MVA_project.R

Saiprasad Laxmeshwar

2019-10-17

```
#Importing the Hepatatis c Dataset
HCV= read.csv("HCV-Egy-Data.csv")
HCV
```

```
##       Age Gender BMI Fever Nausea.Vomting Headache Diarrhea
## 1      56      1  35     2               1        1        1
## 2      46      1  29     1               2        2        1
## 3      57      1  33     2               2        2        2
## 4      49      2  33     1               2        1        2
## 5      59      1  32     1               1        2        1
## 6      58      2  22     2               2        2        1
## 7      42      2  26     1               1        2        2
## 8      48      2  30     1               1        2        2
## 9      44      1  23     1               1        2        2
## 10     45      1  30     2               1        2        2
## 11     37      2  24     2               1        2        1
## 12     36      1  22     2               2        1        1
## 13     45      2  25     2               1        1        1
## 14     34      1  22     1               2        1        1
## 15     40      2  32     2               2        2        1
## 16     58      1  34     2               1        1        1
## 17     61      1  35     1               2        2        2
## 18     55      2  24     2               1        2        2
## 19     56      1  27     1               2        2        2
## 20     35      2  23     2               2        1        1
## 21     57      2  23     1               1        2        2
## 22     33      1  25     2               1        2        2
## 23     41      1  23     1               2        2        2
## 24     39      2  29     1               2        1        2
## 25     33      2  24     1               2        2        2
## 26     43      2  34     2               2        2        1
## 27     51      1  34     2               1        2        2
## 28     39      2  33     2               1        2        1
## 29     57      2  26     1               2        2        1
## 30     47      2  29     1               1        2        1
## 31     55      2  33     1               2        2        1
## 32     58      2  35     2               2        2        2
## 33     47      2  25     2               1        2        2
## 34     61      1  33     1               2        2        2
## 35     37      1  27     2               2        1        2
## 36     41      1  29     1               2        1        1
```

```
## 37    60       2  32     2                2         1         2
```

```
#Summary
attach(HCV)
summary(HCV)
```

```
##       Age             Gender           BMI             Fever
##  Min.   :32.00   Min.   :1.00   Min.   :22.00   Min.   :1.000
##  1st Qu.:39.00   1st Qu.:1.00   1st Qu.:25.00   1st Qu.:1.000
##  Median :46.00   Median :1.00   Median :29.00   Median :2.000
##  Mean   :46.32   Mean   :1.49   Mean   :28.61   Mean   :1.516
##  3rd Qu.:54.00   3rd Qu.:2.00   3rd Qu.:32.00   3rd Qu.:2.000
##  Max.   :61.00   Max.   :2.00   Max.   :35.00   Max.   :2.000
##  Nausea.Vomting     Headache        Diarrhea
##  Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1.000
##  Median :2.000   Median :1.000   Median :2.000
##  Mean   :1.503   Mean   :1.496   Mean   :1.503
##  3rd Qu.:2.000   3rd Qu.:2.000   3rd Qu.:2.000
##  Max.   :2.000   Max.   :2.000   Max.   :2.000
##  Fatigue...generalized.bone.ache    Jaundice     Epigastric.pain
##  Min.   :1.000                   Min.   :1.000   Min.   :1.000
##  1st Qu.:1.000                   1st Qu.:1.000   1st Qu.:1.000
##  Median :1.000                   Median :2.000   Median :2.000
##  Mean   :1.499                   Mean   :1.501   Mean   :1.504
##  3rd Qu.:2.000                   3rd Qu.:2.000   3rd Qu.:2.000
##  Max.   :2.000                   Max.   :2.000   Max.   :2.000
##       WBC             RBC               HGB             Plat
##  Min.   : 2991   Min.   :3816422   Min.   :10.00   Min.   : 93013
##  1st Qu.: 5219   1st Qu.:4121374   1st Qu.:11.00   1st Qu.:124479
##  Median : 7498   Median :4438465   Median :13.00   Median :157916
##  Mean   : 7533   Mean   :4422130   Mean   :12.59   Mean   :158348
##  3rd Qu.: 9902   3rd Qu.:4721279   3rd Qu.:14.00   3rd Qu.:190314
##  Max.   :12101   Max.   :5018451   Max.   :15.00   Max.   :226464
##      AST.1            ALT.1             ALT4            ALT.12
##  Min.   : 39.00   Min.   : 39.00   Min.   : 39.00   Min.   : 39.00
##  1st Qu.: 60.00   1st Qu.: 62.00   1st Qu.: 61.00   1st Qu.: 60.00
##  Median : 83.00   Median : 83.00   Median : 82.00   Median : 84.00
##  Mean   : 82.77   Mean   : 83.92   Mean   : 83.41   Mean   : 83.51
##  3rd Qu.:105.00   3rd Qu.:106.00   3rd Qu.:107.00   3rd Qu.:106.00
##  Max.   :128.00   Max.   :128.00   Max.   :128.00   Max.   :128.00
##      ALT.24           ALT.36           ALT.48        ALT.after.24.w
##  Min.   : 39.00   Min.   :  5.00   Min.   :  5.00   Min.   : 5.00
##  1st Qu.: 61.00   1st Qu.: 61.00   1st Qu.: 61.00   1st Qu.:28.00
##  Median : 83.00   Median : 84.00   Median : 83.00   Median :34.00
##  Mean   : 83.71   Mean   : 83.12   Mean   : 83.63   Mean   :33.44
##  3rd Qu.:107.00   3rd Qu.:106.00   3rd Qu.:106.00   3rd Qu.:40.00
##  Max.   :128.00   Max.   :128.00   Max.   :128.00   Max.   :45.00
##     RNA.Base            RNA.4             RNA.12           RNA.EOT
##  Min.   :     11   Min.   :      5   Min.   :      5   Min.   :     5
##  1st Qu.: 269253   1st Qu.: 270893   1st Qu.:      5   1st Qu.:     5
```

```
##  Median : 593103   Median : 597869   Median : 234359   Median :251376
##  Mean   : 590951   Mean   : 600896   Mean   : 288754   Mean   :287660
##  3rd Qu.: 886791   3rd Qu.: 909093   3rd Qu.: 524819   3rd Qu.:517806
##  Max.   :1201086   Max.   :1201715   Max.   :3731527   Max.   :808450
##      RNA.EF       Baseline.histological.Grading
##  Min.   :     5   Min.   : 3.000
##  1st Qu.:     5   1st Qu.: 6.000
##  Median :244049   Median :10.000
##  Mean   :291378   Mean   : 9.762
##  3rd Qu.:527864   3rd Qu.:13.000
##  Max.   :810333   Max.   :16.000
##  Baselinehistological.staging
##  Min.   :1.000
##  1st Qu.:2.000
##  Median :3.000
##  Mean   :2.536
##  3rd Qu.:4.000
##  Max.   :4.000

#Dimensions of the data set
NROW(HCV)

## [1] 1385

NCOL(HCV)

## [1] 29

#Displaying the column names of the dataset
colnames(HCV)

##  [1] "Age"                                  "Gender"
##  [3] "BMI"                                  "Fever"
##  [5] "Nausea.Vomting"                       "Headache"
##  [7] "Diarrhea"
"Fatigue...generalized.bone.ache"
##  [9] "Jaundice"                             "Epigastric.pain"
## [11] "WBC"                                  "RBC"
## [13] "HGB"                                  "Plat"
## [15] "AST.1"                                "ALT.1"
## [17] "ALT4"                                 "ALT.12"
## [19] "ALT.24"                               "ALT.36"
## [21] "ALT.48"                               "ALT.after.24.w"
## [23] "RNA.Base"                             "RNA.4"
## [25] "RNA.12"                               "RNA.EOT"
## [27] "RNA.EF"                               "Baseline.histological.Grading"
## [29] "Baselinehistological.staging"

#Another menthod for dimensions
dim(HCV)

## [1] 1385   29

#Preprocessing data was done but did'nt find any discrepancies.
na= is.na(HCV)
na
```

```
##           Age Gender   BMI Fever Nausea.Vomting Headache Diarrhea
##    [1,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##    [2,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##    [3,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##    [4,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##    [5,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##    [6,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##    [7,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##    [8,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##    [9,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [10,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [11,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [12,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [13,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [14,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [15,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [16,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [17,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [18,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [19,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [20,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [21,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [22,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [23,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [24,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [25,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [26,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [27,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [28,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [29,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [30,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [31,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [32,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [33,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [34,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [35,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [36,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [37,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [38,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [39,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [40,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [41,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [42,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [43,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [44,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [45,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [46,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [47,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [48,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
```

```
##   [49,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [50,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [51,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [52,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
##   [53,] FALSE  FALSE FALSE FALSE          FALSE    FALSE    FALSE
```

```
#Displaying the first six rows of the datasets
head(HCV)
```

```
##   Age Gender BMI Fever Nausea.Vomting Headache Diarrhea
## 1  56      1  35     2              1        1        1
## 2  46      1  29     1              2        2        1
## 3  57      1  33     2              2        2        2
## 4  49      2  33     1              2        1        2
## 5  59      1  32     1              1        2        1
## 6  58      2  22     2              2        2        1
##   Fatigue...generalized.bone.ache Jaundice Epigastric.pain   WBC
RBC
## 1                               2        2               2  7425
4248807
## 2                               2        2               1 12101
4429425
## 3                               1        1               1  4178
4621191
## 4                               1        2               1  6490
4794631
## 5                               2        2               2  3661
4606375
## 6                               2        2               1 11785
3882456
##   HGB   Plat AST.1 ALT.1 ALT4 ALT.12 ALT.24 ALT.36 ALT.48
ALT.after.24.w
## 1  14 112132    99    84   52    109     81      5      5
5
## 2  10 129367    91   123   95     75    113     57    123
44
## 3  12 151522   113    49   95    107    116      5      5
5
## 4  10 146457    43    64  109     80     88     48     77
33
## 5  11 187684    99   104   67     48    120     94     90
30
## 6  15 131228    66   104  121     96     65     73    114
29
##   RNA.Base   RNA.4  RNA.12 RNA.EOT RNA.EF Baseline.histological.Grading
## 1   655330  634536  288194       5      5                            13
## 2    40620  538635  637056  336804  31085                             4
## 3   571148  661346       5  735945 558829                             4
## 4  1041941  449939  585688  744463 582301                            10
## 5   660410  738756 3731527  338946 242861                            11
## 6  1157452 1086852       5       5      5                             4
```

```
##   Baselinehistological.staging
## 1                            2
## 2                            2
## 3                            4
## 4                            3
## 5                            1
## 6                            4
```

```
tail(HCV)
```

```
##      Age Gender BMI Fever Nausea.Vomting Headache Diarrhea
## 1380  53      1  31     2              2        1        2
## 1381  44      1  29     1              2        2        2
## 1382  55      1  34     1              2        2        1
## 1383  42      1  26     2              2        1        1
## 1384  52      1  29     2              1        1        2
## 1385  55      2  26     1              2        2        2
##      Fatigue...generalized.bone.ache Jaundice Epigastric.pain   WBC
## 1380                               2        2               1  4196
## 1381                               1        1               1  7044
## 1382                               1        1               1  6207
## 1383                               1        2               1  4913
## 1384                               2        2               1  7257
## 1385                               1        2               1 11832
##          RBC HGB   Plat AST.1 ALT.1 ALT4 ALT.12 ALT.24 ALT.36 ALT.48
## 1380 4076324  12 150065    89   113   52     39     54     86     78
## 1381 4957174  15 202520   122    59   78    106    127     63     44
## 1382 4636951  10 115776   128   102   65     99    108     97     64
## 1383 4122038  14 128354    61    93  123     61    116     87     39
## 1384 4241990  10 205908    70    97  104     74     47     48     81
## 1385 4059176  14 136615    51   126   39     68    115     64     71
##      ALT.after.24.w RNA.Base  RNA.4 RNA.12 RNA.EOT RNA.EF
## 1380             36   886656 460080 591040  621014  20645
## 1381             45   387795  55938      5       5      5
## 1382             41   481378 152961 393339   73574 236273
## 1383             24   612664 572756 806109  343719 160457
## 1384             43   139872  76161 515730    2460 696074
## 1385             34  1190577 628730      5       5      5
##      Baseline.histological.Grading Baselinehistological.staging
## 1380                            13                            4
## 1381                            15                            4
## 1382                            10                            2
## 1383                             6                            2
## 1384                            15                            3
## 1385                            13                            3
```

