4 3 3 4 2 3 3 1 4 1 3 3 2 2 3 2 1 3 4 3 2 3 4 3 3 1 3
## [1463] 3 3 3 3 2 2 1 3 4 3 4 4 4 2 2 3 2 3 2 3 4 3 3 3 1 3 4 1 4 2 4 3 1 4
## [1497] 4 4 3 1 4 3 3 4 1 3 4 3 3 2 2 3 4 2 2 4 2 2 3 3 3 4 2 3 4 2 4 2 2 2
## [1531] 2 1 4 4 4 3 2 4 3 2 2 4 2 2 1 2 3 1 4 2 2 4 2 2 2 1 4 3 2 1 3 4 3 1
## [1565] 2 2 4 2 2 2 4 4 1 2 3 4 3 2 3 2 1 4 2 1 2 2 2 2 2 2 1 2 2 2 1 2 2 4
## [1599] 3 3 2 3 3 3 3 4 3 2 3 2 3 3 2 3 3 1 1 1 3 3 4 4 1 3 4 3 4 4 4 3 1 2 4
## [1633] 1 3 1 3 1 4 2 3 4 1 2 2 2 1 4 2 2 2 1 1 1 2 2 2 1 3 1 2 2 2 1 2 3
## [1667] 2 2 1 2 4 3 4 4 1 1 1 3 1 1 4 1 1 4 4 2 3 3 3 3 3 4 4 3 4 2 1 1 1
## [1701] 3 4 4 3 1 3 4 2 3 1 4 3 1 2 3 4 3 2 3 2 3 4 4 2 1 2 3 2 1 1 3 4 1 2
## [1735] 4 4 1 4 4 4 3 4 4 4 4 4 3 4 4 2 3 4 2 4 4 4 4 4 4 4 4 4 2 4 4 3 4 4
## [1769] 2 4 4 4 4 4 3 2 4 4 4 4 4 3 4 4 4 4 2 2 3 2 4 2 4 4 3 4 4 4 2 2 2
## [1803] 2 4 4 4 4 4 4 4 4 4 4 4 3 3 1 4 4 3 4 4 4 1 2 4 4 4 4 2 2 1 2 2 2
## [1837] 1 2 1 4 3 2 2 4 4 4 2 1 3 2 2 4 3 4 2 3 4 1 3 3 3 1 3 4 2 3 4 3 3 3
## [1871] 1 4 4 1 2 1 4 4 1 4 2 4 2 1 1 4 4 1 3 4 4 3 2 3 3 4 3 3 3 4 4 4 3 3
## [1905] 1 3 1 3 1 3 1 4 4 4 4 1 3 4 4 3 2 3 1 1 4 3 3 4 3 3 4 2 3 1 2 3 4 1
## [1939] 1 1 2 3 4 3 3 1 1 4 1 4 3 3 1 3 1 1 1 1 3 3 1 1 2 2 3 1 4 3 3 1 1 3
```
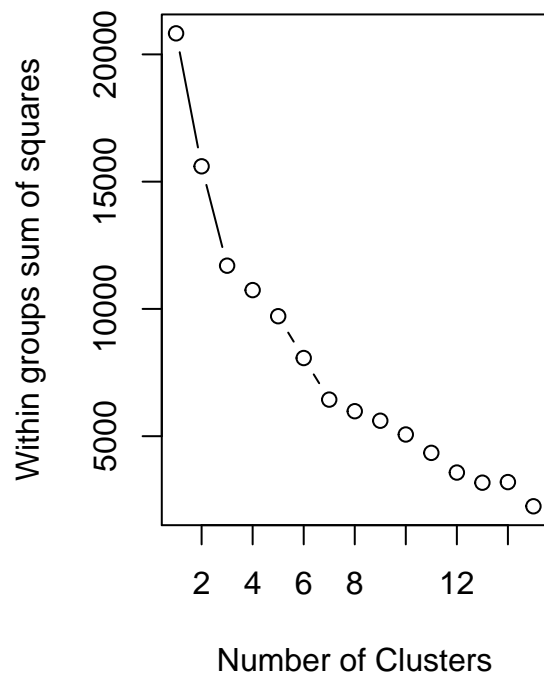
21

