Breast Cancer Coimbra

Group 13

Authors: Adjei-Mosi Angela; Bakare Olaleye; Lucas Okimi

Introduction-Lucas Okimi, Angela,

The dataset was collated from the women's population of early diagnosed breast cancer from the Gynecology department of the University Hospital Centre of Coimbra between 2009 and 2013. It comprised of 116 respondents, out of which 64 were diagnosed with breast cancer and the remaining 52 women are healthy volunteers. Ten different variables were recorded from the respondents which are:

- Age refers to the age of an individual patient or health volunteer.
- BMI refers to the body mass index which is a ratio of weight of the patient concerning the height. It is recorded in (kg/m2).
- Glucose refers to the glucose level of the patient which was determined by an automatic analyzer using a commercial kit, and it is denoted in (mmol/L).
- Insulin refers to insulin level which were measured with the aid of ELSA kit and recorded in (uL/mg)
- Homa refers to the homeostasis model assessment and it is derived to evaluate insulin resistance. It is calculated as the logarithm of insulin fasting and glucose level divided by 22.5.
- Leptin refers to a hormone made by adipose and enterocytes in the small intestine that helps to regulate energy balance.
- Adiponectin refers to a protein hormone that is involved in regulating glucose levels as well as fatty acid breakdown.
- Resistin refers to as a cysteine-rich peptide hormone.
- MCP1 refers to as Monocyte Chemoattractant Protein-1
- Classification is used to classify the respondents as health volunteers or breast cancer patients.

Data Source: The data was called from the UCL machine learning website. The data is obtainable from the following link:

https://archive.ics.uci.edu/ml/machine-learning-databases/00451/dataR2.csv.

Data Description and Representation-Angela Adjei-Mosi

R software is used for data representation to provide basic descriptive statistics of the variables. The data is called into R program using the code defined in the appendix, and first few values are read using the head function.

The summary helps to provide the minimum value, maximum value, mean, median, first quartile value, and third quartile value of each of the variables. The complete dataset summary is generated along with the histogram for each of the variables. The histogram shows that the variable (Age) is appropriate for the test considering the respondents' minimum age is 24 years old and the maximum age is 89 years. The

variable (BMI) shows the body mass index with the minimum BMI as 18.37 which is a reference of underweight respondent based on the standard BMI categories listed below:

Underweight = <18.5
Normal weight = 18.5–24.9
Overweight = 25–29.9
Obesity = BMI of 30 or greater

This shows that there is a likelihood of other health challenges some respondents are having along with the breast cancer. Other variables in the dataset show some skewness present in the data. In the dataset, the classification or factor variable will be dropped since the instructor advised against the presence of factor or classification data. The correlation matrix between pairs is very important to understand how the data are related within the dataset. Adicponetin is negatively correlated with the remaining 6 variables. It is the only variable that is completely relating negatively with the others. Adicponetin is a protein hormone that is involved in regulating glucose levels as well as fatty acid breakdown in humans.

Histogram plots are generated on each of the variables as shown in the appendix. It can be seen from the charts that some of the variables are right-skewed. Glucose, Insulin, Homa, and Resistin are right-skewed compared to the remaining variables indicating that further analysis needs to be conducted on the data set so that the skewness will not create some biasedness to the analysis. This pictorial representation of each of the variables helps to see the spread of the data which allows for understanding the skewness of some of the variables.

The head function is generated to display the first 6 respondents with the variables.

The variable "classification" can be classified as a factor variable, and this will be dropped from the data set as instructed since it is a categorical variable. In this project, the group performed a simple linear regression so that more leanings will be placed on significant variables. This is carried out using 8 independent variables can be adequately selected for further analysis with the variable resistin standing out as the most non-significant when regressed against BMI on the dataset. Hence, classification was dropped before performing additional analyses on the dataset.

Simple Linear Regression

A simple linear regression model is generated from the data set, and it is shown that Resistin is not a significant variable to BMI. The final model is generated as shown with the summary detailed in the appendix.

This model can be written as:

BMI = 21.4402 - 0.0168* Age +0.0284 * Glucose + 0.2864* Insulin – 1.0619*HOMA + 0.1546*Leptin - 0.1616* Adiponectin + 0.0035*MCP.1

The R-squared value is about 0.43 indicating that only 43 percent of the data is explained in the model.

It can be seen that remaining variables are significant variables excluding age and glucose after removing classification and resistin from the dataset.

Data Cleaning

The quality of the selected dataset was checked using R software. A thorough summary of the dataset is checked to provide key analysis of each of the variables. It is observed that the data is clean as there is no missing data available in the dataset.

Outlier Removal

The model "cancerlm" is checked for any possible outliers using the cook distance function in R. This helps in identifying any possible outlier observations that can be removed from the dataset. This is achieved using the codes stated in the appendix with the aid of cooks distance and cut off line is set indicating the outliers about this threshold. Halfnormal distribution is generated to show the outliers as observations within the dataset.

Another function halfnorm is generated to identify the observations that are removed from the data set. Observations IDs 47,75,109 and 116 can be seen to be identified as outliers in the dataset. These observations are removed from the dataset before other analyses are carried out. As shown in the model summary, classification and resistin have been removed from the dataset using the function below:

The identified outliers are removed from the dataset using the code shown in the appendix, and a new dataset is generated called cancernew.

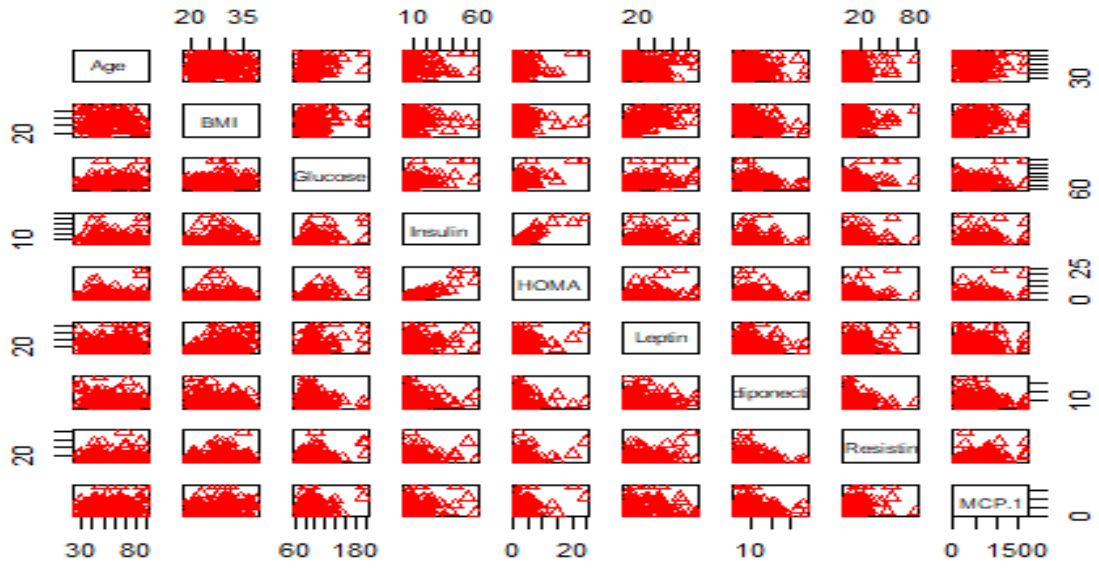
Correlation Matrix-Ola Bakare

A correlation matrix is generated from the new dataset as presented below using the appropriate f unction in R, and a correlation plot is generated as shown below. The matrix shows that Adipone ctin is negatively correlated with other variables in the dataset implying that as Adiponectin decr eases there is likely negative impact on each of the variables.

Also, the lesser the adiponectin present in a person, the higher the tendency to accumulate fat, which will lead to increase in other health conditions. For instance, if MCP.1 is high, which is a receptor of chemical, it encourages creation of wild cells which can easily influence various diseases.
The high correlation between glucose and insulin is an evidence that, a normal body tend to produce more insulin with the increase in glucose presence in the body. HOMA is a derivative of glucose and insulin, it explains why it is of strong positive correlations with glucose and insulin.

	Age	BMI	Glucose	Insulin	HOMA	Leptin	Adiponectin	MCP.1	Resistin
Age	1.00	0.01	0.23	0.03	0.13	0.10	-0.22	0.01	0.00
BMI	0.01	1.00	0.14	0.15	0.11	0.57	-0.30	0.22	0.20
Glucose	0.23	0.14	1.00	0.50	0.70	0.31	-0.12	0.26	0.29
Insulin	0.03	0.15	0.50	1.00	0.93	0.30	-0.03	0.17	0.15
HOMA	0.13	0.11	0.70	0.93	1.00	0.33	-0.06	0.26	0.23
Leptin	0.10	0.57	0.31	0.30	0.33	1.00	-0.10	0.01	0.26
Adiponectin	-0.22	-0.30	-0.12	-0.03	-0.06	-0.10	1.00	-0.20	-0.25
MCP.1	0.01	0.22	0.26	0.17	0.26	0.01	-0.20	1.00	0.37
Resistin	0.00	0.20	0.29	0.15	0.23	0.26	-0.25	0.37	1.00

The correlation matrix plot is shown below



Data Dimension Reduction Analysis-Angela, Ola

In order to carry out data reduction analysis on the dataset, two distinct approaches are explored which are the Principal Component Analysis (PCA) and Exploratory Factor Analysis (EFA).

Principal Component Analysis- ALL

This method is using new set of predictors that are orthogonal normalized eigenvectors with their corresponding eigenvalues. It shows the decomposition of the variations among the existing predictors with predictors with small variations being discarded. The summary of the principal component analysis indicates that the first 4 components explain about 76 percent of the variations with the remaining components discarded. Component 1(Age) accounts for 33.98 percent, component 2(BMI) accounts for 16.91 percent, component 3(Glucose) accounts for 12.97 percent and the component 4 (Insulin) accounts for 12.29 percent. Several analyses are carried out including principal component using covariance matrix, and Kmeans with plots on the loadings and scores. The results are shown in the appendix.

Variance inflator factor analysis is carried out on the components to check for multicollinearity and it can be noticed from the variance inflator factor plot that variable insulin and Homa are of great concern and should be critically looked into for collinearity problems on the dataset.

Age has the highest variation in the data. This is understandable since the data is derived from both health volunteers and breast cancer patients with different ages ranging from age 24 to age 89.

Exploratory Factor Analysis (EFA)-ALl

This method is carried out on the new dataset to show factors that are really of importance in the dataset. It helps to reveal the relationships between assumed latent variables and the manifest variables regarded as factor. 5 factors are defined in the dataset to evaluate critical factors that are contributing to incidence of breast cancer among women using the heath volunteers and patients attending to quality care in Coimbra. The result indicates that factor 1 has strong loadings from insulin and HOMA which can be referred to as sugar related factor of breast cancer since HOMA is a derivative of glucose and insulin. Factor 2 has strong score on BMI and leptin indicating weight factor of patients, factor 3 has strong score in MCP.1 and resistin indicating responsible chemical imbalance factor, factor 4 has strong score on adiponectin indicating protein factor of respondents and factor 5 can be regarded as age factor since age and glucose have strong scores. The root mean square error is calculated as 0.0191 showing that the model is acceptable along with the p-value of 0.0348 which is less than 5 percent. The R codes and results are shown in the appendix.

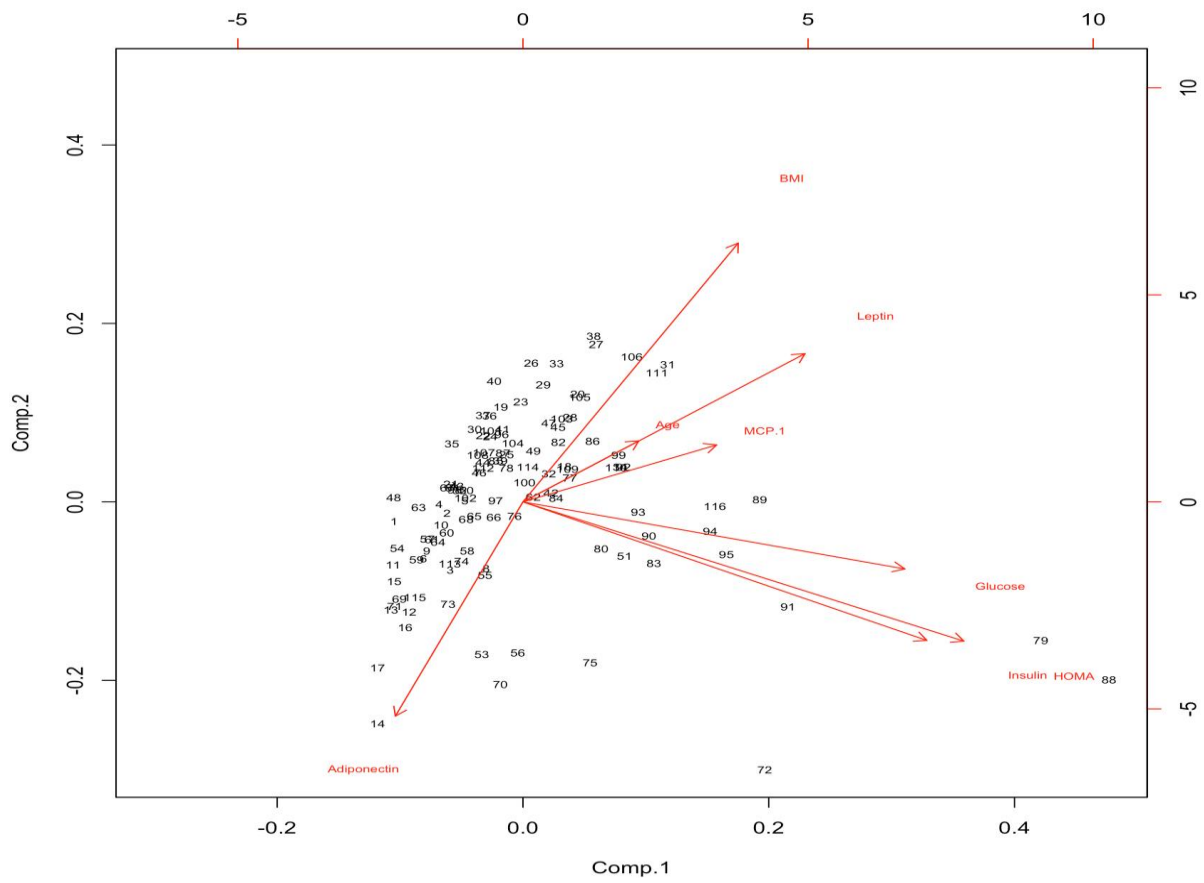
	Factor1	Factor2	Factor3	Factor4	Factor5
Age				-0.143	0.337
BMI		0.620	0.200	-0.215	
Glucose	0.502		0.240		0.593
Insulin	0.986	0.136			
HOMA	0.931	0.100	0.167		0.299
Leptin	0.184	0.955			0.210
Adiponectin		-0.127	-0.153	0.947	-0.239
Resistin		0.226	0.372	-0.127	0.203
MCP.1	0.119		0.920		

Multidimensional Scaling- ALL

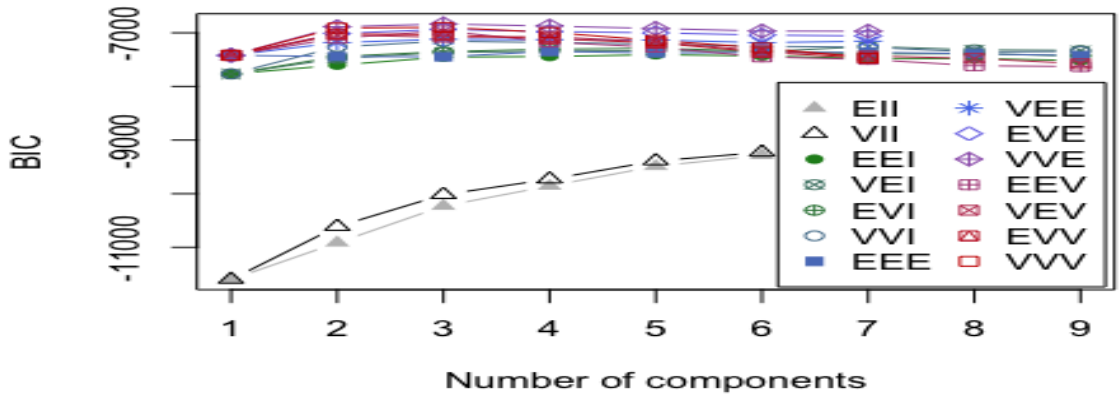
This is generated to show Euclidean distances among the dataset based on scaled distances, and the results are shown in the appendix.

Cluster Analysis-ALL

Different cluster analysis are performed on the dataset which include the complete hierarchical analysis, single linkage and the average linkage with the dendograms presented below, and the codes are shown in the appendix. The analysis allows generation of four (4) distinct cluster groups based on the dataset with good proportion of the respondents belonging to the cluster 1 precisely over 88 percent falls into cluster 1 while the remaining 12 percent are shared among three(3) cluster groups.



Complete Hierarchical Dendogram



Cluster Plot

Confirmatory Factor Analysis-ALL

The confirmatory factor analysis is carried out on the dataset to evaluate how manifest variables relate to certain factors while constraining other manifest variables with zero

correlation. The confidence intervals are between 0.0177 and 0.55 The results are shown in the Appendiz.

Conclusion-ALL

The dataset is a very good data with no missing data and several analyses are carried out from simple regression model to different multivariate analysis. The data has presented a detailed relationship on the factors that are responsible for inducing breast cancer among women with statistics showing respondents ranging from age 24 through 89. Several information can be deduced from the analysis showing that future data and research can be carried out to identify more critical factors that are related to quality health of women.

APPENDIX

Project _Group13

Lucas Okimi

12/5/2020

```
library(MASS)
library(e1071)
library(faraway)

## Warning: package 'faraway' was built under R version 4.0.3

library(HistogramTools)

## Warning: package 'HistogramTools' was built under R version 4.0.3

library(sem)

## Warning: package 'sem' was built under R version 4.0.3

library(semPlot)

## Warning: package 'semPlot' was built under R version 4.0.3

## Registered S3 methods overwritten by 'huge':
##   method    from
##   plot.sim  BDgraph
##   print.sim BDgraph

require(MASS)
require(e1071)
require(faraway)
require(HistogramTools)
require(sem)
require(semPlot)
```

```r
#Reading the data
cancer <-read.csv("https://archive.ics.uci.edu/ml/machine-learning-databases/
00451/dataR2.csv")

#display the first 6 rows of the data
round(head(cancer), 2)
```

```
##   Age   BMI Glucose Insulin HOMA Leptin Adiponectin Resistin  MCP.1
## 1  48 23.50      70    2.71 0.47   8.81        9.70     8.00 417.11
## 2  83 20.69      92    3.12 0.71   8.84        5.43     4.06 468.79
## 3  82 23.12      91    4.50 1.01  17.94       22.43     9.28 554.70
## 4  68 21.37      77    3.23 0.61   9.88        7.17    12.77 928.22
## 5  86 21.11      92    3.55 0.81   6.70        4.82    10.58 773.92
## 6  49 22.85      92    3.23 0.73   6.83       13.68    10.32 530.41
##   Classification
## 1              1
## 2              1
## 3              1
## 4              1
## 5              1
## 6              1
```

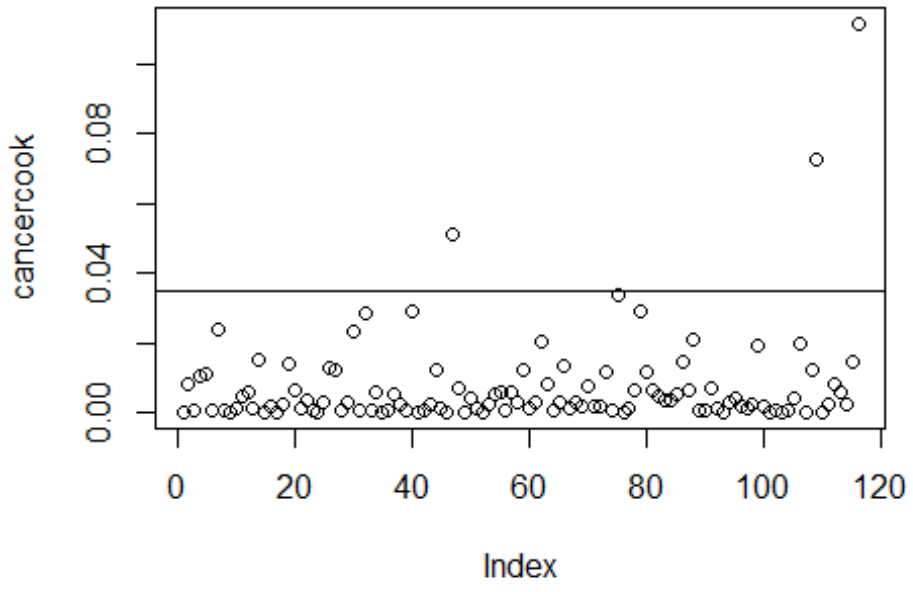
```r
#Checking the dimension of the data
dim(cancer)
```

```
## [1] 116  10
```

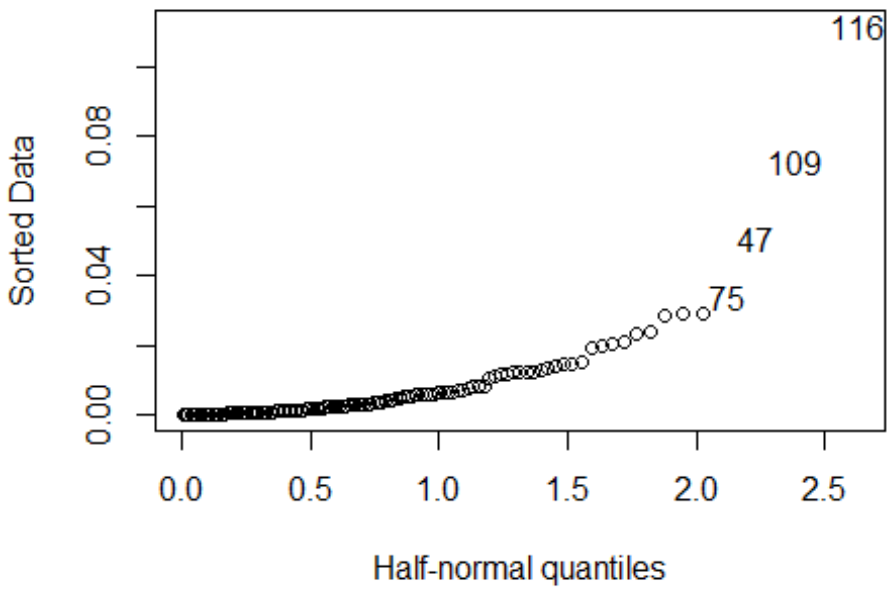
```r
# Defineing a simple linear regression to show impact on BMI

cancerlm<-lm(BMI~.-Classification, data = cancer)
summary(cancerlm)
```

```
##
## Call:
## lm(formula = BMI ~ . - Classification, data = cancer)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.8222 -2.3517 -0.0606  2.2932 10.3177
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 21.682435   3.088296   7.021 2.11e-10 ***
## Age         -0.018817   0.023629  -0.796  0.42759
## Glucose      0.029528   0.026673   1.107  0.27076
## Insulin      0.277143   0.120723   2.296  0.02364 *
## HOMA        -1.035371   0.401013  -2.582  0.01118 *
## Leptin       0.157687   0.020272   7.779 4.86e-12 ***
## Adiponectin -0.169188   0.055460  -3.051  0.00288 **
## Resistin    -0.022960   0.033003  -0.696  0.48814
## MCP.1        0.003684   0.001154   3.193  0.00185 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.784 on 107 degrees of freedom
## Multiple R-squared:  0.4715, Adjusted R-squared:  0.432
## F-statistic: 11.93 on 8 and 107 DF,  p-value: 4.851e-12

cancercook<-cooks.distance(cancerlm)
plot(cancercook)
abline(h=0.035)
identify(1:116,cancercook)
```