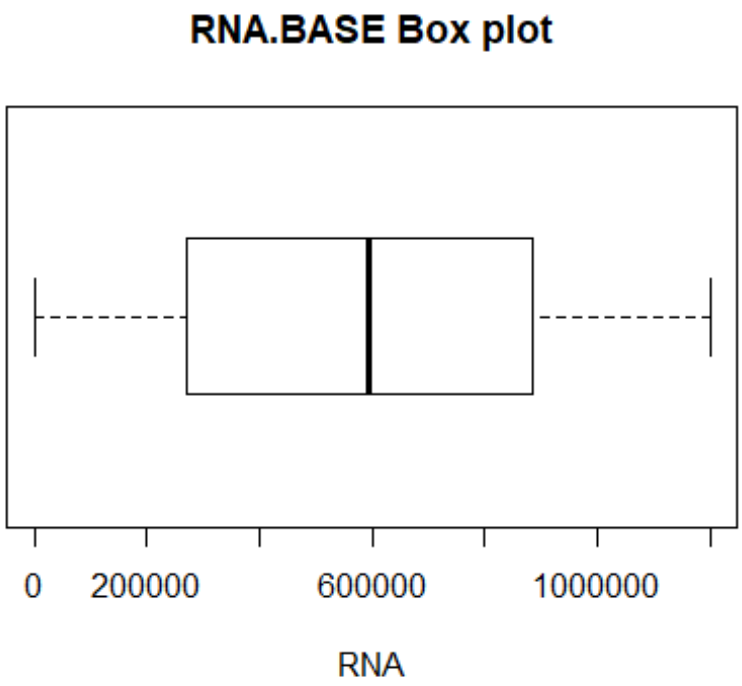
```
#CORRELATION, COVARIANCE AND DISTANCE
covariance<-cov(HCV[,c(11:16,23)]) #variamce-covariance matrix created
correlation<-cor(HCV[,c(11:16,23)]) #standardized
#colmeans
cm<-colMeans(HCV[,c(11:16,23)])
distance<-dist(scale(HCV[,c(11:16,23)],center=FALSE))
#Calculating di(generalized distance for all observations of our data)
```

```r
#before that first extract all numeric variable in a dataframe
x<-HCV[,c(11:16,23)]
d <- apply(x, MARGIN = 1, function(x) + t(x - cm) %*% solve(covariance)
%*% (x - cm))



#Exlporation of the data for high chances of HCV Infection
#Here RNA.base value if it is more than 700000 units then virus is
detected in high quantity.
#Here ALT.1 if value is greater than 57 then it is not normal.
#we sorted the data on these two components.

library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

HCV_male =  HCV %>% filter(Gender == 1 & RNA.Base>= 700000 & ALT.1 >= 57)
HCV_male
```

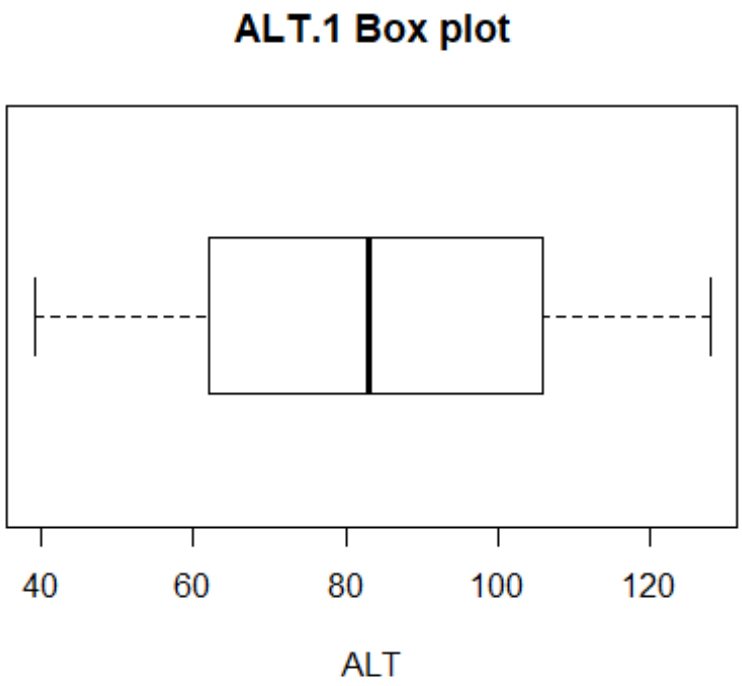
```
##      Age Gender BMI Fever Nausea.Vomting Headache Diarrhea
## 1     45      1  30     2               1        2        2
## 2     61      1  35     1               2        2        2
## 3     33      1  25     2               1        2        2
## 4     41      1  29     1               2        1        1
## 5     32      1  31     1               2        1        1
## 6     44      1  31     1               1        1        1
## 7     61      1  25     1               1        2        2
## 8     59      1  25     1               2        1        1
## 9     32      1  26     1               2        2        2
## 10    56      1  26     1               2        2        2
## 11    40      1  31     1               2        2        1
## 12    33      1  28     2               2        1        2
## 13    39      1  30     1               2        2        1
## 14    52      1  26     2               2        1        1
## 15    49      1  25     1               2        1        2
## 16    53      1  28     2               2        1        2
## 17    54      1  31     2               2        1        2
## 18    55      1  30     2               2        2        1
## 19    56      1  34     1               1        1        2
## 20    58      1  33     1               1        2        1
## 21    33      1  30     2               2        1        2
## 22    34      1  28     2               1        2        2
## 23    60      1  32     1               1        1        1
```

```
## 24   34       1  32      1                2         2         2
## 25   37       1  34      2                1         1         2
## 26   53       1  22      1                1         1         2
## 27   59       1  28      1                1         2         2
## 28   56       1  34      1                2         2         1
## 29   61       1  34      2                1         1         1
## 30   58       1  29      2                2         1         1
## 31   33       1  30      2                2         2         2
## 32   48       1  30      1                1         1         1
## 33   35       1  26      2                2         1         2
## 34   55       1  31      1                1         2         2
## 35   51       1  30      1                2         1         1
## 36   49       1  30      2                1         2         1
## 37   57       1  24      2                2         1         2
## 38   45       1  34      2                1         1         1
## 39   46       1  22      1                2         1         1
## 40   59       1  28      1                2         2         1
## 41   52       1  26      1                2         2         1
## 42   60       1  22      1                2         1         1
## 43   39       1  35      1                1         2         2
## 44   34       1  28      1                2         2         2
## 45   37       1  35      2                1         1         1
## 46   38       1  25      2                2         2         2
## 47   58       1  35      1                1         2         2
## 48   58       1  26      2                1         2         2
## 49   38       1  32      2                1         1         2
## 50   35       1  22      2                1         2         1
## 51   35       1  25      2                1         1         2
## 52   57       1  26      1                1         1         1
## 53   52       1  34      1                2         2         2
## 54   54       1  25      2                1         2         2
```

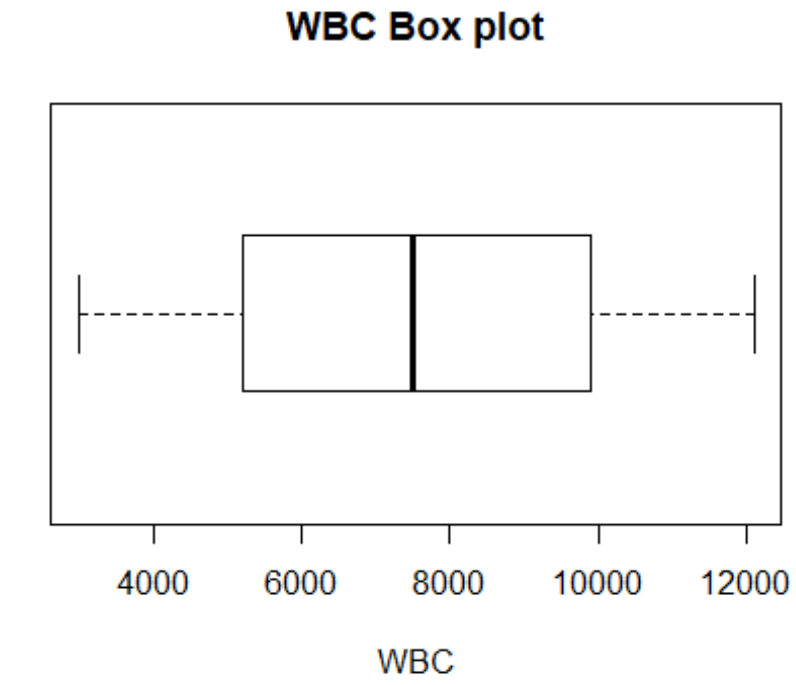
```r
#Box Plot
boxplot(RNA.Base, main="RNA.BASE Box plot",yaxt="n", xlab="RNA",
horizontal=TRUE)
```



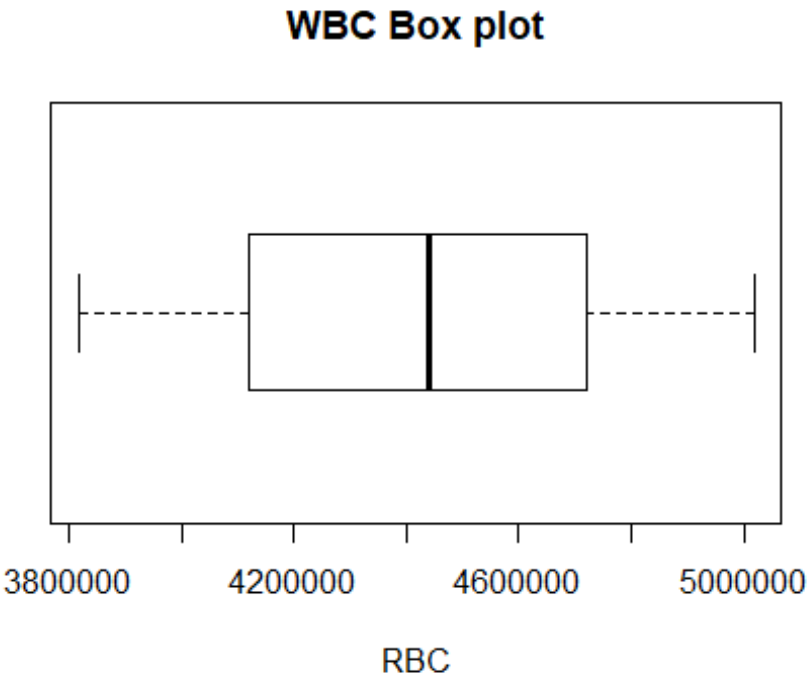

```
boxplot(ALT.1, main="ALT.1 Box plot",yaxt="n", xlab="ALT",
horizontal=TRUE)
```



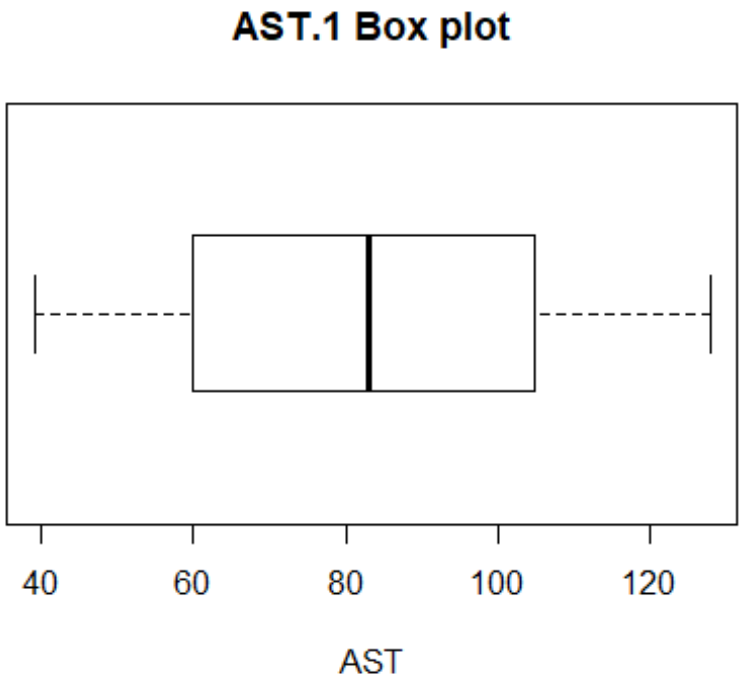

```
boxplot(WBC, main="WBC Box plot",yaxt="n", xlab="WBC", horizontal=TRUE)
```



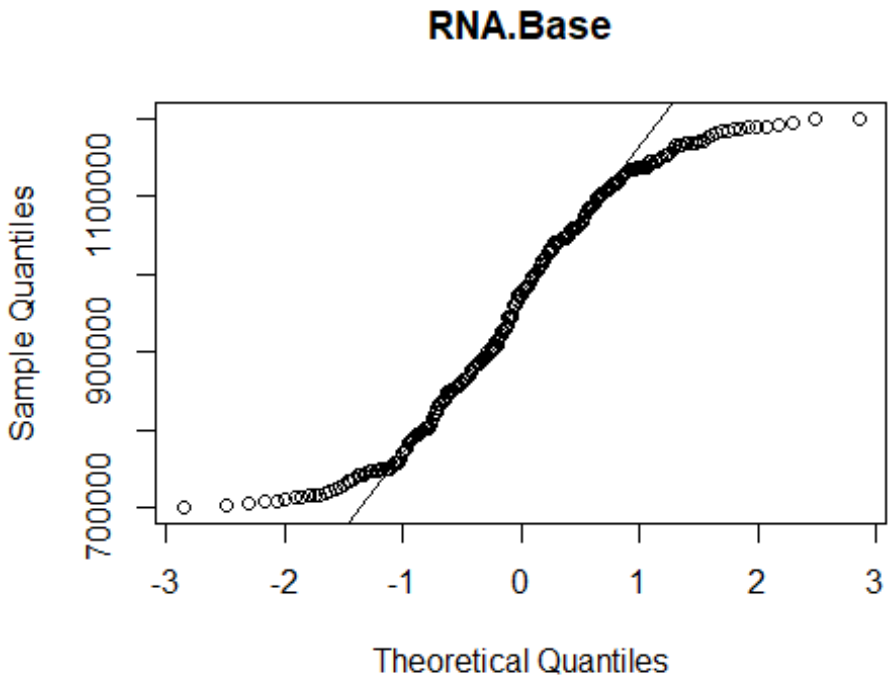

```
boxplot(RBC, main="WBC Box plot",yaxt="n", xlab="RBC", horizontal=TRUE)
```



