Bioinformatics Statistics Practical

Pierre Wensel

06 September, 2024

Contents

```
# install.packages("epitools")
# install.packages("dplyr")
# BiocManager::install("multtest") # install bioconductor package "multtest" that contains "golub" data
# install.packages("outliers")
# install.packages("lmtest")
# install.packages("sandwich")
# install.packages("glmnet")
# BiocManager::install("ROCR")
# BiocManager::install("CMA")
# install.packages("randomForest")
# install.packages("survival")
# install.packages("KMsurv")
# install.packages("glmnet")
# install.packages("penalized")
# install.packages("PerformanceAnalytics")
# install.packages("corrr")
# install.packages("dplyr")
# install.packages("psych")
# install.packages("corrplot")
# install.packages("GGally")
# install.packages("ggcorrplot")
# BiocManager::install("multtest")
# install.packages(c("factoextra", "dendextend"))
# BiocManager::install("ComplexHeatmap")
# install.packages("caret")
# install.packages("FactoMineR")
# install.packages("klaR")
# install.packages("cba")
# install.packages("factoextra")
# install.packages("lmtest")
# install.packages("tidyverse")
# BiocManager::install("CMA")
# install.packages("randomForest")
# BiocManager::install("Biobase")
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
```

```
library(factoextra)
```

```
## Loading required package: ggplot2

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
library(glmnet)
```

```
## Loading required package: Matrix

## Loaded glmnet 4.1-8
```

```
library(survival)
library(KMsurv)
library(caret)
```

```
## Warning: package 'caret' was built under R version 4.4.1

## Loading required package: lattice

##
## Attaching package: 'caret'

## The following object is masked from 'package:survival':
##
##     cluster
```

```
library(klaR) #For kmode
```

```
## Warning: package 'klaR' was built under R version 4.4.1

## Loading required package: MASS

##
## Attaching package: 'MASS'

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

```r
library(cba) #For ROCK
```

```
## Warning: package 'cba' was built under R version 4.4.1
```

```
## Loading required package: grid
```

```
## Loading required package: proxy
```

```
##
## Attaching package: 'proxy'
```

```
## The following object is masked from 'package:Matrix':
##
##     as.matrix
```

```
## The following objects are masked from 'package:stats':
##
##     as.dist, dist
```

```
## The following object is masked from 'package:base':
##
##     as.matrix
```

```r
library(CMA)
```

```
## Loading required package: Biobase
```

```
## Loading required package: BiocGenerics
```

```
##
## Attaching package: 'BiocGenerics'
```

```
## The following objects are masked from 'package:dplyr':
##
##     combine, intersect, setdiff, union
```

```
## The following objects are masked from 'package:stats':
##
##     IQR, mad, sd, var, xtabs
```

```
## The following objects are masked from 'package:base':
##
##     anyDuplicated, aperm, append, as.data.frame, basename, cbind,
##     colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,
##     get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,
##     match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,
##     Position, rank, rbind, Reduce, rownames, sapply, setdiff, table,
##     tapply, union, unique, unsplit, which.max, which.min
```

```
## Welcome to Bioconductor
##
##     Vignettes contain introductory material; view with
##     'browseVignettes()'. To cite Bioconductor, see
##     'citation("Biobase")', and for packages 'citation("pkgname")'.

##
## Attaching package: 'CMA'

## The following objects are masked from 'package:caret':
##
##     best, rfe
```

```
library(Biobase)
library(randomForest)
```

```
## randomForest 4.7-1.1

## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'

## The following object is masked from 'package:Biobase':
##
##     combine

## The following object is masked from 'package:BiocGenerics':
##
##     combine

## The following object is masked from 'package:ggplot2':
##
##     margin

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

```
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 4.4.1

## corrplot 0.94 loaded
```

```
library(tinytex)
```

```
## Warning: package 'tinytex' was built under R version 4.4.1
```