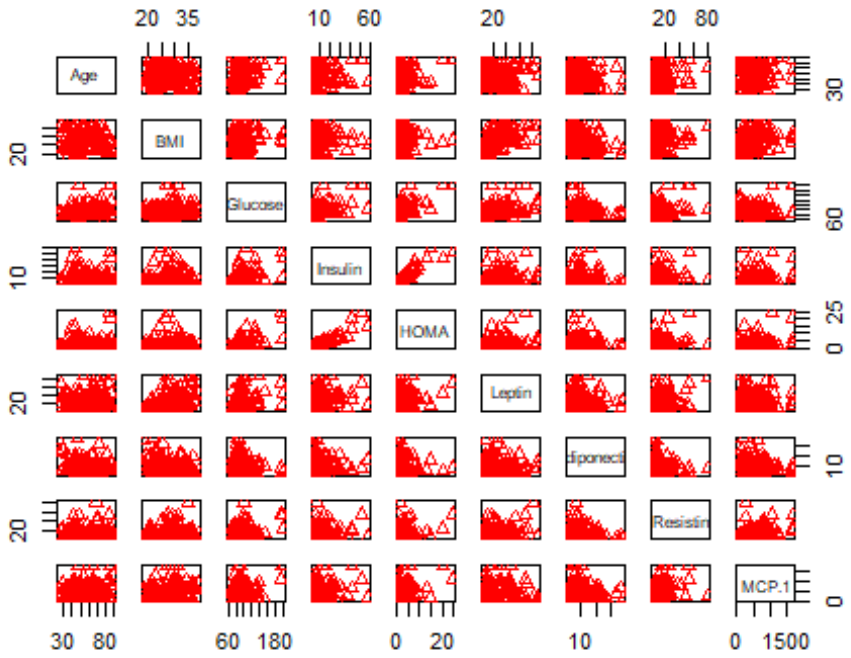

```
## integer(0)

halfnorm(cancercook,4)
```




```
#Determining the correlation of the first seven variables rounded to 2 decima
l
cancer1<-cancer[,-c(10)]
head(cancer1)

##   Age      BMI Glucose Insulin      HOMA  Leptin Adiponectin Resistin   MC
P.1
## 1  48 23.50000      70   2.707 0.4674087  8.8071    9.702400  7.99585 417.
114
## 2  83 20.69049      92   3.115 0.7068973  8.8438    5.429285  4.06405 468.
786
## 3  82 23.12467      91   4.498 1.0096511 17.9393   22.432040  9.27715 554.
697
## 4  68 21.36752      77   3.226 0.6127249  9.8827    7.169560 12.76600 928.
220
## 5  86 21.11111      92   3.549 0.8053864  6.6994    4.819240 10.57635 773.
920
## 6  49 22.85446      92   3.226 0.7320869  6.8317   13.679750 10.31760 530.
410

cancernew<-cancer1[-c(47,75,109,116)]
corr = cor(cancernew)
round(corr, 2)

##               Age   BMI Glucose Insulin  HOMA Leptin Adiponectin Resistin
MCP.1
## Age          1.00  0.01    0.23    0.03  0.13   0.10       -0.22     0.00
```

```
0.01
## BMI          0.01  1.00    0.14    0.15  0.11   0.57       -0.30     0.20
0.22
## Glucose      0.23  0.14    1.00    0.50  0.70   0.31       -0.12     0.29
0.26
## Insulin      0.03  0.15    0.50    1.00  0.93   0.30       -0.03     0.15
0.17
## HOMA         0.13  0.11    0.70    0.93  1.00   0.33       -0.06     0.23
0.26
## Leptin       0.10  0.57    0.31    0.30  0.33   1.00       -0.10     0.26
0.01
## Adiponectin -0.22 -0.30   -0.12   -0.03 -0.06  -0.10        1.00    -0.25
-0.20
## Resistin     0.00  0.20    0.29    0.15  0.23   0.26       -0.25     1.00
0.37
## MCP.1        0.01  0.22    0.26    0.17  0.26   0.01       -0.20     0.37
1.00
```

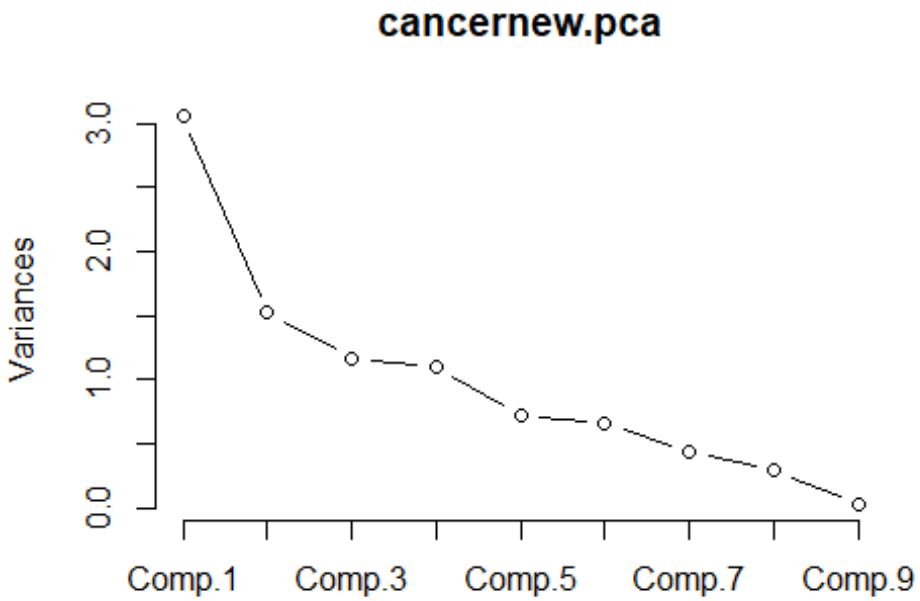
```
plot(cancernew,col="red",pch=24)
```




```
cancernew.pca<-princomp(cancernew,cor=T)

summary(cancernew.pca)

## Importance of components:
##                           Comp.1    Comp.2    Comp.3    Comp.4     Comp.5
## Standard deviation     1.7488681 1.2337752 1.0805294 1.0515218 0.85002491
```

```
## Proportion of Variance 0.3398377 0.1691335 0.1297271 0.1228553 0.08028248
## Cumulative Proportion  0.3398377 0.5089712 0.6386983 0.7615536 0.84183613
##                            Comp.6     Comp.7     Comp.8      Comp.9
## Standard deviation     0.81072756 0.66449266 0.54094725 0.178944741
## Proportion of Variance 0.07303102 0.04906117 0.03251377 0.003557913
## Cumulative Proportion  0.91486715 0.96392832 0.99644209 1.000000000
```

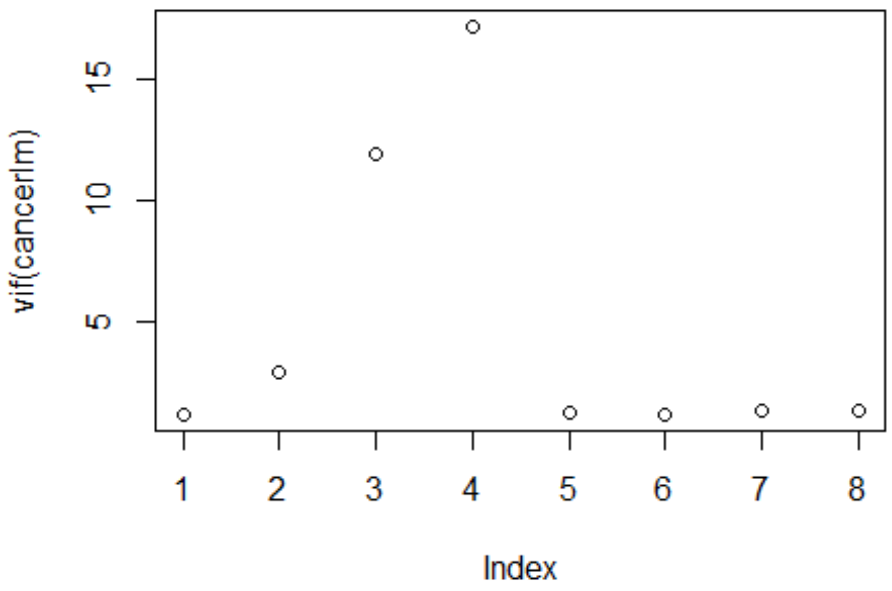
```
plot(cancernew.pca, type="l")
```



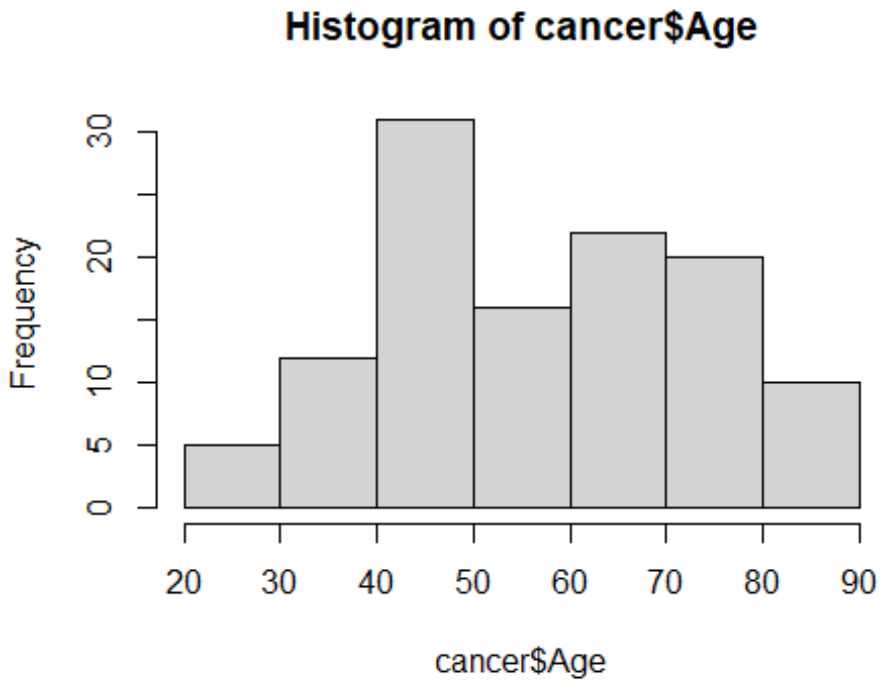

```
print(cancernew.pca$loadings, cut = 0.3)
```

```
##
## Loadings:
##             Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9
## Age                                 0.821                0.308
## BMI                  0.499 -0.426                              0.599
## Glucose       0.439                                     -0.807
## Insulin       0.444 -0.386                               0.391         0.614
## HOMA          0.493 -0.375                                              -0.758
## Leptin        0.331        -0.583                                -0.636
## Adiponectin         -0.481                0.529  0.488
## Resistin             0.304        -0.303  0.598 -0.421
## MCP.1                       0.497 -0.359         0.633        -0.301
##
##                Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Com
p.9
## SS loadings     1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.
```

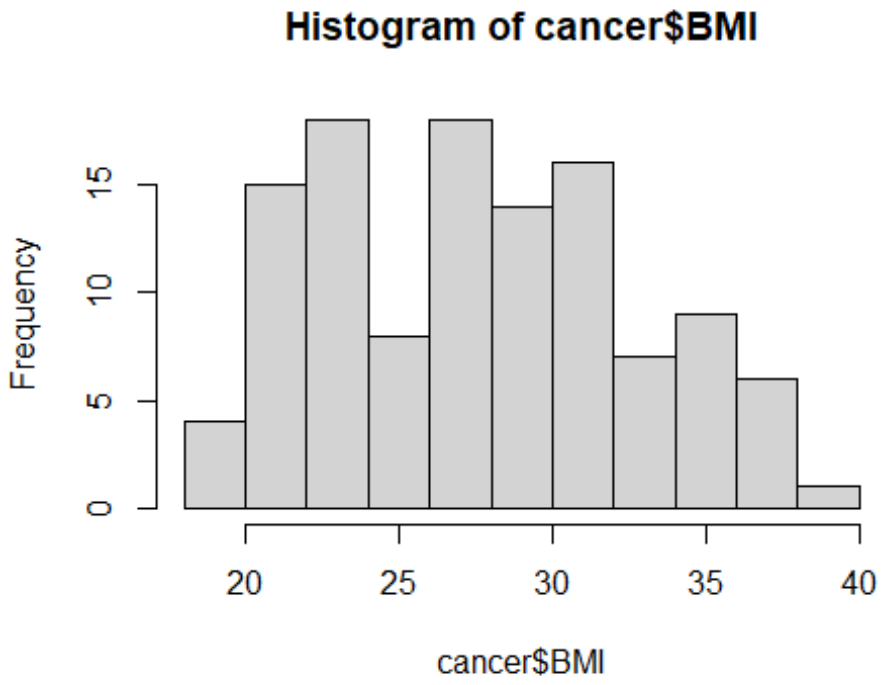
```
000
## Proportion Var  0.111  0.111  0.111  0.111  0.111  0.111  0.111  0.111  0.
111
## Cumulative Var  0.111  0.222  0.333  0.444  0.556  0.667  0.778  0.889  1.
000

score <- cancernew.pca$score
head(score)

##          Comp.1     Comp.2    Comp.3     Comp.4      Comp.5      Comp.6
## [1,] -1.9927846 -0.1256122 0.3628691 -0.2741197 -0.49265203 -0.34467055
## [2,] -1.3288300 -0.2479302 1.3367401  1.8943923 -0.17535832  0.08598345
## [3,] -1.2172439 -0.9923928 0.3630650  0.9136718  1.31476235  1.38764696
## [4,] -1.1992026  0.2724091 1.7595832  0.2717166 -0.16270670  0.50100442
## [5,] -0.8984674  0.1430886 2.0191848  1.5792614 -0.03384181  0.44944415
## [6,] -1.5273050 -0.5940507 0.6685777 -0.4369111  0.07583587  0.06809827
##           Comp.7     Comp.8      Comp.9
## [1,]  0.27867803 -0.1998847 -0.25002048
## [2,]  0.08434951 -0.4409006  0.01888667
## [3,]  0.54708226  0.1567938  0.04434073
## [4,]  0.65016620 -0.7395776 -0.04857189
## [5,]  0.35043610 -0.4495838  0.07041023
## [6,] -0.32769907 -0.0325212 -0.04875596

vif(cancerlm)

##         Age     Glucose     Insulin        HOMA      Leptin Adiponectin
##    1.164422    2.899819   11.866925   17.135545    1.214816    1.157153
##    Resistin       MCP.1
##    1.343371    1.279714

plot(vif(cancerlm))
```



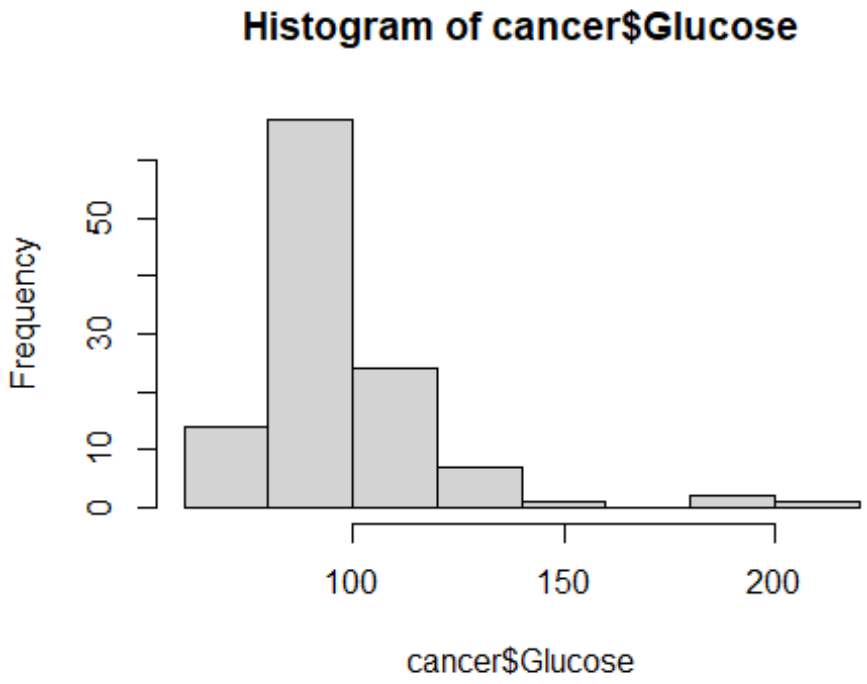

```
# It can be seen that the variance inflation factor for insulin and HOMA are
of great concern

# To run the histogram of the first 7 variables along with the saummary of th
e dataset Cancer



hist(cancer$Age)
```