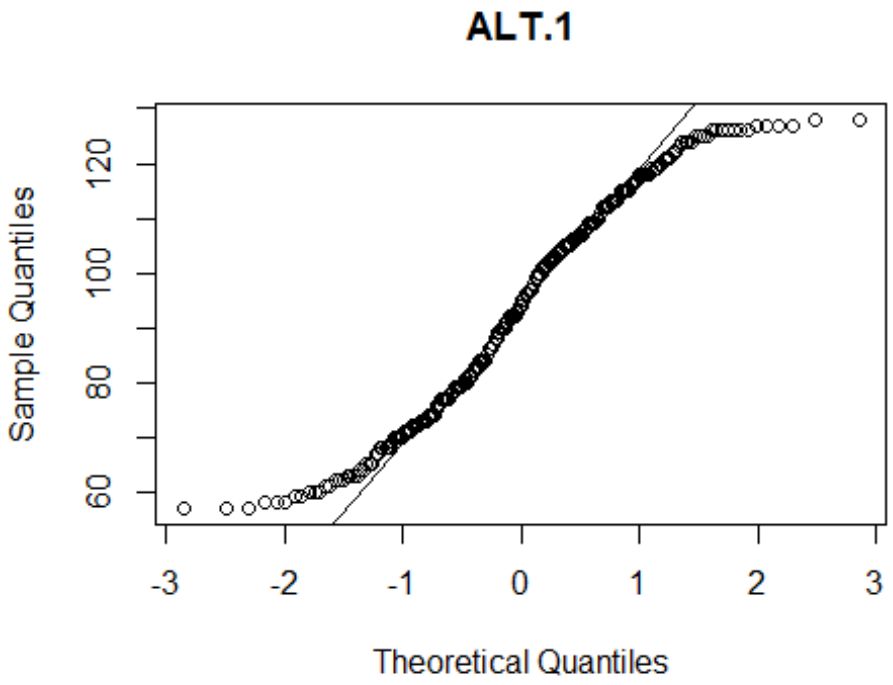

```
boxplot(AST.1, main="AST.1 Box plot",yaxt="n", xlab="AST",
horizontal=TRUE)
```



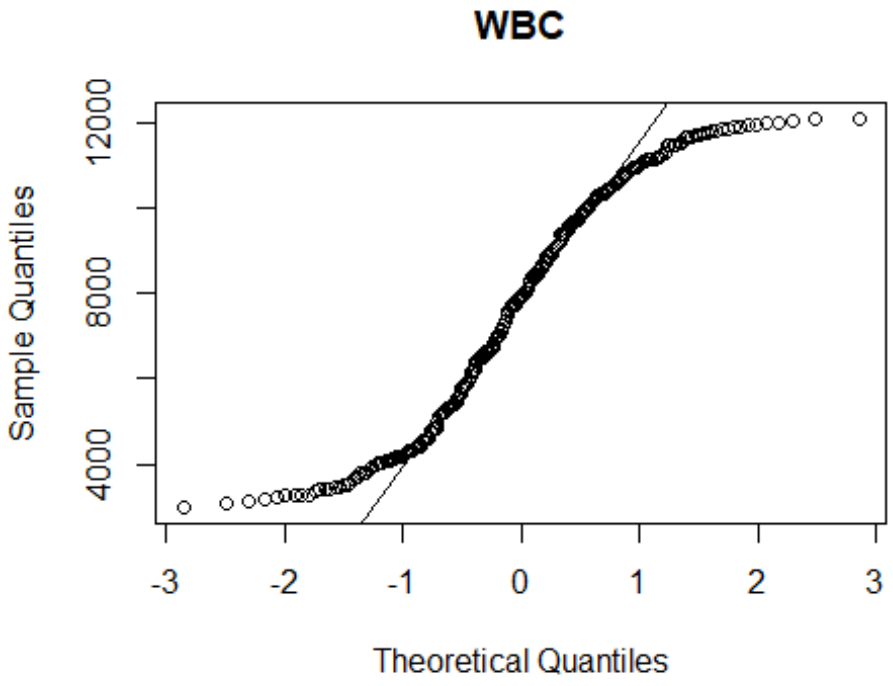

```r
#plotting, Are they in a straight line.
#Male Plotting of the dataset is done for five different attributes.
qqnorm(HCV_male[,"RNA.Base"], main = "RNA.Base");
qqline(HCV_male[,"RNA.Base"])
```



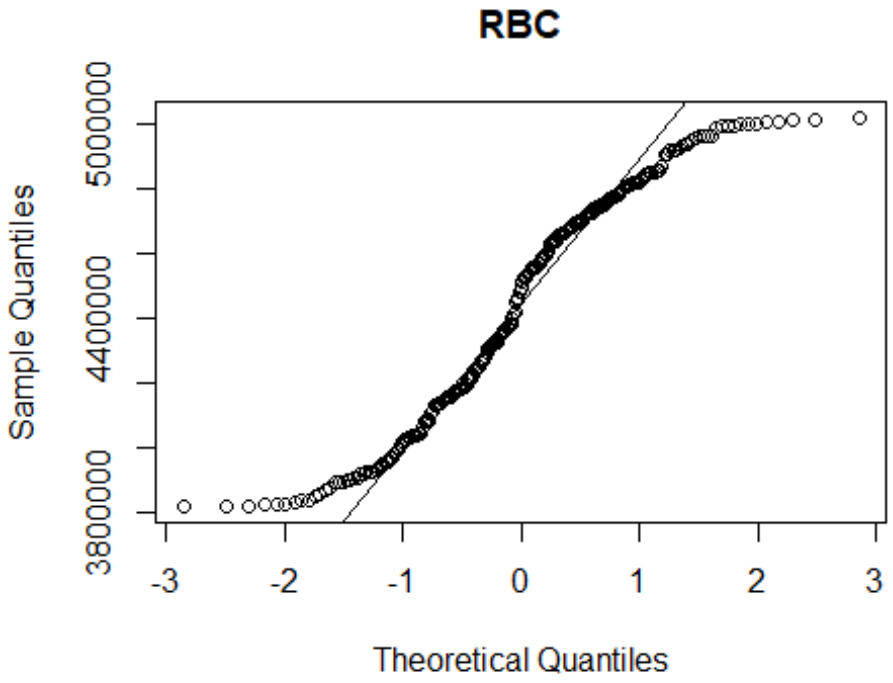

```r
qqnorm(HCV_male[,"ALT.1"], main = "ALT.1"); qqline(HCV_male[,"ALT.1"])
```



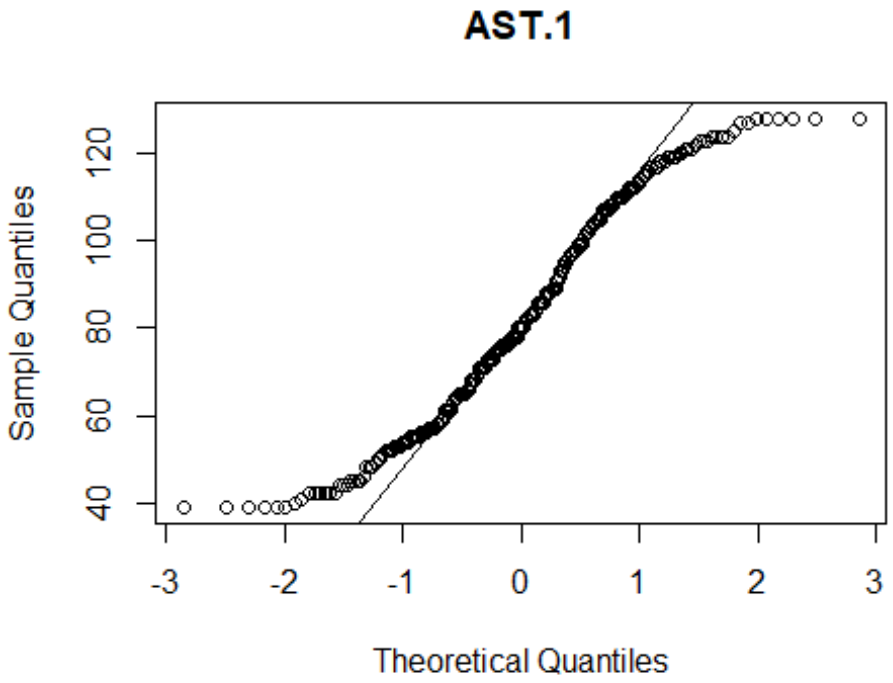

```r
qqnorm(HCV_male[,"WBC"], main = "WBC"); qqline(HCV_male[,"WBC"])
```



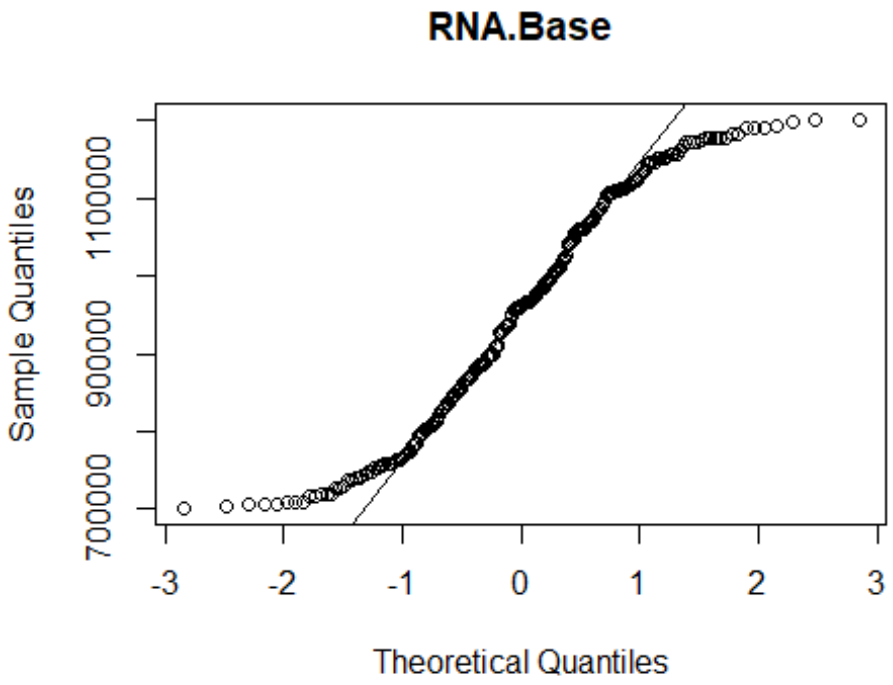

```r
qqnorm(HCV_male[,"RBC"], main = "RBC"); qqline(HCV_male[,"RBC"])
```



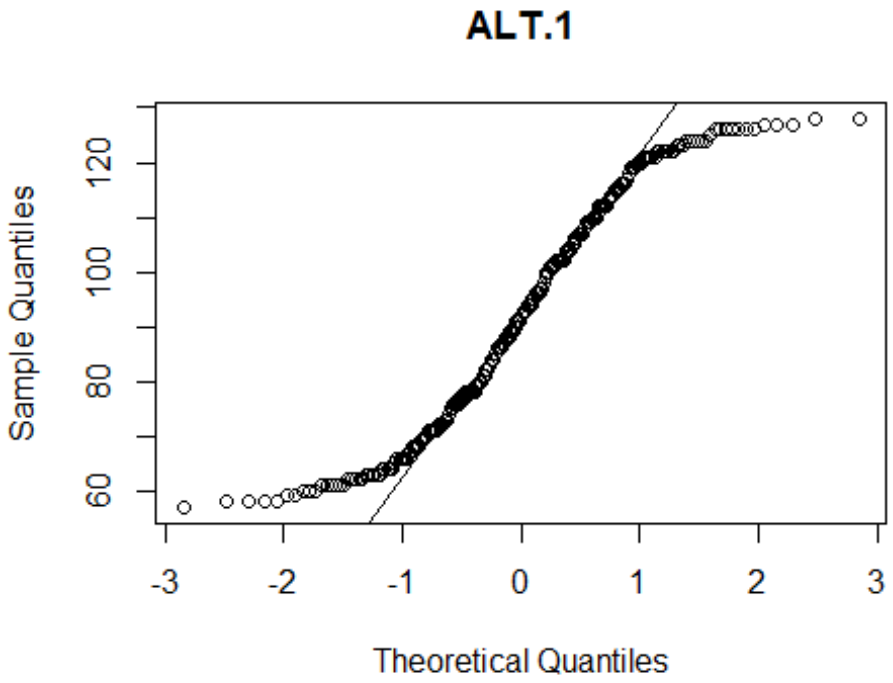

```
qqnorm(HCV_male[,"AST.1"], main = "AST.1"); qqline(HCV_male[,"AST.1"])
```