QUESTION 1 Describe the main characteristics of the dataset: perform a univariate descriptive analysis of the first 6 variables

```r
#Input the text file using read.table, assigning the input to a variable pdata.
viral34 <- read.table("viral.34.txt", header=T, sep="")

#Dataframe confirmed
class(viral34)
```

```
## [1] "data.frame"
```

```r
str(viral34)
```

```
## 'data.frame':    140 obs. of  57 variables:
##  $ infection     : int  0 1 0 1 1 1 0 0 1 0 ...
##  $ stime         : num  7.3 6.72 7 9.33 3.44 ...
##  $ sind          : int  0 0 1 0 1 0 0 1 1 0 ...
##  $ gender        : int  0 1 1 0 1 0 1 1 0 0 ...
##  $ hosp          : int  1 0 1 1 0 0 0 0 1 0 ...
##  $ age           : int  47 47 38 45 31 41 48 47 38 44 ...
##  $ ancestry      : chr  "A" "A" "B" "A" ...
##  $ GSTM3         : num  0.1465 -0.0354 -0.2626 0.3379 0.0966 ...
##  $ RP5.860F19.3  : num  -0.098 -0.021 0.0108 0.3417 0.0782 ...
##  $ BBC3          : num  0.3082 -0.0964 0.0885 0.3277 -0.4061 ...
##  $ MMP9          : num  -0.2064 0.2415 -0.0258 -0.3414 -0.1786 ...
##  $ Contig35251_RC: num  -0.519 -0.532 -0.34 -0.626 -0.643 ...
##  $ Contig40831_RC: num  -0.11515 -0.00337 -0.08097 -0.12634 0.08476 ...
##  $ ALDH4A1       : num  0.1667 0.0985 0.0671 0.603 0.1597 ...
##  $ SERF1A        : num  -0.0838 0.1067 0.0627 -0.5238 -0.0179 ...
##  $ SCUBE2        : num  0.00606 0.09757 -0.12583 0.08587 -0.14412 ...
##  $ MTDH          : num  -0.1378 0.4942 0.0723 -0.577 -0.5194 ...
##  $ DCK           : num  -0.336 -0.58 0.139 -0.525 -0.197 ...
##  $ FLT1          : num  -0.0559 0.169 0.067 -0.0304 0.0635 ...
##  $ PECI.1        : num  -0.0878 -0.0968 -0.1134 0.1088 -0.1864 ...
##  $ QSCN6L1       : num  -0.1246 0.2656 0.0957 -0.0828 -0.082 ...
##  $ DIAPH3        : num  0.0808 0.1249 0.234 -0.1774 0.1007 ...
##  $ SLC2A3        : num  0.3686 0.4642 -0.0776 -0.2203 -0.0472 ...
##  $ GPR180        : num  -0.04521 -0.18754 -0.00541 0.11594 -0.11199 ...
##  $ RTN4RL1       : num  -0.1629 0.2001 0.1219 -0.6442 0.0282 ...
##  $ Contig32125_RC: num  -0.0133 -0.0133 -0.0788 -0.0236 -0.1042 ...
##  $ STK32B        : num  0.0278 0.1297 -0.062 0.0129 -0.1214 ...
##  $ EXT1          : num  -0.1345 -0.1956 -0.1134 0.0219 -0.3168 ...
##  $ COL4A2        : num  -0.0299 -0.2267 -0.2083 0.1027 -0.2578 ...
##  $ PECI          : num  0.1735 0.212 0.0423 0.4796 0.1005 ...
##  $ GNAZ          : num  0.0705 -0.0317 0.063 0.3349 -0.1467 ...
##  $ AYTL2         : num  0.2393 0.0157 -0.1276 0.5336 0.1171 ...
##  $ Contig63649_RC: num  0.02962 0.00419 0.0504 0.29442 0.02671 ...
##  $ RAB6B         : num  0.4614 0.0186 -0.1425 0.3254 -0.0873 ...
##  $ AA555029_RC   : num  -0.0481 0.1593 0.1142 -0.3106 -0.2201 ...
##  $ GPR126        : num  -0.1002 0.2812 0.0571 0.1912 -0.0826 ...
##  $ ECT2          : num  0.0354 -0.0377 -0.1813 0.0334 0.3287 ...
##  $ NUSAP1        : num  0.1098 0.0323 -0.0482 0.6553 0.0795 ...
##  $ GMPS          : num  0.2181 0.1857 0.0404 0.2371 0.1784 ...
##  $ UCHL5         : num  -0.0381 -0.2708 -0.0432 -0.1923 -0.1409 ...
##  $ ORC6L         : num  0.173 0.102 -0.15 0.193 0.126 ...
##  $ TSPYL5        : num  0.1559 -0.0588 0.0909 0.541 0.0456 ...
```

```
##  $ MELK          : num  -0.3458 -0.0108 -0.1366 -0.3397 -0.2484 ...
##  $ RUNDC1        : num  0.55836 -0.35885 0.12992 -0.07181 -0.00765 ...
##  $ DIAPH3.1      : num  -0.4446 -0.2426 -0.0564 -0.4456 -0.0968 ...
##  $ C16orf61      : num  0.0591 -0.0502 -0.2737 0.1355 0.2048 ...
##  $ TGFB3         : num  0.0818 0.1187 0.1132 -0.0205 0.1275 ...
##  $ FGF18         : num  -0.0482 0.2738 0.0347 0.1039 0.1898 ...
##  $ CDC42BPA      : num  0.192 0.1254 -0.2281 0.0328 -0.0739 ...
##  $ DTL           : num  -1.012 -0.146 0.448 -1.077 -0.843 ...
##  $ WISP1         : num  -0.00498 0.21792 0.07126 -0.44042 0.11942 ...
##  $ DIAPH3.2      : num  -0.29778 0.02057 -0.14414 0.05123 0.00824 ...
##  $ OXCT1         : num  0.0314 0.1633 0.0569 -0.2054 -0.139 ...
##  $ ZNF533        : num  0.8648 0.0158 -0.1476 -0.2065 0.2885 ...
##  $ RFC4          : num  0.0619 0.0169 -0.0543 0.3945 0.0624 ...
##  $ KNTC2         : num  0.5975 -0.3272 0.0965 0.046 -0.0926 ...
##  $ FBXO31        : num  -0.0414 -0.1352 0.0352 0.0607 0.264 ...
```

```
#There are 140 observations (rows), 57 variables (columns)
dim(viral34)
```

```
## [1] 140  57
```

```
nrow(viral34)
```

```
## [1] 140
```

```
ncol(viral34)
```

```
## [1] 57
```

```
names(viral34)
```

```
##  [1] "infection"      "stime"          "sind"           "gender"
##  [5] "hosp"           "age"            "ancestry"       "GSTM3"
##  [9] "RP5.860F19.3"   "BBC3"           "MMP9"           "Contig35251_RC"
## [13] "Contig40831_RC" "ALDH4A1"        "SERF1A"         "SCUBE2"
## [17] "MTDH"           "DCK"            "FLT1"           "PECI.1"
## [21] "QSCN6L1"        "DIAPH3"         "SLC2A3"         "GPR180"
## [25] "RTN4RL1"        "Contig32125_RC" "STK32B"         "EXT1"
## [29] "COL4A2"         "PECI"           "GNAZ"           "AYTL2"
## [33] "Contig63649_RC" "RAB6B"          "AA555029_RC"    "GPR126"
## [37] "ECT2"           "NUSAP1"         "GMPS"           "UCHL5"
## [41] "ORC6L"          "TSPYL5"         "MELK"           "RUNDC1"
## [45] "DIAPH3.1"       "C16orf61"       "TGFB3"          "FGF18"
## [49] "CDC42BPA"       "DTL"            "WISP1"          "DIAPH3.2"
## [53] "OXCT1"          "ZNF533"         "RFC4"           "KNTC2"
## [57] "FBXO31"
```

```
head(viral34)
```

```
##   infection     stime sind gender hosp age ancestry       GSTM3 RP5.860F19.3
## 1         0  7.296372    0      0    1  47        A  0.14647630  -0.09803689
```

```
## 2         1  6.718686    0      1    0  47        A -0.03543524  -0.02103562
## 3         0  6.995209    1      1    1  38        B -0.26258909   0.01080372
## 4         1  9.330595    0      0    1  45        A  0.33787726   0.34173748
## 5         1  3.438741    1      1    0  31        A  0.09657176   0.07818674
## 6         1 15.329227    0      0    0  41        A -0.21568976  -0.02222821
##          BBC3        MMP9 Contig35251_RC Contig40831_RC     ALDH4A1      SERF1A
## 1  0.30821656 -0.20635196     -0.5190545   -0.115149133  0.16674915 -0.08378990
## 2 -0.09643536  0.24147416     -0.5319210   -0.003368632  0.09845308  0.10674758
## 3  0.08854258 -0.02584139     -0.3400320   -0.080972535  0.06714804  0.06265814
## 4  0.32773524 -0.34135513     -0.6258146   -0.126344331  0.60304554 -0.52384308
## 5 -0.40614429 -0.17861930     -0.6432336    0.084764382  0.15974585 -0.01786414
## 6  0.25438095  0.22220903      0.5438846    0.155627613 -0.11070614 -0.15845941
##         SCUBE2        MTDH        DCK        FLT1      PECI.1     QSCN6L1
## 1  0.006056118 -0.13776980 -0.3355029 -0.05592013 -0.08777178 -0.12461618
## 2  0.097569969  0.49423377 -0.5800370  0.16900027 -0.09680589  0.26558747
## 3 -0.125834721  0.07226344  0.1389293  0.06697535 -0.11336268  0.09573915
## 4  0.085869554 -0.57695111 -0.5250435 -0.03036967  0.10875631 -0.08282874
## 5 -0.144122253 -0.51943819 -0.1974320  0.06349948 -0.18639378 -0.08196639
## 6  0.003581842  0.01179057 -0.5446215 -0.01398411 -0.29762197  0.25657503
##        DIAPH3      SLC2A3       GPR180     RTN4RL1 Contig32125_RC      STK32B
## 1  0.08084737  0.36857538 -0.045212204 -0.16287206    -0.01326674  0.02778352
## 2  0.12485686  0.46424057 -0.187539926  0.20005740    -0.01326124  0.12973975
## 3  0.23404700 -0.07758043 -0.005409022  0.12189144    -0.07876585 -0.06204672
## 4 -0.17743971 -0.22028177  0.115940757 -0.64421170    -0.02356354  0.01292946
## 5  0.10069529 -0.04723501 -0.111992419  0.02815844    -0.10416021 -0.12139572
## 6 -0.03541076 -0.05243454  0.117939467  0.38596644    -0.16397529 -0.11000363
##          EXT1      COL4A2        PECI        GNAZ       AYTL2 Contig63649_RC
## 1 -0.13445797 -0.02990369  0.17349207  0.07047200  0.23928807    0.029619073
## 2 -0.19559405 -0.22673707  0.21204866 -0.03168261  0.01566120    0.004188184
## 3 -0.11337939 -0.20833506  0.04232249  0.06296012 -0.12755049    0.050402671
## 4  0.02194074  0.10271000  0.47963136  0.33487679  0.53361497    0.294420521
## 5 -0.31679313 -0.25780916  0.10048933 -0.14666447  0.11709402    0.026705563
## 6 -0.12637540 -0.35700047 -0.10518681 -0.20496584 -0.05130563   -0.303730416
##         RAB6B AA555029_RC      GPR126        ECT2      NUSAP1        GMPS
## 1  0.46141386 -0.04808210 -0.10022007  0.03544526  0.10981625  0.21805322
## 2  0.01856611  0.15926624  0.28115470 -0.03772432  0.03225047  0.18573594
## 3 -0.14251272  0.11420782  0.05710594 -0.18130437 -0.04820767  0.04043471
## 4  0.32544391 -0.31064082  0.19116150  0.03338104  0.65528392  0.23712422
## 5 -0.08731065 -0.22007577 -0.08256859  0.32874172  0.07952344  0.17836256
## 6 -0.27515282  0.05199264 -0.16971181 -0.21921949 -0.27156456 -0.33843161
##         UCHL5      ORC6L      TSPYL5        MELK       RUNDC1    DIAPH3.1
## 1 -0.03809348  0.1728180  0.15589646 -0.34581318  0.558363335 -0.44455761
## 2 -0.27078994  0.1017376 -0.05882551 -0.01081727 -0.358850367 -0.24259102
## 3 -0.04321627 -0.1501144  0.09089323 -0.13659874  0.129923754 -0.05644101
## 4 -0.19231373  0.1926975  0.54098979 -0.33968909 -0.071808178 -0.44559038
## 5 -0.14087134  0.1256798  0.04560219 -0.24841783 -0.007654665 -0.09683179
## 6  0.13085033 -0.3269674 -0.21852037 -0.13351906 -0.495218907  0.20965383
##      C16orf61       TGFB3       FGF18    CDC42BPA        DTL        WISP1
## 1  0.05912505  0.08180754 -0.04819787  0.19203352 -1.0115741 -0.004976858
## 2 -0.05018147  0.11869773  0.27382112  0.12535125 -0.1460709  0.217921322
## 3 -0.27369996  0.11315389  0.03470079 -0.22807824  0.4482309  0.071255809
## 4  0.13548833 -0.02046981  0.10391225  0.03281893 -1.0765161 -0.440424509
## 5  0.20482626  0.12745233  0.18982466 -0.07390664 -0.8427116  0.119424302
## 6 -0.17109518  0.26659311 -0.19909908 -0.26596006  0.7104907 -0.166179324
```