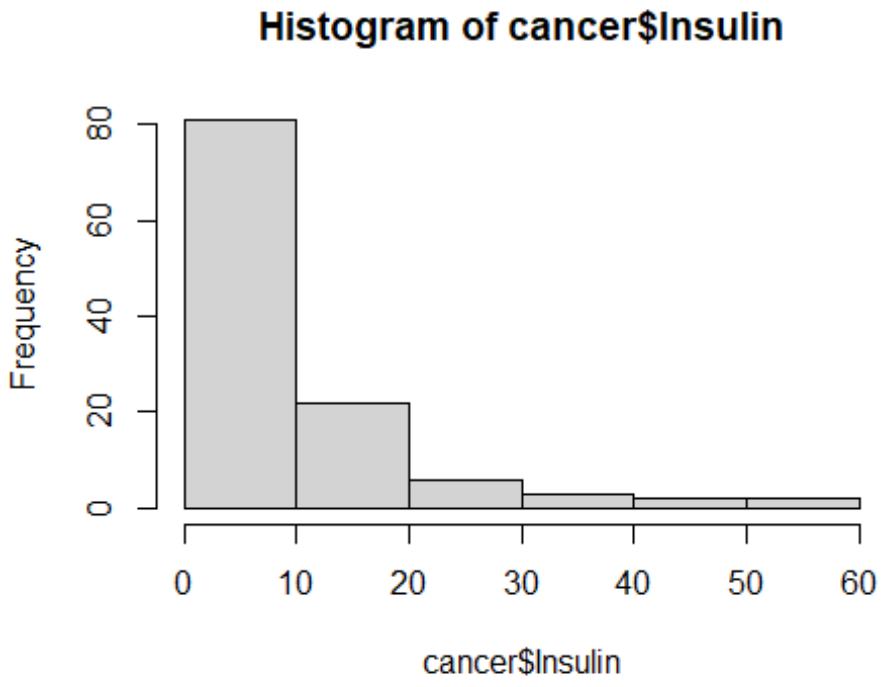

```r
hist(cancer$BMI)
```



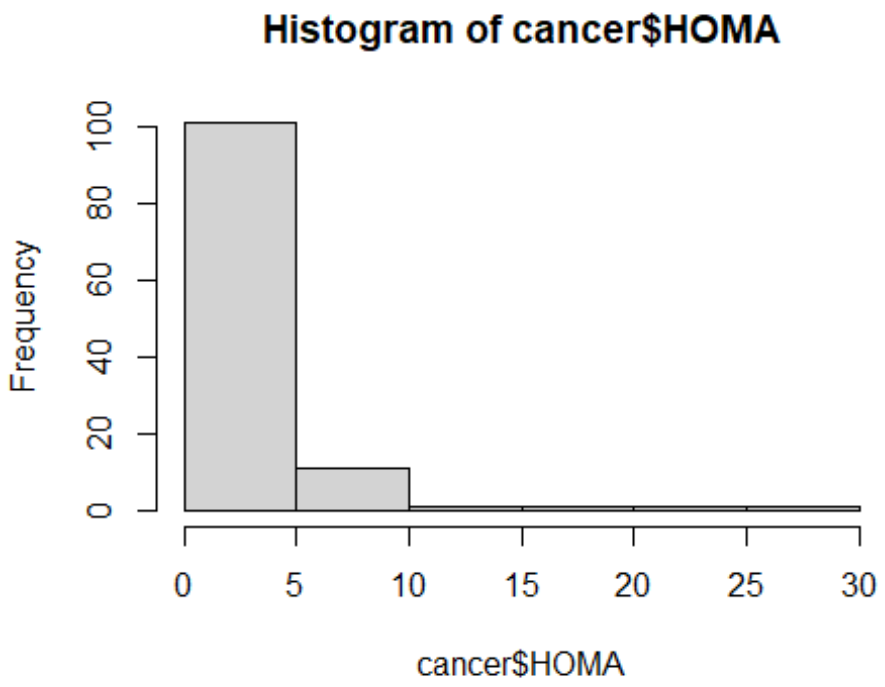

```r
hist(cancer$Glucose)
```



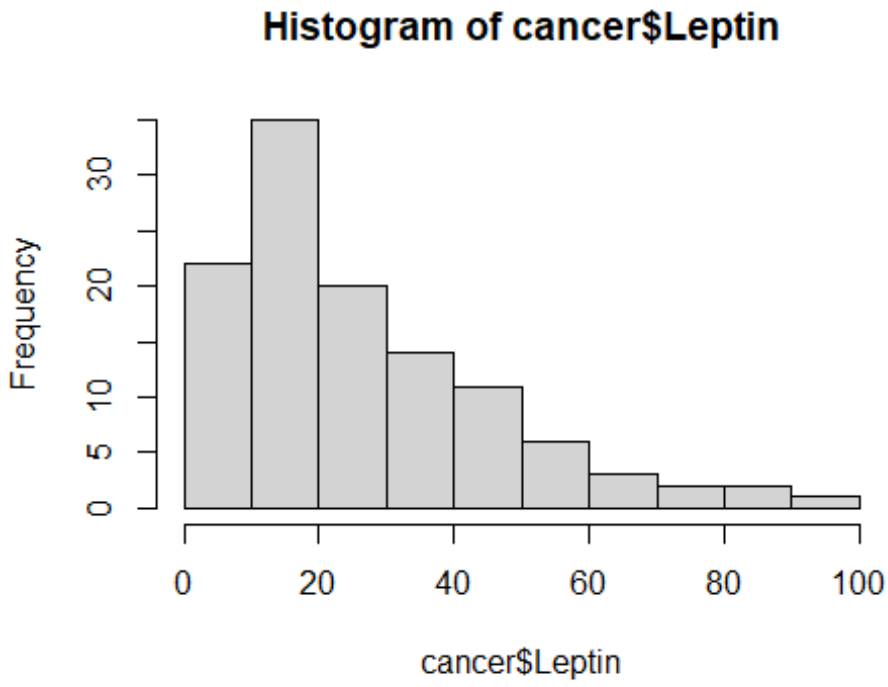

```r
hist(cancer$Insulin)
```



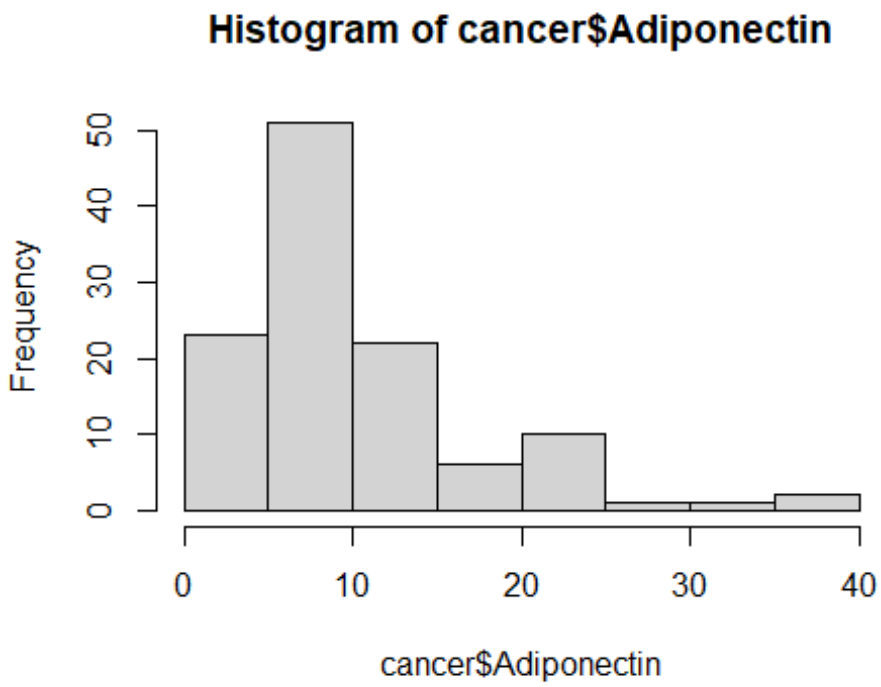

```r
hist(cancer$HOMA)
```



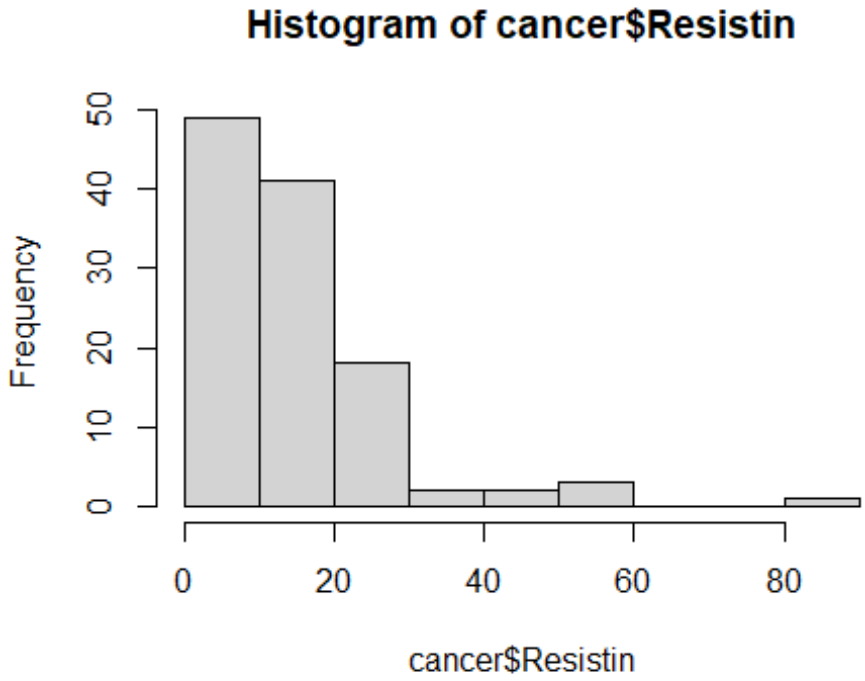

```r
hist(cancer$Leptin)
```



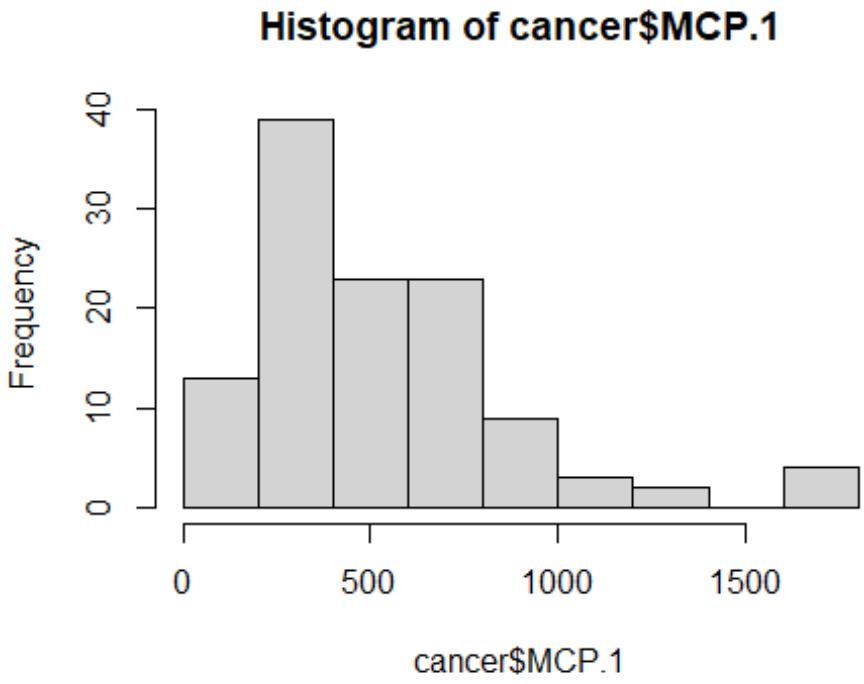

```r
hist(cancer$Adiponectin)
```



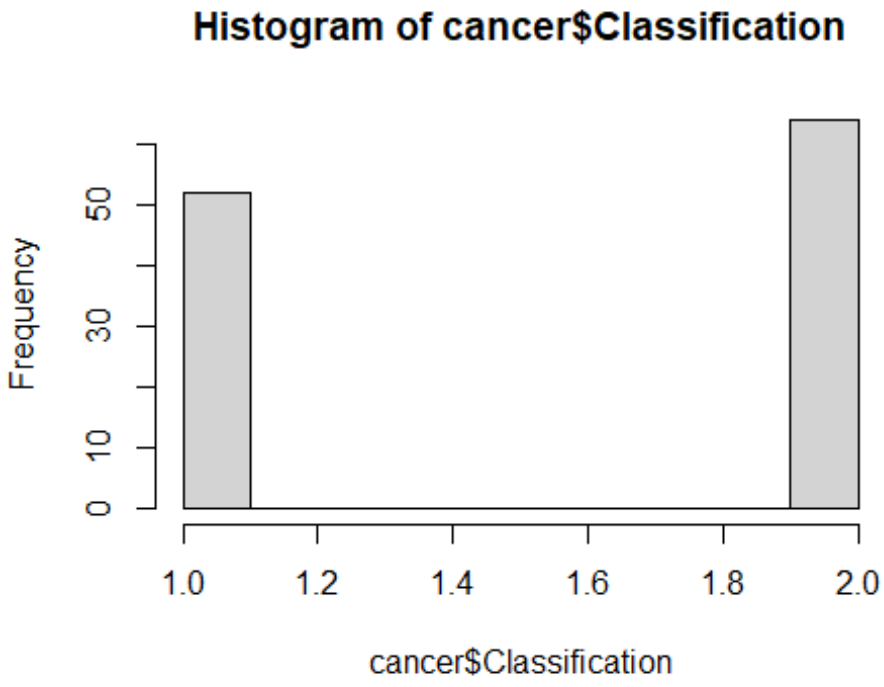

```r
hist(cancer$Resistin)
```




```r
hist(cancer$MCP.1)
```




```r
hist(cancer$Classification)
```




```r
#install.packages("ggplot2")
```

```r
#Cleaning data
#Check for null data
table(is.na(cancer))
```

```
##
## FALSE
##  1160
```

```r
#There is no null value

#Visualising the mahalanobis Chi-square plot for normality

library(MVA)
```

```
## Warning: package 'MVA' was built under R version 4.0.3
```

```
## Loading required package: HSAUR2
```

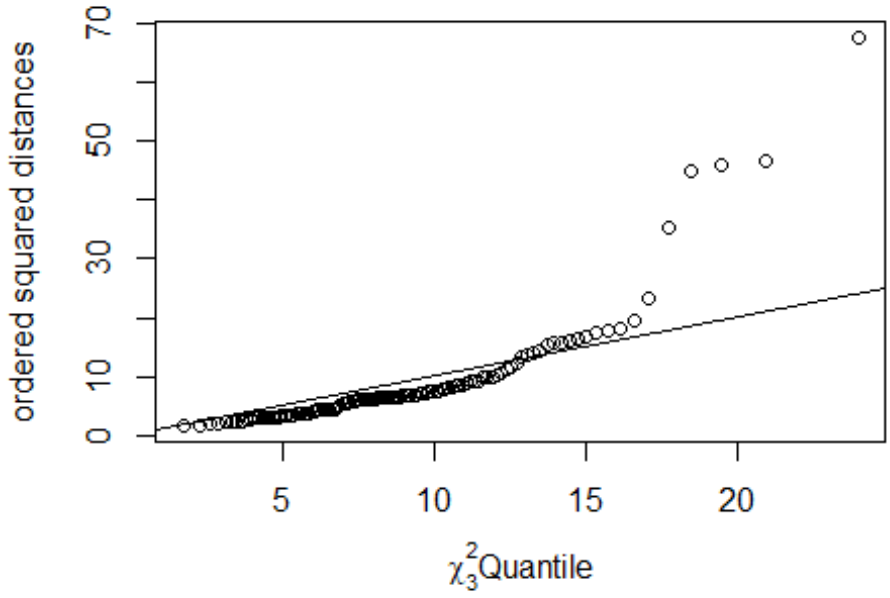
```
## Warning: package 'HSAUR2' was built under R version 4.0.3
```

```
## Loading required package: tools
```

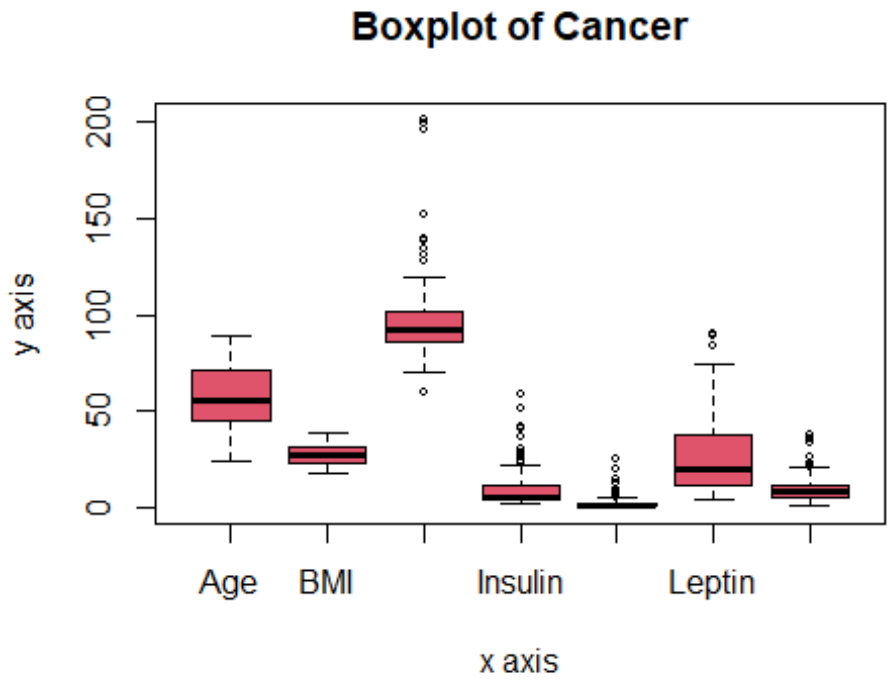
```
##
## Attaching package: 'HSAUR2'
```

```
## The following objects are masked from 'package:faraway':
##
##     epilepsy, toenail
```

```r
xbar <- colMeans(cancernew)
S <- cov(cancernew)
d2 <- mahalanobis(cancernew, xbar, S)

# Chi-Square plot:

quantiles <- qchisq((1:nrow(cancernew) - 1/2) / nrow(cancernew), df = ncol(ca
ncernew))
sd2 <- sort(d2)

# You can do the plot using
plot(quantiles, sd2,
    xlab = expression(paste(chi[3]^2, "Quantile")),
    ylab = "ordered squared distances")
abline(a=0, b=1) #a 45 degree angle
```

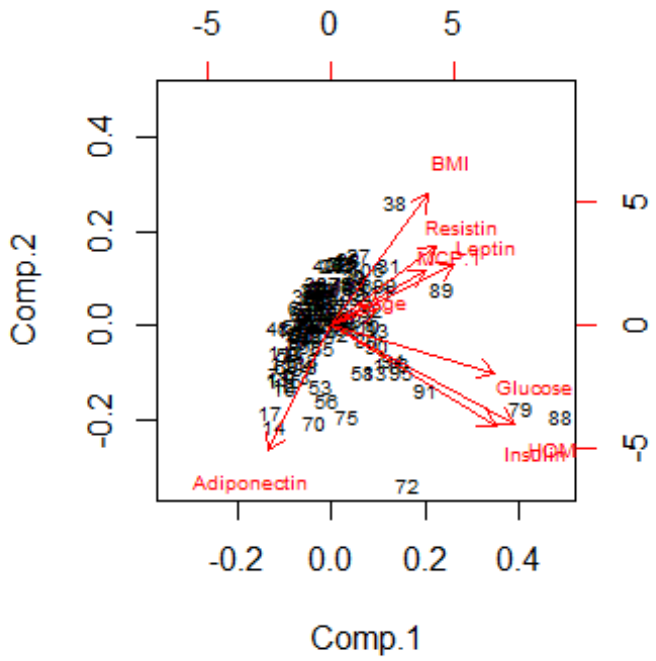



```
#The data is normaly distributed.

#creating boxplot
boxplot(cancernew[1:7], xlab = "x axis", ylab = "y axis", main = "Boxplot of
Cancer", col = 2, cex = 0.6)
```



Principal Component Analysis

```
#To find principal components of the standardized data


#The first 4 components explained 76.2% of the data


# Looking at the pc1
head(pc1 <- cancernew.pca$score[,1])

## [1] -1.9927846 -1.3288300 -1.2172439 -1.1992026 -0.8984674 -1.5273050

pc1[1]

## [1] -1.992785

biplot(cancernew.pca, col=c("black", "red"), cex = 0.6)
```




```
#There was highest variation in the HOMA data. This is understandable since t
he data is derived from both healthy volunteer and breast cancer patient. Thi
s also apply to the BMI.
#The BMI and Adiponectin are not correlated. This is because, the lesser the
adiponectin present in a person, the higher the tendency to accumulate fat, w
hich will lead to increase in BMI.
#With negative correlation between Resistin(promotes growth of cells) and adi
ponectin, the indicates that if adiponectin decreases in person with high res
istin, there is high tendency to develop cancer.
#If MCP is high,being a receptor of chemical, it encourages creation of wild
cells and it induced various deceases. It has negative correlation with adipo
nectin which breakdown protein and fat.

# The high correlation between glucose and insulin is an evidence that, a nor
mal body tend to produce more insulin with the increase in glucose presence i
n the body.
#Since HOMA is a derivative of glucose and insulin, it is therefore correlate
d to the the two.
```

Multidimesional Scaling

```
options(digits = 3)
#The dist(cancer1) first convert the data to Euclidean distances
d = dist(cancernew)

#Scaling the data
```