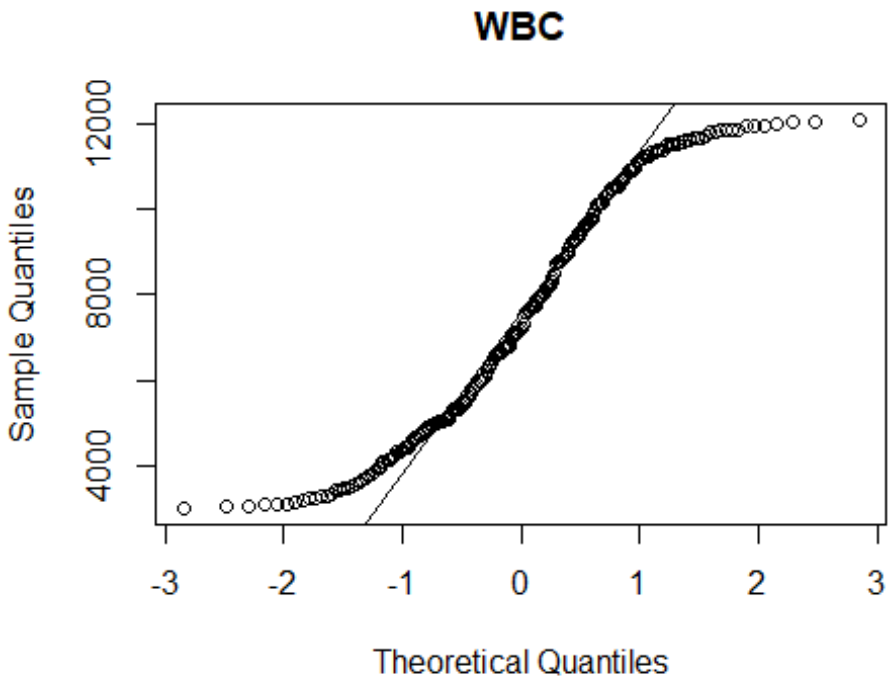

```
#Female, Are they in a straight line.
#FeMale Plotting of the dataset is done for five different attributes.
qqnorm(HCV_female[,"RNA.Base"], main = "RNA.Base");
qqline(HCV_female[,"RNA.Base"])
```



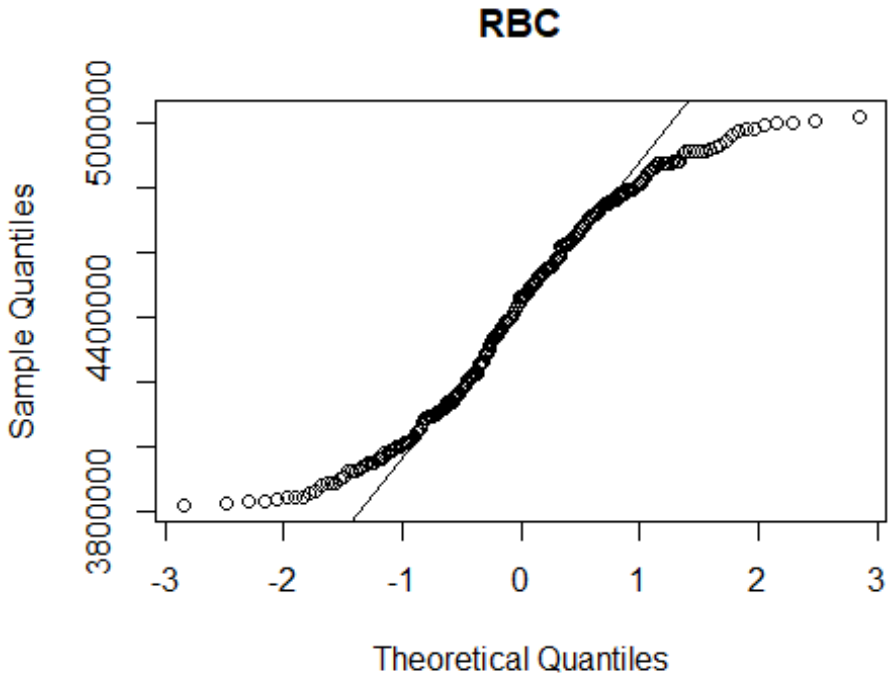

```
qqnorm(HCV_female[,"ALT.1"], main = "ALT.1"); qqline(HCV_female[,"ALT.1"])
```



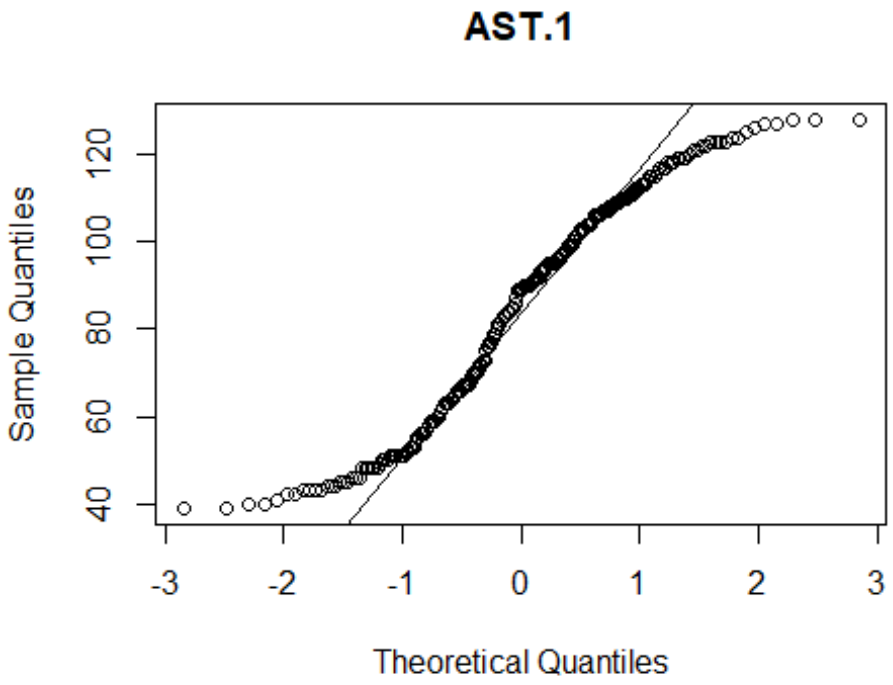

```
qqnorm(HCV_female[,"WBC"], main = "WBC"); qqline(HCV_female[,"WBC"])
```



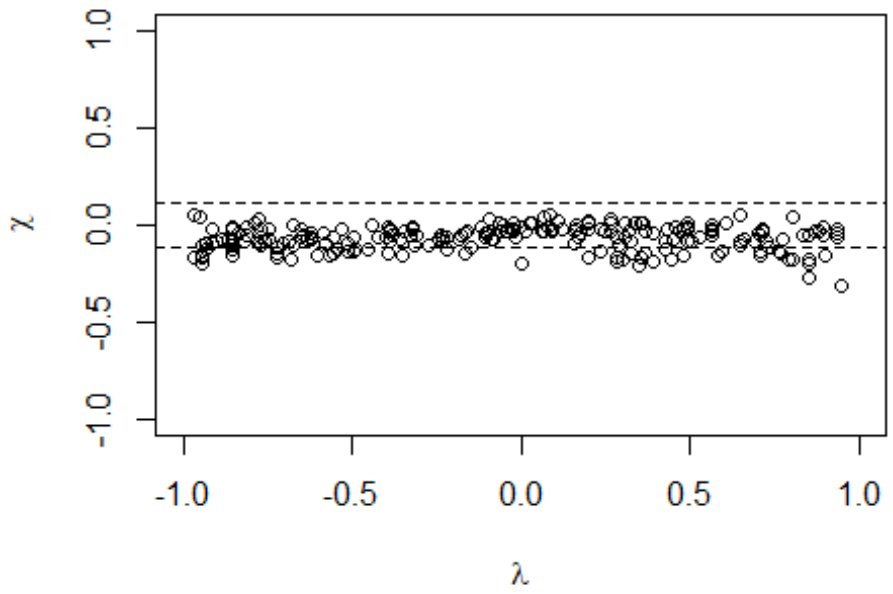

```
qqnorm(HCV_female[,"RBC"], main = "RBC"); qqline(HCV_female[,"RBC"])
```




```
qqnorm(HCV_female[,"AST.1"], main = "AST.1"); qqline(HCV_female[,"AST.1"])
```




```
#Visualisatiom
#Chiplot
library(HSAUR2)

## Loading required package: tools

library(tools)
library(MVA)

#Chiplot
#For male data
with(HCV_male, chiplot(RNA.Base, ALT.1))
```




```
#For Female Data
with(HCV_female, chiplot(RNA.Base, ALT.1))
```

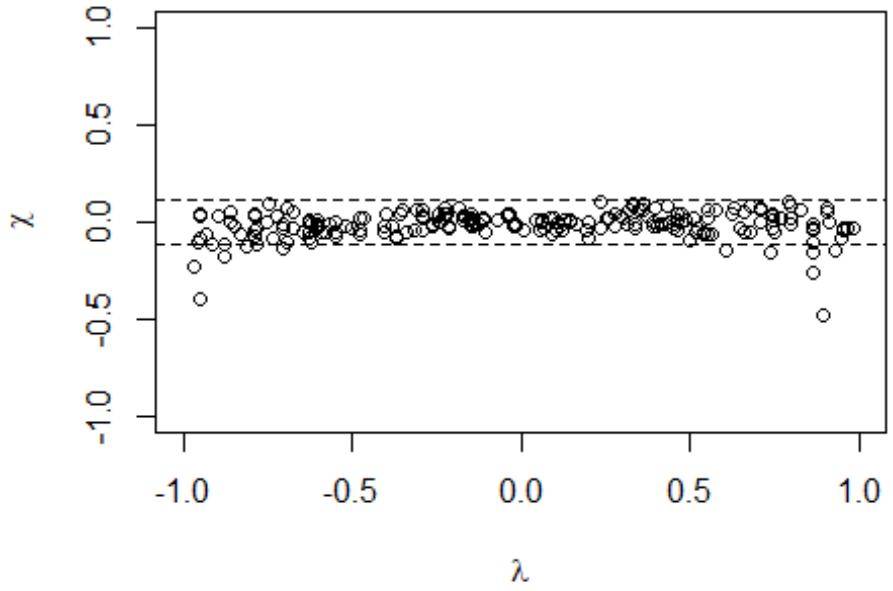
```
library(GGally)
```

```
## Loading required package: ggplot2
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

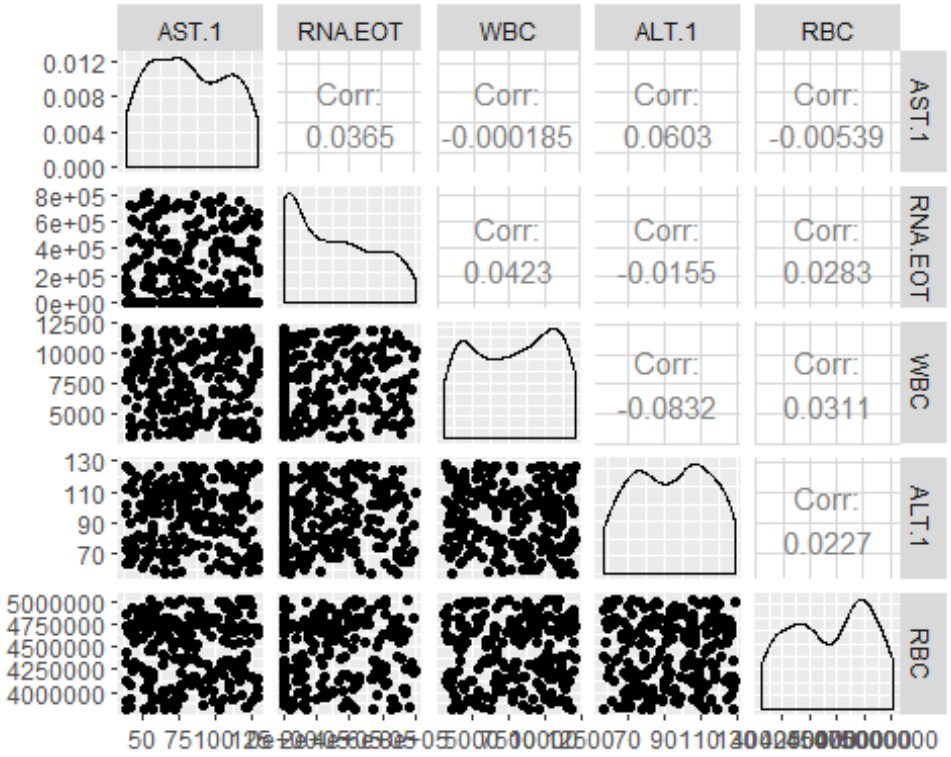
```
##
## Attaching package: 'GGally'
```