```
##       DIAPH3.2       OXCT1      ZNF533        RFC4       KNTC2      FBXO31
## 1 -0.297776741  0.03135030  0.86482037  0.06185603  0.59748058 -0.04140661
## 2  0.020572007  0.16334775  0.01575178  0.01687964 -0.32724674 -0.13521580
## 3 -0.144136303  0.05694880 -0.14760060 -0.05427720  0.09654722  0.03522958
## 4  0.051227778 -0.20543872 -0.20651322  0.39446170  0.04598343  0.06070769
## 5  0.008235576 -0.13898008  0.28849255  0.06241235 -0.09261277  0.26401621
## 6  0.121529650  0.08704478 -0.15575775 -0.19665862 -0.18365899 -0.01361086
```

```
summary(viral34)
```

```
##    infection          stime              sind            gender
##  Min.   :0.0000   Min.   : 0.05476   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.: 4.69541   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :1.0000   Median : 6.96235   Median :0.0000   Median :1.0000
##  Mean   :0.5071   Mean   : 7.35621   Mean   :0.3357   Mean   :0.5571
##  3rd Qu.:1.0000   3rd Qu.:10.05681   3rd Qu.:1.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :17.65914   Max.   :1.0000   Max.   :1.0000
##       hosp             age          ancestry            GSTM3
##  Min.   :0.0000   Min.   :26.00   Length:140         Min.   :-0.359446
##  1st Qu.:0.0000   1st Qu.:41.00   Class :character   1st Qu.:-0.145519
##  Median :0.0000   Median :45.00   Mode  :character   Median :-0.020332
##  Mean   :0.4786   Mean   :44.25                      Mean   : 0.005313
##  3rd Qu.:1.0000   3rd Qu.:49.00                      3rd Qu.: 0.123288
##  Max.   :1.0000   Max.   :53.00                      Max.   : 0.556137
##   RP5.860F19.3           BBC3               MMP9          Contig35251_RC
##  Min.   :-0.424157   Min.   :-1.08275   Min.   :-0.49427   Min.   :-0.91770
##  1st Qu.:-0.107249   1st Qu.:-0.33332   1st Qu.:-0.16053   1st Qu.:-0.59254
##  Median : 0.008689   Median :-0.09531   Median :-0.04761   Median :-0.40266
##  Mean   : 0.015576   Mean   :-0.11296   Mean   :-0.03699   Mean   :-0.25165
##  3rd Qu.: 0.103068   3rd Qu.: 0.11098   3rd Qu.: 0.08797   3rd Qu.: 0.04371
##  Max.   : 0.593821   Max.   : 0.60179   Max.   : 0.51679   Max.   : 0.99436
##  Contig40831_RC         ALDH4A1              SERF1A              SCUBE2
##  Min.   :-0.471530   Min.   :-0.767944   Min.   :-0.556292   Min.   :-0.51521
##  1st Qu.:-0.125633   1st Qu.:-0.174898   1st Qu.:-0.098369   1st Qu.:-0.12915
##  Median : 0.027046   Median :-0.004138   Median : 0.004863   Median :-0.02263
##  Mean   : 0.005541   Mean   :-0.027698   Mean   :-0.007046   Mean   :-0.02425
##  3rd Qu.: 0.122544   3rd Qu.: 0.137834   3rd Qu.: 0.089994   3rd Qu.: 0.07491
##  Max.   : 0.418517   Max.   : 0.603046   Max.   : 0.356074   Max.   : 0.43717
##       MTDH               DCK               FLT1               PECI.1
##  Min.   :-0.67564   Min.   :-0.9087   Min.   :-0.4825872   Min.   :-0.43361
##  1st Qu.:-0.29327   1st Qu.:-0.5287   1st Qu.:-0.1008469   1st Qu.:-0.13963
##  Median :-0.08343   Median :-0.3398   Median : 0.0188510   Median :-0.04026
##  Mean   :-0.08674   Mean   :-0.3213   Mean   :-0.0005165   Mean   :-0.03362
##  3rd Qu.: 0.07384   3rd Qu.:-0.1596   3rd Qu.: 0.0896944   3rd Qu.: 0.05882
##  Max.   : 0.64056   Max.   : 0.5985   Max.   : 0.5082785   Max.   : 0.51284
##     QSCN6L1              DIAPH3              SLC2A3
##  Min.   :-0.379444   Min.   :-0.449314   Min.   :-0.3715558
##  1st Qu.:-0.046621   1st Qu.:-0.112040   1st Qu.:-0.0777269
##  Median : 0.007762   Median :-0.005755   Median : 0.0005181
##  Mean   : 0.021679   Mean   :-0.010889   Mean   : 0.0113789
##  3rd Qu.: 0.098100   3rd Qu.: 0.099199   3rd Qu.: 0.0805766
##  Max.   : 0.540118   Max.   : 0.354887   Max.   : 0.4642406
##      GPR180            RTN4RL1          Contig32125_RC          STK32B
##  Min.   :-0.35519   Min.   :-0.664571   Min.   :-0.532111   Min.   :-0.48045
```

```
##  1st Qu.:-0.08033   1st Qu.:-0.205543   1st Qu.:-0.113477   1st Qu.:-0.14288
##  Median :-0.02057   Median : 0.004592   Median :-0.009005   Median :-0.02346
##  Mean   :-0.01371   Mean   :-0.041391   Mean   :-0.011050   Mean   :-0.04124
##  3rd Qu.: 0.05980   3rd Qu.: 0.131846   3rd Qu.: 0.073370   3rd Qu.: 0.04487
##  Max.   : 0.33055   Max.   : 0.428095   Max.   : 0.456306   Max.   : 0.45805
##       EXT1              COL4A2              PECI               GNAZ
##  Min.   :-0.47784   Min.   :-0.59870   Min.   :-0.44234   Min.   :-0.31745
##  1st Qu.:-0.16753   1st Qu.:-0.19791   1st Qu.:-0.19421   1st Qu.:-0.09565
##  Median :-0.05578   Median :-0.05285   Median :-0.06374   Median :-0.01636
##  Mean   :-0.05193   Mean   :-0.05964   Mean   :-0.03729   Mean   : 0.01008
##  3rd Qu.: 0.06052   3rd Qu.: 0.06271   3rd Qu.: 0.09660   3rd Qu.: 0.08337
##  Max.   : 0.37411   Max.   : 0.56018   Max.   : 0.60898   Max.   : 0.43061
##      AYTL2          Contig63649_RC          RAB6B            AA555029_RC
##  Min.   :-0.69430   Min.   :-0.365412   Min.   :-0.56918   Min.   :-0.430735
##  1st Qu.:-0.13194   1st Qu.:-0.098367   1st Qu.:-0.14308   1st Qu.:-0.159998
##  Median :-0.04600   Median :-0.024872   Median :-0.05221   Median :-0.001041
##  Mean   :-0.02517   Mean   :-0.009363   Mean   :-0.01720   Mean   :-0.020952
##  3rd Qu.: 0.06544   3rd Qu.: 0.090043   3rd Qu.: 0.08955   3rd Qu.: 0.107535
##  Max.   : 0.53361   Max.   : 0.320536   Max.   : 0.49465   Max.   : 0.820083
##      GPR126              ECT2              NUSAP1              GMPS
##  Min.   :-0.37971   Min.   :-0.50768   Min.   :-0.586304   Min.   :-0.59153
##  1st Qu.:-0.13606   1st Qu.:-0.23113   1st Qu.:-0.160713   1st Qu.:-0.28408
##  Median :-0.01046   Median :-0.08127   Median :-0.009314   Median :-0.04513
##  Mean   :-0.01639   Mean   :-0.05000   Mean   :-0.002911   Mean   :-0.06046
##  3rd Qu.: 0.09784   3rd Qu.: 0.09838   3rd Qu.: 0.150407   3rd Qu.: 0.15284
##  Max.   : 0.43925   Max.   : 0.77567   Max.   : 0.676529   Max.   : 0.55193
##      UCHL5              ORC6L              TSPYL5              MELK
##  Min.   :-0.45852   Min.   :-0.79678   Min.   :-0.67892   Min.   :-0.78982
##  1st Qu.:-0.13107   1st Qu.:-0.21396   1st Qu.:-0.17860   1st Qu.:-0.18946
##  Median :-0.03862   Median :-0.02437   Median :-0.02444   Median :-0.06113
##  Mean   :-0.02417   Mean   :-0.05166   Mean   :-0.03200   Mean   :-0.04928
##  3rd Qu.: 0.09208   3rd Qu.: 0.15011   3rd Qu.: 0.13126   3rd Qu.: 0.07438
##  Max.   : 0.56070   Max.   : 0.50672   Max.   : 0.61785   Max.   : 0.81893
##      RUNDC1           DIAPH3.1           C16orf61            TGFB3
##  Min.   :-0.8704   Min.   :-0.76818   Min.   :-0.61186   Min.   :-0.415229
##  1st Qu.:-0.3306   1st Qu.:-0.25637   1st Qu.:-0.18891   1st Qu.:-0.092384
##  Median :-0.1184   Median :-0.06829   Median :-0.09306   Median :-0.005316
##  Mean   :-0.1059   Mean   :-0.05389   Mean   :-0.05912   Mean   :-0.002289
##  3rd Qu.: 0.1037   3rd Qu.: 0.11787   3rd Qu.: 0.05866   3rd Qu.: 0.082730
##  Max.   : 0.7527   Max.   : 0.70489   Max.   : 0.59408   Max.   : 0.439666
##      FGF18             CDC42BPA             DTL               WISP1
##  Min.   :-0.597786   Min.   :-0.44439   Min.   :-1.2645   Min.   :-0.44042
##  1st Qu.:-0.140422   1st Qu.:-0.15187   1st Qu.:-0.6506   1st Qu.:-0.08759
##  Median : 0.001504   Median :-0.04357   Median :-0.1533   Median : 0.02402
##  Mean   :-0.023152   Mean   :-0.02640   Mean   :-0.2095   Mean   : 0.01312
##  3rd Qu.: 0.106955   3rd Qu.: 0.08044   3rd Qu.: 0.2034   3rd Qu.: 0.12234
##  Max.   : 0.482246   Max.   : 0.48422   Max.   : 0.8919   Max.   : 0.37552
##     DIAPH3.2              OXCT1              ZNF533
##  Min.   :-0.4510200   Min.   :-0.427838   Min.   :-0.51090
##  1st Qu.:-0.1220947   1st Qu.:-0.090491   1st Qu.:-0.26128
##  Median : 0.0088287   Median : 0.009548   Median :-0.13802
##  Mean   :-0.0009119   Mean   : 0.016115   Mean   :-0.05926
##  3rd Qu.: 0.1126542   3rd Qu.: 0.123381   3rd Qu.: 0.03807
##  Max.   : 0.3668805   Max.   : 0.649058   Max.   : 0.86482
```

```
##       RFC4                KNTC2              FBXO31
##  Min.   :-0.5635877   Min.   :-0.43109   Min.   :-0.42152
##  1st Qu.:-0.0824637   1st Qu.:-0.18407   1st Qu.:-0.13880
##  Median :-0.0009982   Median :-0.06158   Median :-0.04505
##  Mean   : 0.0080165   Mean   :-0.03585   Mean   :-0.02535
##  3rd Qu.: 0.1044821   3rd Qu.: 0.07221   3rd Qu.: 0.08601
##  Max.   : 0.4790691   Max.   : 0.59748   Max.   : 0.55562
```

```
#Based on these preliminary function calls, original data frame described in the pdata in roblem statem
#The data correspond to a follow-up study of 140 patients suffering from acute diarrhea of different in
#The main goal of this study is to identify a biomarker signature for discriminating viral from bacteri
# Variables:
# infection: Indicator of viral infection:
# (1 = viral infection; 0 = bacterial infection)
# stime: Time with symptoms (days).
# sind: Indicator of symptoms:
#   (1 = symptoms finished; 0 = symptoms remain)
# Gender: (1= male, 0 = female).
# hosp: Indicator of hospitalization (1= hospitalization, 0 = no hospitalization).
# Age: Patient age at diagnosis (years).
# Ancestry: Three different ancestry groups (A, B, C)
# Columms from 8 to 57: Gene expression measurements of 50 genes
# Use a significance level alpha=0.05 for each individual test and multiple testing
# correction whenever necessary.

#Check numeric columns
is.numeric(viral34$Age)
```