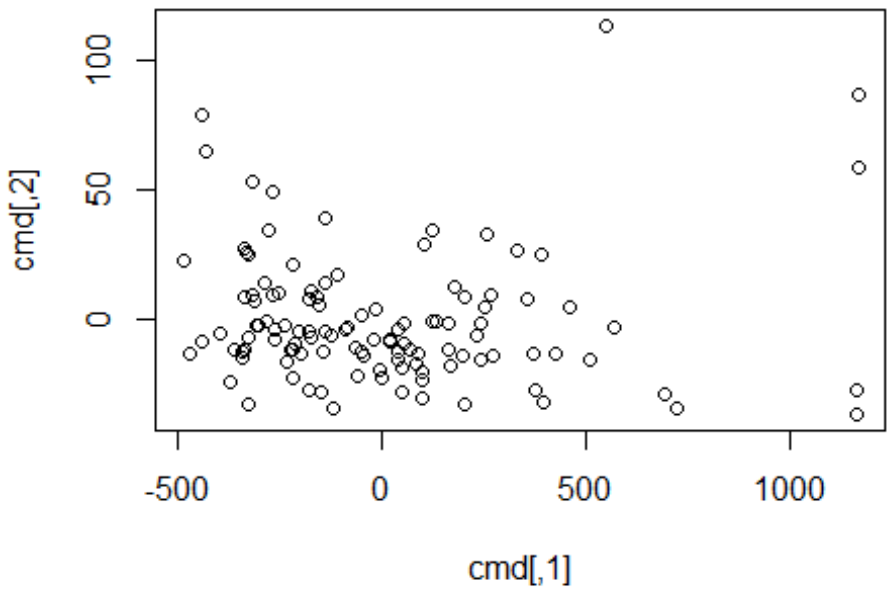
```r
cmd = cmdscale(d)
cmd # this gives 2D like X and Y coordinates
```

```
##            [,1]     [,2]
##   [1,] -118.147 -33.7453
##   [2,]  -66.126 -10.3555
##   [3,]   19.768  -7.5463
##   [4,]  393.027 -31.5569
##   [5,]  239.019 -15.0452
##   [6,]   -4.485 -19.2517
##   [7,]  720.834 -33.6783
##   [8,] -253.712  10.0635
##   [9,] -397.876  -4.8065
##  [10,] -216.691 -11.3422
##  [11,] -180.540 -26.8350
##  [12,] -360.276 -11.0782
##  [13,] -221.388 -22.4302
##  [14,]   97.083 -29.6214
##  [15,] -370.086 -23.6999
##  [16,] -471.305 -12.7503
##  [17,] -343.402 -11.8010
##  [18,]  -90.389  -3.2312
##  [19,] -282.268  -0.4602
##  [20,]   53.716  -1.1412
##  [21,]   -0.748 -21.9265
##  [22,]   37.798 -15.4912
##  [23,]  370.129 -12.8021
##  [24,]  198.856 -13.5742
##  [25,]  692.854 -28.0319
##  [26,]   51.212  -9.1354
##  [27,]  352.572   8.4464
##  [28,]  567.318  -2.8086
##  [29,]  133.230  -0.0418
##  [30,]   46.352 -27.8748
##  [31,]  330.503  26.5340
##  [32,]  161.050  -1.5100
##  [33,] -176.021  11.2538
##  [34,]  425.326 -12.7511
##  [35,]  -61.207 -21.6903
##  [36,]   50.198 -17.9410
##  [37,]   99.502 -20.1575
##  [38,] -270.069  49.1170
##  [39,] -155.448   6.0609
##  [40,]   83.265 -16.9921
##  [41,]  163.756 -11.1533
##  [42,] -157.483   8.7942
##  [43,] -199.543 -13.0758
##  [44,] -264.576  -3.6896
##  [45,] -333.606  25.8697
##  [46,] -308.934  -1.6575
```

```
##  [47,] -327.907  25.1776
##  [48,] -325.649 -32.4949
##  [49,] -319.041   9.8340
##  [50,] -302.683  -1.6865
##  [51,] -488.422  22.9451
##  [52,]  -46.100 -13.2376
##  [53,]   17.909  -8.2201
##  [54,] -151.959 -27.8558
##  [55,]   38.997 -11.9504
##  [56,]  -52.834 -12.1237
##  [57,] -212.790  -8.8795
##  [58,]  -85.809  -3.0260
##  [59,] -444.192  -8.4314
##  [60,]  169.058 -17.6633
##  [61,] -335.588 -11.0367
##  [62,]  178.801  12.6894
##  [63,]  202.675 -31.9218
##  [64,] -175.393  -6.4154
##  [65,] -179.390  -4.1942
##  [66,]  -15.901   4.4124
##  [67,]  100.074 -23.0779
##  [68,] -138.670  -4.6679
##  [69,] -233.672 -15.6050
##  [70,] -313.807   7.1432
##  [71,] -340.989 -14.3656
##  [72,] -289.568  14.1311
##  [73,]  -21.089  -7.6674
##  [74,] -338.693   8.9002
##  [75,] -222.796 -11.4953
##  [76,]  272.069 -13.8055
##  [77,]  -51.401   2.0358
##  [78,]  240.974  -1.4505
##  [79,] 1165.659  86.6630
##  [80,]  249.297   5.4000
##  [81,]  375.536 -26.5645
##  [82,] -108.176  17.6058
##  [83,]  203.381   8.9817
##  [84,]  265.812   9.8816
##  [85,]  507.540 -14.8662
##  [86,] 1163.371 -26.8112
##  [87,] 1163.308 -36.0261
##  [88,]  546.138 112.9294
##  [89,] 1165.790  58.3515
##  [90,]  389.883  24.8823
##  [91,]  104.357  28.7384
##  [92,]  459.732   4.6364
##  [93,]  230.035  -5.6144
##  [94,]  122.335  34.8205
##  [95,] -137.905  39.4045
##  [96,]   67.427 -11.1834
```

```
##  [97,] -127.282  -5.7157
##  [98,] -241.273  -1.8962
##  [99,] -277.898  34.2995
## [100,] -180.818   7.7822
## [101,]   37.609  -3.1989
## [102,] -265.260  -7.0611
## [103,] -137.843  13.9620
## [104,] -302.756  -2.1473
## [105,]  121.117  -0.3887
## [106,]  254.767  33.2206
## [107,]   86.457 -12.6673
## [108,] -325.564  -6.8967
## [109,] -335.711  27.8744
## [110,] -434.156  64.9392
## [111,] -315.946  53.2365
## [112,] -266.530   9.6499
## [113,] -204.579  -3.9160
## [114,] -220.636  21.0761
## [115,] -142.708 -11.8411
## [116,] -443.763  79.0049
```

```
plot(cmd)
```




```
cmdscale(dist(cancernew), k= 5, eig = T) # K is the number of coordinates ins
tead of the default 2D
```

```
## $points
##            [,1]     [,2]     [,3]     [,4]     [,5]
##   [1,] -118.147 -33.7453  -0.0535   1.7082  -0.1769
##   [2,]  -66.126 -10.3555 -23.0719 -23.0978   2.6876
##   [3,]   19.768  -7.5463 -13.5368 -22.6901   3.4910
##   [4,]  393.027 -31.5569  -6.3134 -17.1099   0.2820
##   [5,]  239.019 -15.0452 -23.2045 -26.8810  -0.5204
##   [6,]   -4.485 -19.2517 -11.2543   8.5846  -0.0776
##   [7,]  720.834 -33.6783 -14.0022 -37.3979   2.5777
##   [8,] -253.712  10.0635 -36.6526  -4.6168   1.9881
##   [9,] -397.876  -4.8065 -26.6189  -8.8161  -2.4258
##  [10,] -216.691 -11.3422  -9.6651 -18.6263   1.0767
##  [11,] -180.540 -26.8350   4.9147  16.6817   2.5697
##  [12,] -360.276 -11.0782  20.3291  20.9897   9.4156
##  [13,] -221.388 -22.4302  10.5931  26.1987   7.1981
##  [14,]   97.083 -29.6214   0.3546  31.2825   9.8502
##  [15,] -370.086 -23.6999   5.0011  14.1112  -0.6183
##  [16,] -471.305 -12.7503  -4.2766  13.8485   0.4691
##  [17,] -343.402 -11.8010   5.8893   7.8682   9.9268
##  [18,]  -90.389  -3.2312   7.7603  -7.0763   1.6007
##  [19,] -282.268  -0.4602  -6.8483  -6.2731   1.3064
##  [20,]   53.716  -1.1412  33.6444  14.0738   3.0756
##  [21,]   -0.748 -21.9265   3.2651  15.9024   2.3250
##  [22,]   37.798 -15.4912  14.4705  14.6343   8.4537
##  [23,]  370.129 -12.8021  35.4038  13.8957  -2.7786
##  [24,]  198.856 -13.5742  25.9151  10.4874   1.8296
##  [25,]  692.854 -28.0319  -2.9395  -2.4258   8.5734
##  [26,]   51.212  -9.1354  22.5410   1.1539   7.1368
##  [27,]  352.572   8.4464  17.7417   1.3405   9.2282
##  [28,]  567.318  -2.8086  20.6946 -17.6271   0.8082
##  [29,]  133.230  -0.0418  33.2669  11.0196  -1.8132
##  [30,]   46.352 -27.8748   4.1474  14.7720  -4.3120
##  [31,]  330.503  26.5340  37.4826 -17.6446   5.3296
##  [32,]  161.050  -1.5100   2.0646   4.1943   8.2992
##  [33,] -176.021  11.2538  49.5452  14.6654  -1.4992
##  [34,]  425.326 -12.7511  14.5930   7.4039  16.1525
##  [35,]  -61.207 -21.6903   5.8165  -1.7194  -0.5570
##  [36,]   50.198 -17.9410   0.7117 -16.7132   6.5718
##  [37,]   99.502 -20.1575  -4.4024 -14.1535   7.1968
##  [38,] -270.069  49.1170  48.5791 -16.0546 -58.9388
##  [39,] -155.448   6.0609  -5.3125  -0.4399  -9.7847
##  [40,]   83.265 -16.9921   1.4319 -25.9586  -5.1718
##  [41,]  163.756 -11.1533   2.0941 -24.7643   5.8671
##  [42,] -157.483   8.7942   0.6851 -18.7312   5.6431
##  [43,] -199.543 -13.0758 -15.1310 -17.0597  -0.5386
##  [44,] -264.576  -3.6896 -11.9656  -9.1086  -3.3406
##  [45,] -333.606  25.8697  14.0652 -16.0054  10.0633
##  [46,] -308.934  -1.6575  -3.5371  -9.5724   1.9463
##  [47,] -327.907  25.1776  23.1917 -24.5346  10.2226
##  [48,] -325.649 -32.4949  -8.9828 -26.8749  -0.8100
```

```
##  [49,] -319.041   9.8340  11.8155 -16.0017   8.3441
##  [50,] -302.683  -1.6865 -26.0539 -21.2123  -4.4852
##  [51,] -488.422  22.9451 -19.8789  -6.7792   1.9004
##  [52,]  -46.100 -13.2376 -11.9580 -20.1995  -2.3438
##  [53,]   17.909  -8.2201 -12.3356  18.6520  -7.7316
##  [54,] -151.959 -27.8558   0.5342   7.7311 -19.3698
##  [55,]   38.997 -11.9504  -7.1116   9.9657  -9.8454
##  [56,]  -52.834 -12.1237   1.6538  24.3001   5.4294
##  [57,] -212.790  -8.8795   0.0659  14.0113  -7.7024
##  [58,]  -85.809  -3.0260 -20.5057  -4.8666  -4.4569
##  [59,] -444.192  -8.4314 -12.6579   8.2855  -1.0325
##  [60,]  169.058 -17.6633 -12.4478  -4.9579   1.1552
##  [61,] -335.588 -11.0367  -8.6414  21.0353  -9.2031
##  [62,]  178.801  12.6894  -5.6338  -8.9117   0.1244
##  [63,]  202.675 -31.9218  -2.5258   2.4150 -11.9904
##  [64,] -175.393  -6.4154 -16.8154  10.9134  -3.5718
##  [65,] -179.390  -4.1942 -12.4519   0.6406   0.2969
##  [66,]  -15.901   4.4124 -12.8581  17.0957   6.0824
##  [67,]  100.074 -23.0779  -8.0258   0.0506  -0.5208
##  [68,] -138.670  -4.6679 -15.3860  -3.2987  -5.4729
##  [69,] -233.672 -15.6050  -4.6635  10.9870  -3.8872
##  [70,] -313.807   7.1432 -17.8979  23.7776   7.5635
##  [71,] -340.989 -14.3656 -11.7475  14.4233  -0.7329
##  [72,] -289.568  14.1311 -10.0686  26.4086  16.6096
##  [73,]  -21.089  -7.6674 -17.5329  10.8067   7.5700
##  [74,] -338.693   8.9002 -15.9415  -8.2713   4.3392
##  [75,] -222.796 -11.4953  -7.3669  14.8279   9.0442
##  [76,]  272.069 -13.8055  -8.2673  15.3649  -0.9724
##  [77,]  -51.401   2.0358  11.0371  -1.1886   9.9635
##  [78,]  240.974  -1.4505   7.2552  13.9419 -20.1072
##  [79,] 1165.659  86.6630 -34.3076   3.4130  15.9289
##  [80,]  249.297   5.4000  23.1849  12.3377  -2.7852
##  [81,]  375.536 -26.5645   6.1109  -8.4548 -14.5284
##  [82,] -108.176  17.6058   2.4480 -21.4612 -13.9707
##  [83,]  203.381   8.9817  18.8846   7.3243  16.4063
##  [84,]  265.812   9.8816 -12.2324  -6.1745 -25.6807
##  [85,]  507.540 -14.8662   2.4840  14.5702 -34.7937
##  [86,] 1163.371 -26.8112  10.2749 -17.9358   8.4525
##  [87,] 1163.308 -36.0261   4.7634  -0.6529   8.3993
##  [88,]  546.138 112.9294 -12.3909   5.9326 -13.6009
##  [89,] 1165.790  58.3515 -37.8755  39.0540 -22.2621
##  [90,]  389.883  24.8823 -20.0982  12.6103   7.1592
##  [91,]  104.357  28.7384  -3.6311  29.2407   7.3215
##  [92,]  459.732   4.6364  -2.0844 -23.8229   1.9776
##  [93,]  230.035  -5.6144  11.5975   2.5079  -2.2681
##  [94,]  122.335  34.8205  -1.5308  17.5870  17.4312
##  [95,] -137.905  39.4045  -8.4698  10.3370   8.5418
##  [96,]   67.427 -11.1834  36.8136  -7.2649  -1.5111
##  [97,] -127.282  -5.7157  -6.3771  15.8381 -12.0527
##  [98,] -241.273  -1.8962  -5.6170  20.3831 -16.8836
```

```
##  [99,] -277.898  34.2995  16.8451 -11.0847 -31.2477
## [100,] -180.818   7.7822 -18.7395  -4.1347  -6.9313
## [101,]   37.609  -3.1989   3.2480 -20.6584  -5.1363
## [102,] -265.260  -7.0611  -9.7123  -7.4605  -6.8228
## [103,] -137.843  13.9620  11.9745 -10.1284  -5.2988
## [104,] -302.756  -2.1473   0.4234 -16.9732  -5.4347
## [105,]  121.117  -0.3887   7.5521  -3.7323   6.3072
## [106,]  254.767  33.2206  -8.7318  -8.7169  11.2856
## [107,]   86.457 -12.6673   6.6507   7.1306  -1.7534
## [108,] -325.564  -6.8967  -1.3151  10.4199  -0.1767
## [109,] -335.711  27.8744 -28.0300   4.4109   5.3368
## [110,] -434.156  64.9392 -14.5259  -0.3819   2.2234
## [111,] -315.946  53.2365  39.0267  -1.2633  17.1190
## [112,] -266.530   9.6499  27.2104   3.9972   4.4198
## [113,] -204.579  -3.9160 -15.9763   0.4469   2.8661
## [114,] -220.636  21.0761  24.1725 -12.6646   8.4531
## [115,] -142.708 -11.8411  -1.6711 -16.7397  10.2672
## [116,] -443.763  79.0049  17.6394 -20.6259  18.4467
##
## $eig
##   [1]  1.38e+07  7.25e+04  3.42e+04  2.73e+04  1.39e+04  8.14e+03  4.80e+0
3
##   [8]  1.51e+03  7.72e+01  5.37e-09  3.22e-09  2.55e-09  1.26e-09  4.96e-1
0
##  [15]  4.69e-10  4.41e-10  3.93e-10  3.09e-10  2.82e-10  2.56e-10  2.32e-1
0
##  [22]  2.10e-10  1.90e-10  1.78e-10  1.09e-10  1.03e-10  9.58e-11  9.56e-1
1
##  [29]  9.55e-11  9.00e-11  8.86e-11  7.46e-11  6.63e-11  5.04e-11  4.99e-1
1
##  [36]  4.79e-11  4.67e-11  4.33e-11  2.97e-11  2.45e-11  2.37e-11  2.35e-1
1
##  [43]  2.28e-11  1.99e-11  1.75e-11  1.51e-11  1.48e-11  1.21e-11  1.16e-1
1
##  [50]  1.14e-11  1.06e-11  7.05e-12  6.96e-12  6.86e-12  6.16e-12  4.84e-1
2
##  [57]  4.10e-12  7.41e-13  1.99e-13  1.18e-13 -4.82e-14 -6.52e-13 -7.94e-1
3
##  [64] -1.55e-12 -1.68e-12 -1.73e-12 -2.19e-12 -2.37e-12 -3.37e-12 -3.43e-1
2
##  [71] -7.30e-12 -7.95e-12 -8.42e-12 -9.47e-12 -9.89e-12 -1.11e-11 -1.17e-1
1
##  [78] -1.22e-11 -1.29e-11 -1.47e-11 -1.79e-11 -1.93e-11 -2.45e-11 -2.56e-1
1
##  [85] -2.82e-11 -2.83e-11 -3.29e-11 -3.69e-11 -3.78e-11 -3.88e-11 -3.94e-1
1
##  [92] -4.21e-11 -5.82e-11 -6.67e-11 -6.69e-11 -7.17e-11 -7.42e-11 -7.65e-1
1
##  [99] -8.51e-11 -9.90e-11 -1.06e-10 -1.19e-10 -1.21e-10 -1.43e-10 -1.61e-1
0
```

```
## [106] -1.61e-10 -1.70e-10 -1.87e-10 -2.29e-10 -2.95e-10 -3.59e-10 -3.65e-1
0
## [113] -4.38e-10 -1.59e-09 -1.85e-09 -5.12e-09
##
## $x
## NULL
##
## $ac
## [1] 0
##
## $GOF
## [1] 0.999 0.999

# Comparing results with the principle component scores using cov matrix:
princomp(cancernew)$scores

##          Comp.1   Comp.2   Comp.3   Comp.4   Comp.5   Comp.6  Comp.7  Comp
.8
##   [1,] -118.147 -33.7453   0.0535  -1.7082   0.1769   0.8231 -1.9295 -2.15
39
##   [2,]  -66.126 -10.3555  23.0719  23.0978  -2.6876  -1.7696 -2.4526 -4.54
24
##   [3,]   19.768  -7.5463  13.5368  22.6901  -3.4910  -0.9170 14.3825  0.75
16
##   [4,]  393.027 -31.5569   6.3134  17.1099  -0.2820   0.9499  0.4795 -5.36
17
##   [5,]  239.019 -15.0452  23.2045  26.8810   0.5204  -0.5005 -1.0435 -4.41
74
##   [6,]   -4.485 -19.2517  11.2543  -8.5846   0.0776  -4.2289  2.2846 -1.22
55
##   [7,]  720.834 -33.6783  14.0022  37.3979  -2.5777   3.9431  1.5918 -4.14
06
##   [8,] -253.712  10.0635  36.6526   4.6168  -1.9881  -5.3002  2.1648  1.98
42
##   [9,] -397.876  -4.8065  26.6189   8.8161   2.4258  -2.2685 -0.1408 -0.55
49
##  [10,] -216.691 -11.3422   9.6651  18.6263  -1.0767   1.4159  2.4778 -1.87
78
##  [11,] -180.540 -26.8350  -4.9147 -16.6817  -2.5697  -2.6896  0.1499 -4.36
08
##  [12,] -360.276 -11.0782 -20.3291 -20.9897  -9.4156  -4.7108 12.1595 -2.31
36
##  [13,] -221.388 -22.4302 -10.5931 -26.1987  -7.1981  -5.1560  8.6449 -1.44
58
##  [14,]   97.083 -29.6214  -0.3546 -31.2825  -9.8502  -4.1624 21.7139 -1.10
36
##  [15,] -370.086 -23.6999  -5.0011 -14.1112   0.6183   1.4484  4.4605 -0.59
44
##  [16,] -471.305 -12.7503   4.2766 -13.8485  -0.4691   0.9957  7.0981 -1.56
91
```

```
##  [17,] -343.402 -11.8010  -5.8893  -7.8682  -9.9268  -5.3718 24.4271  1.67
25
##  [18,]  -90.389  -3.2312  -7.7603   7.0763  -1.6007  10.8103 -3.3798  3.32
50
##  [19,] -282.268  -0.4602   6.8483   6.2731  -1.3064  -4.2723 -7.6625  6.74
34
##  [20,]   53.716  -1.1412 -33.6444 -14.0738  -3.0756   1.3583 -7.9798  2.47
99
##  [21,]   -0.748 -21.9265  -3.2651 -15.9024  -2.3250  -4.2445 -5.0204  0.92
83
##  [22,]   37.798 -15.4912 -14.4705 -14.6343  -8.4537  -6.9455 -6.7432  1.42
10
##  [23,]  370.129 -12.8021 -35.4038 -13.8957   2.7786  -4.9090 -3.8268 -1.60
30
##  [24,]  198.856 -13.5742 -25.9151 -10.4874  -1.8296  -5.8345 -2.0732 -1.43
74
##  [25,]  692.854 -28.0319   2.9395   2.4258  -8.5734  -3.7993 -0.6495  2.29
55
##  [26,]   51.212  -9.1354 -22.5410  -1.1539  -7.1368  -3.5624 -8.6702  4.19
43
##  [27,]  352.572   8.4464 -17.7417  -1.3405  -9.2282 -10.3679 -7.1690  4.45
16
##  [28,]  567.318  -2.8086 -20.6946  17.6271  -0.8082  -0.1945  4.3030 -0.36
32
##  [29,]  133.230  -0.0418 -33.2669 -11.0196   1.8132  -5.4507 -4.2396  1.56
08
##  [30,]   46.352 -27.8748  -4.1474 -14.7720   4.3120   0.8545 -8.9046  6.21
46
##  [31,]  330.503  26.5340 -37.4826  17.6446  -5.3296  -0.0526  0.6408 -0.13
76
##  [32,]  161.050  -1.5100  -2.0646  -4.1943  -8.2992  -3.0210  6.1903 10.43
27
##  [33,] -176.021  11.2538 -49.5452 -14.6654   1.4992  -4.9811 -7.2301 -0.65
12
##  [34,]  425.326 -12.7511 -14.5930  -7.4039 -16.1525   9.9868 -5.2272  3.30
12
##  [35,]  -61.207 -21.6903  -5.8165   1.7194   0.5570  -0.8022 -7.8734 -1.44
61
##  [36,]   50.198 -17.9410  -0.7117  16.7132  -6.5718   0.5174 -8.3138 -0.08
22
##  [37,]   99.502 -20.1575   4.4024  14.1535  -7.1968  -1.7000 -7.1741  2.67
94
##  [38,] -270.069  49.1170 -48.5791  16.0546  58.9388   1.9422  9.0626  1.70
50
##  [39,] -155.448   6.0609   5.3125   0.4399   9.7847  -4.0718 -6.5343 -2.08
97
##  [40,]   83.265 -16.9921  -1.4319  25.9586   5.1718   2.4086 -0.9580  7.97
38
##  [41,]  163.756 -11.1533  -2.0941  24.7643  -5.8671  -0.4656 -1.1429  0.10
87
```

```
##  [42,] -157.483   8.7942  -0.6851  18.7312  -5.6431   5.0698  0.1951 -1.09
74
##  [43,] -199.543 -13.0758  15.1310  17.0597   0.5386   1.7382 -1.3619  2.59
93
##  [44,] -264.576  -3.6896  11.9656   9.1086   3.3406  -0.7038 -0.1304  7.80
20
##  [45,] -333.606  25.8697 -14.0652  16.0054 -10.0633  -5.7590 -2.2142 -1.60
91
##  [46,] -308.934  -1.6575   3.5371   9.5724  -1.9463  -1.2236 -3.4711  0.38
05
##  [47,] -327.907  25.1776 -23.1917  24.5346 -10.2226  -4.2378 -4.0335 -8.58
43
##  [48,] -325.649 -32.4949   8.9828  26.8749   0.8100   7.5319  0.4253  1.59
76
##  [49,] -319.041   9.8340 -11.8155  16.0017  -8.3441   1.0666 -2.4499 -0.91
04
##  [50,] -302.683  -1.6865  26.0539  21.2123   4.4852   0.0765 -1.4285  3.00
31
##  [51,] -488.422  22.9451  19.8789   6.7792  -1.9004  15.1260 -6.4669  1.36
39
##  [52,]  -46.100 -13.2376  11.9580  20.1995   2.3438   1.3580  1.2396  0.75
53
##  [53,]   17.909  -8.2201  12.3356 -18.6520   7.7316   4.9320 10.7098 -0.09
96
##  [54,] -151.959 -27.8558  -0.5342  -7.7311  19.3698   5.0370 -0.0921 -3.85
25
##  [55,]   38.997 -11.9504   7.1116  -9.9657   9.8454   5.6854 -0.2466 -4.35
61
##  [56,]  -52.834 -12.1237  -1.6538 -24.3001  -5.4294  11.0296  7.9199  0.75
65
##  [57,] -212.790  -8.8795  -0.0659 -14.0113   7.7024  -4.8957 -1.8252 -5.25
13
##  [58,]  -85.809  -3.0260  20.5057   4.8666   4.4569  -1.4674  0.0552 -3.31
44
##  [59,] -444.192  -8.4314  12.6579  -8.2855   1.0325  -3.0685 -5.8414 -6.30
08
##  [60,]  169.058 -17.6633  12.4478   4.9579  -1.1552  -3.3434  2.1041 -2.50
30
##  [61,] -335.588 -11.0367   8.6414 -21.0353   9.2031  -2.4582 -7.3520 -3.24
74
##  [62,]  178.801  12.6894   5.6338   8.9117  -0.1244  -6.5035 -2.0077 -8.13
19
##  [63,]  202.675 -31.9218   2.5258  -2.4150  11.9904   0.8840 -3.7434 -5.36
18
##  [64,] -175.393  -6.4154  16.8154 -10.9134   3.5718  -6.8730 -2.0184 -1.70
24
##  [65,] -179.390  -4.1942  12.4519  -0.6406  -0.2969  -2.0285 -6.2591 -3.99
73
##  [66,]  -15.901   4.4124  12.8581 -17.0957  -6.0824 -11.5364 -7.7738 -5.09
53
```

```
##  [67,]  100.074 -23.0779   8.0258  -0.0506   0.5208  -2.0491 -6.4736 -3.02
11
##  [68,] -138.670  -4.6679  15.3860   3.2987   5.4729  -1.4058 -4.1391 -3.68
51
##  [69,] -233.672 -15.6050   4.6635 -10.9870   3.8872  -2.7229  7.3226 -2.17
89
##  [70,] -313.807   7.1432  17.8979 -23.7776  -7.5635   0.7800  6.5507 -2.50
85
##  [71,] -340.989 -14.3656  11.7475 -14.4233   0.7329  -3.9683  4.0259 -2.32
19
##  [72,] -289.568  14.1311  10.0686 -26.4086 -16.6096  43.2004 -0.0755  1.32
36
##  [73,]  -21.089  -7.6674  17.5329 -10.8067  -7.5700  -6.1390  1.4346 -6.09
20
##  [74,] -338.693   8.9002  15.9415   8.2713  -4.3392  -6.0235  7.1146  0.09
14
##  [75,] -222.796 -11.4953   7.3669 -14.8279  -9.0442  28.6362 -4.1112 -2.32
37
##  [76,]  272.069 -13.8055   8.2673 -15.3649   0.9724  -0.4083 -4.7514 -0.51
45
##  [77,]  -51.401   2.0358 -11.0371   1.1886  -9.9635   4.9732 -1.5444  2.50
37
##  [78,]  240.974  -1.4505  -7.2552 -13.9419  20.1072  -4.9691  4.5232  4.11
07
##  [79,] 1165.659  86.6630  34.3076  -3.4130 -15.9289   3.9567  1.6636 -5.23
78
##  [80,]  249.297   5.4000 -23.1849 -12.3377   2.7852   9.5589  4.9329 -3.29
07
##  [81,]  375.536 -26.5645  -6.1109   8.4548  14.5284   2.1097  9.6990  3.46
45
##  [82,] -108.176  17.6058  -2.4480  21.4612  13.9707   2.8103  4.4833  4.03
59
##  [83,]  203.381   8.9817 -18.8846  -7.3243 -16.4063  12.6651  8.6204  3.84
72
##  [84,]  265.812   9.8816  12.2324   6.1745  25.6807   3.1244  2.0359 -0.66
76
##  [85,]  507.540 -14.8662  -2.4840 -14.5702  34.7937  -0.2897  1.7397  2.62
60
##  [86,] 1163.371 -26.8112 -10.2749  17.9358  -8.4525   6.1715  1.6928 -1.69
86
##  [87,] 1163.308 -36.0261  -4.7634   0.6529  -8.3993  -7.7784  2.4192 -1.71
64
##  [88,]  546.138 112.9294  12.3909  -5.9326  13.6009  19.8820  7.6370 -3.94
87
##  [89,] 1165.790  58.3515  37.8755 -39.0540  22.2621 -21.2954 -0.4250  2.93
26
##  [90,]  389.883  24.8823  20.0982 -12.6103  -7.1592  -5.3542 -0.4256  1.73
64
##  [91,]  104.357  28.7384   3.6311 -29.2407  -7.3215  20.8465 -7.3364  1.64
26
```

```
##  [92,]  459.732   4.6364   2.0844  23.8229  -1.9776   8.5930  3.4320  2.70
16
##  [93,]  230.035  -5.6144 -11.5975  -2.5079   2.2681  22.1288 -3.4654  1.54
64
##  [94,]  122.335  34.8205   1.5308 -17.5870 -17.4312   0.1054 -3.0045  2.07
56
##  [95,] -137.905  39.4045   8.4698 -10.3370  -8.5418   9.4569 -4.0810  2.52
67
##  [96,]   67.427 -11.1834 -36.8136   7.2649   1.5111   2.9414  2.1206 -2.36
00
##  [97,] -127.282  -5.7157   6.3771 -15.8381  12.0527   0.8605 -5.7188  1.51
44
##  [98,] -241.273  -1.8962   5.6170 -20.3831  16.8836  -6.4965 -4.3903  1.57
87
##  [99,] -277.898  34.2995 -16.8451  11.0847  31.2477  11.4388 -1.1133 -2.84
89
## [100,] -180.818   7.7822  18.7395   4.1347   6.9313  -0.5881 -4.6532  2.61
88
## [101,]   37.609  -3.1989  -3.2480  20.6584   5.1363  -2.3873  0.4665  0.04
19
## [102,] -265.260  -7.0611   9.7123   7.4605   6.8228   1.9872 -1.5669  1.32
76
## [103,] -137.843  13.9620 -11.9745  10.1284   5.2988   0.7421 -4.7446 -0.27
79
## [104,] -302.756  -2.1473  -0.4234  16.9732   5.4347   7.1051 -6.5195  0.77
81
## [105,]  121.117  -0.3887  -7.5521   3.7323  -6.3072   0.9232 -9.4028  3.25
15
## [106,]  254.767  33.2206   8.7318   8.7169 -11.2856 -16.4629 -7.4285  4.90
84
## [107,]   86.457 -12.6673  -6.6507  -7.1306   1.7534  -4.1483 -4.4920  0.28
77
## [108,] -325.564  -6.8967   1.3151 -10.4199   0.1767  -3.2638 -5.6184  6.33
73
## [109,] -335.711  27.8744  28.0300  -4.4109  -5.3368  -9.9613 -2.9522 10.77
87
## [110,] -434.156  64.9392  14.5259   0.3819  -2.2234 -17.8164 -0.2824  1.57
19
## [111,] -315.946  53.2365 -39.0267   1.2633 -17.1190 -12.4633 -4.0886 -1.67
41
## [112,] -266.530   9.6499 -27.2104  -3.9972  -4.4198  -9.2673  1.2621 -4.73
37
## [113,] -204.579  -3.9160  15.9763  -0.4469  -2.8661  -4.6687  8.9572  4.80
21
## [114,] -220.636  21.0761 -24.1725  12.6646  -8.4531  -7.5886 12.0576  2.61
13
## [115,] -142.708 -11.8411   1.6711  16.7397 -10.2672  -2.7557 22.5756  4.54
90
## [116,] -443.763  79.0049 -17.6394  20.6259 -18.4467  -5.8652  6.2116 -6.61
69
```