```
## The following object is masked from 'package:dplyr':
##
##     nasa
```

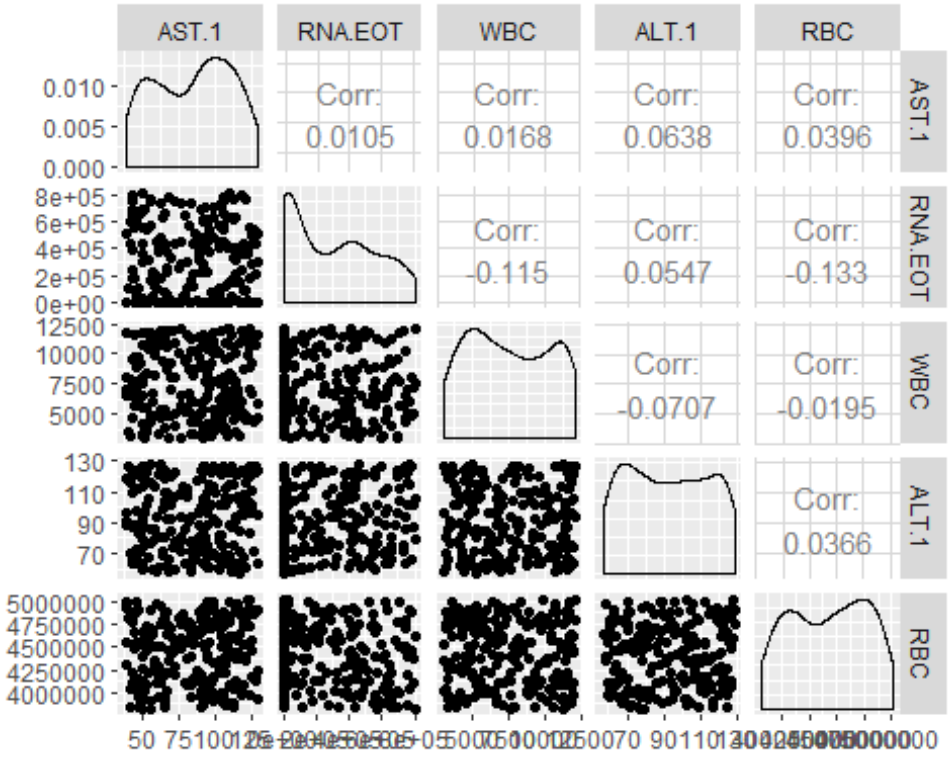



```
ggpairs(HCV_male, columns=c("AST.1","RNA.EOT","WBC","ALT.1", "RBC"),
color="Survivorship")
```

```
## Warning in warn_if_args_exist(list(...)): Extra arguments: "color" are
## being ignored. If these are meant to be aesthetics, submit them using
the
## 'mapping' variable within ggpairs with ggplot2::aes or
ggplot2::aes_string.
```




```
ggpairs(HCV_female, columns=c("AST.1","RNA.EOT","WBC","ALT.1", "RBC"),
color="Survivorship")

## Warning in warn_if_args_exist(list(...)): Extra arguments: "color" are
## being ignored. If these are meant to be aesthetics, submit them using
the
## 'mapping' variable within ggpairs with ggplot2::aes or
ggplot2::aes_string.
```




```
summary(lm(data = HCV , RNA.EOT~Age))

##
## Call:
## lm(formula = RNA.EOT ~ Age, data = HCV)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -308665 -271984  -35364  226006  533719
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 355622.3    38146.1   9.323   <2e-16 ***
## Age          -1467.3      809.1  -1.813     0.07 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 264300 on 1383 degrees of freedom
## Multiple R-squared:  0.002372,   Adjusted R-squared:  0.001651
## F-statistic: 3.288 on 1 and 1383 DF,  p-value: 0.07

summary(lm(data = HCV , RNA.EOT~Gender))

##
## Call:
## lm(formula = RNA.EOT ~ Gender, data = HCV)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -297181 -278520  -34151  223816  529617
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   259865      22338  11.634   <2e-16 ***
## Gender         18661      14217   1.313     0.19
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 264500 on 1383 degrees of freedom
## Multiple R-squared:  0.001244,   Adjusted R-squared:  0.000522
## F-statistic: 1.723 on 1 and 1383 DF,  p-value: 0.1896
```

```
summary(lm(data = HCV , RNA.EOT~WBC))
```

```
##
## Call:
## lm(formula = RNA.EOT ~ WBC, data = HCV)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -296698 -280505  -38809  230123  527965
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 302720.284  21302.853   14.21   <2e-16 ***
## WBC             -1.999      2.666   -0.75    0.453
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 264600 on 1383 degrees of freedom
## Multiple R-squared:  0.0004065,  Adjusted R-squared:  -0.0003163
## F-statistic: 0.5624 on 1 and 1383 DF,  p-value: 0.4534
```

```
summary(lm(data = HCV , RNA.EOT~ALT.1))
```

```
##
## Call:
## lm(formula = RNA.EOT ~ ALT.1, data = HCV)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -305707 -274761  -32847  226046  529732
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 321386.0    24083.4  13.345   <2e-16 ***
## ALT.1         -401.9      274.2  -1.466    0.143
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 264400 on 1383 degrees of freedom
## Multiple R-squared:  0.001551,   Adjusted R-squared:  0.0008288
## F-statistic: 2.148 on 1 and 1383 DF,  p-value: 0.143
```

```
cor(HCV)
```

```
##                                          Age        Gender           BMI
## Age                                1.000000000 -0.0099335141 -0.025353485
## Gender                            -0.009933514  1.0000000000  0.006495822
## BMI                               -0.025353485  0.0064958221  1.000000000
## Fever                             -0.021208025  0.0216038570 -0.017581296
## Nausea.Vomting                    -0.025023686 -0.0367257983  0.005801303
## Headache                           0.017561878 -0.0239990326 -0.007850790
## Diarrhea                           0.043574033  0.0181605819 -0.026094093
## Fatigue...generalized.bone.ache -0.007817311  0.0454520816 -0.007295820
## Jaundice                           0.010284829  0.0007675462 -0.071378670
## Epigastric.pain                   -0.010817131 -0.0251107556  0.008205190
## WBC                                0.014310232  0.0272897099  0.036721434
## RBC                               -0.002161729 -0.0009591287 -0.004922280
## HGB                               -0.012810104 -0.0012462736  0.057987192
## Plat                              -0.002225244  0.0192990586 -0.003093342
## AST.1                             -0.016632599 -0.0130777126  0.001806482
## ALT.1                              0.005722864  0.0224504101  0.034036533
## ALT4                               0.030259565 -0.0087203690  0.001416015
## ALT.12                             0.019046099  0.0099728383 -0.059658524
## ALT.24                             0.002711517 -0.0159186857  0.007909135
## ALT.36                            -0.008720825  0.0060865959 -0.029995451
## ALT.48                             0.027556595 -0.0203239599 -0.007866295
## ALT.after.24.w                     0.003865167  0.0101835990 -0.021412112
## RNA.Base                           0.022775156 -0.0133707332 -0.016418030
## RNA.4                             -0.012699117 -0.0233148008  0.036863599
## RNA.12                             0.001578231 -0.0313789591 -0.009189913
## RNA.EOT                           -0.048702492  0.0352722375 -0.021427851
## RNA.EF                            -0.030296896 -0.0152924344 -0.044901060
## Baseline.histological.Grading     -0.040705291  0.0145587520 -0.023045099
## Baselinehistological.staging      -0.019599169  0.0119553382 -0.057258857
```

Principal Component Analysis

```
> hcv_pca <-prcomp(HCV[,1:29],scale=TRUE)
> hcv_pca
```

Standard deviations (1, .., p=29):
[1] 1.3769419 1.1174363 1.1113462 1.0916269 1.0782624 1.0707535 1.0626130
1.0521221 1.0306635 1.0292621
[11] 1.0172925 1.0141324 1.0121337 1.0049132 0.9926278 0.9870471 0.9805851
0.9792502 0.9631161 0.9530072
[21] 0.9405476 0.9304068 0.9283507 0.9201871 0.9127624 0.8919466 0.8754930
0.7514594 0.7276547

OUTPUT:

```
Rotation (n x k) = (29 x 29):
                                          PC1          PC2         PC3         PC4          PC5
Age                                0.043992545 -0.035109552  0.15542945 -0.21681173  0.1846420065
Gender                             0.008756739 -0.031538265  0.01225109  0.11432799 -0.2272447661
BMI                                0.059438630  0.004938750 -0.21133076  0.43507918  0.2105290684
Fever                             -0.018638143 -0.016959013  0.14647417  0.11272529 -0.0194418421
Nausea.Vomting                    -0.018052038  0.523815972  0.02931082  0.03619724  0.0100755834
Headache                          -0.007027392  0.124936603  0.20495296 -0.22055035  0.0328136750
Diarrhea                          -0.037779605  0.112284699  0.31055207  0.09505129  0.1941946991
Fatigue...generalized.bone.ache   -0.039589819 -0.113283136  0.02431225  0.03182207 -0.5018034613
Jaundice                          -0.060924334 -0.039975480  0.15722114 -0.28037289  0.0021353121
Epigastric.pain                   -0.091593158 -0.314657318  0.08262704  0.13208431 -0.1562642350
WBC                                0.079856632 -0.207631406 -0.09056514  0.08357598  0.0941100088
RBC                                0.062330494  0.242506719 -0.14413534 -0.04346368 -0.0591444285
HGB                                0.012151709  0.030902639 -0.39565562  0.16325012  0.0113733042
Plat                              -0.051058807  0.378757510  0.01108177  0.28487255 -0.1382597553
AST.1                              0.028793449  0.013713351  0.04919642  0.04294564 -0.1320669862
ALT.1                              0.018974095  0.262467291  0.39485581  0.25382822 -0.0368534659
ALT4                              -0.040780344  0.034435407 -0.08938033 -0.27015280  0.4057377396
ALT.12                             0.047599395 -0.157661272 -0.16534036 -0.38085696 -0.2283790159
ALT.24                            -0.018567737  0.212800327 -0.32270954 -0.24975390  0.0210555092
ALT.36                            -0.054830139  0.088193685  0.08694682 -0.12844262 -0.2103541059
ALT.48                            -0.026185451 -0.255155588  0.31754417  0.06703057  0.0024901089
ALT.after.24.w                    -0.019391592 -0.145425335  0.17697872  0.01737588  0.0255244283
RNA.Base                          -0.013683156 -0.002907261  0.30619803 -0.11191128  0.1894906165
RNA.4                              0.056900757 -0.183428054 -0.04195740  0.16586510  0.3981169569
RNA.12                            -0.552679994  0.001687528 -0.02336655  0.03313714  0.0093530466
RNA.EOT                           -0.563868495 -0.039526114 -0.07119686  0.03930584  0.0771614060
RNA.EF                            -0.570538298 -0.011466450 -0.04305053 -0.00047804  0.0008019717
Baseline.histological.Grading      0.051971451 -0.140448509  0.00159462  0.13066428 -0.1768391380
Baselinehistological.staging      -0.032348367  0.213485043  0.10947878 -0.19354490 -0.1009923507
```

```
> summary(hcv_pca)
Importance of components:
                          PC1     PC2     PC3     PC4     PC5     PC6     PC7     PC8     PC9    PC10
Standard deviation     1.37694 1.11744 1.11135 1.09163 1.07826 1.07075 1.06261 1.05212 1.03066 1.02926
Proportion of Variance 0.06538 0.04306 0.04259 0.04109 0.04009 0.03953 0.03894 0.03817 0.03663 0.03653
Cumulative Proportion  0.06538 0.10844 0.15102 0.19212 0.23221 0.27174 0.31068 0.34885 0.38548 0.42201
                          PC11    PC12    PC13    PC14    PC15   PC16    PC17    PC18    PC19    PC20
Standard deviation     1.01729 1.01413 1.01213 1.00491 0.99263 0.9870 0.98059 0.97925 0.96312 0.95301
Proportion of Variance 0.03569 0.03546 0.03532 0.03482 0.03398 0.0336 0.03316 0.03307 0.03199 0.03132
Cumulative Proportion  0.45770 0.49316 0.52848 0.56331 0.59728 0.6309 0.66404 0.69710 0.72909 0.76041
                          PC21    PC22    PC23   PC24    PC25    PC26    PC27    PC28    PC29
Standard deviation     0.9405 0.93041 0.92835 0.9202 0.91276 0.89195 0.87549 0.75146 0.72765
Proportion of Variance 0.0305 0.02985 0.02972 0.0292 0.02873 0.02743 0.02643 0.01947 0.01826
Cumulative Proportion  0.7909 0.82076 0.85048 0.8797 0.90841 0.93584 0.96227 0.98174 1.00000
>
```

```
> (eigen_hcv <- hcv_pca$sdev^2)
 [1] 1.8959690 1.2486640 1.2350903 1.1916494 1.1626498 1.1465131 1.1291464 1.1069609 1.0622673 1.0593804
[11] 1.0348841 1.0284646 1.0244145 1.0098506 0.9853100 0.9742619 0.9615471 0.9589310 0.9275925 0.9082227
[21] 0.8846297 0.8656567 0.8618350 0.8467443 0.8331352 0.7955688 0.7664880 0.5646913 0.5294814
> names(eigen_hcv) <- paste("PC",1:29,sep="")
> eigen_hcv
      PC1       PC2       PC3       PC4       PC5       PC6       PC7       PC8       PC9      PC10
1.8959690 1.2486640 1.2350903 1.1916494 1.1626498 1.1465131 1.1291464 1.1069609 1.0622673 1.0593804
     PC11      PC12      PC13      PC14      PC15      PC16      PC17      PC18      PC19      PC20
1.0348841 1.0284646 1.0244145 1.0098506 0.9853100 0.9742619 0.9615471 0.9589310 0.9275925 0.9082227
     PC21      PC22      PC23      PC24      PC25      PC26      PC27      PC28      PC29
0.8846297 0.8656567 0.8618350 0.8467443 0.8331352 0.7955688 0.7664880 0.5646913 0.5294814
> sumlambdas <- sum(eigen_hcv)
> sumlambdas
[1] 29
> propvar <- eigen_hcv/sumlambdas
> propvar
       PC1        PC2        PC3        PC4        PC5        PC6        PC7        PC8        PC9
0.06537824 0.04305738 0.04258932 0.04109136 0.04009137 0.03953493 0.03893608 0.03817107 0.03662991
      PC10       PC11       PC12       PC13       PC14       PC15       PC16       PC17       PC18
0.03653036 0.03568566 0.03546430 0.03532464 0.03482243 0.03397621 0.03359524 0.03315680 0.03306659
      PC19       PC20       PC21       PC22       PC23       PC24       PC25       PC26       PC27
0.03198595 0.03131802 0.03050447 0.02985023 0.02971845 0.02919808 0.02872880 0.02743341 0.02643062
      PC28       PC29
0.01947211 0.01825798
> cumvar_hcv <- cumsum(propvar)
> cumvar_hcv
       PC1        PC2        PC3        PC4        PC5        PC6        PC7        PC8        PC9
0.06537824 0.10843562 0.15102494 0.19211630 0.23220767 0.27174261 0.31067869 0.34884975 0.38547966
      PC10       PC11       PC12       PC13       PC14       PC15       PC16       PC17       PC18
0.42201002 0.45769568 0.49315997 0.52848461 0.56330705 0.59728325 0.63087849 0.66403529 0.69710187
      PC19       PC20       PC21       PC22       PC23       PC24       PC25       PC26       PC27
0.72908782 0.76040585 0.79091032 0.82076055 0.85047900 0.87967708 0.90840588 0.93583929 0.96226991
      PC28       PC29
0.98174202 1.00000000
>
```