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

```
#Check for missing values in data
is.na (viral34)
```

```
##        infection stime  sind gender  hosp   age ancestry GSTM3 RP5.860F19.3
##   [1,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##   [2,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##   [3,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##   [4,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##   [5,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##   [6,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##   [7,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##   [8,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##   [9,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [10,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [11,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [12,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [13,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [14,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [15,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [16,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [17,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [18,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [19,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
```

```
##  [20,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [21,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [22,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [23,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [24,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [25,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [26,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [27,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [28,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [29,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [30,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [31,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [32,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [33,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [34,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [35,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [36,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [37,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [38,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [39,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [40,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [41,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [42,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [43,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [44,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [45,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [46,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [47,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [48,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [49,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [50,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [51,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [52,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [53,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [54,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [55,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [56,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [57,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [58,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [59,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [60,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [61,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [62,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [63,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [64,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [65,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [66,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [67,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [68,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [69,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [70,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [71,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [72,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [73,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
```

```
##  [74,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [75,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [76,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [77,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [78,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [79,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [80,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [81,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [82,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [83,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [84,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [85,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [86,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [87,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [88,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [89,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [90,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [91,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [92,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [93,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [94,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [95,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [96,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [97,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [98,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##  [99,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [100,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [101,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [102,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [103,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [104,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [105,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [106,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [107,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [108,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [109,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [110,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [111,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [112,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [113,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [114,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [115,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [116,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [117,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [118,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [119,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [120,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [121,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [122,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [123,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [124,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [125,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [126,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [127,]    FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
```

```
## [128,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [129,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [130,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [131,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [132,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [133,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [134,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [135,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [136,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [137,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [138,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [139,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
## [140,]     FALSE FALSE FALSE  FALSE FALSE FALSE    FALSE FALSE        FALSE
##         BBC3  MMP9 Contig35251_RC Contig40831_RC ALDH4A1 SERF1A SCUBE2  MTDH
##   [1,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##   [2,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##   [3,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##   [4,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##   [5,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##   [6,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##   [7,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##   [8,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##   [9,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [10,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [11,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [12,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [13,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [14,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [15,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [16,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [17,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [18,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [19,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [20,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [21,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [22,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [23,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [24,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [25,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [26,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [27,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [28,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [29,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [30,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [31,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [32,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [33,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [34,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [35,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [36,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [37,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [38,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [39,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [40,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
```

```
##  [41,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [42,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [43,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [44,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [45,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [46,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [47,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [48,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [49,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [50,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [51,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [52,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [53,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [54,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [55,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [56,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [57,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [58,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [59,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [60,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [61,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [62,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [63,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [64,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [65,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [66,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [67,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [68,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [69,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [70,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [71,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [72,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [73,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [74,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [75,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [76,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [77,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [78,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [79,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [80,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [81,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [82,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [83,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [84,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [85,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [86,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [87,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [88,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [89,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [90,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [91,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [92,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [93,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
##  [94,] FALSE FALSE         FALSE         FALSE   FALSE  FALSE  FALSE FALSE
```

```
##  [95,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [96,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [97,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [98,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##  [99,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [100,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [101,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [102,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [103,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [104,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [105,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [106,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [107,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [108,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [109,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [110,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [111,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [112,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [113,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [114,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [115,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [116,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [117,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [118,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [119,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [120,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [121,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [122,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [123,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [124,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [125,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [126,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [127,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [128,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [129,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [130,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [131,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [132,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [133,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [134,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [135,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [136,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [137,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [138,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [139,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
## [140,] FALSE FALSE          FALSE          FALSE   FALSE  FALSE  FALSE FALSE
##          DCK  FLT1 PECI.1 QSCN6L1 DIAPH3 SLC2A3 GPR180 RTN4RL1 Contig32125_RC
##   [1,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##   [2,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##   [3,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##   [4,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##   [5,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##   [6,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##   [7,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
```

```
##   [8,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##   [9,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [10,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [11,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [12,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [13,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [14,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [15,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [16,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [17,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [18,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [19,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [20,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [21,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [22,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [23,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [24,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [25,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [26,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [27,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [28,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [29,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [30,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [31,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [32,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [33,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [34,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [35,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [36,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [37,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [38,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [39,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [40,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [41,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [42,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [43,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [44,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [45,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [46,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [47,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [48,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [49,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [50,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [51,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [52,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [53,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [54,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [55,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [56,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [57,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [58,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [59,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [60,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [61,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
```

```
##  [62,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [63,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [64,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [65,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [66,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [67,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [68,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [69,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [70,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [71,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [72,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [73,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [74,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [75,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [76,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [77,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [78,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [79,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [80,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [81,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [82,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [83,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [84,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [85,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [86,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [87,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [88,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [89,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [90,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [91,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [92,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [93,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [94,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [95,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [96,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [97,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [98,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##  [99,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [100,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [101,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [102,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [103,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [104,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [105,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [106,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [107,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [108,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [109,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [110,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [111,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [112,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [113,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [114,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [115,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
```

```
## [116,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [117,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [118,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [119,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [120,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [121,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [122,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [123,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [124,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [125,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [126,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [127,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [128,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [129,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [130,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [131,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [132,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [133,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [134,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [135,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [136,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [137,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [138,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [139,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
## [140,] FALSE FALSE  FALSE   FALSE  FALSE  FALSE  FALSE   FALSE          FALSE
##        STK32B  EXT1 COL4A2  PECI  GNAZ AYTL2 Contig63649_RC RAB6B AA555029_RC
##   [1,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##   [2,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##   [3,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##   [4,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##   [5,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##   [6,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##   [7,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##   [8,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##   [9,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##  [10,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##  [11,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##  [12,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##  [13,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##  [14,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##  [15,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##  [16,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##  [17,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##  [18,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##  [19,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##  [20,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##  [21,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##  [22,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##  [23,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##  [24,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##  [25,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##  [26,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##  [27,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##  [28,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
```

```
##  [29,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [30,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [31,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [32,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [33,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [34,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [35,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [36,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [37,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [38,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [39,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [40,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [41,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [42,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [43,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [44,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [45,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [46,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [47,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [48,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [49,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [50,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [51,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [52,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [53,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [54,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [55,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [56,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [57,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [58,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [59,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [60,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [61,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [62,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [63,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [64,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [65,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [66,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [67,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [68,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [69,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [70,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [71,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [72,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [73,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [74,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [75,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [76,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [77,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [78,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [79,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [80,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [81,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [82,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
```

```
##  [83,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [84,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [85,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [86,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [87,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [88,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [89,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [90,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [91,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [92,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [93,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [94,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [95,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [96,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [97,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [98,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
##  [99,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [100,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [101,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [102,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [103,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [104,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [105,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [106,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [107,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [108,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [109,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [110,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [111,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [112,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [113,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [114,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [115,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [116,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [117,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [118,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [119,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [120,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [121,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [122,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [123,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [124,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [125,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [126,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [127,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [128,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [129,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [130,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [131,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [132,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [133,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [134,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [135,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
## [136,]  FALSE FALSE  FALSE FALSE FALSE FALSE         FALSE FALSE       FALSE
```

```
## [137,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
## [138,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
## [139,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
## [140,]  FALSE FALSE  FALSE FALSE FALSE FALSE          FALSE FALSE       FALSE
##        GPR126  ECT2 NUSAP1  GMPS UCHL5 ORC6L TSPYL5  MELK RUNDC1 DIAPH3.1
##   [1,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##   [2,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##   [3,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##   [4,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##   [5,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##   [6,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##   [7,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##   [8,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##   [9,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [10,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [11,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [12,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [13,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [14,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [15,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [16,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [17,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [18,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [19,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [20,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [21,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [22,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [23,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [24,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [25,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [26,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [27,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [28,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [29,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [30,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [31,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [32,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [33,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [34,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [35,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [36,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [37,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [38,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [39,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [40,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [41,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [42,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [43,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [44,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [45,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [46,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [47,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [48,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [49,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
```

```
##  [50,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [51,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [52,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [53,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [54,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [55,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [56,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [57,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [58,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [59,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [60,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [61,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [62,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [63,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [64,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [65,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [66,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [67,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [68,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [69,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [70,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [71,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [72,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [73,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [74,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [75,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [76,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [77,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [78,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [79,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [80,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [81,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [82,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [83,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [84,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [85,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [86,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [87,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [88,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [89,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [90,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [91,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [92,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [93,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [94,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [95,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [96,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [97,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [98,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##  [99,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [100,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [101,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [102,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [103,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
```

```
## [104,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [105,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [106,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [107,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [108,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [109,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [110,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [111,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [112,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [113,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [114,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [115,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [116,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [117,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [118,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [119,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [120,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [121,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [122,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [123,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [124,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [125,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [126,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [127,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [128,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [129,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [130,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [131,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [132,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [133,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [134,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [135,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [136,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [137,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [138,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [139,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
## [140,]  FALSE FALSE  FALSE FALSE FALSE FALSE  FALSE FALSE  FALSE    FALSE
##        C16orf61 TGFB3 FGF18 CDC42BPA   DTL WISP1 DIAPH3.2 OXCT1 ZNF533  RFC4
##   [1,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##   [2,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##   [3,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##   [4,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##   [5,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##   [6,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##   [7,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##   [8,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##   [9,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [10,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [11,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [12,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [13,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [14,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [15,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [16,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
```

```
##  [17,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [18,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [19,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [20,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [21,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [22,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [23,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [24,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [25,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [26,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [27,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [28,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [29,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [30,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [31,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [32,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [33,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [34,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [35,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [36,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [37,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [38,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [39,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [40,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [41,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [42,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [43,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [44,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [45,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [46,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [47,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [48,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [49,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [50,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [51,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [52,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [53,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [54,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [55,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [56,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [57,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [58,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [59,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [60,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [61,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [62,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [63,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [64,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [65,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [66,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [67,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [68,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [69,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [70,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
```

```
##  [71,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [72,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [73,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [74,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [75,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [76,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [77,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [78,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [79,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [80,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [81,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [82,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [83,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [84,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [85,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [86,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [87,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [88,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [89,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [90,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [91,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [92,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [93,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [94,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [95,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [96,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [97,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [98,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##  [99,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [100,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [101,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [102,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [103,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [104,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [105,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [106,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [107,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [108,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [109,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [110,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [111,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [112,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [113,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [114,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [115,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [116,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [117,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [118,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [119,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [120,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [121,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [122,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [123,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [124,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
```

```
## [125,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [126,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [127,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [128,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [129,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [130,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [131,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [132,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [133,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [134,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [135,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [136,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [137,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [138,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [139,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
## [140,]    FALSE FALSE FALSE    FALSE FALSE FALSE    FALSE FALSE  FALSE FALSE
##        KNTC2 FBXO31
##   [1,] FALSE  FALSE
##   [2,] FALSE  FALSE
##   [3,] FALSE  FALSE
##   [4,] FALSE  FALSE
##   [5,] FALSE  FALSE
##   [6,] FALSE  FALSE
##   [7,] FALSE  FALSE
##   [8,] FALSE  FALSE
##   [9,] FALSE  FALSE
##  [10,] FALSE  FALSE
##  [11,] FALSE  FALSE
##  [12,] FALSE  FALSE
##  [13,] FALSE  FALSE
##  [14,] FALSE  FALSE
##  [15,] FALSE  FALSE
##  [16,] FALSE  FALSE
##  [17,] FALSE  FALSE
##  [18,] FALSE  FALSE
##  [19,] FALSE  FALSE
##  [20,] FALSE  FALSE
##  [21,] FALSE  FALSE
##  [22,] FALSE  FALSE
##  [23,] FALSE  FALSE
##  [24,] FALSE  FALSE
##  [25,] FALSE  FALSE
##  [26,] FALSE  FALSE
##  [27,] FALSE  FALSE
##  [28,] FALSE  FALSE
##  [29,] FALSE  FALSE
##  [30,] FALSE  FALSE
##  [31,] FALSE  FALSE
##  [32,] FALSE  FALSE
##  [33,] FALSE  FALSE
##  [34,] FALSE  FALSE
##  [35,] FALSE  FALSE
##  [36,] FALSE  FALSE
##  [37,] FALSE  FALSE
```