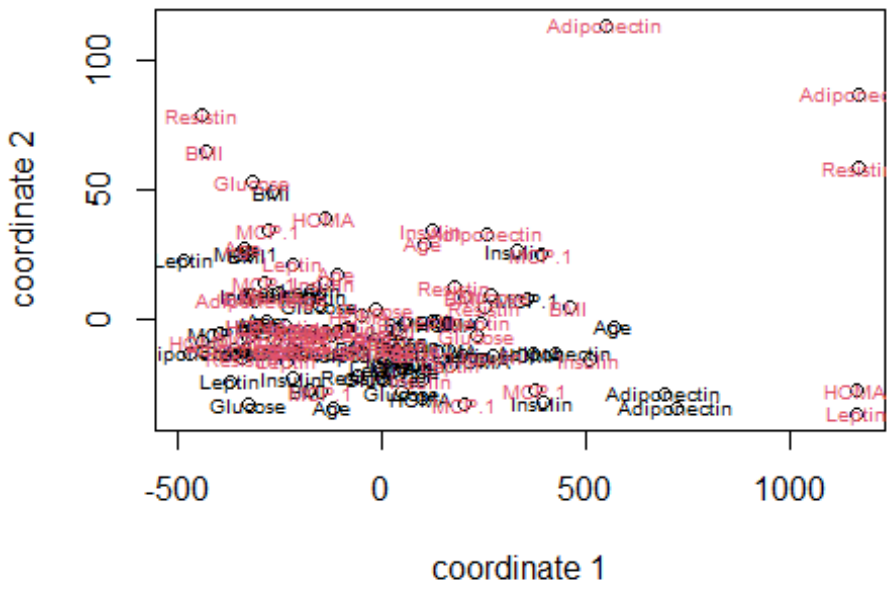
```
##           Comp.9
##   [1,] -1.11323
##   [2,]  0.11314
##   [3,]  0.23015
##   [4,] -0.18900
##   [5,]  0.34821
##   [6,] -0.18757
##   [7,]  0.21323
##   [8,]  0.61931
##   [9,] -0.07762
##  [10,] -0.23547
##  [11,] -0.74791
##  [12,] -0.47533
##  [13,] -0.66843
##  [14,]  0.03358
##  [15,] -0.78566
##  [16,] -0.23771
##  [17,] -0.45678
##  [18,]  0.35815
##  [19,] -0.67665
##  [20,] -0.14112
##  [21,] -0.69311
##  [22,] -0.74363
##  [23,] -0.33907
##  [24,] -0.48469
##  [25,] -0.03736
##  [26,] -0.78577
##  [27,]  0.01464
##  [28,]  0.50710
##  [29,] -0.35517
##  [30,] -0.99510
##  [31,]  0.73876
##  [32,] -0.11325
##  [33,] -0.45963
##  [34,]  0.68262
##  [35,] -0.85541
##  [36,] -0.55813
##  [37,] -0.71959
##  [38,]  0.53549
##  [39,]  0.19934
##  [40,] -0.96505
##  [41,] -0.22014
##  [42,]  0.51374
##  [43,] -0.45441
##  [44,] -0.51957
##  [45,]  0.21713
##  [46,] -0.31001
##  [47,]  0.39307
##  [48,] -1.38887
##  [49,]  0.03391
```

```
##  [50,] -0.12651
##  [51,]  0.45688
##  [52,] -0.25068
##  [53,]  0.65914
##  [54,] -0.58214
##  [55,]  0.56575
##  [56,]  0.71927
##  [57,] -0.14130
##  [58,]  0.47061
##  [59,] -0.16327
##  [60,]  0.07044
##  [61,] -0.25009
##  [62,]  1.08161
##  [63,] -0.45933
##  [64,]  0.07832
##  [65,]  0.15716
##  [66,]  0.59774
##  [67,] -0.34457
##  [68,]  0.22256
##  [69,] -0.27807
##  [70,]  0.58516
##  [71,] -0.27083
##  [72,]  1.23640
##  [73,]  0.55829
##  [74,]  0.31550
##  [75,]  1.80294
##  [76,]  0.31433
##  [77,]  0.27678
##  [78,]  0.16876
##  [79,] -3.08250
##  [80,]  1.09248
##  [81,] -0.48821
##  [82,]  0.37545
##  [83,]  0.95831
##  [84,]  1.00332
##  [85,]  0.23271
##  [86,]  1.06964
##  [87,]  0.23115
##  [88,] -4.82138
##  [89,]  2.22824
##  [90,]  0.75653
##  [91,] -0.37685
##  [92,]  0.95066
##  [93,]  1.41165
##  [94,]  0.06512
##  [95,] -0.09826
##  [96,] -0.30009
##  [97,] -0.02865
##  [98,] -0.19997
##  [99,]  0.90978
```

```
## [100,]  0.25265
## [101,] -0.20939
## [102,] -0.24766
## [103,]  0.21794
## [104,] -0.10822
## [105,]  0.00317
## [106,]  0.90488
## [107,] -0.36148
## [108,] -0.78654
## [109,]  0.21625
## [110,]  1.37477
## [111,]  0.38247
## [112,] -0.20601
## [113,] -0.06474
## [114,]  0.04586
## [115,] -0.43961
## [116,]  0.90457
```

```
#The result is thesame. This shows another way to represent the data.
#After about the first 9 eignen values, remaining ones is very cose to 0.
```

```
plot(cmdscale(dist(cancernew)), xlab = "coordinate 1", ylab = "coordinate 2")
text(cmdscale(dist(cancernew)), labels=(colnames(cancernew)), cex = 0.6,col =
cancer$Classification)
```



```r
# 2d representation
cmd2 <- cmdscale(dist(cancernew), k=2)
dist(cmd2[1:9])
```

```
##        1      2      3      4      5      6      7      8
## 2   52.0
## 3  137.9   85.9
## 4  511.2  459.2  373.3
## 5  357.2  305.1  219.3  154.0
## 6  113.7   61.6   24.3  397.5  243.5
## 7  839.0  787.0  701.1  327.8  481.8  725.3
## 8  135.6  187.6  273.5  646.7  492.7  249.2  974.5
## 9  279.7  331.8  417.6  790.9  636.9  393.4 1118.7  144.2
```

```r
dist(cancernew[(1:9),])
```

```
##        1      2      3      4      5      6      7      8
## 2   66.5
## 3  144.1   88.3
## 4  511.6  460.0  374.5
## 5  359.6  305.2  220.3  156.1
## 6  115.5   71.2   43.4  398.6  246.5
## 7  840.0  787.5  701.9  328.6  482.4  727.0
## 8  147.5  190.3  275.9  649.0  494.1  252.6  976.4
## 9  282.7  332.2  418.4  791.7  637.3  394.4 1119.6  145.5
```

```r
# comparing this with original dist matrix.
# There are some errors because 2 dimension cannot fully explain the data.
```

Exploratory Factor Analysis of the Data

```r
#The null hypothesis is that k factors is sufficient for presenting the data
cancernew.fa <- factanal(cancernew, factors = 5) # 5 is the max number of fac
tor this data with 9 variables can accept.
cancernew.fa
```

```
##
## Call:
## factanal(x = cancernew, factors = 5)
##
## Uniquenesses:
##         Age         BMI     Glucose     Insulin        HOMA      Leptin
##       0.864       0.527       0.326       0.005       0.005       0.005
## Adiponectin    Resistin       MCP.1
##       0.005       0.744       0.133
##
## Loadings:
##             Factor1 Factor2 Factor3 Factor4 Factor5
## Age                                 -0.143   0.337
## BMI                  0.620   0.200  -0.215
## Glucose      0.502           0.240           0.593
```

```
## Insulin      0.986   0.136
## HOMA         0.931   0.100   0.167           0.299
## Leptin       0.184   0.955                   0.210
## Adiponectin         -0.127  -0.153   0.947  -0.239
## Resistin             0.226   0.372  -0.127   0.203
## MCP.1        0.119           0.920
##
##                  Factor1 Factor2 Factor3 Factor4 Factor5
## SS loadings        2.150   1.402   1.138   0.996   0.700
## Proportion Var     0.239   0.156   0.126   0.111   0.078
## Cumulative Var     0.239   0.395   0.521   0.632   0.710
##
## Test of the hypothesis that 5 factors are sufficient.
## The chi square statistic is 4.46 on 1 degree of freedom.
## The p-value is 0.0348
```

```
# Since p-value < 0.05 for k = 5

## The null hypothesis is that k factors are sufficient for presenting the da
ta. We assumed k=5 factors are sufficient.
# Though p-value < 0.05 for k = 5,which will reject the null hypothesis. It i
s arguable that p-value may not be reliable in determining the validity of a
model.
# There is a serious debate between statisticians in terms of whether to rely
on p-value or not. In this data the sample size is large, any small discrepan
cies between the estimated correlation matrix (corhat) and the original corre
lation matrix will be detected as significant, and the chance of rejection (p
-value<0.05) is very high. (B.Everitt et.al.2011)


######REMOVE TO REFERENCE PAGE  *B. Everitt & T. Hothorn, An Introduction to
Applied Multivariate Analysis with R: Use R! 2011*####
cancernew.fa$loadings
```

```
##
## Loadings:
##              Factor1 Factor2 Factor3 Factor4 Factor5
## Age                                  -0.143   0.337
## BMI                   0.620   0.200  -0.215
## Glucose      0.502            0.240           0.593
## Insulin      0.986   0.136
## HOMA         0.931   0.100   0.167           0.299
## Leptin       0.184   0.955                   0.210
## Adiponectin         -0.127  -0.153   0.947  -0.239
## Resistin             0.226   0.372  -0.127   0.203
## MCP.1        0.119           0.920
##
##                  Factor1 Factor2 Factor3 Factor4 Factor5
## SS loadings        2.150   1.402   1.138   0.996   0.700
```

```
## Proportion Var   0.239   0.156   0.126   0.111   0.078
## Cumulative Var   0.239   0.395   0.521   0.632   0.710

#the low uniqueness of Insulin, HOMA, Leptin, Adiponectin and MCP.1 shows tha
t high percentage of the data can be explained by the 5 factors but will be m
ore difficult to resolve for Age and Resistin

###### INPUT the uniqueness table into the report#########
```

To check the quality of the model, we check the RMSE

```
f.loading = cancernew.fa$loadings[, 1:5]

corHat = f.loading %*% t(f.loading) + diag(cancernew.fa$uniquenesses)
corr = cor(cancernew)

# discrepancy, the root-mean-square error (RMSE)
rmse = sqrt(mean((corHat-corr)^2))
rmse

## [1] 0.0191

#Less than 2% discrepancy is good for the validity of the data

# Focussing on the loading of the first 5 factors.
cancernew.fa$loadings

##
## Loadings:
##             Factor1 Factor2 Factor3 Factor4 Factor5
## Age                                 -0.143   0.337
## BMI                  0.620   0.200  -0.215
## Glucose      0.502           0.240           0.593
## Insulin      0.986   0.136
## HOMA         0.931   0.100   0.167           0.299
## Leptin       0.184   0.955                   0.210
## Adiponectin         -0.127  -0.153   0.947  -0.239
## Resistin             0.226   0.372  -0.127   0.203
## MCP.1        0.119           0.920
##
##                Factor1 Factor2 Factor3 Factor4 Factor5
## SS loadings      2.150   1.402   1.138   0.996   0.700
## Proportion Var   0.239   0.156   0.126   0.111   0.078
## Cumulative Var   0.239   0.395   0.521   0.632   0.710

# Dropping off some of the loadings below a certain level for easier interpre
tation.
print(cancernew.fa$loadings, cut = 0.25)

##
## Loadings:
##             Factor1 Factor2 Factor3 Factor4 Factor5
```

```
## Age                                          0.337
## BMI                  0.620
## Glucose      0.502                           0.593
## Insulin      0.986
## HOMA         0.931                           0.299
## Leptin               0.955
## Adiponectin                         0.947
## Resistin                    0.372
## MCP.1                       0.920
##
##                Factor1 Factor2 Factor3 Factor4 Factor5
## SS loadings      2.150   1.402   1.138   0.996   0.700
## Proportion Var   0.239   0.156   0.126   0.111   0.078
## Cumulative Var   0.239   0.395   0.521   0.632   0.710
```

Factor1: Insulin resistance indicator Factor 2: Body fatness Factor 3: Cell growth and physiology regulator Factor 4: Protein and fat regulator Factpor5: Aging factor

Adding rotation to the factor analysis

```
# Factor analysis without rotation
faNR <- factanal(cancernew, factors = 4, rotation = "none")
faLNR <- faNR$loadings[,1:4] #this is loading matrix
faLNR
```

```
##             Factor1 Factor2  Factor3  Factor4
## Age           0.091  0.0781 -0.00206  0.30630
## BMI           0.303 -0.2145  0.92263  0.00101
## Glucose       0.629  0.2861  0.00969  0.51529
## Insulin       0.760  0.5634  0.03949 -0.26065
## HOMA          0.840  0.5366 -0.02688  0.02725
## Leptin        0.293  0.1716  0.56088  0.15266
## Adiponectin  -0.176  0.1626 -0.23180 -0.11093
## Resistin      0.372 -0.1575  0.05489  0.19752
## MCP.1         0.729 -0.6643 -0.15074 -0.00427
```

```
#getting variance for f1, f2, f3, f4 and f5
varLNR = var(faLNR[,1]^2) + var(faLNR[,2]^2) + var(faLNR[,3]^2)+var(faLNR[,4]
^2)
varLNR
```

```
## [1] 0.187
```

```
# Factor analysis with rotation (by default, the varimax rotation)
faR <- factanal(cancernew, factors = 4)
faLR <- faR$loadings[,1:4]
faLR
```

```
##             Factor1 Factor2 Factor3 Factor4
## Age          0.0229  0.0293  0.0154  0.3265
## BMI          0.0336  0.9792  0.1557 -0.0693
```

```
## Glucose      0.4754  0.1380  0.2190  0.6714
## Insulin      0.9738  0.1066  0.0620  0.0211
## HOMA         0.9315  0.0813  0.1600  0.3083
## Leptin       0.2434  0.5948 -0.0488  0.1943
## Adiponectin  0.0219 -0.2851 -0.1880 -0.0799
## Resistin     0.1040  0.1568  0.3525  0.2130
## MCP.1        0.1092  0.0665  0.9891 -0.0190
```

```
print(faR$loadings[,1:4], cut = 0.3)
```

```
##             Factor1 Factor2 Factor3 Factor4
## Age          0.0229  0.0293  0.0154  0.3265
## BMI          0.0336  0.9792  0.1557 -0.0693
## Glucose      0.4754  0.1380  0.2190  0.6714
## Insulin      0.9738  0.1066  0.0620  0.0211
## HOMA         0.9315  0.0813  0.1600  0.3083
## Leptin       0.2434  0.5948 -0.0488  0.1943
## Adiponectin  0.0219 -0.2851 -0.1880 -0.0799
## Resistin     0.1040  0.1568  0.3525  0.2130
## MCP.1        0.1092  0.0665  0.9891 -0.0190
```

```
# Varimax rotation maximizes the sum squared variance of loadings.
varLR = var(faLR[,1]^2) + var(faLR[,2]^2) + var(faLR[,3]^2)+var(faLNR[,4]^2)
varLR
```

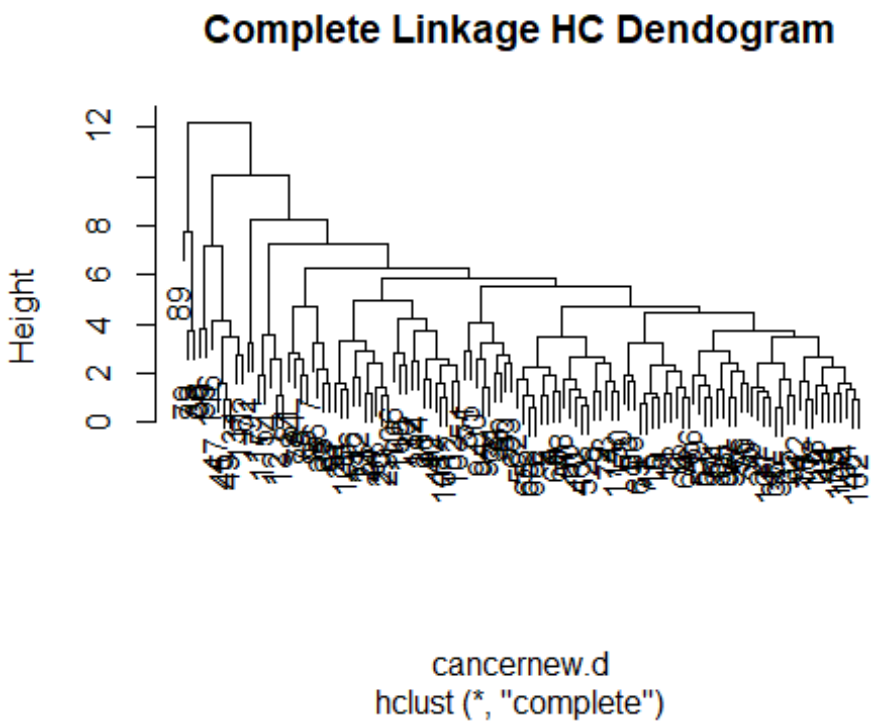
```
## [1] 0.36
```

```
#The varimax rotation of 0.376 is bigger than the one without rotation of 0.3
04 because it is maximized.
```

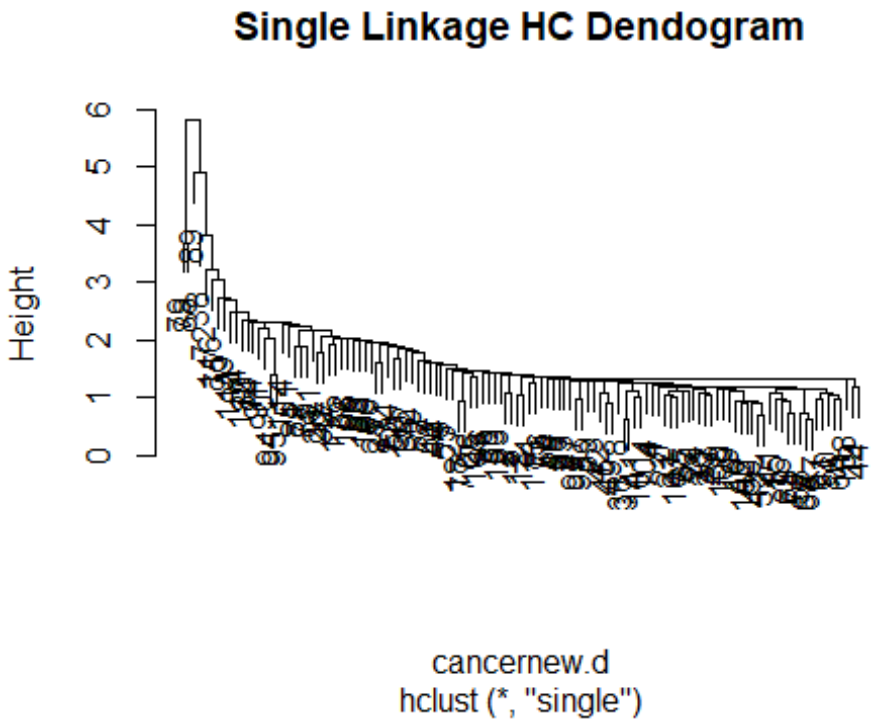
The factor loadings are the approximate correlations of the manifest variables and the factors. This shows a lot of correlation between the principal component and the factors.