```
> matlambdas <- rbind(eigen_hcv,propvar,cumvar_hcv)
> rownames(matlambdas) <- c("Eigenvalues","Prop. variance","Cum. prop. variance")
> round(matlambdas,29)
                           PC1        PC2        PC3        PC4        PC5        PC6        PC7        PC8
Eigenvalues          1.89596904 1.24866396 1.23509034 1.19164939 1.16264975 1.14651308 1.12914639 1.10696093
Prop. variance       0.06537824 0.04305738 0.04258932 0.04109136 0.04009137 0.03953493 0.03893608 0.03817107
Cum. prop. variance  0.06537824 0.10843562 0.15102494 0.19211630 0.23220767 0.27174261 0.31067869 0.34884975
                           PC9       PC10       PC11       PC12       PC13       PC14       PC15       PC16
Eigenvalues          1.06226726 1.05938037 1.03488407 1.0284646 1.02441453 1.00985059 0.98530999 0.97426192
Prop. variance       0.03662991 0.03653036 0.03568566 0.0354643 0.03532464 0.03482243 0.03397621 0.03359524
Cum. prop. variance  0.38547966 0.42201002 0.45769568 0.4931600 0.52848461 0.56330705 0.59728325 0.63087849
                          PC17       PC18       PC19       PC20       PC21       PC22       PC23       PC24
Eigenvalues          0.9615471 0.95893098 0.92759253 0.90822269 0.88462975 0.86565673 0.86183502 0.84674429
Prop. variance       0.0331568 0.03306659 0.03198595 0.03131802 0.03050447 0.02985023 0.02971845 0.02919808
Cum. prop. variance  0.6640353 0.69710187 0.72908782 0.76040585 0.79091032 0.82076055 0.85047900 0.87967708
                          PC25       PC26       PC27       PC28       PC29
Eigenvalues          0.8331352 0.79556880 0.76648805 0.56469127 0.52948135
Prop. variance       0.0287288 0.02743341 0.02643062 0.01947211 0.01825798
Cum. prop. variance  0.9084059 0.93583929 0.96226991 0.98174202 1.00000000
> summary(hcv_pca)
Importance of components:
                          PC1     PC2     PC3     PC4     PC5     PC6     PC7     PC8     PC9    PC10
Standard deviation     1.37694 1.11744 1.11135 1.09163 1.07826 1.07075 1.06261 1.05212 1.03066 1.02926
Proportion of Variance 0.06538 0.04306 0.04259 0.04109 0.04009 0.03953 0.03894 0.03817 0.03663 0.03653
Cumulative Proportion  0.06538 0.10844 0.15102 0.19212 0.23221 0.27174 0.31068 0.34885 0.38548 0.42201
                         PC11    PC12    PC13    PC14    PC15   PC16    PC17    PC18    PC19    PC20
Standard deviation     1.01729 1.01413 1.01213 1.00491 0.99263 0.9870 0.98059 0.97925 0.96312 0.95301
Proportion of Variance 0.03569 0.03546 0.03532 0.03482 0.03398 0.0336 0.03316 0.03307 0.03199 0.03132
Cumulative Proportion  0.45770 0.49316 0.52848 0.56331 0.59728 0.6309 0.66404 0.69710 0.72909 0.76041
                         PC21    PC22    PC23   PC24    PC25    PC26    PC27    PC28    PC29
Standard deviation     0.9405 0.93041 0.92835 0.9202 0.91276 0.89195 0.87549 0.75146 0.72765
Proportion of Variance 0.0305 0.02985 0.02972 0.0292 0.02873 0.02743 0.02643 0.01947 0.01826
Cumulative Proportion  0.7909 0.82076 0.85048 0.8797 0.90841 0.93584 0.96227 0.98174 1.00000
>
```

```
> hcv_pca$rotation
                                          PC1          PC2         PC3         PC4          PC5
Age                                0.043992545 -0.035109552  0.15542945 -0.21681173  0.1846420065
Gender                             0.008756739 -0.031538265  0.01225109  0.11432799 -0.2272447661
BMI                                0.059438630  0.004938750 -0.21133076  0.43507918  0.2105290684
Fever                             -0.018638143 -0.016959013  0.14647417  0.11272529 -0.0194418421
Nausea.Vomting                    -0.018052038  0.523815972  0.02931082  0.03619724  0.0100755834
Headache                          -0.007027392  0.124936603  0.20495296 -0.22055035  0.0328136750
Diarrhea                          -0.037779605  0.112284699  0.31055207  0.09505129  0.1941946991
Fatigue...generalized.bone.ache   -0.039589819 -0.113283136  0.02431225  0.03182207 -0.5018034613
Jaundice                          -0.060924334 -0.039975480  0.15722114 -0.28037289  0.0021353121
Epigastric.pain                   -0.091593158 -0.314657318  0.08262704  0.13208431 -0.1562642350
WBC                                0.079856632 -0.207631406 -0.09056514  0.08357598  0.0941100088
RBC                                0.062330494  0.242506719 -0.14413534 -0.04346368 -0.0591444285
HGB                                0.012151709  0.030902639 -0.39565562  0.16325012  0.0113733042
Plat                              -0.051058807  0.378757510  0.01108177  0.28487255 -0.1382597553
AST.1                              0.028793449  0.013713351  0.04919642  0.04294564 -0.1320669862
ALT.1                              0.018974095  0.262467291  0.39485581  0.25382822 -0.0368534659
ALT4                              -0.040780344  0.034435407 -0.08938033 -0.27015280  0.4057377396
ALT.12                             0.047599395 -0.157661272 -0.16534036 -0.38085696 -0.2283790159
ALT.24                            -0.018567737  0.212800327 -0.32270954 -0.24975390  0.0210555092
ALT.36                            -0.054830139  0.088193685  0.08694682 -0.12844262 -0.2103541059
ALT.48                            -0.026185451 -0.255155588  0.31754417  0.06703057  0.0024901089
ALT.after.24.w                    -0.019391592 -0.145425335  0.17697872  0.01737588  0.0255244283
RNA.Base                          -0.013683156 -0.002907261  0.30619803 -0.11191128  0.1894906165
RNA.4                              0.056900757 -0.183428054 -0.04195740  0.16586510  0.3981169569
RNA.12                            -0.552679994  0.001687528 -0.02336655  0.03313714  0.0093530466
RNA.EOT                           -0.563868495 -0.039526114 -0.07119686  0.03930584  0.0771614060
RNA.EF                            -0.570538298 -0.011466450 -0.04305053 -0.00047804  0.0008019717
Baseline.histological.Grading      0.051971451 -0.140448509  0.00159462  0.13066428 -0.1768391380
Baselinehistological.staging      -0.032348367  0.213485043  0.10947878 -0.19354490 -0.1009923507
```

```
> tabmeansPC <- aggregate(hcvtyp_pca[,2:30],by=list(Survivorship=HCV$Survivorship),mean)
> tabmeansPC
  Survivorship        PC1         PC2         PC3         PC4         PC5         PC6         PC7
1            C  0.6123744  0.02475479  0.06579593 -0.03430532 -0.04507313  0.01425565 -0.04281583
2           NC -1.1472491 -0.04637671 -0.12326498  0.06426909  0.08444199 -0.02670717  0.08021306
           PC8         PC9        PC10         PC11        PC12         PC13        PC14        PC15
1 -0.004890105  0.06224251 -0.05949025 -0.004544576 -0.03920054 -0.002755935 -0.02361107  0.02266102
2  0.009161338 -0.11660786  0.11145164  0.008514008  0.07344002  0.005163090  0.04423401 -0.04245415
         PC16        PC17         PC18        PC19         PC20        PC21        PC22         PC23
1  0.02326687  0.05588213  0.002161106  0.02975834 -0.002715703 -0.04697310 -0.01407614 -0.001335249
2 -0.04358918 -0.10469204 -0.004048710 -0.05575059  0.005087717  0.08800146  0.02637085  0.002501515
          PC24        PC25        PC26        PC27       PC28       PC29
1 -0.008750238  0.01021780 -0.02426390 -0.02104547  0.2054995 -0.2452732
2  0.016393081 -0.01914247  0.04545705  0.03942750 -0.3849919  0.4595055
> |
```

```
> tabmeansPC <- tabmeansPC[rev(order(tabmeansPC$Survivorship)),]
> tabmeansPC
  Survivorship        PC1         PC2         PC3         PC4         PC5         PC6         PC7
2           NC -1.1472491 -0.04637671 -0.12326498  0.06426909  0.08444199 -0.02670717  0.08021306
1            C  0.6123744  0.02475479  0.06579593 -0.03430532 -0.04507313  0.01425565 -0.04281583
           PC8         PC9        PC10         PC11        PC12         PC13        PC14         PC15
2  0.009161338 -0.11660786  0.11145164  0.008514008  0.07344002  0.005163090  0.04423401 -0.04245415
1 -0.004890105  0.06224251 -0.05949025 -0.004544576 -0.03920054 -0.002755935 -0.02361107  0.02266102
         PC16        PC17         PC18        PC19         PC20        PC21        PC22         PC23
2 -0.04358918 -0.10469204 -0.004048710 -0.05575059  0.005087717  0.08800146  0.02637085  0.002501515
1  0.02326687  0.05588213  0.002161106  0.02975834 -0.002715703 -0.04697310 -0.01407614 -0.001335249
          PC24        PC25        PC26        PC27       PC28       PC29
2  0.016393081 -0.01914247  0.04545705  0.03942750 -0.3849919  0.4595055
1 -0.008750238  0.01021780 -0.02426390 -0.02104547  0.2054995 -0.2452732
>
```

```
> tabfmeans <- t(tabmeansPC[,-1])
> tabfmeans
                2             1
PC1  -1.147249113  0.612374388
PC2  -0.046376709  0.024754788
PC3  -0.123264984  0.065795927
PC4   0.064269094 -0.034305319
PC5   0.084441992 -0.045073134
PC6  -0.026707165  0.014255652
PC7   0.080213059 -0.042815830
PC8   0.009161338 -0.004890105
PC9  -0.116607855  0.062242510
PC10  0.111451643 -0.059490246
PC11  0.008514008 -0.004544576
PC12  0.073440022 -0.039200543
PC13  0.005163090 -0.002755935
PC14  0.044234009 -0.023611066
PC15 -0.042454146  0.022661017
PC16 -0.043589184  0.023266873
PC17 -0.104692043  0.055882132
PC18 -0.004048710  0.002161106
PC19 -0.055750587  0.029758342
PC20  0.005087717 -0.002715703
PC21  0.088001462 -0.046973095
PC22  0.026370853 -0.014076137
PC23  0.002501515 -0.001335249
PC24  0.016393081 -0.008750238
PC25 -0.019142468  0.010217796
PC26  0.045457051 -0.024263896
PC27  0.039427504 -0.021045467
PC28 -0.384991894  0.205499549
PC29  0.459505511 -0.245273152
>
```

```
> colnames(tabfmeans) <- t(as.vector(tabmeansPC[1]))
> tabfmeans
                NC             C
PC1  -1.147249113  0.612374388
PC2  -0.046376709  0.024754788
PC3  -0.123264984  0.065795927
PC4   0.064269094 -0.034305319
PC5   0.084441992 -0.045073134
PC6  -0.026707165  0.014255652
PC7   0.080213059 -0.042815830
PC8   0.009161338 -0.004890105
PC9  -0.116607855  0.062242510
PC10  0.111451643 -0.059490246
PC11  0.008514008 -0.004544576
PC12  0.073440022 -0.039200543
PC13  0.005163090 -0.002755935
PC14  0.044234009 -0.023611066
PC15 -0.042454146  0.022661017
PC16 -0.043589184  0.023266873
PC17 -0.104692043  0.055882132
PC18 -0.004048710  0.002161106
PC19 -0.055750587  0.029758342
PC20  0.005087717 -0.002715703
PC21  0.088001462 -0.046973095
PC22  0.026370853 -0.014076137
PC23  0.002501515 -0.001335249
PC24  0.016393081 -0.008750238
PC25 -0.019142468  0.010217796
PC26  0.045457051 -0.024263896
PC27  0.039427504 -0.021045467
PC28 -0.384991894  0.205499549
PC29  0.459505511 -0.245273152
>
```