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

```
##  [92,] FALSE  FALSE
##  [93,] FALSE  FALSE
##  [94,] FALSE  FALSE
##  [95,] FALSE  FALSE
##  [96,] FALSE  FALSE
##  [97,] FALSE  FALSE
##  [98,] FALSE  FALSE
##  [99,] FALSE  FALSE
## [100,] FALSE  FALSE
## [101,] FALSE  FALSE
## [102,] FALSE  FALSE
## [103,] FALSE  FALSE
## [104,] FALSE  FALSE
## [105,] FALSE  FALSE
## [106,] FALSE  FALSE
## [107,] FALSE  FALSE
## [108,] FALSE  FALSE
## [109,] FALSE  FALSE
## [110,] FALSE  FALSE
## [111,] FALSE  FALSE
## [112,] FALSE  FALSE
## [113,] FALSE  FALSE
## [114,] FALSE  FALSE
## [115,] FALSE  FALSE
## [116,] FALSE  FALSE
## [117,] FALSE  FALSE
## [118,] FALSE  FALSE
## [119,] FALSE  FALSE
## [120,] FALSE  FALSE
## [121,] FALSE  FALSE
## [122,] FALSE  FALSE
## [123,] FALSE  FALSE
## [124,] FALSE  FALSE
## [125,] FALSE  FALSE
## [126,] FALSE  FALSE
## [127,] FALSE  FALSE
## [128,] FALSE  FALSE
## [129,] FALSE  FALSE
## [130,] FALSE  FALSE
## [131,] FALSE  FALSE
## [132,] FALSE  FALSE
## [133,] FALSE  FALSE
## [134,] FALSE  FALSE
## [135,] FALSE  FALSE
## [136,] FALSE  FALSE
## [137,] FALSE  FALSE
## [138,] FALSE  FALSE
## [139,] FALSE  FALSE
## [140,] FALSE  FALSE
```

```
sum(is.na(viral34))
```

```
## [1] 0
```

```r
# Exclude rows that have missing data in ANY variable (na.exclude())
#viral34_o <- na.omit(viral34)

#Scale Gene Expression Values
viral34_s<-scale(viral34[,8:57])
dim(viral34_s)
```

```
## [1] 140  50
```

```r
#ncol(viral34_s)
#head(viral34_s)

# # TEST SCALING
# gs<-(viral34$GSTM3-mean(viral34$GSTM3))/sd(viral34$GSTM3)
# head(gs)
# bs<-(viral34$BBC3-mean(viral34$BBC3))/sd(viral34$BBC3)
# head(bs)

#Log transformation of Gene Expression Values
viral34_l <- log(3+viral34[8:57])
dim(viral34_l)
```

```
## [1] 140  50
```

```r
head(viral34_l)
```

```
##      GSTM3 RP5.860F19.3      BBC3      MMP9 Contig35251_RC Contig40831_RC
## 1 1.146283     1.065387 1.1964092 1.0273483      0.9086397       1.059473
## 2 1.086730     1.091576 1.0659392 1.1760282      0.9034401       1.097489
## 3 1.007013     1.102207 1.1276993 1.0899612      0.9783141       1.071251
## 4 1.205335     1.206491 1.2022920 0.9778165      0.8646544       1.055585
## 5 1.130296     1.124341 0.9531455 1.0372264      0.8572905       1.126475
## 6 1.024000     1.091175 1.1800021 1.1700672      1.2652235       1.149187
##    ALDH4A1    SERF1A   SCUBE2      MTDH       DCK     FLT1    PECI.1  QSCN6L1
## 1 1.152706 1.0702849 1.100629 1.0516011 0.9800153 1.079796 1.0689185 1.056186
## 2 1.130903 1.1335764 1.130618 1.2511141 0.8837522 1.153416 1.0658115 1.183440
## 3 1.120748 1.1192832 1.055762 1.1224146 1.1438818 1.120692 1.0600923 1.130027
## 4 1.281779 0.9067077 1.126833 0.8850266 0.9062228 1.088437 1.1342227 1.070614
## 5 1.150492 1.0926398 1.049379 0.9084851 1.0305361 1.119558 1.0344670 1.070910
## 6 1.061012 1.0443464 1.099806 1.1025348 0.8982809 1.093940 0.9941321 1.180676
##     DIAPH3   SLC2A3   GPR180   RTN4RL1 Contig32125_RC   STK32B      EXT1
## 1 1.125205 1.214490 1.083427 1.0427923       1.094180 1.107831 1.0527575
## 2 1.139388 1.242493 1.034060 1.1631687       1.094182 1.140950 1.0311917
## 3 1.173734 1.072412 1.096808 1.1384390       1.072006 1.077713 1.0600865
## 4 1.037644 1.022350 1.136531 0.8568754       1.090727 1.102913 1.1058993
## 5 1.131626 1.082742 1.060567 1.1079547       1.063275 1.057306 0.9870127
## 6 1.086738 1.080980 1.137172 1.2196394       1.042403 1.061255 1.0555742
##      COL4A2     PECI     GNAZ    AYTL2 Contig63649_RC    RAB6B AA555029_RC
## 1 1.0885944 1.154833 1.121831 1.175354      1.1084369 1.241677   1.0824551
## 2 1.0200246 1.166909 1.087995 1.103819      1.1000074 1.104782   1.1503398
## 3 1.0266382 1.112621 1.119382 1.055165      1.1152736 1.049943   1.1359748
## 4 1.1322759 1.246926 1.204436 1.262321      1.1922303 1.201603   0.9893029
```

```
## 5 1.0087572 1.131560 1.048489 1.136901      1.1074748 1.069077   1.0224237
## 6 0.9719145 1.062921 1.027844 1.081362      0.9918692 1.002412   1.1157947
##     GPR126     ECT2   NUSAP1      GMPS    UCHL5     ORC6L   TSPYL5      MELK
## 1 1.064635 1.110358 1.134564 1.1687766 1.085833 1.1546201 1.149273 0.9761383
## 2 1.188195 1.085958 1.109305 1.1586833 1.004012 1.1319625 1.078809 1.0950000
## 3 1.117469 1.036274 1.082413 1.1120005 1.084102 1.0472789 1.128460 1.0520102
## 4 1.160385 1.109678 1.296174 1.1746853 1.032361 1.1608662 1.264406 0.9784430
## 5 1.070704 1.202594 1.124775 1.1563661 1.050517 1.1396518 1.113699 1.0121761
## 6 1.040379 1.022732 1.003728 0.9789156 1.141305 0.9832136 1.022983 1.0530851
##      RUNDC1  DIAPH3.1 C16orf61    TGFB3    FGF18 CDC42BPA       DTL     WISP1
## 1 1.2693007 0.9382254 1.118129 1.125516 1.082416 1.160658 0.6873433 1.0969520
## 2 0.9712143 1.0142915 1.081744 1.137416 1.185958 1.139547 1.0486967 1.1687356
## 3 1.1410086 1.0796194 1.002945 1.135636 1.110113 1.019541 1.2378613 1.1220865
## 4 1.0743851 0.9378211 1.142785 1.091766 1.132663 1.109493 0.6541381 0.9398414
## 5 1.0960575 1.0658026 1.164658 1.140219 1.159966 1.073668 0.7688520 1.1376485
## 6 0.9182013 1.1661631 1.039890 1.183748 1.029941 1.005780 1.3111641 1.0416259
##    DIAPH3.2    OXCT1   ZNF533     RFC4     KNTC2   FBXO31
## 1 0.9940749 1.109008 1.351915 1.119021 1.2802338 1.084714
## 2 1.1054462 1.151631 1.103849 1.104223 0.9831091 1.052493
## 3 1.0493743 1.117417 1.048161 1.080354 1.1302877 1.110287
## 4 1.1155441 1.027675 1.027291 1.222145 1.1138238 1.118646
## 5 1.1013537 1.051178 1.190429 1.119203 1.0672548 1.182958
## 6 1.1383232 1.127214 1.045297 1.030812 1.0354385 1.094065
```

```
#Had challenges with log transformations as negative numbers were obtained and processed, leading to er
#viral34[, 8:57] <- log1p(viral34[8:57]) #computes log(1+x)
#viral34[,8:57] <- lapply(viral34[,8:57], log)
#warnings()
#50: In lapply(X = x, FUN = .Generic, ...) : NaNs produced

# #Test Log transformation
# log_col1<-log(3+viral34$GSTM3)
# head(log_col1)
# log_col2<-log(3+viral34$BBC3)
# head(log_col2)

#Column bind the Original dataframe, Standardized G.E dataframe, and Log-Transformed G.E. dataframe
viral34_c <- cbind(viral34,viral34_s, viral34_l)
dim(viral34_c)
```

```
## [1] 140 157
```

```
head(viral34_c)
```

```
##   infection     stime sind gender hosp age ancestry       GSTM3 RP5.860F19.3
## 1         0  7.296372    0      0    1  47        A  0.14647630  -0.09803689
## 2         1  6.718686    0      1    0  47        A -0.03543524  -0.02103562
## 3         0  6.995209    1      1    1  38        B -0.26258909   0.01080372
## 4         1  9.330595    0      0    1  45        A  0.33787726   0.34173748
## 5         1  3.438741    1      1    0  31        A  0.09657176   0.07818674
## 6         1 15.329227    0      0    0  41        A -0.21568976  -0.02222821
##          BBC3        MMP9 Contig35251_RC Contig40831_RC     ALDH4A1      SERF1A
## 1  0.30821656 -0.20635196     -0.5190545   -0.115149133  0.16674915 -0.08378990
```

```