Hierachial Cluster Analysis

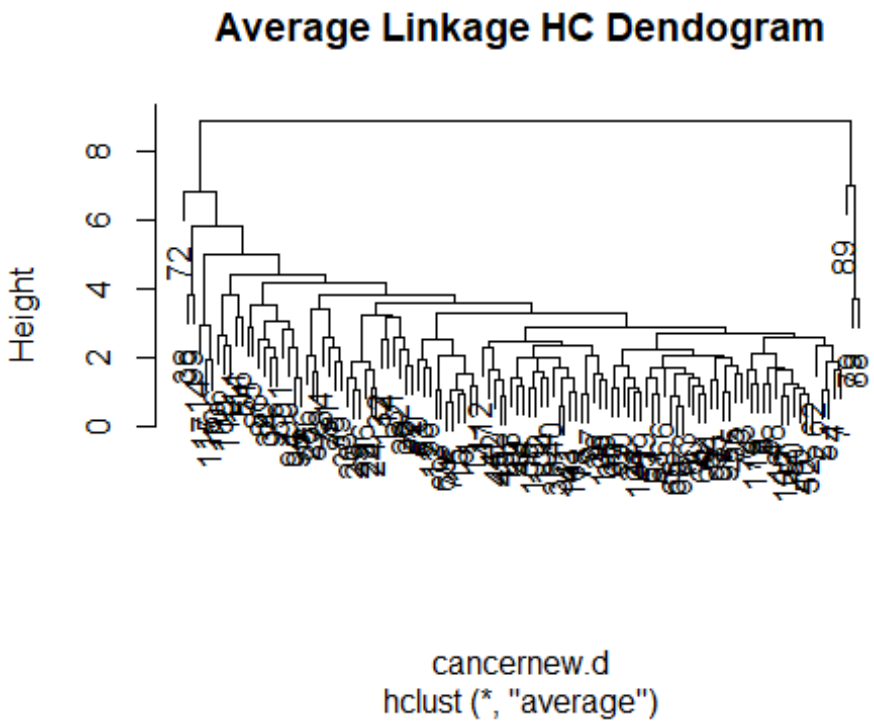
```
cancernew.s = scale(cancernew)
cancernew.d = dist(cancernew.s)

#Using complete linkage
hc1 <- hclust(cancernew.d, "complete")
plot(hc1, main = "Complete Linkage HC Dendogram")
```




```r
#Using single linkage
hc2 <- hclust(cancernew.d, "single")
plot(hc2, main = "Single Linkage HC Dendogram")
```



```r
#Using average linkage
hc3 <- hclust(cancernew.d, "average")
plot(hc3, main = "Average Linkage HC Dendogram")
```



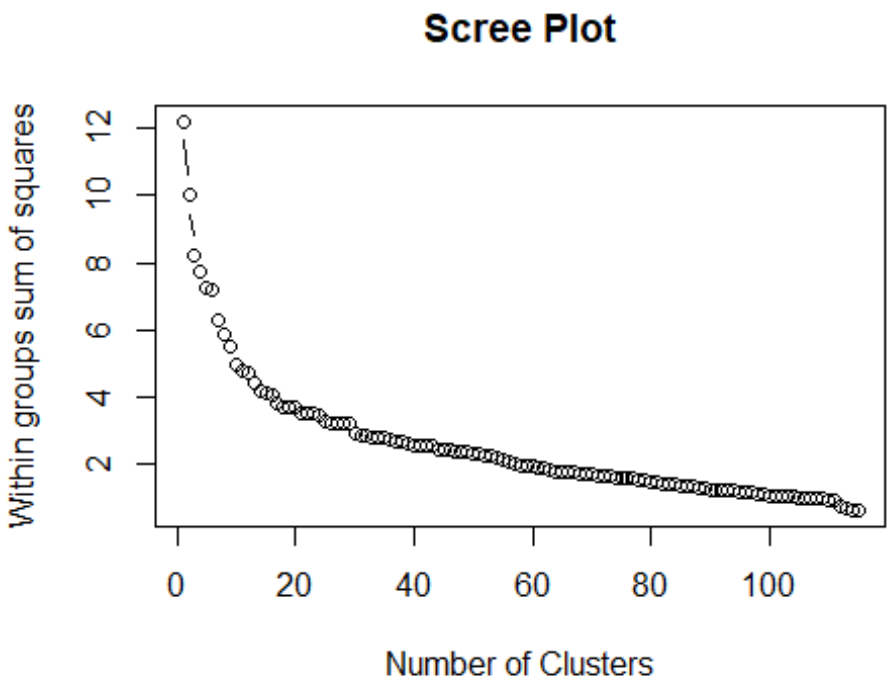

```r
#The "complete linkage" looks like the better linkage and it is showing about
5 possible grouping.

## checking the height of each observation in the dendrogram.
hc1$height
```

```
##   [1]  0.604  0.619  0.673  0.712  0.887  0.890  0.939  0.957  0.960  0.96
8
##  [11]  0.980  0.995  0.997  1.006  1.039  1.043  1.058  1.104  1.117  1.13
4
##  [21]  1.140  1.186  1.187  1.197  1.205  1.224  1.243  1.256  1.334  1.33
9
##  [31]  1.352  1.363  1.375  1.395  1.427  1.445  1.488  1.525  1.543  1.55
5
##  [41]  1.582  1.582  1.606  1.610  1.636  1.673  1.683  1.693  1.740  1.76
8
##  [51]  1.769  1.777  1.810  1.881  1.883  1.912  1.912  1.934  1.973  2.03
0
##  [61]  2.092  2.168  2.203  2.248  2.312  2.314  2.349  2.361  2.374  2.39
6
##  [71]  2.422  2.439  2.515  2.532  2.538  2.550  2.567  2.628  2.678  2.71
1
##  [81]  2.750  2.782  2.789  2.853  2.857  2.869  3.208  3.212  3.217  3.22
```

```
3
##  [91]  3.269  3.442  3.474  3.492  3.518  3.683  3.689  3.697  3.805  4.05
8
## [101]  4.124  4.193  4.418  4.733  4.744  4.944  5.496  5.878  6.255  7.16
7
## [111]  7.272  7.736  8.242 10.030 12.211
```

Determining the number of clusters

```
# hc1 is the hierarchical clustering object of the cancer data. hc1 contains
following information.
names(hc1)

## [1] "merge"       "height"      "order"       "labels"      "method"
## [6] "call"        "dist.method"

plot((rev(hc1$height)),xlab="Number of Clusters",
  ylab="Within groups sum of squares",type = "b", main="Scree Plot")
```




```
# Using the drop up point, the number of clusters is 4
```

The number of clusters will be 4.

```
# Getting the 4 clusters solution?
cancernew.ct <- cutree(hc1, 4)
cancernew.ct
```

```
##   [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1
1 1 1
##  [38] 2 1 1 1 1 1 1 2 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
3 1 1
##  [75] 1 1 1 1 4 1 1 1 1 1 1 1 1 4 4 1 3 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1
1 1 2
## [112] 1 1 1 1 2
```

```
cancernew.clust <- data.frame(cancernew.ct)
cancernew.clust
```

```
##     cancernew.ct
## 1              1
## 2              1
## 3              1
## 4              1
## 5              1
## 6              1
## 7              1
## 8              1
## 9              1
## 10             1
## 11             1
## 12             1
## 13             1
## 14             1
## 15             1
## 16             1
## 17             1
## 18             1
## 19             1
## 20             1
## 21             1
## 22             1
## 23             1
## 24             1
## 25             1
## 26             1
## 27             1
## 28             1
## 29             1
## 30             1
## 31             2
## 32             1
## 33             1
## 34             1
## 35             1
## 36             1
## 37             1
## 38             2
```

```
## 39              1
## 40              1
## 41              1
## 42              1
## 43              1
## 44              1
## 45              2
## 46              1
## 47              2
## 48              1
## 49              2
## 50              1
## 51              1
## 52              1
## 53              1
## 54              1
## 55              1
## 56              1
## 57              1
## 58              1
## 59              1
## 60              1
## 61              1
## 62              1
## 63              1
## 64              1
## 65              1
## 66              1
## 67              1
## 68              1
## 69              1
## 70              1
## 71              1
## 72              3
## 73              1
## 74              1
## 75              1
## 76              1
## 77              1
## 78              1
## 79              4
## 80              1
## 81              1
## 82              1
## 83              1
## 84              1
## 85              1
## 86              1
## 87              1
## 88              4
```

```
## 89             4
## 90             1
## 91             3
## 92             1
## 93             1
## 94             1
## 95             1
## 96             1
## 97             1
## 98             1
## 99             2
## 100            1
## 101            1
## 102            1
## 103            1
## 104            1
## 105            1
## 106            1
## 107            1
## 108            1
## 109            1
## 110            1
## 111            2
## 112            1
## 113            1
## 114            1
## 115            1
## 116            2
```

```
#Above tells which cluster is in which row.

#summarizing the clustering information using the table of counts.
table(cancernew.ct) #This gives how many items in each cluster
```

```
## cancernew.ct
##   1   2   3   4
## 103   8   2   3
```

```
# finding the content of each group
cancer1.s = scale(cancernew)

# Then by looking at the average z-score value of the group data, we can find
a meaning for that group.
cluster1 = subset(rownames(cancernew), cancernew.ct==1)
index1 = match(cluster1, rownames(cancernew))
colMeans(cancer1.s[index1, ])
```

```
##         Age         BMI     Glucose     Insulin        HOMA      Leptin
##     -0.0718     -0.0589     -0.1753     -0.1679     -0.1911     -0.2018
## Adiponectin    Resistin       MCP.1
##      0.0489     -0.1105     -0.0203
```

```
cluster2 = subset(rownames(cancernew),cancernew.ct==2)
index2 = match(cluster2, rownames(cancernew))
colMeans(cancer1.s[index2, ])
```

```
##         Age         BMI     Glucose     Insulin        HOMA      Leptin
##       0.796       0.677       0.364       0.226       0.171       2.255
## Adiponectin    Resistin       MCP.1
##      -0.423       0.591      -0.709
```

```
cluster3 = subset(rownames(cancernew), cancernew.ct==3)
index3 = match(cluster3, rownames(cancernew))
colMeans(cancer1.s[index3, ])
```

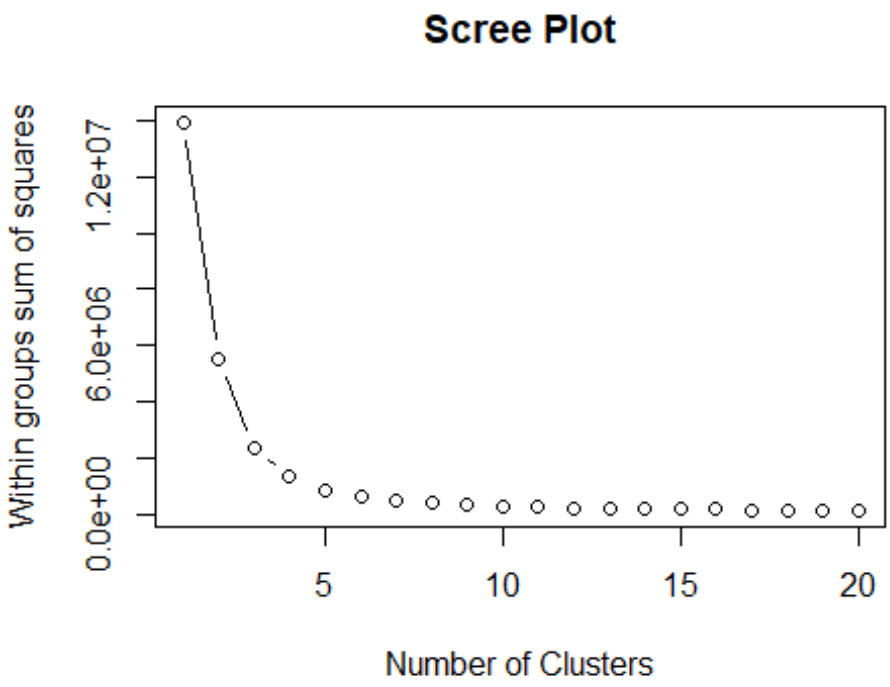
```
##         Age         BMI     Glucose     Insulin        HOMA      Leptin
##     -0.9497      0.0411      0.8527      3.9895      3.1744     -0.1051
## Adiponectin    Resistin       MCP.1
##      0.1388     -0.2658     -0.2693
```

```
cluster4 = subset(rownames(cancernew), cancernew.ct==4)
index4 = match(cluster4, rownames(cancernew))
colMeans(cancernew.s[index4, ])
```

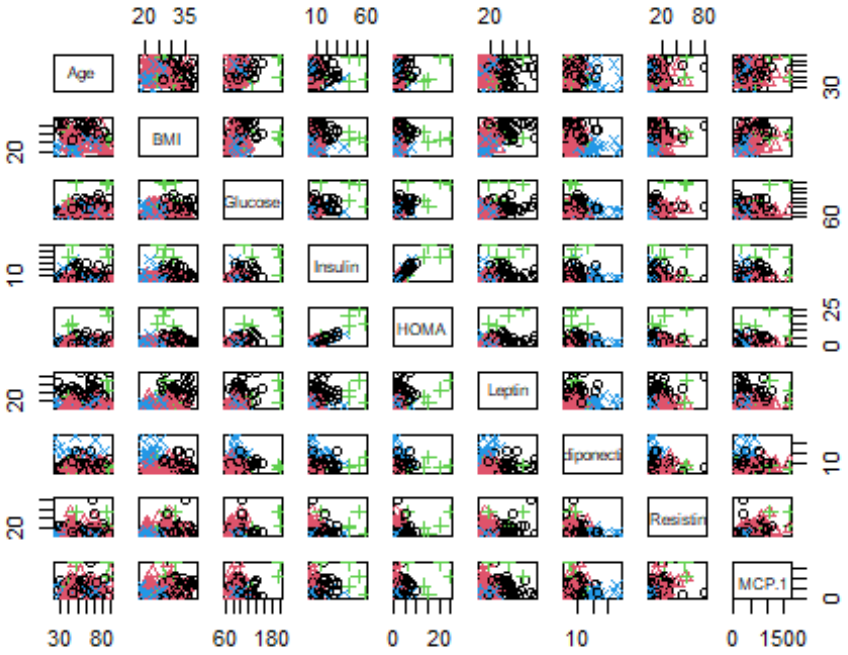
```
##         Age         BMI     Glucose     Insulin        HOMA      Leptin
##       0.974       0.190       4.478       2.501       3.987       0.987
## Adiponectin    Resistin       MCP.1
##      -0.642       2.395       2.767
```

#Checking the clustering using K-Means Cluster**

```
set.seed(123)
#Identifying the appropriate number of cluster for the k-Means

plot.wgss = function(cancernew, maxc) {
  wss = numeric(maxc)
  for (i in 1:maxc)
    wss[i] = kmeans(cancernew,centers=i, nstart = 10)$tot.withinss
  plot(1:maxc, wss, type="b", xlab="Number of Clusters",
  ylab="Within groups sum of squares", main="Scree Plot")
}

#plotting the scree plot
plot.wgss(cancernew, 20)
```




```
#from the K-Means scree plot, the number of clusters is 4

km <- kmeans(cancernew.s, centers = 4, nstart = 10)


table(km$cluster, cancer$Classification)

##
##      1  2
##   1 18 22
##   2 26 27
##   3  0  5
##   4  8 10

# We see that Mclust is more accurate.
plot(cancernew, col = km$cluster, pch = km$cluster)
```




```
km.cluster1 = subset(cancernew.s, km$cluster == 1)
km.cluster1
```

```
##           Age     BMI Glucose Insulin    HOMA   Leptin Adiponectin Resisti
n
##  [1,]  0.2295  0.8878 -0.5679  0.8011  0.3007  0.21674     -0.3508  -0.084
1
##  [2,] -1.5703  1.7789 -0.4792  0.3987  0.0865  1.18632     -0.7425   0.191
8
##  [3,] -1.7565  0.9340 -0.6123 -0.4174 -0.4094  0.99068     -0.5803   0.797
2
##  [4,] -1.3841  0.5368 -0.6123 -0.5598 -0.4910  0.65673     -0.1655   0.138
1
##  [5,] -0.7635  1.8831 -0.6567 -0.3230 -0.3600  0.69671     -0.8130  -0.486
0
##  [6,] -0.4532  2.1905  0.3643 -0.3287 -0.2587  1.04388     -0.8056  -0.237
4
##  [7,]  0.5398  0.7698 -0.3460 -0.0762 -0.1764  1.00855      0.0255   0.698
6
##  [8,] -1.3841  1.5276 -0.3460 -0.3174 -0.3244  1.25079     -0.4697   0.590
0
##  [9,]  0.5398  1.7191  0.1424  0.5484  0.3226  2.50696     -0.3860   0.612
9
## [10,] -0.2670  1.8342  0.1424  0.0162 -0.0439  0.02966      1.4392  -0.360
2
## [11,] -1.8185  1.6481 -0.4792 -0.1426 -0.2346  2.18394     -0.7871   0.542
2
```

```
## [12,] -0.8876  1.3625 -0.3904  1.3093  0.6581  0.23967     -0.2747  -0.646
9
## [13,]  0.7260  1.4961  0.1424 -0.4337 -0.3538  2.96440     -0.4945   5.437
5
## [14,]  1.1605 -0.0761 -0.1684  0.4031  0.1558  0.48354     -0.1219  -0.509
3
## [15,]  0.8502  0.5414  0.1868 -0.1661 -0.1638  1.55797     -0.2997  -0.841
5
## [16,]  1.0984 -0.3749 -0.1684 -0.1920 -0.2256  2.04923     -0.9410  -0.825
6
## [17,]  0.7260  0.3621 -0.3904  0.0687 -0.0948  0.97256     -0.2768  -0.822
9
## [18,]  1.1605 -0.0960  0.5419  1.6090  1.2128 -0.25215     -0.7665  -0.503
0
## [19,] -0.1429  0.8751 -0.2572  0.6578  0.2966  0.55300      0.1224  -0.610
2
## [20,] -0.8876  0.7306  0.2312 -0.5646 -0.4380 -0.04345      0.3709   1.931
1
## [21,] -1.0117 -0.1811 -0.0352  1.1940  0.7075  0.94305      0.4843   1.057
8
## [22,]  1.4708  0.8084  0.0980 -0.0341 -0.0851  0.63553      0.0666   1.197
1
## [23,] -0.5773  0.9720  0.0536  1.8539  1.1828  1.01447      1.6643  -0.368
7
## [24,]  0.4778  0.4152 -0.5679  0.4606  0.1034 -0.00513     -0.4235   0.382
3
## [25,]  0.0433  0.3132  1.8294  0.6526  0.8211 -0.19427      0.0120  -0.060
7
## [26,]  1.5328  0.7242  0.0980  0.8011  0.4843  0.26222     -0.0376   0.421
4
## [27,] -0.3290  0.6412 -0.4792  2.0064  1.0402  0.13860     -0.5717   0.768
3
## [28,] -0.5152  0.9720  1.6074  1.4775  1.5187  0.82240      0.0896  -0.723
0
## [29,]  0.1675  0.7269  1.4742  1.9982  1.9333  0.58530     -0.2596  -0.260
4
## [30,] -0.5152  0.4374 -1.2339 -0.1605 -0.3419  1.28881      0.0805   0.487
6
## [31,]  0.8502  0.0664  0.2756  0.8133  0.5420  1.40146     -1.2457   2.785
6
## [32,]  0.4778  0.6640 -0.0352  0.0476 -0.0508  0.90738     -0.9456   0.463
5
## [33,] -0.0187  1.4454 -0.1240  0.2519  0.0674  0.34124     -1.1421  -0.385
1
## [34,]  0.9743  1.8978  1.6074 -0.4347 -0.2285  0.77105     -1.0003  -0.632
2
## [35,]  0.6640  1.5892  1.4742 -0.1850 -0.0169 -0.45587      0.2512  -0.850
3
## [36,]  1.0984  0.5773  2.4065 -0.2982 -0.0183  1.24665     -0.0177  -0.241
8
```

```
## [37,] -0.2049  1.6868  0.9415  0.1885  0.2199  3.26612     -0.3172  -0.780
1
## [38,] -0.7635 -0.1458 -0.2572 -0.6637 -0.5325  1.46299      0.2804  -0.303
9
## [39,]  0.4778  0.8900 -0.0352 -0.4253 -0.3635  1.81746      1.8060  -0.354
8
## [40,]  1.7811 -0.0801  1.7850  0.9831  1.1209  3.31877      0.5742  -0.837
4
##        MCP.1
##  [1,] -0.261
##  [2,]  0.156
##  [3,]  1.071
##  [4,]  0.576
##  [5,]  0.149
##  [6,]  1.019
##  [7,]  1.640
##  [8,]  0.385
##  [9,]  0.955
## [10,]  0.466
## [11,] -0.509
## [12,]  1.230
## [13,] -0.784
## [14,] -0.455
## [15,] -0.965
## [16,] -0.948
## [17,] -0.922
## [18,] -1.413
## [19,] -0.148
## [20,]  0.696
## [21,]  0.720
## [22,] -0.314
## [23,]  0.588
## [24,]  3.364
## [25,]  1.125
## [26,]  1.329
## [27,]  0.665
## [28,]  0.352
## [29,] -0.401
## [30,]  0.196
## [31,] -0.806
## [32,] -0.399
## [33,]  0.350
## [34,]  0.735
## [35,] -0.972
## [36,] -1.258
## [37,] -0.915
## [38,] -0.770
## [39,] -0.638
## [40,] -1.285
```

```
km.cluster2 = subset(cancernew.s, km$cluster == 2)
km.cluster2
```

```
##          Age     BMI  Glucose Insulin    HOMA  Leptin Adiponectin Resistin
##  [1,] -0.577 -0.8131 -1.23387 -0.7256 -0.6116 -0.9283    -0.06992 -0.54316
##  [2,]  1.595 -1.3728 -0.25718 -0.6851 -0.5459 -0.9264    -0.69434 -0.86048
##  [3,]  0.664 -1.2379 -0.92311 -0.6740 -0.5717 -0.8722    -0.44004 -0.15818
##  [4,]  1.781 -1.2890 -0.25718 -0.6420 -0.5188 -1.0382    -0.78348 -0.33490
##  [5,] -0.515 -0.9417 -0.25718 -0.6740 -0.5390 -1.0313     0.51128 -0.35578
##  [6,]  1.967 -0.9725 -0.92311 -0.5286 -0.4954 -1.0244    -0.67087 -0.14445
##  [7,]  1.160 -0.7534  0.89708 -0.3518 -0.2229 -1.1627     0.44868 -0.77653
##  [8,]  0.974 -1.1119 -0.03521 -0.6617 -0.5199 -1.1544     0.02599 -0.68128
##  [9,]  1.098 -0.9127 -0.65674 -0.5026 -0.4616 -0.4946     0.20430 -0.61616
## [10,]  0.416  1.3839 -0.12400 -0.5547 -0.4551 -0.2817    -0.68947 -0.64759
## [11,] -1.322  0.1981 -0.52355 -0.5629 -0.4869 -0.5990    -0.23101 -0.44970
## [12,] -1.446  0.8751 -0.47916 -0.5445 -0.4730  0.1113    -0.37089 -0.73443
## [13,] -0.205  0.5779 -0.34597 -0.4445 -0.4025 -0.7446    -0.06568 -0.36584
## [14,] -1.322  1.3133 -0.78992 -0.3399 -0.3829 -0.8515    -0.74744  0.08038
## [15,] -0.391  0.0212 -0.92311 -0.6116 -0.5389 -0.3400    -1.02125 -0.35113
## [16,]  0.602  0.4033 -0.83432 -0.4165 -0.4286 -0.2456    -1.16706 -0.84890
## [17,]  0.540  0.7284 -0.70113 -0.5792 -0.5078 -0.5416    -0.86416 -0.92281
## [18,]  0.167 -0.2456  0.23116 -0.4841 -0.3815 -0.1207    -1.16706  0.44611
## [19,]  1.223  1.5947 -0.96750 -0.6090 -0.5402 -0.2517    -0.30034  0.20463
## [20,]  1.160  0.3259 -0.65674 -0.4605 -0.4378  0.1015    -0.41075 -0.53930
## [21,]  1.098 -0.0562 -0.56795 -0.4783 -0.4408 -0.8458    -0.17244 -0.57699
## [22,]  0.726  0.9796 -0.21279 -0.4551 -0.3979 -0.5979     0.23484 -0.23711
## [23,]  0.540  0.0235 -0.34597 -0.3943 -0.3717 -0.0922    -0.36953 -0.64732
## [24,]  1.285 -0.4546 -1.67782 -0.6460 -0.5974 -1.0416     0.05647 -0.81208
## [25,]  1.719 -0.1956 -0.07960 -0.5513 -0.4499 -0.9782    -0.32867 -0.41261
## [26,]  1.223 -0.3351 -0.56795 -0.5396 -0.4763 -0.6712    -0.06249 -0.23824
## [27,] -0.763 -1.3450 -1.05629 -0.5415 -0.5114 -0.9833    -0.28399  1.07390
## [28,] -0.515 -1.3198 -0.16839  0.2277  0.0434 -0.8015    -0.25846  0.67726
## [29,] -0.950 -1.2394 -0.21279 -0.6966 -0.5511 -0.3927    -0.25104  0.21389
## [30,]  0.664 -1.2946  0.18676 -0.3786 -0.3117 -0.8818    -0.23471 -0.07938
## [31,] -0.391 -1.6831 -0.21279 -0.5610 -0.4651 -0.8097    -0.63905 -0.73890
## [32,]  0.292 -0.9812 -0.25718 -0.6486 -0.5230 -0.8732     0.15422 -0.32528
## [33,] -1.198 -1.0124 -0.12400 -0.4719 -0.4015 -0.9475    -0.79040  0.08152
## [34,]  0.726 -1.2088  0.63071 -0.3307 -0.2330  0.3109    -0.88303  0.07851
## [35,] -0.515 -1.2379 -0.87871 -0.7322 -0.6005 -1.0572    -0.91983  0.66313
## [36,] -0.391 -0.9341  0.23116 -0.7223 -0.5488 -0.9695    -0.12145 -0.25592
## [37,]  0.105 -0.9460  0.00919 -0.3129 -0.2845 -0.6105    -0.86957 -0.52629
## [38,] -0.763 -0.8848  0.80829 -0.5076 -0.3548 -0.4492    -0.86013 -0.76369
## [39,] -0.205 -0.6700 -0.52355 -0.6240 -0.5227 -0.9345    -0.94627 -0.35361
## [40,]  0.416 -1.0677  0.00919 -0.4283 -0.3616 -0.7519    -0.78863 -0.06566
## [41,] -0.888 -0.2031  0.14237  0.0539 -0.0180 -0.8765    -0.54952  0.11089
## [42,]  0.105  0.2172 -0.92311 -0.6778 -0.5737 -0.5001     0.91470  1.36913
## [43,]  0.850 -0.4127  0.63071  0.0380  0.0485 -0.3936    -0.68604  2.26128
## [44,] -0.950  0.3416  0.00919 -0.5801 -0.4631 -0.7482    -0.50930  3.14316
## [45,] -0.577  0.1081 -0.34597 -0.7422 -0.5851 -0.5777     0.00605  0.11173
## [46,] -0.826  0.0609  0.05358 -0.0799 -0.1226 -0.7266    -0.68719  0.67043
```

```
## [47,] -1.074  0.0107  0.23116 -0.7529 -0.5703 -0.6408    -0.49640  0.91098
## [48,]  0.726  0.1718  0.45313 -0.1196 -0.0957 -0.6186    -0.71498  0.14197
## [49,]  1.036  0.2127 -0.43476 -0.6953 -0.5604  0.2350    -0.36951  0.29294
## [50,]  0.540 -0.2031 -0.39037 -0.3465 -0.3467 -0.6103    -0.25586  0.01560
## [51,]  0.912  0.3096 -0.65674  0.0931 -0.1245  0.0101    -1.08075  0.00353
## [52,] -0.763  0.3591 -0.34597 -0.5263 -0.4527 -0.1442    -0.51680  0.06701
## [53,] -0.701  1.1151 -0.25718 -0.4233 -0.3817 -0.4131    -0.14918 -0.47100
##          MCP.1
##  [1,] -0.33978
##  [2,] -0.19040
##  [3,]  1.13778
##  [4,]  0.69172
##  [5,] -0.01225
##  [6,]  2.08560
##  [7,] -0.73415
##  [8,] -1.14998
##  [9,] -0.62543
## [10,] -0.81581
## [11,] -0.00122
## [12,]  0.11025
## [13,]  2.00416
## [14,]  0.13491
## [15,] -0.17573
## [16,]  0.14645
## [17,]  0.28896
## [18,] -0.44997
## [19,]  0.24175
## [20,]  0.47452
## [21,] -0.57602
## [22,] -0.76466
## [23,] -0.89262
## [24,] -0.93925
## [25,] -0.87491
## [26,] -0.13246
## [27,] -0.43853
## [28,]  0.11270
## [29,] -0.61498
## [30,] -0.24818
## [31,] -1.28370
## [32,]  0.48951
## [33,] -0.97016
## [34,]  0.51629
## [35,]  0.58693
## [36,] -0.50711
## [37,] -0.51845
## [38,] -0.04643
## [39,]  0.29025
## [40,] -0.40088
## [41,]  0.78655
## [42,]  1.08652
```

```
## [43,]  0.76681
## [44,]  1.46625
## [45,]  3.36441
## [46,] -0.36842
## [47,] -0.69822
## [48,] -0.52348
## [49,]  0.10914
## [50,] -0.76655
## [51,] -0.87487
## [52,]  0.25043
## [53,] -0.94086
```

```
km.cluster3 = subset(cancernew.s, km$cluster == 3)
km.cluster3
```

```
##         Age     BMI Glucose Insulin  HOMA Leptin Adiponectin Resistin  MCP
.1
## [1,] -0.826 -0.5661   0.364   4.812 3.457 -0.441       0.865   -0.760 -0.8
38
## [2,]  1.781 -0.1824   4.582   3.139 4.925  1.096      -0.705    0.778  3.3
64
## [3,]  1.719  0.0212   4.360   4.152 6.138  2.308      -0.333    3.268  1.5
72
## [4,] -0.577  0.7306   4.493   0.214 0.899 -0.442      -0.888    3.140  3.3
64
## [5,] -1.074  0.6483   1.341   3.167 2.892  0.231      -0.587    0.228  0.3
00
```