```
> tabsdsPC <- aggregate(hcvtyp_pca[,2:30],by=list(Survivorship=HCV$Survivorship),sd)
> tabfsds <- t(tabsdsPC[,-1])
> colnames(tabfsds) <- t(as.vector(tabsdsPC[1]))
> tabfsds
              C         NC
PC1  1.2298718 0.7719668
PC2  1.0967902 1.1548515
PC3  1.1099905 1.1045144
PC4  1.1015816 1.0709045
PC5  1.0739757 1.0823164
PC6  1.0599013 1.0914031
PC7  1.0790286 1.0274616
PC8  1.0597309 1.0387495
PC9  1.0180357 1.0450316
PC10 1.0153498 1.0468206
PC11 1.0196557 1.0138527
PC12 1.0407117 0.9591108
PC13 1.0061530 1.0242748
PC14 1.0201189 0.9752947
PC15 1.0217320 0.9352267
PC16 0.9916090 0.9779775
PC17 0.9823685 0.9696058
PC18 0.9770937 0.9842820
PC19 0.9740629 0.9407323
PC20 0.9417524 0.9747108
PC21 0.9223920 0.9685001
PC22 0.9380579 0.9162855
PC23 0.8997361 0.9806692
PC24 0.9241690 0.9134110
PC25 0.9266039 0.8868826
PC26 0.8808408 0.9115743
PC27 0.8760836 0.8739294
PC28 0.6719146 0.7418754
PC29 0.5891777 0.7402801
```

T-Tests

```
> t.test(PC1~HCV$Survivorship,data=hcvtyp_pca)

        Welch Two Sample t-test

data:  PC1 by HCV$Survivorship
t = 32.611, df = 1347.9, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.653773 1.865474
sample estimates:
 mean in group C mean in group NC
       0.6123744       -1.1472491

> t.test(PC2~HCV$Survivorship,data=hcvtyp_pca)

        Welch Two Sample t-test

data:  PC2 by HCV$Survivorship
t = 1.111, df = 939.52, p-value = 0.2669
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.0545162  0.1967792
sample estimates:
 mean in group C mean in group NC
      0.02475479      -0.04637671

> t.test(PC3~HCV$Survivorship,data=hcvtyp_pca)

        Welch Two Sample t-test

data:  PC3 by HCV$Survivorship
t = 3.0292, df = 986.5, p-value = 0.002516
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.06658236 0.31153946
sample estimates:
 mean in group C mean in group NC
      0.06579593       0.12326498
```

```
> t.test(PC4~HCV$Survivorship,data=hcvtyp_pca)

        Welch Two Sample t-test

data:  PC4 by HCV$Survivorship
t = -1.6155, df = 1006.5, p-value = 0.1065
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.21831093  0.02116211
sample estimates:
 mean in group C mean in group NC
     -0.03430532       0.06426909

> t.test(PC5~HCV$Survivorship,data=hcvtyp_pca)

        Welch Two Sample t-test

data:  PC5 by HCV$Survivorship
t = -2.127, df = 975.75, p-value = 0.03367
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.24900621 -0.01002404
sample estimates:
 mean in group C mean in group NC
     -0.04507313       0.08444199

>
```

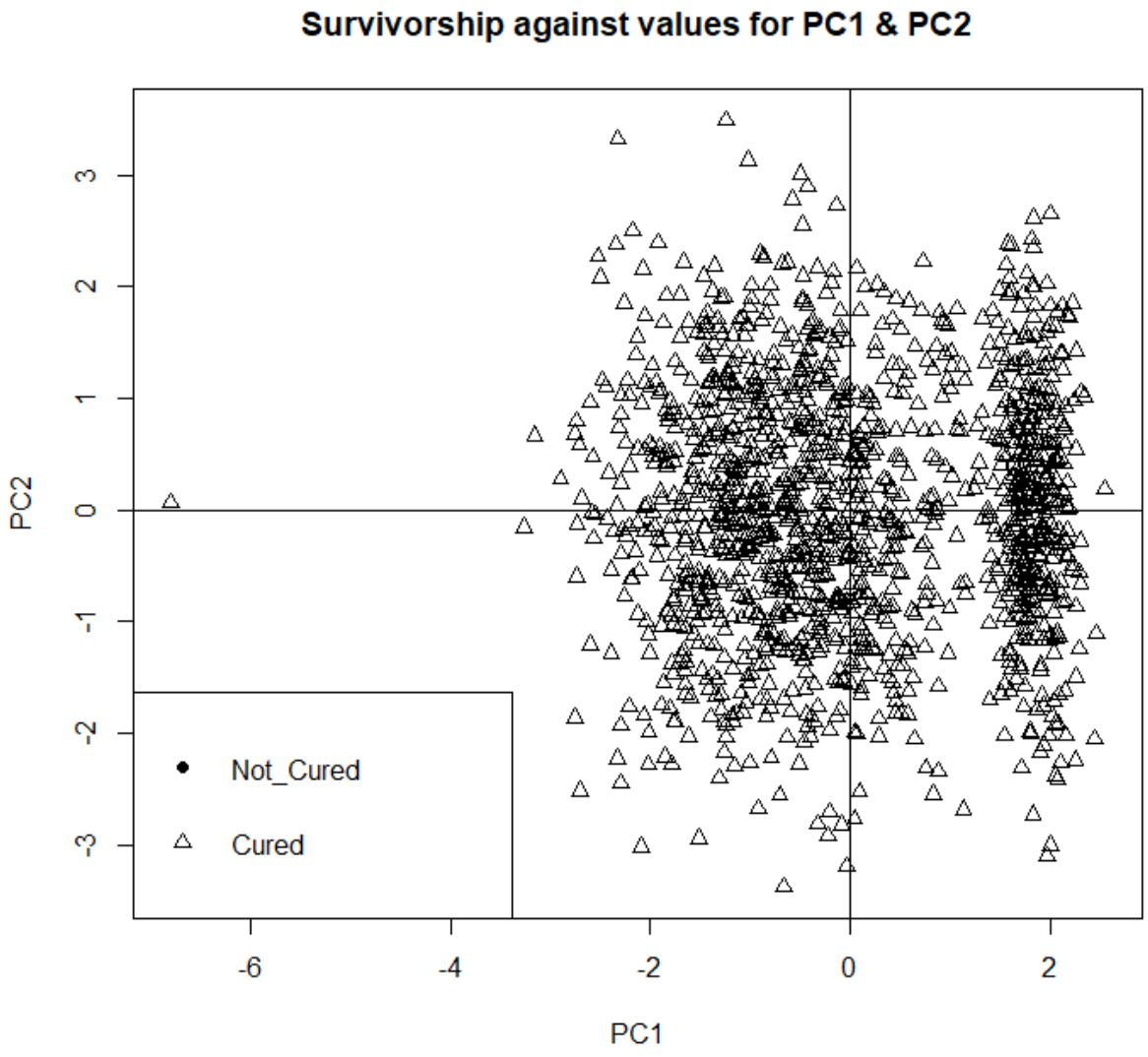
F-Ratio Tests

```
> # F ratio tests
> var.test(PC1~HCV$Survivorship,data=hcvtyp_pca)

        F test to compare two variances

data:  PC1 by HCV$Survivorship
F = 2.5382, num df = 902, denom df = 481, p-value < 2.2e-16
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 2.166058 2.962996
sample estimates:
ratio of variances
           2.53818

> var.test(PC2~HCV$Survivorship,data=hcvtyp_pca)

        F test to compare two variances

data:  PC2 by HCV$Survivorship
F = 0.90198, num df = 902, denom df = 481, p-value = 0.1915
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.7697373 1.0529399
sample estimates:
ratio of variances
         0.9019758

> var.test(PC3~HCV$Survivorship,data=hcvtyp_pca)

        F test to compare two variances

data:  PC3 by HCV$Survivorship
F = 1.0099, num df = 902, denom df = 481, p-value = 0.9079
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.8618732 1.1789746
sample estimates:
ratio of variances
          1.00994
```

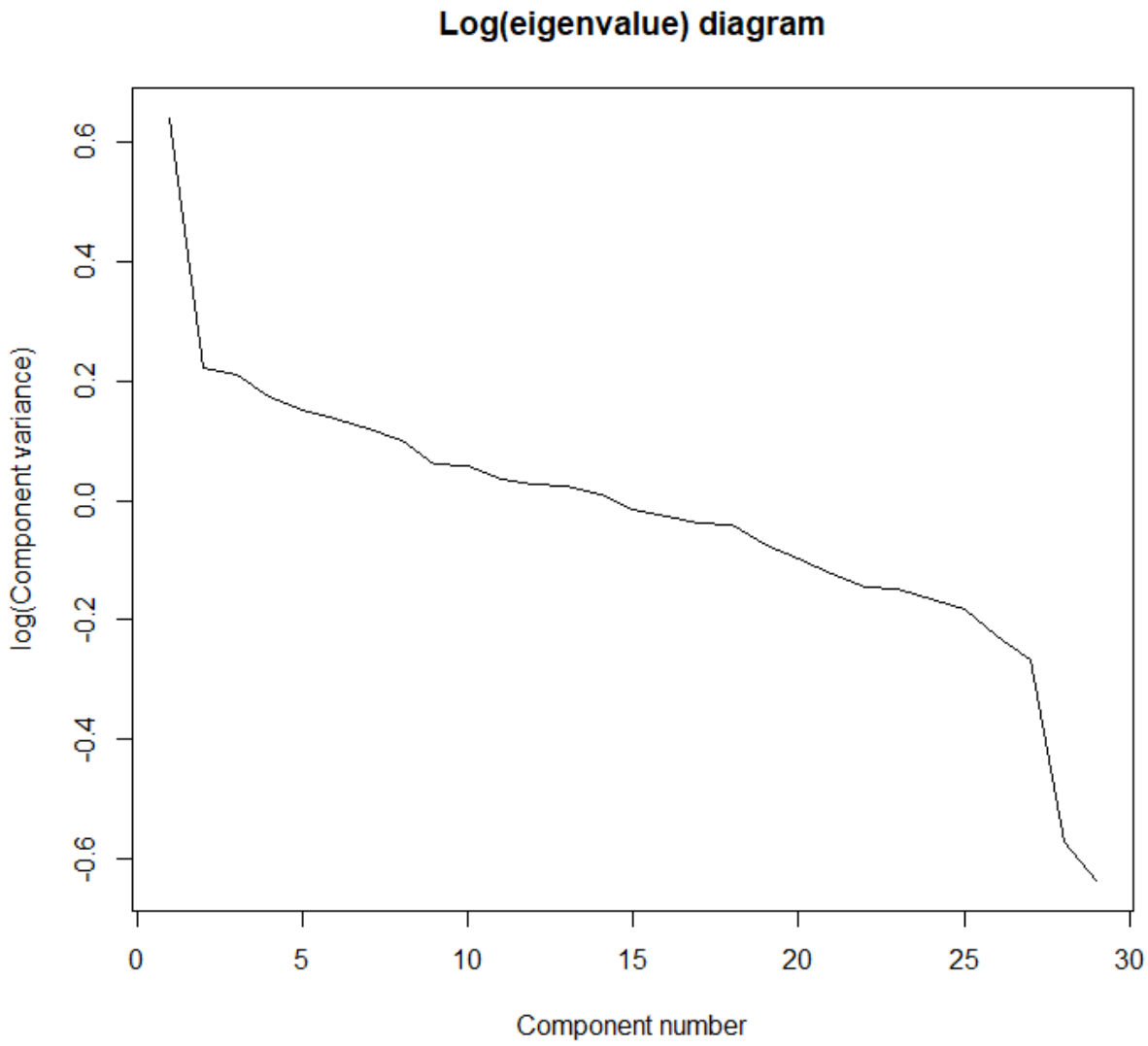
```
> var.test(PC4~HCV$Survivorship,data=hcvtyp_pca)

        F test to compare two variances

data:  PC4 by HCV$Survivorship
F = 1.0581, num df = 902, denom df = 481, p-value = 0.4858
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.9029828 1.2352093
sample estimates:
ratio of variances
          1.058113

> var.test(PC5~HCV$Survivorship,data=hcvtyp_pca)

        F test to compare two variances

data:  PC5 by HCV$Survivorship
F = 0.98465, num df = 902, denom df = 481, p-value = 0.8399
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.8402878 1.1494474
sample estimates:
ratio of variances
         0.9846467