## 2 -0.09643536  0.24147416     -0.5319210   -0.003368632  0.09845308  0.10674758
## 3  0.08854258 -0.02584139     -0.3400320   -0.080972535  0.06714804  0.06265814
## 4  0.32773524 -0.34135513     -0.6258146   -0.126344331  0.60304554 -0.52384308
## 5 -0.40614429 -0.17861930     -0.6432336    0.084764382  0.15974585 -0.01786414
## 6  0.25438095  0.22220903      0.5438846    0.155627613 -0.11070614 -0.15845941
##         SCUBE2        MTDH        DCK        FLT1      PECI.1     QSCN6L1
## 1  0.006056118 -0.13776980 -0.3355029 -0.05592013 -0.08777178 -0.12461618
## 2  0.097569969  0.49423377 -0.5800370  0.16900027 -0.09680589  0.26558747
## 3 -0.125834721  0.07226344  0.1389293  0.06697535 -0.11336268  0.09573915
## 4  0.085869554 -0.57695111 -0.5250435 -0.03036967  0.10875631 -0.08282874
## 5 -0.144122253 -0.51943819 -0.1974320  0.06349948 -0.18639378 -0.08196639
## 6  0.003581842  0.01179057 -0.5446215 -0.01398411 -0.29762197  0.25657503
##        DIAPH3      SLC2A3       GPR180     RTN4RL1 Contig32125_RC      STK32B
## 1  0.08084737  0.36857538 -0.045212204 -0.16287206    -0.01326674  0.02778352
## 2  0.12485686  0.46424057 -0.187539926  0.20005740    -0.01326124  0.12973975
## 3  0.23404700 -0.07758043 -0.005409022  0.12189144    -0.07876585 -0.06204672
## 4 -0.17743971 -0.22028177  0.115940757 -0.64421170    -0.02356354  0.01292946
## 5  0.10069529 -0.04723501 -0.111992419  0.02815844    -0.10416021 -0.12139572
## 6 -0.03541076 -0.05243454  0.117939467  0.38596644    -0.16397529 -0.11000363
##          EXT1      COL4A2        PECI        GNAZ       AYTL2 Contig63649_RC
## 1 -0.13445797 -0.02990369  0.17349207  0.07047200  0.23928807    0.029619073
## 2 -0.19559405 -0.22673707  0.21204866 -0.03168261  0.01566120    0.004188184
## 3 -0.11337939 -0.20833506  0.04232249  0.06296012 -0.12755049    0.050402671
## 4  0.02194074  0.10271000  0.47963136  0.33487679  0.53361497    0.294420521
## 5 -0.31679313 -0.25780916  0.10048933 -0.14666447  0.11709402    0.026705563
## 6 -0.12637540 -0.35700047 -0.10518681 -0.20496584 -0.05130563   -0.303730416
##         RAB6B AA555029_RC      GPR126        ECT2      NUSAP1        GMPS
## 1  0.46141386 -0.04808210 -0.10022007  0.03544526  0.10981625  0.21805322
## 2  0.01856611  0.15926624  0.28115470 -0.03772432  0.03225047  0.18573594
## 3 -0.14251272  0.11420782  0.05710594 -0.18130437 -0.04820767  0.04043471
## 4  0.32544391 -0.31064082  0.19116150  0.03338104  0.65528392  0.23712422
## 5 -0.08731065 -0.22007577 -0.08256859  0.32874172  0.07952344  0.17836256
## 6 -0.27515282  0.05199264 -0.16971181 -0.21921949 -0.27156456 -0.33843161
##         UCHL5      ORC6L      TSPYL5        MELK       RUNDC1    DIAPH3.1
## 1 -0.03809348  0.1728180  0.15589646 -0.34581318  0.558363335 -0.44455761
## 2 -0.27078994  0.1017376 -0.05882551 -0.01081727 -0.358850367 -0.24259102
## 3 -0.04321627 -0.1501144  0.09089323 -0.13659874  0.129923754 -0.05644101
## 4 -0.19231373  0.1926975  0.54098979 -0.33968909 -0.071808178 -0.44559038
## 5 -0.14087134  0.1256798  0.04560219 -0.24841783 -0.007654665 -0.09683179
## 6  0.13085033 -0.3269674 -0.21852037 -0.13351906 -0.495218907  0.20965383
##      C16orf61       TGFB3       FGF18    CDC42BPA        DTL        WISP1
## 1  0.05912505  0.08180754 -0.04819787  0.19203352 -1.0115741 -0.004976858
## 2 -0.05018147  0.11869773  0.27382112  0.12535125 -0.1460709  0.217921322
## 3 -0.27369996  0.11315389  0.03470079 -0.22807824  0.4482309  0.071255809
## 4  0.13548833 -0.02046981  0.10391225  0.03281893 -1.0765161 -0.440424509
## 5  0.20482626  0.12745233  0.18982466 -0.07390664 -0.8427116  0.119424302
## 6 -0.17109518  0.26659311 -0.19909908 -0.26596006  0.7104907 -0.166179324
##       DIAPH3.2       OXCT1      ZNF533        RFC4       KNTC2      FBXO31
## 1 -0.297776741  0.03135030  0.86482037  0.06185603  0.59748058 -0.04140661
## 2  0.020572007  0.16334775  0.01575178  0.01687964 -0.32724674 -0.13521580
## 3 -0.144136303  0.05694880 -0.14760060 -0.05427720  0.09654722  0.03522958
## 4  0.051227778 -0.20543872 -0.20651322  0.39446170  0.04598343  0.06070769
## 5  0.008235576 -0.13898008  0.28849255  0.06241235 -0.09261277  0.26401621
## 6  0.121529650  0.08704478 -0.15575775 -0.19665862 -0.18365899 -0.01361086
```

```
##        GSTM3 RP5.860F19.3        BBC3        MMP9 Contig35251_RC Contig40831_RC
## 1  0.7138813  -0.57440211  1.26618208 -0.81209763     -0.6159014    -0.67108894
## 2 -0.2060675  -0.18510052  0.04967577  1.33518865     -0.6455369    -0.04954233
## 3 -1.3548121  -0.02412778  0.60577550  0.05343442     -0.2035604    -0.48105268
## 4  1.6818192   1.64900082  1.32486112 -1.45942582     -0.8618013    -0.73333895
## 5  0.4615079   0.31654599 -0.88140315 -0.67912202     -0.9019223     0.44051430
## 6 -1.1176365  -0.19112999  1.10433590  1.24281408      1.8323579     0.83454368
##      ALDH4A1      SERF1A     SCUBE2       MTDH         DCK        FLT1
## 1  0.8726701 -0.44912514  0.1942805 -0.1821611 -0.05112308 -0.35139159
## 2  0.5661609  0.66594479  0.7808869  2.0740241 -0.93365927  1.07514078
## 3  0.4256655  0.40792311 -0.6511435  0.5676351  1.66112734  0.42805919
## 4  2.8307454 -3.02441939  0.7058869 -1.7499914 -0.73518462 -0.18934069
## 5  0.8412396 -0.06331218 -0.7683671 -1.5446764  0.44718207  0.40601389
## 6 -0.3725348 -0.88610863  0.1784203  0.3517534 -0.80584276 -0.08541711
##       PECI.1    QSCN6L1     DIAPH3     SLC2A3      GPR180    RTN4RL1
## 1 -0.3325357 -1.0997776  0.5392279  2.2788093 -0.24330929 -0.4879262
## 2 -0.3880154  1.8335965  0.7979150  2.8891253 -1.34271285  0.9697711
## 3 -0.4896928  0.5567539  1.4397330 -0.5675343  0.06414844  0.6558194
## 4  0.8743696 -0.7856385 -0.9789797 -1.4779272  1.00150889 -2.4212151
## 5 -0.9381866 -0.7791558  0.6558937 -0.3739394 -0.75914981  0.2793431
## 6 -1.6212538  1.7658451 -0.1441357 -0.4071109  1.01694783  1.7164702
##   Contig32125_RC     STK32B       EXT1     COL4A2       PECI       GNAZ
## 1    -0.01428243  0.4260294 -0.4813926  0.1461046  0.9860751  0.3995737
## 2    -0.01424700  1.0552938 -0.8379918 -0.8211227  1.1664517 -0.2763188
## 3    -0.43623247 -0.1283947 -0.3584438 -0.7306964  0.3724338  0.3498723
## 4    -0.08061514  0.3343515  0.4308618  0.7977604  2.4182648  2.1489730
## 5    -0.59982477 -0.4946913 -1.5449311 -0.9738091  0.6445516 -1.0370811
## 6    -0.98515784 -0.4243804 -0.4342479 -1.4612293 -0.3176484 -1.4228244
##        AYTL2 Contig63649_RC      RAB6B AA555029_RC     GPR126       ECT2
## 1  1.5729248     0.28173399  2.3501669  -0.1449979 -0.5035802  0.3578867
## 2  0.2428707     0.09793693  0.1756376   0.9631957  1.7873898  0.0514164
## 3 -0.6089022     0.43194361 -0.6153129   0.7223765  0.4414985 -0.5499677
## 4  3.3234779     2.19553759  1.6825094  -1.5482689  1.2467884  0.3492407
## 5  0.8461575     0.26067713 -0.3442524  -1.0642351 -0.3975453  1.5863570
## 6 -0.1554245    -2.12748792 -1.2666197   0.3898614 -0.9210265 -0.7087750
##       NUSAP1       GMPS       UCHL5      ORC6L     TSPYL5       MELK     RUNDC1
## 1  0.4808374  1.0797140 -0.08519916  0.8392967  0.7914942 -1.2786512  2.1740843
## 2  0.1499810  0.9544301 -1.50870499  0.5735413 -0.1130234  0.1658503 -0.8279555
## 3 -0.1932127  0.3911430 -0.11653751 -0.3680828  0.5176677 -0.3765191  0.7718018
## 4  2.8075265  1.1536463 -1.02863177  0.9136225  2.4137021 -1.2522442  0.1115334
## 5  0.3516236  0.9258458 -0.71393620  0.6630564  0.3268789 -0.8586827  0.3215078
## 6 -1.1459402 -1.0776023  0.94830393 -1.0293010 -0.7857389 -0.3632395 -1.2742896
##       DIAPH3.1    C16orf61      TGFB3      FGF18   CDC42BPA        DTL
## 1 -1.455677597  0.61186055  0.5877739 -0.1281211  1.2597867 -1.5254135
## 2 -0.703133863  0.04624479  0.8456102  1.5191381  0.8751973  0.1207287
## 3 -0.009523967 -1.11037043  0.8068628  0.2959395 -1.1632042  1.2510601
## 4 -1.459525754  1.00700876 -0.1270723  0.6499845  0.3415180 -1.6489300
## 5 -0.160023276  1.36580373  0.9067987  1.0894618 -0.2740209 -1.2042460
## 6  0.981966756 -0.57943337  1.8792943 -0.9000427 -1.3816873  1.7498645
##        WISP1    DIAPH3.2      OXCT1     ZNF533        RFC4      KNTC2
## 1 -0.1128211 -1.82122539  0.0835812  3.2029626  0.34098731  3.1695391
## 2  1.2764815  0.13180112  0.8077063  0.2600142  0.05613357 -1.4582792
## 3  0.3623297 -0.87866209  0.2240121 -0.3061799 -0.39453158  0.6626070
## 4 -2.8269239  0.31987013 -1.2154201 -0.5103764  2.44751394  0.4095594
```

```