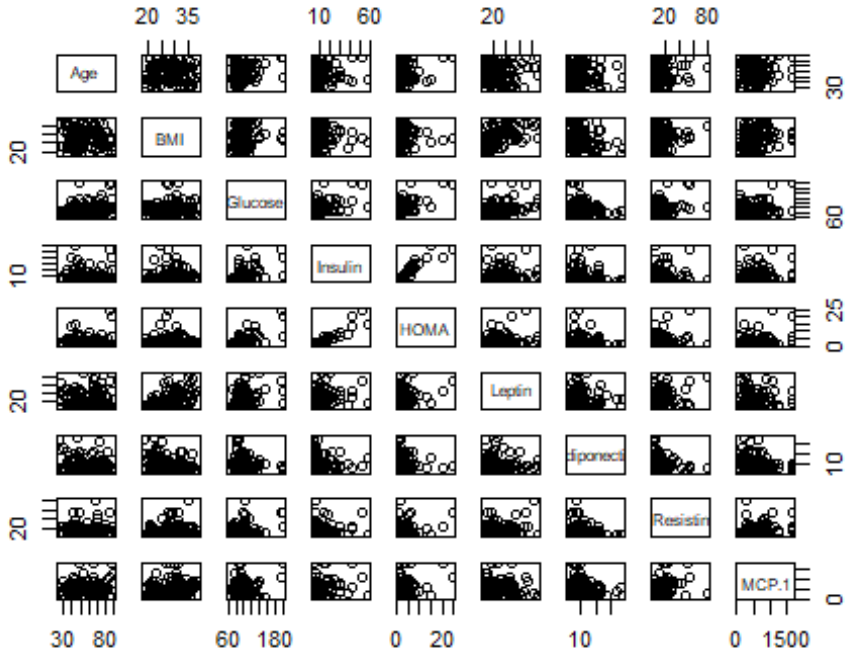
```
km.cluster4 = subset(cancernew.s, km$cluster == 4)
km.cluster4
```

```
##          Age    BMI Glucose Insulin   HOMA   Leptin Adiponectin Resistin
##  [1,]  1.533 -0.888  -0.302  -0.548 -0.463 -0.45226      1.7902  -0.4398
##  [2,] -1.446 -1.218  -0.879  -0.650 -0.557 -0.62789      0.4280  -0.6300
##  [3,] -1.756 -0.911  -0.701  -0.432 -0.425  0.46785      2.4168  -0.8188
##  [4,] -2.005 -0.941  -0.701  -0.588 -0.513 -0.32138      1.9711  -0.7736
##  [5,] -2.067 -1.775  -0.435  -0.388 -0.375 -0.92451      3.7817  -0.6356
##  [6,] -1.198 -0.845  -1.012  -0.420 -0.446 -0.59193      1.1353  -0.4339
##  [7,] -0.826 -1.359  -0.524  -0.244 -0.301 -0.65292      1.4816  -0.5719
##  [8,] -0.639 -1.106  -0.612  -0.710 -0.578  0.00182      4.0710  -0.9205
##  [9,] -0.763 -1.251   0.187   0.381  0.217 -0.98875      1.5892   0.6705
## [10,] -1.446 -0.665  -0.257   1.161  0.612 -0.51502      1.7013  -0.2147
## [11,] -0.701 -1.345  -0.435  -0.655 -0.536 -0.71651      1.2230  -0.0941
## [12,] -0.826 -1.598   0.720   0.584  0.487 -0.70557      1.4889  -0.8156
## [13,] -0.763 -1.459  -0.257  -0.653 -0.526 -0.98862      0.9482  -0.5557
## [14,] -0.391 -1.835   0.320  -0.396 -0.311 -0.88593      0.3769  -0.9294
## [15,]  0.912 -0.789   0.320  -0.555 -0.426 -0.25205      1.1221  -0.7995
## [16,] -0.701 -1.070  -0.524   2.675  1.412 -0.85778     -0.0615  -0.7301
## [17,]  0.292 -0.148   0.098  -0.545 -0.433 -0.73841      1.6423  -0.5977
## [18,]  0.912 -0.397  -0.701  -0.714 -0.583 -0.08628      3.4441  -0.9246
##          MCP.1
```

```
##  [1,]  0.0580
##  [2,] -0.5205
##  [3,] -1.0403
##  [4,] -0.6386
##  [5,]  0.2821
##  [6,] -1.0686
##  [7,] -1.3617
##  [8,] -0.9914
##  [9,]  0.0514
## [10,] -0.1523
## [11,] -0.6748
## [12,] -0.9077
## [13,] -0.9852
## [14,] -0.0607
## [15,] -0.9792
## [16,] -0.6437
## [17,] -0.5912
## [18,] -0.4110
```

Using Model based Cluster Analysis

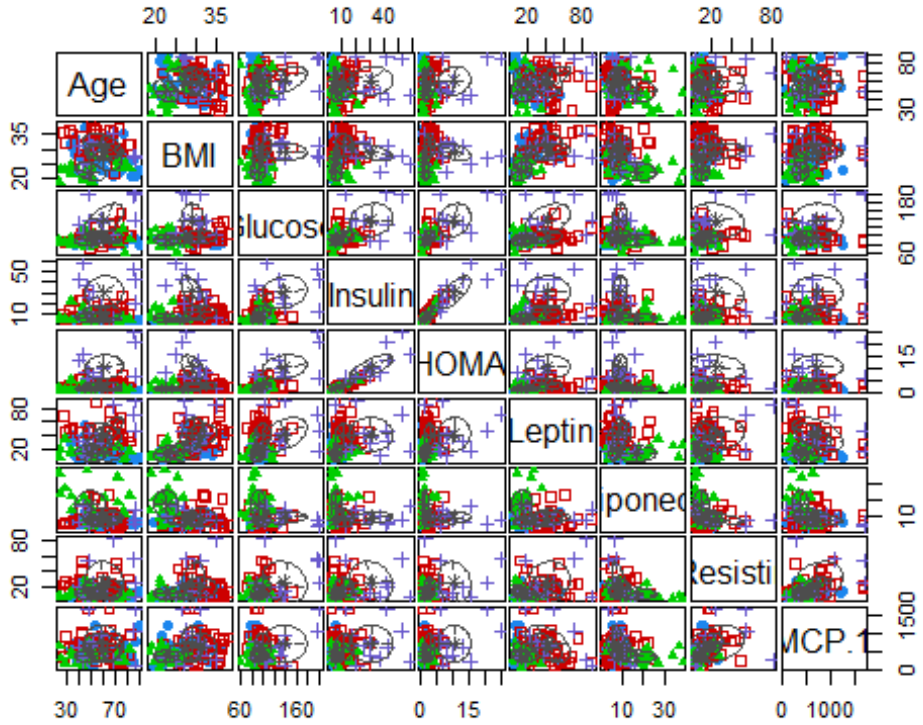
```
plot(cancernew) #the original plot does not show the clusters clearly
```



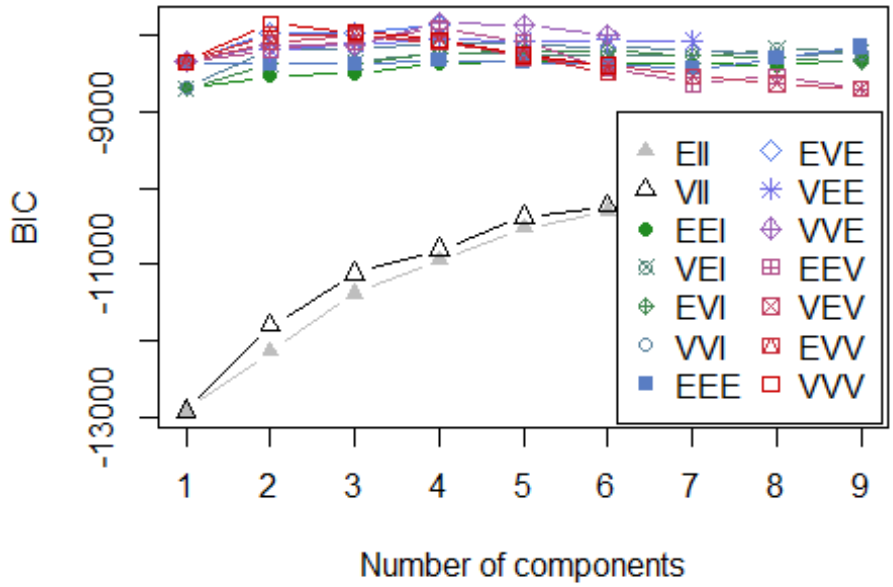

```
library(mclust)
```

```
## Warning: package 'mclust' was built under R version 4.0.3
```

```
## Package 'mclust' version 5.4.6
## Type 'citation("mclust")' for citing this R package in publications.

##
## Attaching package: 'mclust'

## The following object is masked from 'package:faraway':
##
##     diabetes

# we have expected clusters; male and female. Mclust optimally recognizes tha
t:
mc <- Mclust(cancernew) #number of cluster is not specified. Let the model-ba
sed algorithm choose the number.
#It find it to be 4 clusters
table(mc$classification, cancer$Classification)

##
##      1  2
##   1 23 17
##   2 15 22
##   3 13 14
##   4  1 11

plot(mc, what = "classification") #There is a couple of overlaps, but it is e
vidence of 4 clusters
```



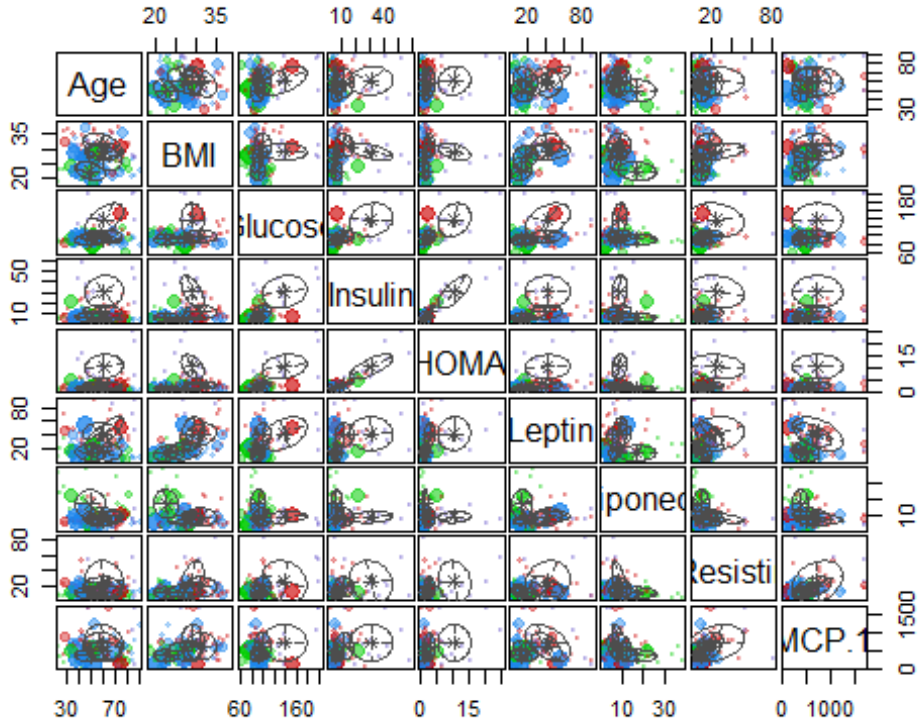
```r
# The number of clusters is based on maximum BIC
plot(mc, what = "BIC") # This shows 4 to have the highest BIC making 4 cluste
r to be the number of clusters needed.
```




```r
mc$modelName  #It is using VVE which assumed ellipsoidal distribution, equal
volme, equal shape and variable orientation.
```

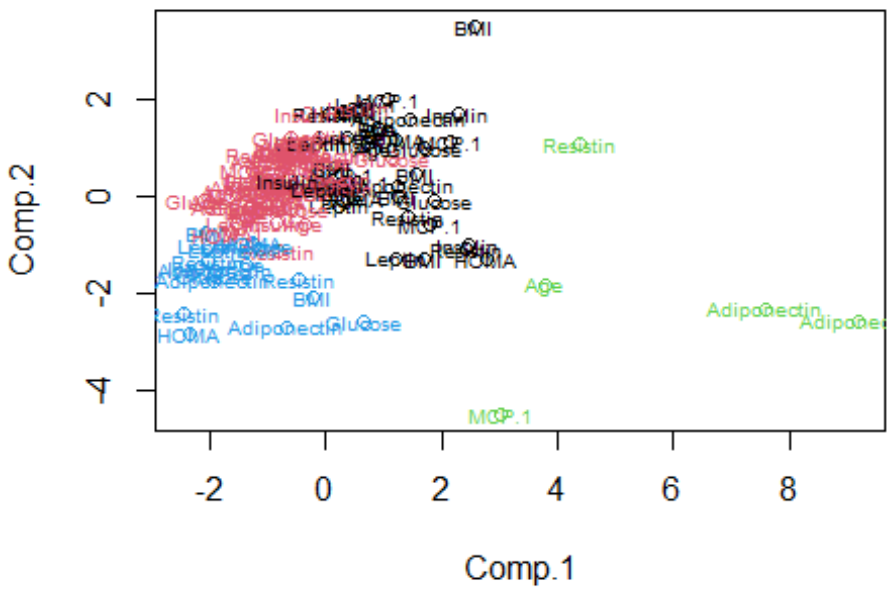
```
## [1] "VVE"
```

```r
# Find uncertain points, darker and larger points are more uncertain

plot(mc, what = "uncertainty")
```