```
## [1973] 4 3 4 3 3 3 1 3 4 3 4 4 1 2 3 1 2 3 1 3 1 3 3 3 1 1 1 1 3 4 3 3 3 3
## [2007] 4 3 3 4 2 1 4 1 4 2 2 2 3 3 3 3 2 2 2 3 2 2 2 2 2 3 2 3 2 4 3 4 3 2
## [2041] 3 2 2 3 4 2 2 4 4 4 3 4 4 1 2 2 1 4 3 3 2 2 1 4 1 1 2 1 4 4 2 4 3 4
## [2075] 2 1 4 1 4 4 4 1 2 4 4 4 1 2 1 4 2 4 2 4 3 4 1 4 3 1 3 2 1 1 2 2 1 4
## [2109] 3 1 1 2 1 1 3 1 3 4 4 4 4 4 4 4 2 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 3 4
## [2143] 2 4 2 4 4 4 4 4 4 4 4 2 4 4 4 4 4 4 4 4 2 2 1 2 1 2 4 3 2 1 4 3
## [2177] 2 4 2 4 3 2 4 2 2 1 4 1 2 2 2 4 4 2 3 4 2 4 4 3 2 4 1 4 1 2 4 4 4 1
## [2211] 4 1 3 3 4 4 3 2 3 4 1 3 4 1 3 3 3 3 4 4 3 3 4 4 3 3 3 3 4 3 4 1 4 4
## [2245] 4 4 3 1 4 3 1 4 3 3 4 3 3 4 4 3 4 4 4 4 3 4 1 4 3 4 4 3 4 4 4 4 3 3
## [2279] 4 3 1 3 4 3 4 3 3 4 3 1 3 4 1 4 1 4 3 1 3 3 1 3 3 3 1 3 2 4 3 2 3 4
## [2313] 4 3 4 3 1 4 1 2 3 2 2 2 3 2 2 4 4 2 3 4 4 3 3 1 4 2 1 4 4 3 1 4 4 2
## [2347] 2 3 1 2 2 4 1 1 3 3 2 1 4 1 2 1 1 2 4 1 1 3 3 2 2 1 2 3 3 1 2 2 3 3
## [2381] 3 2 3 3 1 4 1 2 3 2 2 2 3 3 3 3 1 4 3 3 3 3 1 1 3 1 2 3 3 2 1 1 3 1
## [2415] 4 1 1 1 3 3 3 2 4 2 2 3 3 1 4 3 3 2 2 1 2 3 3 3 2 3 1 3 3 3 2 1 4 3
## [2449] 4 3 2 1 4 3 2 4 4 3 3 3 3 3 3 3 4 4 2 3 2 2 2 4 1 3 2 3 1 2 3 4 1 2
## [2483] 4 2 4 4 4 2 2 3 1 2 1 3 2 2 2 3 4 1 2 2 2 1 3 3 1 3 2 1 4 2 2 2 2 2
## [2517] 2 3 2 2 2 1 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 1 4 4 2 2 2 2 2 2 4 4 1
## [2551] 4 2 4 2 4 2 1 2 4 3 4 4 4 2 4 3 3 3 4 2 2 2 3 4 2 2 2 2 4 3 4 4 4 2
## [2585] 3 4 4 3 4 4 3 4 2 4 4 3 3 2 4 4 2 4 4 3 2 1 4 4 4 4 4 4 2 4 3 4 4 2
## [2619] 4 1 4 2 1 2 2 3 2 3 3 2 2 4 2 4 3 3 4 3 1 1 2 2 3 3 4 4 3 3 4 4 3 2
## [2653] 4 4 3 4 2 4 4 3 4 1 4 4 4 3 4 3 4 4 3 3 4 4 4 4 2 4 3 2 2 3 3 2 4
## [2687] 3 1 3 2 2 2 3 2 3 3 1 3 2 3 4 3 2 3 4 3 2 1 3 2 2 2 2 4 4 3 3 3 3 4
## [2721] 3 3 3 3 1 2 1 4 3 3 4 3 3 1 4 3 3 3 1 3 3 1 3 3 2 4 3 3 1 3 2 4 3 1
## [2755] 1 1 1 4 3 1 3 1 3 1 4 3 1 3 3 3 2 4 1 2 2 2 2 2 2 1 2 1 2 3 2 2 2 4
## [2789] 1 2 2 1 4 4 1 3 3 2 2 1 1 4 2 2 3 3 3 4 1 2 2 2 2 1 3 2 2 1 4 2 4 2
## [2823] 2 4 3 3 2 1 1 1 3 1 4 2 2 3 1 2 2 2 3 3 4 4 2 4 2 3 2 2 2 2 4 2 4 4
## [2857] 2 2 2 3
##
## Within cluster sum of squares by cluster:
## [1] 1069 2287 3494 3231
##  (between_SS / total_SS =  51.6 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"
## [5] "tot.withinss" "betweenss"    "size"         "iter"
## [9] "ifault"
```

```r
#plot the 4 groups of rhidata data frame
par (mfrow=c(1,1))
```

```
plot(rhidataK4, data = rhidata)
```