>
```

Plotting the scores for the first and second components

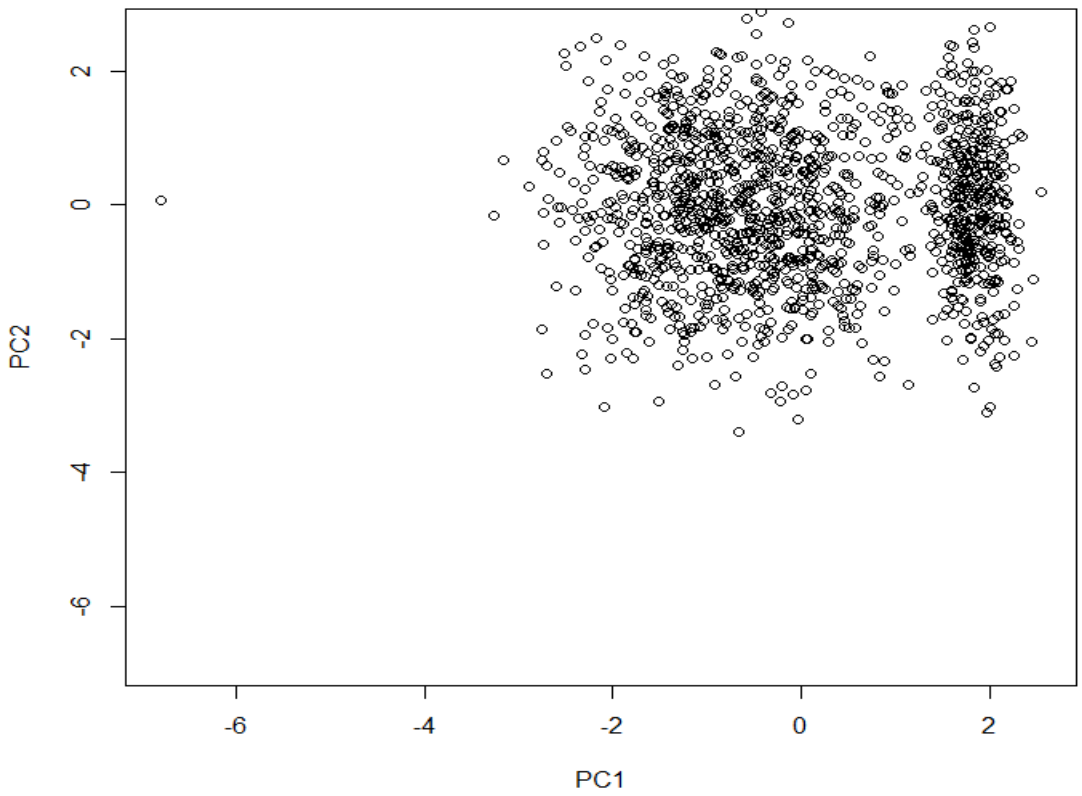
```
plot(hcvtyp_pca$PC1, hcvtyp_pca$PC2,pch=ifelse(hcvtyp_pca$Survivorship ==
"S",16,2),xlab="PC1", ylab="PC2", main="49 HCV against values for PC1 & PC2")
> abline(h=0)
> abline(v=0)
> legend("bottomleft", legend=c("Cured","Not_Cured"), pch=c(16,2),)
```



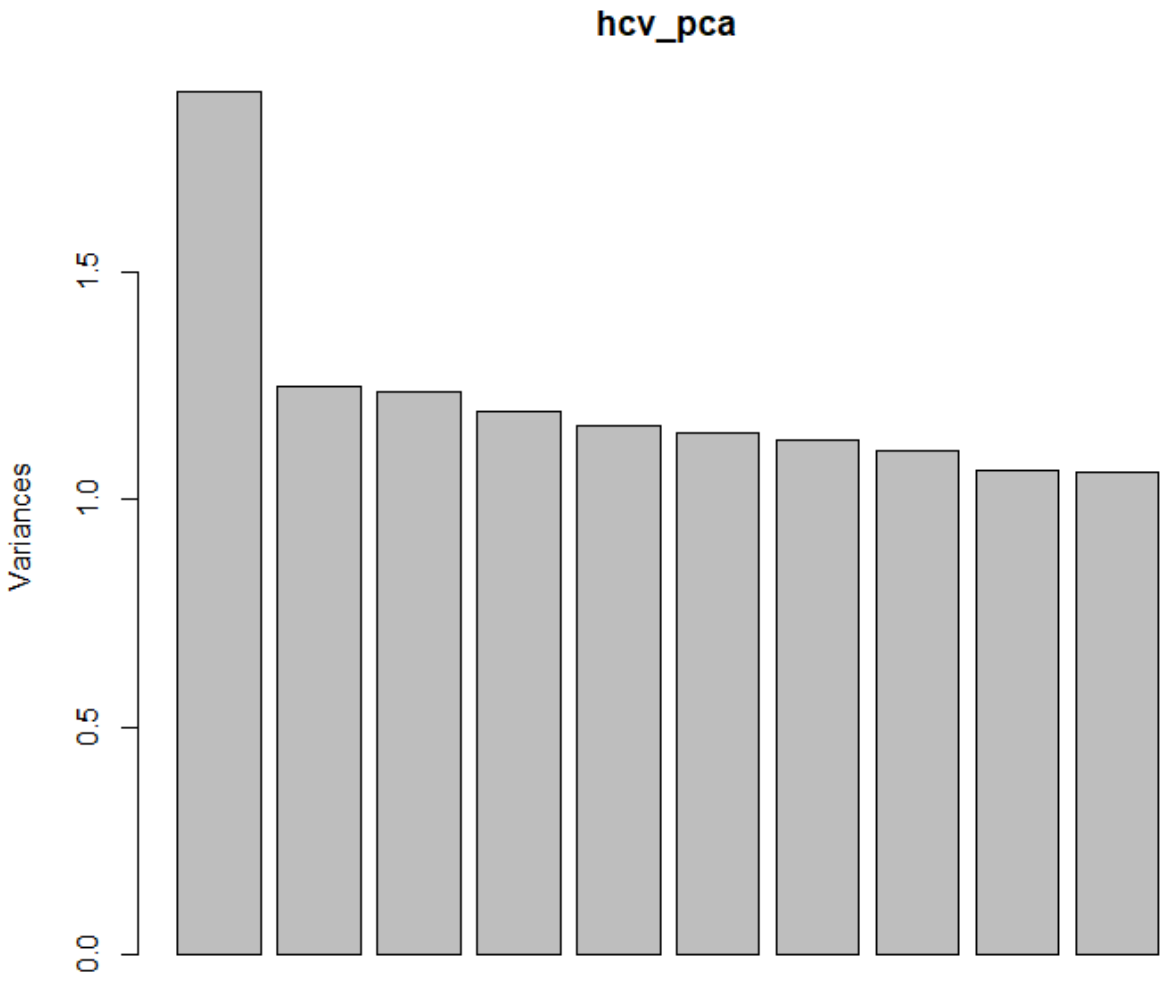
```
> plot(eigen_hcv, xlab = "Component number", ylab = "Component variance", type = "l", main = "Scree diagram")
> plot(log(eigen_hcv), xlab = "Component number",ylab = "log(Component variance)", type="l",main = "Log(eigenva
lue) diagram")
> print(summary(hcv_pca))
Importance of components:
                          PC1     PC2     PC3     PC4     PC5     PC6     PC7     PC8     PC9    PC10
Standard deviation     1.37694 1.11744 1.11135 1.09163 1.07826 1.07075 1.06261 1.05212 1.03066 1.02926
Proportion of Variance 0.06538 0.04306 0.04259 0.04109 0.04009 0.03953 0.03894 0.03817 0.03663 0.03653
Cumulative Proportion  0.06538 0.10844 0.15102 0.19212 0.23221 0.27174 0.31068 0.34885 0.38548 0.42201
                         PC11    PC12    PC13    PC14    PC15   PC16    PC17    PC18    PC19    PC20
Standard deviation     1.01729 1.01413 1.01213 1.00491 0.99263 0.9870 0.98059 0.97925 0.96312 0.95301
Proportion of Variance 0.03569 0.03546 0.03532 0.03482 0.03398 0.0336 0.03316 0.03307 0.03199 0.03132
Cumulative Proportion  0.45770 0.49316 0.52848 0.56331 0.59728 0.6309 0.66404 0.69710 0.72909 0.76041
                          PC21    PC22    PC23   PC24    PC25    PC26    PC27    PC28    PC29
Standard deviation     0.9405 0.93041 0.92835 0.9202 0.91276 0.89195 0.87549 0.75146 0.72765
Proportion of Variance 0.0305 0.02985 0.02972 0.0292 0.02873 0.02743 0.02643 0.01947 0.01826
Cumulative Proportion  0.7909 0.82076 0.85048 0.8797 0.90841 0.93584 0.96227 0.98174 1.00000
> 
```



```
> view(hcv_pca)
> diag(cov(hcv_pca$x))
      PC1       PC2       PC3       PC4       PC5       PC6       PC7       PC8       PC9      PC10
1.8959690 1.2486640 1.2350903 1.1916494 1.1626498 1.1465131 1.1291464 1.1069609 1.0622673 1.0593804
     PC11      PC12      PC13      PC14      PC15      PC16      PC17      PC18      PC19      PC20
1.0348841 1.0284646 1.0244145 1.0098506 0.9853100 0.9742619 0.9615471 0.9589310 0.9275925 0.9082227
     PC21      PC22      PC23      PC24      PC25      PC26      PC27      PC28      PC29
0.8846297 0.8656567 0.8618350 0.8467443 0.8331352 0.7955688 0.7664880 0.5646913 0.5294814
> xlim <- range(hcv_pca$x[,1])
> hcv_pca$x[,1]
   [1]  1.752028782 -0.125120259 -0.582399282 -2.085681979 -6.798188688  1.749113325 -0.590303483
   [8] -1.564717696  0.510283184 -0.588712397 -1.022320922 -0.962514349  1.707989062 -1.528713168
  [15] -0.049587298  0.432247099  1.852374878 -0.660406526 -1.127124155 -1.636602476 -1.088482877
  [22]  1.666271696  1.064152892 -0.798078912 -0.969386323  2.021606955 -2.013528283 -0.017980495
  [29]  1.760727933 -0.985329698  0.996146113  0.333530076 -1.780818574 -0.025899178 -1.417814487
  [36] -1.793894337 -0.390902443 -0.139473954  1.643574847 -0.669603195 -1.631512707  0.502948531
  [43]  0.588390510 -2.586019373  0.116808274 -0.524253243 -0.778339505 -0.320301025 -2.294294180
  [50] -0.281952226 -0.510799747  0.456407698 -0.422122534 -1.374782140 -0.329563302  0.099821119
  [57]  0.934665222 -0.438548488 -0.594170636  0.011562761  0.543354103 -1.773278006 -0.832846177
  [64] -2.127924184 -2.433998741 -0.427582983  0.734606924 -0.172389851 -0.842059433 -1.196245523
  [71] -1.179017676 -0.201355958 -0.579293636 -1.652600524 -1.469329779  1.320019225 -0.864879087
  [78] -1.170746846 -1.323100814 -0.025789209 -1.221369325 -0.488701234 -0.907145939 -0.766911350
  [85] -0.783371949  0.827179651 -0.606542502 -2.056249867 -1.799195517 -0.810106919  0.374473782
  [92] -1.337203727  1.750690841 -1.145751688 -1.534882120  2.075021466 -2.133537704  0.549400134
  [99]  2.045395929  0.925674949 -2.049845810 -1.420305157  1.666908549 -1.107860127  2.079193403
 [106] -0.500317409  1.980382756  1.978614381 -0.970399214  1.939636252  1.489137414 -2.288222057
 [113]  0.096747085 -2.310899960 -0.497826028  1.750726115  2.044411160 -2.295321121  0.823883639
 [120] -0.542571573 -0.033144043  1.657684088 -2.003172967  1.600796059  1.717756417  2.221271450
 [127] -2.604177500  1.817323300 -1.325819301  0.651781756 -1.287569137 -1.418202444 -1.180130962
 [134] -0.566850896 -0.100026377 -2.365308311 -2.072957921 -0.694508820 -1.222450327  0.185452086
 [141]  1.634829190 -1.799663398  0.074848618  1.574529284  1.898653759  0.957344290  0.158573101
 [148]  1.503510943 -1.873590833 -0.539976869  1.931465640 -1.573019968 -0.480779763  2.062468988
 [155]  1.891818100 -1.850684355 -1.914856651  1.972169814 -1.826473077  1.839397182  1.915152600
 [162]  0.135647552  0.112017231  0.521292370  1.681489545  0.110022689 -1.267206207 -0.031278772
 [169] -0.078879118  1.642842120  1.790709740 -2.551695792 -0.638398692  2.045832372  0.866146959
 [176]  2.115080577 -0.848224403 -0.310674491  1.562725642  1.773527100  2.058737931  1.894057037
 [183] -2.243139605 -0.580830939  1.107961882 -0.328978421 -0.056139128  0.436533736 -1.279777144
```



```
> plot(hcv_pca$x,xlim=xlim,ylim=xlim)
> hcv_pca$rotation[,1]
                                Age                              Gender
                       0.043992545                       0.008756739
                              Fever                     Nausea.Vomting
                      -0.018638143                      -0.018052038
                           Diarrhea Fatigue...generalized.bone.ache
                      -0.037779605                      -0.039589819
                     Epigastric.pain                               WBC
                      -0.091593158                       0.079856632
                                HGB                              Plat
                       0.012151709                      -0.051058807
                              ALT.1                              ALT4
                       0.018974095                      -0.040780344
                             ALT.24                            ALT.36
                      -0.018567737                      -0.054830139
                       ALT.after.24.w                          RNA.Base
                      -0.019391592                      -0.013683156
                             RNA.12                           RNA.EOT
                      -0.552679994                      -0.563868495
  Baseline.histological.Grading    Baselinehistological.staging
                       0.051971451                      -0.032348367
```



```
out <- sapply(1:5,
function(i){plot(HCV$Survivorship,hcv_pca$x[,i],xlab=paste("PC",i,sep=""),ylab="Survivorshi
p")})

pairs(hcv_pca$x[,1:5], ylim = c(-6,4),xlim = c(-
6,4),panel=function(x,y,...){text(x,y,HCV$Survivorship)})
```