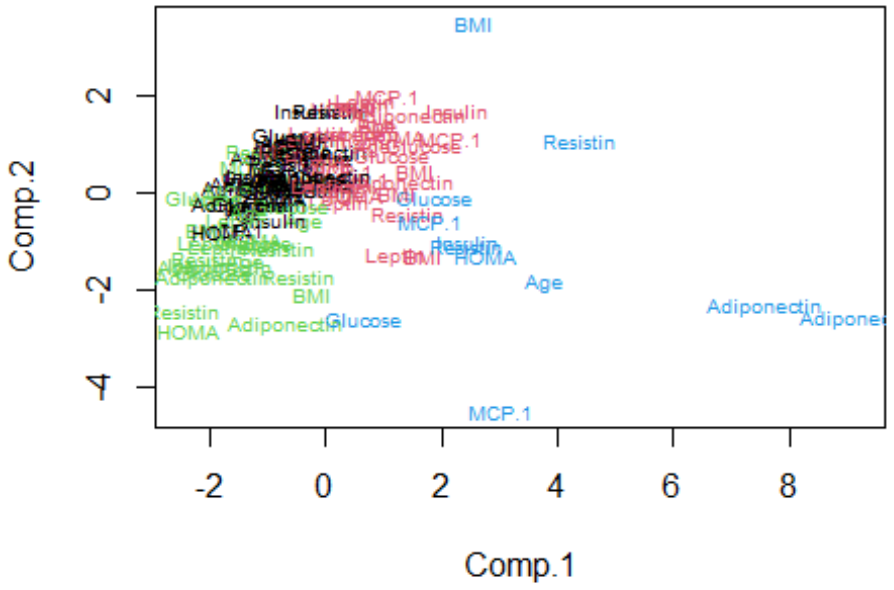


Comparing principal component using kmeans

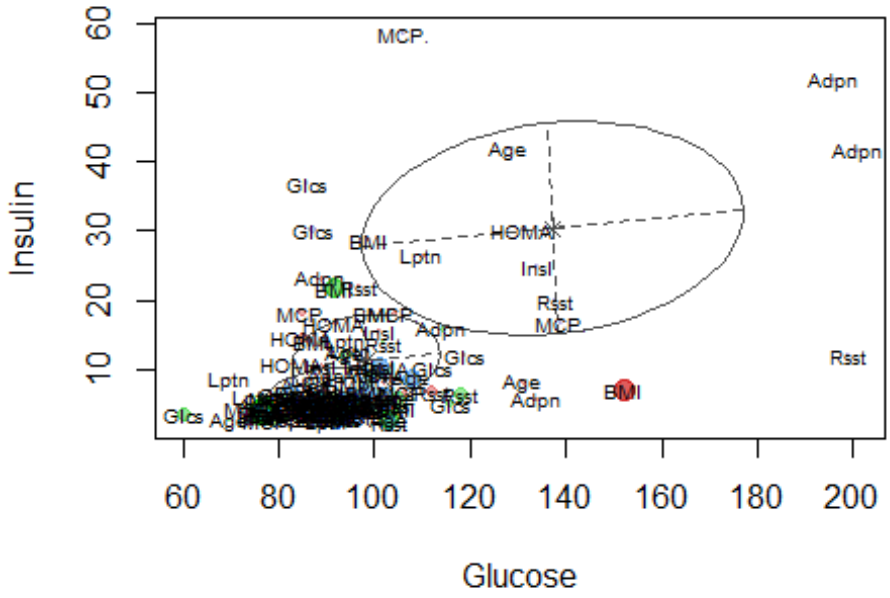
```
set.seed(123)
pc1 <- cancernew.pca$score[,1]

cancernew.pca$loadings[,1:5]

##              Comp.1  Comp.2  Comp.3   Comp.4 Comp.5
## Age           0.125  0.0663  0.2067  0.82139  0.253
## BMI           0.260  0.4993 -0.4257 -0.07092 -0.232
## Glucose       0.439 -0.1859  0.1309  0.12562  0.200
## Insulin       0.444 -0.3863 -0.0937 -0.05977 -0.298
## HOMA          0.493 -0.3747  0.0122 -0.00564 -0.139
## Leptin        0.331  0.2336 -0.5832  0.05835  0.288
## Adiponectin -0.173 -0.4806 -0.2821 -0.27687  0.529
## Resistin      0.282  0.3036  0.2889 -0.30270  0.598
## MCP.1         0.255  0.2104  0.4968 -0.35947 -0.119

plot(cancernew.pca$scores[, 1:2], col = km$cluster)
text(cancernew.pca$scores[, 1:2], labels=(colnames(cancernew)), cex = 0.6, co
l = km$cluster)
```




```
plot(cancernew.pca$scores[, 1:2], col = mc$cluster)
text(cancernew.pca$scores[, 1:2], labels=(colnames(cancernew)), cex = 0.6, co
l = mc$classification)
```



```
plot(mc, dimens = c(3,4),what = "uncertainty")
text(mc$data[,c(3,4)], labels = abbreviate(colnames(cancernew.s)), cex = 0.6)
```




```
#To determine which item with highest uncertainty
clust.data <- cbind(colnames(cancernew), mc$uncertainty)

## Warning in cbind(colnames(cancernew), mc$uncertainty): number of rows of r
esult
## is not a multiple of vector length (arg 1)

clust.data[order(mc$uncertainty, decreasing = T),]

##        [,1]          [,2]
##   [1,] "Resistin"    "0.358385171148007"
##   [2,] "HOMA"        "0.30209515376003"
##   [3,] "Insulin"     "0.280861060621841"
##   [4,] "Adiponectin" "0.254705249452648"
##   [5,] "Glucose"     "0.253264869842006"
##   [6,] "BMI"         "0.203464545912695"
##   [7,] "Glucose"     "0.154358722850188"
##   [8,] "Age"         "0.146797133484552"
##   [9,] "BMI"         "0.143402110947072"
##  [10,] "BMI"         "0.110850705364792"
##  [11,] "Insulin"     "0.109774973680123"
##  [12,] "Insulin"     "0.108170009095194"
##  [13,] "Insulin"     "0.10779066779114"
##  [14,] "Age"         "0.0972460106611442"
```

```
##  [15,] "Adiponectin" "0.0960597474439051"
##  [16,] "Insulin"     "0.084794276053852"
##  [17,] "Resistin"    "0.0782999106771507"
##  [18,] "MCP.1"       "0.0773910130993141"
##  [19,] "Leptin"      "0.076088299710369"
##  [20,] "Glucose"     "0.0677630876130614"
##  [21,] "HOMA"        "0.0626328430455609"
##  [22,] "Leptin"      "0.0598701077307973"
##  [23,] "Glucose"     "0.0594586348779029"
##  [24,] "Adiponectin" "0.0564240799932693"
##  [25,] "Age"         "0.0558012769069084"
##  [26,] "Resistin"    "0.0552090134702723"
##  [27,] "Glucose"     "0.0512950510159567"
##  [28,] "Resistin"    "0.0468970652161095"
##  [29,] "BMI"         "0.0451666827322276"
##  [30,] "Leptin"      "0.0428563044784475"
##  [31,] "HOMA"        "0.0355526245733498"
##  [32,] "Resistin"    "0.0284618210910773"
##  [33,] "HOMA"        "0.0241778327424561"
##  [34,] "Age"         "0.020930585435624"
##  [35,] "Insulin"     "0.0208909353143742"
##  [36,] "Insulin"     "0.020277292360374"
##  [37,] "Age"         "0.0161449925887064"
##  [38,] "BMI"         "0.0130817848008632"
##  [39,] "BMI"         "0.0121929890119157"
##  [40,] "Glucose"     "0.011666190435826"
##  [41,] "MCP.1"       "0.0116514829519372"
##  [42,] "Insulin"     "0.00974615698931625"
##  [43,] "Adiponectin" "0.00937480745001362"
##  [44,] "HOMA"        "0.00834652709018713"
##  [45,] "Glucose"     "0.00829833941579394"
##  [46,] "Glucose"     "0.00829758301537575"
##  [47,] "Glucose"     "0.0072339712027587"
##  [48,] "BMI"         "0.00701402870705758"
##  [49,] "MCP.1"       "0.00482029741212664"
##  [50,] "MCP.1"       "0.00435616949197903"
##  [51,] "MCP.1"       "0.00369008627987566"
##  [52,] "Leptin"      "0.00328138763361696"
##  [53,] "Age"         "0.00323531780022868"
##  [54,] "MCP.1"       "0.00304301217564962"
##  [55,] "Adiponectin" "0.00285740175846472"
##  [56,] "HOMA"        "0.00209018082061818"
##  [57,] "HOMA"        "0.0019711089425245"
##  [58,] "Age"         "0.0019286706831394"
##  [59,] "Age"         "0.00185056272102369"
##  [60,] "Resistin"    "0.00167553652871499"
##  [61,] "MCP.1"       "0.00141626228681302"
##  [62,] "MCP.1"       "0.00139588542319191"
##  [63,] "Resistin"    "0.00135208659860397"
##  [64,] "Adiponectin" "0.00089311833376704"
```

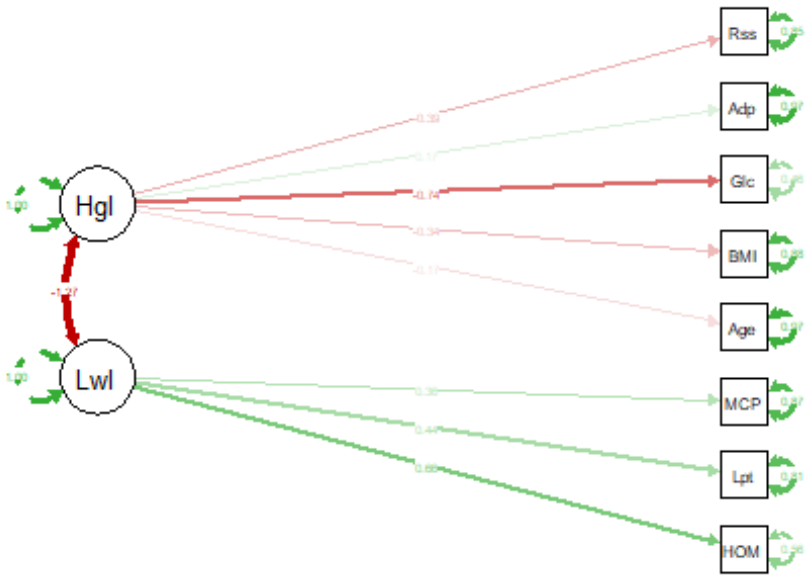
```
##  [65,] "MCP.1"       "0.000877276479713585"
##  [66,] "Adiponectin" "0.000808054156474891"
##  [67,] "Leptin"      "0.000651098138868544"
##  [68,] "Age"         "0.000627817955325183"
##  [69,] "HOMA"        "0.000606934624245437"
##  [70,] "Age"         "0.000520700886634251"
##  [71,] "HOMA"        "0.000477907057392368"
##  [72,] "Adiponectin" "0.000473801855997102"
##  [73,] "Resistin"    "0.000470984911574934"
##  [74,] "Insulin"     "0.000384831106433392"
##  [75,] "Leptin"      "0.000354340905387218"
##  [76,] "HOMA"        "0.000291747139974152"
##  [77,] "Resistin"    "0.000262747949009046"
##  [78,] "Leptin"      "0.000239408315961143"
##  [79,] "Leptin"      "0.00023249683420079"
##  [80,] "HOMA"        "0.000178448466185865"
##  [81,] "Age"         "0.000142247023001918"
##  [82,] "Leptin"      "0.00013019575588491"
##  [83,] "MCP.1"       "0.000121524980860599"
##  [84,] "Insulin"     "0.000113163297104868"
##  [85,] "Adiponectin" "8.5988434458617e-05"
##  [86,] "Insulin"     "3.94795172201645e-05"
##  [87,] "Adiponectin" "3.22001975077146e-05"
##  [88,] "Leptin"      "3.10986844778149e-05"
##  [89,] "Leptin"      "2.56582121367366e-05"
##  [90,] "BMI"         "1.74059338866606e-05"
##  [91,] "Adiponectin" "1.27814889795408e-05"
##  [92,] "BMI"         "1.25248008155943e-05"
##  [93,] "Glucose"     "9.66560073067946e-06"
##  [94,] "Resistin"    "8.00615803253635e-06"
##  [95,] "Glucose"     "5.66383447475616e-06"
##  [96,] "BMI"         "2.35064824793962e-06"
##  [97,] "Age"         "2.20602782563528e-06"
##  [98,] "BMI"         "1.40565525275083e-06"
##  [99,] "Insulin"     "1.16093939883655e-06"
## [100,] "Glucose"     "2.52652066490278e-07"
## [101,] "Glucose"     "2.41775497733521e-07"
## [102,] "Leptin"      "2.2291297541166e-07"
## [103,] "Leptin"      "2.02309509367282e-07"
## [104,] "BMI"         "9.03773279548048e-08"
## [105,] "MCP.1"       "7.15854331367893e-09"
## [106,] "BMI"         "5.75314884621037e-09"
## [107,] "Resistin"    "5.1264972267262e-09"
## [108,] "HOMA"        "9.22901532973697e-10"
## [109,] "Resistin"    "3.99680288865056e-14"
## [110,] "MCP.1"       "0"
## [111,] "Adiponectin" "0"
## [112,] "Adiponectin" "0"
## [113,] "Resistin"    "0"
## [114,] "Age"         "0"
```

```
## [115,] "Insulin"     "0"
## [116,] "HOMA"        "0"

#resistin has the highest uncertainty with about 36%

# To do CFA, we first need a model sem
#Calling the sem library after sem package installation
library(sem)

# Read the model stock_model.txt from file:
cancer_model <- specifyModel(file = "cancer_model.txt")

## NOTE: it is generally simpler to use specifyEquations() or cfa()
##       see ?specifyEquations

#Getting the CFA
cancer_sem <- sem(cancer_model, cor(cancernew), nrow(cancernew))

## Warning in sem.semmod(cancer_model, cor(cancernew), nrow(cancernew)): The
following observed variables are in the input covariance or raw-moment matrix
but do not appear in the model:
## Insulin

#Getting the summary of stock_sem
summary(cancer_sem)

##
##  Model Chisquare =  94.5   Df =  19 Pr(>Chisq) = 5.17e-12
##  AIC =  129
##  BIC =  4.22
##
##  Normalized Residuals
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   -2.62   -0.99    0.00   -0.19    0.39    3.99
##
##  R-square for Endogenous Variables
##         Age         BMI     Glucose    Resistin Adiponectin        HOMA
##      0.0280      0.1184      0.5426      0.1519      0.0288      0.4378
##      Leptin       MCP.1
##      0.1922      0.1322
##
##  Parameter Estimates
##         Estimate Std Error z value Pr(>|z|)
## lambda1 -0.167   0.0986    -1.70   8.96e-02 Age <--- HighIndicator
## lambda2 -0.344   0.0973    -3.54   4.06e-04 BMI <--- HighIndicator
## lambda3 -0.737   0.0976    -7.54   4.57e-14 Glucose <--- HighIndicator
## lambda4 -0.390   0.0969    -4.02   5.74e-05 Resistin <--- HighIndicator
## lambda5  0.170   0.0986     1.72   8.52e-02 Adiponectin <--- HighIndicator
## lambda6  0.662   0.0985     6.72   1.84e-11 HOMA <--- LowIndicator
## lambda7  0.438   0.0959     4.57   4.79e-06 Leptin <--- LowIndicator
## lambda8  0.364   0.0954     3.81   1.38e-04 MCP.1 <--- LowIndicator
```

```
## rho     -1.266   0.1353    -9.36   8.03e-21 LowIndicator <--> HighIndicato
r
## theta1   0.972   0.1287     7.55   4.24e-14 Age <--> Age
## theta2   0.882   0.1189     7.42   1.20e-13 BMI <--> BMI
## theta3   0.457   0.1029     4.45   8.72e-06 Glucose <--> Glucose
## theta4   0.848   0.1155     7.35   2.05e-13 Resistin <--> Resistin
## theta5   0.971   0.1286     7.55   4.26e-14 Adiponectin <--> Adiponectin
## theta6   0.562   0.1029     5.46   4.66e-08 HOMA <--> HOMA
## theta7   0.808   0.1112     7.26   3.79e-13 Leptin <--> Leptin
## theta8   0.868   0.1167     7.44   1.02e-13 MCP.1 <--> MCP.1
##
##  Iterations =  19

# All data are significant from the p value, estimate is approx values for la
mbdas.
```

Making path diagram to show coefficient estimates

```
#Calling the semPlot library
library(semPlot)
##semPaths(fitted.sem.object, "est")

#Generating the plot
semPaths(cancer_sem, rotation = 2, 'std', 'est')
```




```
#The model is as expected based on the set input.
```

```r
#Getting some fit indices
options(fit.indices = c("GFI", "AGFI", "SRMR"))

criteria = summary(cancer_sem)
criteria
```

```
##
##  Model Chisquare =  94.5   Df =  19 Pr(>Chisq) = 5.17e-12
##  Goodness-of-fit index =  0.85
##  Adjusted goodness-of-fit index =  0.716
##  SRMR =  0.113
##
##  Normalized Residuals
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   -2.62   -0.99    0.00   -0.19    0.39    3.99
##
##  R-square for Endogenous Variables
##         Age         BMI     Glucose    Resistin Adiponectin        HOMA
##      0.0280      0.1184      0.5426      0.1519      0.0288      0.4378
##      Leptin       MCP.1
##      0.1922      0.1322
##
##  Parameter Estimates
##         Estimate Std Error z value Pr(>|z|)
## lambda1 -0.167   0.0986    -1.70   8.96e-02 Age <--- HighIndicator
## lambda2 -0.344   0.0973    -3.54   4.06e-04 BMI <--- HighIndicator
## lambda3 -0.737   0.0976    -7.54   4.57e-14 Glucose <--- HighIndicator
## lambda4 -0.390   0.0969    -4.02   5.74e-05 Resistin <--- HighIndicator
## lambda5  0.170   0.0986     1.72   8.52e-02 Adiponectin <--- HighIndicator
## lambda6  0.662   0.0985     6.72   1.84e-11 HOMA <--- LowIndicator
## lambda7  0.438   0.0959     4.57   4.79e-06 Leptin <--- LowIndicator
## lambda8  0.364   0.0954     3.81   1.38e-04 MCP.1 <--- LowIndicator
## rho     -1.266   0.1353    -9.36   8.03e-21 LowIndicator <--> HighIndicato
r
## theta1   0.972   0.1287     7.55   4.24e-14 Age <--> Age
## theta2   0.882   0.1189     7.42   1.20e-13 BMI <--> BMI
## theta3   0.457   0.1029     4.45   8.72e-06 Glucose <--> Glucose
## theta4   0.848   0.1155     7.35   2.05e-13 Resistin <--> Resistin
## theta5   0.971   0.1286     7.55   4.26e-14 Adiponectin <--> Adiponectin
## theta6   0.562   0.1029     5.46   4.66e-08 HOMA <--> HOMA
## theta7   0.808   0.1112     7.26   3.79e-13 Leptin <--> Leptin
## theta8   0.868   0.1167     7.44   1.02e-13 MCP.1 <--> MCP.1
##
##  Iterations =  19
```

```r
criteria$SRMR
```

```
## [1] 0.113
```

```r
criteria$GFI
```

```
## [1] 0.85
```

```
criteria$AGFI
```

```
## [1] 0.716
```

```
criteria$SRMR < 0.05 #Testing if SRMR is less than 0.05
```

```
## [1] FALSE
```

```
criteria$GFI > 0.95  #Comparing GFI to 0.95
```

```
## [1] FALSE
```

```
criteria$AGFI > 0.95 #comparing AGFI against 0.95
```

```
## [1] FALSE
```

```
##Checking if the model is appropriate.comparing the approximated corr matrix
with original corr matrix.

# restricted Cor matrix
cancer_sem$C  #Approximated correlation matrix
```

```
##                 Age     BMI Glucose   HOMA  Leptin Adiponectin Resistin
MCP.1
## Age          1.0000  0.0576   0.123  0.140  0.0929     -0.0284   0.0652  0
.0770
## BMI          0.0576  1.0000   0.253  0.288  0.1910     -0.0584   0.1341  0
.1584
## Glucose      0.1233  0.2535   1.000  0.617  0.4089     -0.1250   0.2871  0
.3390
## HOMA         0.1402  0.2882   0.617  1.000  0.2901     -0.1421   0.3265  0
.2405
## Leptin       0.0929  0.1910   0.409  0.290  1.0000     -0.0942   0.2164  0
.1594
## Adiponectin -0.0284 -0.0584  -0.125 -0.142 -0.0942      1.0000  -0.0661 -0
.0781
## Resistin     0.0652  0.1341   0.287  0.326  0.2164     -0.0661   1.0000  0
.1794
## MCP.1        0.0770  0.1584   0.339  0.241  0.1594     -0.0781   0.1794  1
.0000
```

```
# non-restricted Cor matrix
cancer_sem$S # This is the original correlation matrix: ability.
```

```
##                  Age      BMI Glucose    HOMA  Leptin Adiponectin Resistin
## Age          1.00000  0.00853   0.230  0.1270  0.1026     -0.2198  0.00274
## BMI          0.00853  1.00000   0.139  0.1145  0.5696     -0.3027  0.19535
## Glucose      0.23011  0.13885   1.000  0.6962  0.3051     -0.1221  0.29133
## HOMA         0.12703  0.11448   0.696  1.0000  0.3272     -0.0563  0.23110
## Leptin       0.10263  0.56959   0.305  0.3272  1.0000     -0.0954  0.25623
```

```
## Adiponectin -0.21981 -0.30273  -0.122 -0.0563 -0.0954      1.0000 -0.25236
## Resistin     0.00274  0.19535   0.291  0.2311  0.2562     -0.2524  1.00000
## MCP.1        0.01346  0.22404   0.265  0.2595  0.0140     -0.2007  0.36647
##                MCP.1
## Age           0.0135
## BMI           0.2240
## Glucose       0.2649
## HOMA          0.2595
## Leptin        0.0140
## Adiponectin -0.2007
## Resistin      0.3665
## MCP.1         1.0000
```

```
# the root mean square error
sqrt(mean((cancer_sem$C-cancer_sem$S)^2))
```

```
## [1] 0.12
```

```
# null hypothesis: the restricted cov matrix (of CFA) is similar to the non-r
estricted cov matrix (of original data)

# p-value aprox 0 < 0.05 --> conclusion: reject the null hypothesis, so there
is not enough evidence to say that the restricted cov matrix is equal to the
non-restricted cov matrix.



# All criteria: GFI:0.85, AGFI:0.716 and SRMR:0.113 confirmed the model is no
t good.

# Data does not support the designed CFA model. MODEL IS NOT CONFIRMED.
```

Finding 95% Confidence level correlation between the two factors

```
#To get the 95% confidence interval
parameters = summary(cancer_sem)
parameters$coeff
```

```
##         Estimate Std Error z value Pr(>|z|)
## lambda1   -0.167    0.0986   -1.70 8.96e-02          Age <--- HighIndicato
r
## lambda2   -0.344    0.0973   -3.54 4.06e-04          BMI <--- HighIndicato
r
## lambda3   -0.737    0.0976   -7.54 4.57e-14      Glucose <--- HighIndicato
r
## lambda4   -0.390    0.0969   -4.02 5.74e-05     Resistin <--- HighIndicato
r
## lambda5    0.170    0.0986    1.72 8.52e-02  Adiponectin <--- HighIndicato
r
## lambda6    0.662    0.0985    6.72 1.84e-11          HOMA <--- LowIndicato
r
```

```
## lambda7    0.438    0.0959    4.57 4.79e-06        Leptin <--- LowIndicato
r
## lambda8    0.364    0.0954    3.81 1.38e-04         MCP.1 <--- LowIndicato
r
## rho       -1.266    0.1353   -9.36 8.03e-21 LowIndicator <--> HighIndicato
r
## theta1     0.972    0.1287    7.55 4.24e-14                   Age <--> Ag
e
## theta2     0.882    0.1189    7.42 1.20e-13                   BMI <--> BM
I
## theta3     0.457    0.1029    4.45 8.72e-06           Glucose <--> Glucos
e
## theta4     0.848    0.1155    7.35 2.05e-13          Resistin <--> Resisti
n
## theta5     0.971    0.1286    7.55 4.26e-14    Adiponectin <--> Adiponecti
n
## theta6     0.562    0.1029    5.46 4.66e-08                 HOMA <--> HOM
A
## theta7     0.808    0.1112    7.26 3.79e-13              Leptin <--> Lepti
n
## theta8     0.868    0.1167    7.44 1.02e-13               MCP.1 <--> MCP.
1

# Using Rho, the correlation between the factors
parameters$coeff[8,]$Estimate

## [1] 0.364

#Rho is in row 8

#Finding the confidence interval

conf.L = parameters$coeff[8,]$Estimate - 1.96 * parameters$coeff[8,]$`Std Err
or` #Lower interval

conf.U = parameters$coeff[8,]$Estimate + 1.96 * parameters$coeff[8,]$`Std Err
or`  #Upper interval

conf.L

## [1] 0.177

conf.U

## [1] 0.55

#95% confidence interval for the correlation between factors : High-indicator
s and Low-indicators is between 0.177 and 0.55
```