## 5  0.6625592  0.05611878 -0.8508345  1.2053583  0.34451067 -0.2840483
## 6 -1.1175804  0.75116261  0.3891157 -0.3344533 -1.29629087 -0.7396911
##        FBXO31    GSTM3 RP5.860F19.3      BBC3      MMP9 Contig35251_RC
## 1 -0.09573504 1.146283     1.065387 1.1964092 1.0273483      0.9086397
## 2 -0.65490852 1.086730     1.091576 1.0659392 1.1760282      0.9034401
## 3  0.36107443 1.007013     1.102207 1.1276993 1.0899612      0.9783141
## 4  0.51294316 1.205335     1.206491 1.2022920 0.9778165      0.8646544
## 5  1.72481525 1.130296     1.124341 0.9531455 1.0372264      0.8572905
## 6  0.06994858 1.024000     1.091175 1.1800021 1.1700672      1.2652235
##   Contig40831_RC  ALDH4A1    SERF1A   SCUBE2      MTDH       DCK     FLT1
## 1       1.059473 1.152706 1.0702849 1.100629 1.0516011 0.9800153 1.079796
## 2       1.097489 1.130903 1.1335764 1.130618 1.2511141 0.8837522 1.153416
## 3       1.071251 1.120748 1.1192832 1.055762 1.1224146 1.1438818 1.120692
## 4       1.055585 1.281779 0.9067077 1.126833 0.8850266 0.9062228 1.088437
## 5       1.126475 1.150492 1.0926398 1.049379 0.9084851 1.0305361 1.119558
## 6       1.149187 1.061012 1.0443464 1.099806 1.1025348 0.8982809 1.093940
##      PECI.1  QSCN6L1   DIAPH3   SLC2A3   GPR180   RTN4RL1 Contig32125_RC
## 1 1.0689185 1.056186 1.125205 1.214490 1.083427 1.0427923       1.094180
## 2 1.0658115 1.183440 1.139388 1.242493 1.034060 1.1631687       1.094182
## 3 1.0600923 1.130027 1.173734 1.072412 1.096808 1.1384390       1.072006
## 4 1.1342227 1.070614 1.037644 1.022350 1.136531 0.8568754       1.090727
## 5 1.0344670 1.070910 1.131626 1.082742 1.060567 1.1079547       1.063275
## 6 0.9941321 1.180676 1.086738 1.080980 1.137172 1.2196394       1.042403
##     STK32B      EXT1    COL4A2     PECI     GNAZ    AYTL2 Contig63649_RC
## 1 1.107831 1.0527575 1.0885944 1.154833 1.121831 1.175354      1.1084369
## 2 1.140950 1.0311917 1.0200246 1.166909 1.087995 1.103819      1.1000074
## 3 1.077713 1.0600865 1.0266382 1.112621 1.119382 1.055165      1.1152736
## 4 1.102913 1.1058993 1.1322759 1.246926 1.204436 1.262321      1.1922303
## 5 1.057306 0.9870127 1.0087572 1.131560 1.048489 1.136901      1.1074748
## 6 1.061255 1.0555742 0.9719145 1.062921 1.027844 1.081362      0.9918692
##      RAB6B AA555029_RC   GPR126     ECT2   NUSAP1      GMPS    UCHL5     ORC6L
## 1 1.241677   1.0824551 1.064635 1.110358 1.134564 1.1687766 1.085833 1.1546201
## 2 1.104782   1.1503398 1.188195 1.085958 1.109305 1.1586833 1.004012 1.1319625
## 3 1.049943   1.1359748 1.117469 1.036274 1.082413 1.1120005 1.084102 1.0472789
## 4 1.201603   0.9893029 1.160385 1.109678 1.296174 1.1746853 1.032361 1.1608662
## 5 1.069077   1.0224237 1.070704 1.202594 1.124775 1.1563661 1.050517 1.1396518
## 6 1.002412   1.1157947 1.040379 1.022732 1.003728 0.9789156 1.141305 0.9832136
##     TSPYL5      MELK    RUNDC1  DIAPH3.1 C16orf61    TGFB3    FGF18 CDC42BPA
## 1 1.149273 0.9761383 1.2693007 0.9382254 1.118129 1.125516 1.082416 1.160658
## 2 1.078809 1.0950000 0.9712143 1.0142915 1.081744 1.137416 1.185958 1.139547
## 3 1.128460 1.0520102 1.1410086 1.0796194 1.002945 1.135636 1.110113 1.019541
## 4 1.264406 0.9784430 1.0743851 0.9378211 1.142785 1.091766 1.132663 1.109493
## 5 1.113699 1.0121761 1.0960575 1.0658026 1.164658 1.140219 1.159966 1.073668
## 6 1.022983 1.0530851 0.9182013 1.1661631 1.039890 1.183748 1.029941 1.005780
##         DTL     WISP1  DIAPH3.2    OXCT1   ZNF533     RFC4     KNTC2   FBXO31
## 1 0.6873433 1.0969520 0.9940749 1.109008 1.351915 1.119021 1.2802338 1.084714
## 2 1.0486967 1.1687356 1.1054462 1.151631 1.103849 1.104223 0.9831091 1.052493
## 3 1.2378613 1.1220865 1.0493743 1.117417 1.048161 1.080354 1.1302877 1.110287
## 4 0.6541381 0.9398414 1.1155441 1.027675 1.027291 1.222145 1.1138238 1.118646
## 5 0.7688520 1.1376485 1.1013537 1.051178 1.190429 1.119203 1.0672548 1.182958
## 6 1.3111641 1.0416259 1.1383232 1.127214 1.045297 1.030812 1.0354385 1.094065
```

```
#Rename standardized gene expression column variables by appending "_s"
names(viral34_c)[58:107] <- paste0(names(viral34_c)[58:107], "_s")
```

```r
#Rename standardized gene expression column variables by appending "_l"
names(viral34_c)[108:157] <- paste0(names(viral34_c)[108:157], "_l")

#FACTORING:
#Transform the column variable infection into a factor:
viral34_c$infection <- factor(viral34_c$infection, levels=c(0,1), labels = c("bacterial_infection", "vi

#Transform the column variable sind into a factor:
viral34_c$sind <- factor(viral34_c$sind, levels=c(0,1), labels = c("symptoms_remain", "symptoms_finishe

#Transform the column variable gender into a factor:
viral34_c$gender <- factor(viral34_c$gender, levels=c(0,1), labels = c("female", "male"))

#Transform the column variable gender into a factor:
viral34_c$hosp<- factor(viral34_c$hosp, levels=c(0,1), labels = c("no_hospitalization", "hospitalizatio

summary(viral34_c) #Created baseline dataframe for use of subsequent univariate analysis:
```

```
##                infection      stime                          sind       gender
##  bacterial_infection:69   Min.   : 0.05476   symptoms_remain  :93   female:62
##  viral_infection    :71   1st Qu.: 4.69541   symptoms_finished:47   male  :78
##                           Median : 6.96235
##                           Mean   : 7.35621
##                           3rd Qu.:10.05681
##                           Max.   :17.65914
##                   hosp          age          ancestry             GSTM3
##  no_hospitalization:73   Min.   :26.00   Length:140         Min.   :-0.359446
##  hospitalization   :67   1st Qu.:41.00   Class :character   1st Qu.:-0.145519
##                          Median :45.00   Mode  :character   Median :-0.020332
##                          Mean   :44.25                      Mean   : 0.005313
##                          3rd Qu.:49.00                      3rd Qu.: 0.123288
##                          Max.   :53.00                      Max.   : 0.556137
##   RP5.860F19.3             BBC3               MMP9          Contig35251_RC
##  Min.   :-0.424157   Min.   :-1.08275   Min.   :-0.49427   Min.   :-0.91770
##  1st Qu.:-0.107249   1st Qu.:-0.33332   1st Qu.:-0.16053   1st Qu.:-0.59254
##  Median : 0.008689   Median :-0.09531   Median :-0.04761   Median :-0.40266
##  Mean   : 0.015576   Mean   :-0.11296   Mean   :-0.03699   Mean   :-0.25165
##  3rd Qu.: 0.103068   3rd Qu.: 0.11098   3rd Qu.: 0.08797   3rd Qu.: 0.04371
##  Max.   : 0.593821   Max.   : 0.60179   Max.   : 0.51679   Max.   : 0.99436
##  Contig40831_RC         ALDH4A1              SERF1A              SCUBE2
##  Min.   :-0.471530   Min.   :-0.767944   Min.   :-0.556292   Min.   :-0.51521
##  1st Qu.:-0.125633   1st Qu.:-0.174898   1st Qu.:-0.098369   1st Qu.:-0.12915
##  Median : 0.027046   Median :-0.004138   Median : 0.004863   Median :-0.02263
##  Mean   : 0.005541   Mean   :-0.027698   Mean   :-0.007046   Mean   :-0.02425
##  3rd Qu.: 0.122544   3rd Qu.: 0.137834   3rd Qu.: 0.089994   3rd Qu.: 0.07491
##  Max.   : 0.418517   Max.   : 0.603046   Max.   : 0.356074   Max.   : 0.43717
##       MTDH                DCK                FLT1                PECI.1
##  Min.   :-0.67564   Min.   :-0.9087   Min.   :-0.4825872   Min.   :-0.43361
##  1st Qu.:-0.29327   1st Qu.:-0.5287   1st Qu.:-0.1008469   1st Qu.:-0.13963
##  Median :-0.08343   Median :-0.3398   Median : 0.0188510   Median :-0.04026
##  Mean   :-0.08674   Mean   :-0.3213   Mean   :-0.0005165   Mean   :-0.03362
##  3rd Qu.: 0.07384   3rd Qu.:-0.1596   3rd Qu.: 0.0896944   3rd Qu.: 0.05882
##  Max.   : 0.64056   Max.   : 0.5985   Max.   : 0.5082785   Max.   : 0.51284
```

```
##     QSCN6L1              DIAPH3              SLC2A3
##  Min.   :-0.379444   Min.   :-0.449314   Min.   :-0.3715558
##  1st Qu.:-0.046621   1st Qu.:-0.112040   1st Qu.:-0.0777269
##  Median : 0.007762   Median :-0.005755   Median : 0.0005181
##  Mean   : 0.021679   Mean   :-0.010889   Mean   : 0.0113789
##  3rd Qu.: 0.098100   3rd Qu.: 0.099199   3rd Qu.: 0.0805766
##  Max.   : 0.540118   Max.   : 0.354887   Max.   : 0.4642406
##      GPR180            RTN4RL1          Contig32125_RC          STK32B
##  Min.   :-0.35519   Min.   :-0.664571   Min.   :-0.532111   Min.   :-0.48045
##  1st Qu.:-0.08033   1st Qu.:-0.205543   1st Qu.:-0.113477   1st Qu.:-0.14288
##  Median :-0.02057   Median : 0.004592   Median :-0.009005   Median :-0.02346
##  Mean   :-0.01371   Mean   :-0.041391   Mean   :-0.011050   Mean   :-0.04124
##  3rd Qu.: 0.05980   3rd Qu.: 0.131846   3rd Qu.: 0.073370   3rd Qu.: 0.04487
##  Max.   : 0.33055   Max.   : 0.428095   Max.   : 0.456306   Max.   : 0.45805
##       EXT1              COL4A2              PECI               GNAZ
##  Min.   :-0.47784   Min.   :-0.59870   Min.   :-0.44234   Min.   :-0.31745
##  1st Qu.:-0.16753   1st Qu.:-0.19791   1st Qu.:-0.19421   1st Qu.:-0.09565
##  Median :-0.05578   Median :-0.05285   Median :-0.06374   Median :-0.01636
##  Mean   :-0.05193   Mean   :-0.05964   Mean   :-0.03729   Mean   : 0.01008
##  3rd Qu.: 0.06052   3rd Qu.: 0.06271   3rd Qu.: 0.09660   3rd Qu.: 0.08337
##  Max.   : 0.37411   Max.   : 0.56018   Max.   : 0.60898   Max.   : 0.43061
##      AYTL2           Contig63649_RC          RAB6B            AA555029_RC
##  Min.   :-0.69430   Min.   :-0.365412   Min.   :-0.56918   Min.   :-0.430735
##  1st Qu.:-0.13194   1st Qu.:-0.098367   1st Qu.:-0.14308   1st Qu.:-0.159998
##  Median :-0.04600   Median :-0.024872   Median :-0.05221   Median :-0.001041
##  Mean   :-0.02517   Mean   :-0.009363   Mean   :-0.01720   Mean   :-0.020952
##  3rd Qu.: 0.06544   3rd Qu.: 0.090043   3rd Qu.: 0.08955   3rd Qu.: 0.107535
##  Max.   : 0.53361   Max.   : 0.320536   Max.   : 0.49465   Max.   : 0.820083
##      GPR126              ECT2              NUSAP1              GMPS
##  Min.   :-0.37971   Min.   :-0.50768   Min.   :-0.586304   Min.   :-0.59153
##  1st Qu.:-0.13606   1st Qu.:-0.23113   1st Qu.:-0.160713   1st Qu.:-0.28408
##  Median :-0.01046   Median :-0.08127   Median :-0.009314   Median :-0.04513
##  Mean   :-0.01639   Mean   :-0.05000   Mean   :-0.002911   Mean   :-0.06046
##  3rd Qu.: 0.09784   3rd Qu.: 0.09838   3rd Qu.: 0.150407   3rd Qu.: 0.15284
##  Max.   : 0.43925   Max.   : 0.77567   Max.   : 0.676529   Max.   : 0.55193
##      UCHL5               ORC6L              TSPYL5              MELK
##  Min.   :-0.45852   Min.   :-0.79678   Min.   :-0.67892   Min.   :-0.78982
##  1st Qu.:-0.13107   1st Qu.:-0.21396   1st Qu.:-0.17860   1st Qu.:-0.18946
##  Median :-0.03862   Median :-0.02437   Median :-0.02444   Median :-0.06113
##  Mean   :-0.02417   Mean   :-0.05166   Mean   :-0.03200   Mean   :-0.04928
##  3rd Qu.: 0.09208   3rd Qu.: 0.15011   3rd Qu.: 0.13126   3rd Qu.: 0.07438
##  Max.   : 0.56070   Max.   : 0.50672   Max.   : 0.61785   Max.   : 0.81893
##      RUNDC1            DIAPH3.1           C16orf61            TGFB3
##  Min.   :-0.8704   Min.   :-0.76818   Min.   :-0.61186   Min.   :-0.415229
##  1st Qu.:-0.3306   1st Qu.:-0.25637   1st Qu.:-0.18891   1st Qu.:-0.092384
##  Median :-0.1184   Median :-0.06829   Median :-0.09306   Median :-0.005316
##  Mean   :-0.1059   Mean   :-0.05389   Mean   :-0.05912   Mean   :-0.002289
##  3rd Qu.: 0.1037   3rd Qu.: 0.11787   3rd Qu.: 0.05866   3rd Qu.: 0.082730
##  Max.   : 0.7527   Max.   : 0.70489   Max.   : 0.59408   Max.   : 0.439666
##      FGF18              CDC42BPA             DTL               WISP1
##  Min.   :-0.597786   Min.   :-0.44439   Min.   :-1.2645   Min.   :-0.44042
##  1st Qu.:-0.140422   1st Qu.:-0.15187   1st Qu.:-0.6506   1st Qu.:-0.08759
##  Median : 0.001504   Median :-0.04357   Median :-0.1533   Median : 0.02402
##  Mean   :-0.023152   Mean   :-0.02640   Mean   :-0.2095   Mean   : 0.01312
```

```
##  3rd Qu.: 0.106955   3rd Qu.: 0.08044   3rd Qu.: 0.2034   3rd Qu.: 0.12234
##  Max.   : 0.482246   Max.   : 0.48422   Max.   : 0.8919   Max.   : 0.37552
##     DIAPH3.2              OXCT1               ZNF533
##  Min.   :-0.4510200   Min.   :-0.427838   Min.   :-0.51090
##  1st Qu.:-0.1220947   1st Qu.:-0.090491   1st Qu.:-0.26128
##  Median : 0.0088287   Median : 0.009548   Median :-0.13802
##  Mean   :-0.0009119   Mean   : 0.016115   Mean   :-0.05926
##  3rd Qu.: 0.1126542   3rd Qu.: 0.123381   3rd Qu.: 0.03807
##  Max.   : 0.3668805   Max.   : 0.649058   Max.   : 0.86482
##       RFC4                KNTC2              FBXO31            GSTM3_s
##  Min.   :-0.5635877   Min.   :-0.43109   Min.   :-0.42152   Min.   :-1.8446
##  1st Qu.:-0.0824637   1st Qu.:-0.18407   1st Qu.:-0.13880   1st Qu.:-0.7628
##  Median :-0.0009982   Median :-0.06158   Median :-0.04505   Median :-0.1297
##  Mean   : 0.0080165   Mean   :-0.03585   Mean   :-0.02535   Mean   : 0.0000
##  3rd Qu.: 0.1044821   3rd Qu.: 0.07221   3rd Qu.: 0.08601   3rd Qu.: 0.5966
##  Max.   : 0.4790691   Max.   : 0.59748   Max.   : 0.55562   Max.   : 2.7856
##  RP5.860F19.3_s         BBC3_s              MMP9_s         Contig35251_RC_s
##  Min.   :-2.22320   Min.   :-2.91550   Min.   :-2.19263   Min.   :-1.5341
##  1st Qu.:-0.62098   1st Qu.:-0.66246   1st Qu.:-0.59238   1st Qu.:-0.7852
##  Median :-0.03482   Median : 0.05306   Median :-0.05094   Median :-0.3478
##  Mean   : 0.00000   Mean   : 0.00000   Mean   : 0.00000   Mean   : 0.0000
##  3rd Qu.: 0.44234   3rd Qu.: 0.67322   3rd Qu.: 0.59917   3rd Qu.: 0.6803
##  Max.   : 2.92348   Max.   : 2.14875   Max.   : 2.65532   Max.   : 2.8699
##  Contig40831_RC_s    ALDH4A1_s           SERF1A_s           SCUBE2_s
##  Min.   :-2.6527   Min.   :-3.3222   Min.   :-3.21432   Min.   :-3.14702
##  1st Qu.:-0.7294   1st Qu.:-0.6606   1st Qu.:-0.53444   1st Qu.:-0.67239
##  Median : 0.1196   Median : 0.1057   Median : 0.06969   Median : 0.01043
##  Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.00000   Mean   : 0.00000
##  3rd Qu.: 0.6506   3rd Qu.: 0.7429   3rd Qu.: 0.56790   3rd Qu.: 0.63563
##  Max.   : 2.2963   Max.   : 2.8307   Max.   : 2.12506   Max.   : 2.95771
##      MTDH_s              DCK_s               FLT1_s           PECI.1_s
##  Min.   :-2.10229   Min.   :-2.11967   Min.   :-3.0575   Min.   :-2.45636
##  1st Qu.:-0.73728   1st Qu.:-0.74828   1st Qu.:-0.6363   1st Qu.:-0.65099
##  Median : 0.01181   Median :-0.06646   Median : 0.1228   Median :-0.04077
##  Mean   : 0.00000   Mean   : 0.00000   Mean   : 0.0000   Mean   : 0.00000
##  3rd Qu.: 0.57328   3rd Qu.: 0.58355   3rd Qu.: 0.5722   3rd Qu.: 0.56769
##  Max.   : 2.59638   Max.   : 3.31971   Max.   : 3.2270   Max.   : 3.35592
##    QSCN6L1_s           DIAPH3_s           SLC2A3_s           GPR180_s
##  Min.   :-3.0155   Min.   :-2.57705   Min.   :-2.44301   Min.   :-2.63771
##  1st Qu.:-0.5134   1st Qu.:-0.59456   1st Qu.:-0.56847   1st Qu.:-0.51461
##  Median :-0.1046   Median : 0.03018   Median :-0.06929   Median :-0.05297
##  Mean   : 0.0000   Mean   : 0.00000   Mean   : 0.00000   Mean   : 0.00000
##  3rd Qu.: 0.5745   3rd Qu.: 0.64710   3rd Qu.: 0.44146   3rd Qu.: 0.56789
##  Max.   : 3.8974   Max.   : 2.15003   Max.   : 2.88913   Max.   : 2.65928
##    RTN4RL1_s       Contig32125_RC_s      STK32B_s           EXT1_s
##  Min.   :-2.5030   Min.   :-3.35671   Min.   :-2.7107   Min.   :-2.48431
##  1st Qu.:-0.6593   1st Qu.:-0.65984   1st Qu.:-0.6273   1st Qu.:-0.67431
##  Median : 0.1847   Median : 0.01317   Median : 0.1098   Median :-0.02248
##  Mean   : 0.0000   Mean   : 0.00000   Mean   : 0.0000   Mean   : 0.00000
##  3rd Qu.: 0.6958   3rd Qu.: 0.54383   3rd Qu.: 0.5315   3rd Qu.: 0.65590
##  Max.   : 1.8857   Max.   : 3.01074   Max.   : 3.0816   Max.   : 2.48502
##     COL4A2_s            PECI_s             GNAZ_s           AYTL2_s
##  Min.   :-2.64893   Min.   :-1.8949   Min.   :-2.1671   Min.   :-3.9797
##  1st Qu.:-0.67945   1st Qu.:-0.7341   1st Qu.:-0.6995   1st Qu.:-0.6350
```

```
##  Median : 0.03336   Median :-0.1237   Median :-0.1749   Median :-0.1238
##  Mean   : 0.00000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 0.60120   3rd Qu.: 0.6264   3rd Qu.: 0.4849   3rd Qu.: 0.5389
##  Max.   : 3.04573   Max.   : 3.0234   Max.   : 2.7824   Max.   : 3.3235
##  Contig63649_RC_s     RAB6B_s        AA555029_RC_s        GPR126_s
##  Min.   :-2.5733   Min.   :-2.7104   Min.   :-2.1901   Min.   :-2.18251
##  1st Qu.:-0.6433   1st Qu.:-0.6181   1st Qu.:-0.7431   1st Qu.:-0.71889
##  Median :-0.1121   Median :-0.1719   Median : 0.1064   Median : 0.03564
##  Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.00000
##  3rd Qu.: 0.7184   3rd Qu.: 0.5242   3rd Qu.: 0.6867   3rd Qu.: 0.68622
##  Max.   : 2.3843   Max.   : 2.5134   Max.   : 4.4950   Max.   : 2.73711
##      ECT2_s           NUSAP1_s            GMPS_s           UCHL5_s
##  Min.   :-1.9170   Min.   :-2.48846   Min.   :-2.05879   Min.   :-2.65714
##  1st Qu.:-0.7587   1st Qu.:-0.67310   1st Qu.:-0.86692   1st Qu.:-0.65397
##  Median :-0.1310   Median :-0.02731   Median : 0.05945   Median :-0.08845
##  Mean   : 0.0000   Mean   : 0.00000   Mean   : 0.00000   Mean   : 0.00000
##  3rd Qu.: 0.6215   3rd Qu.: 0.65398   3rd Qu.: 0.82692   3rd Qu.: 0.71113
##  Max.   : 3.4583   Max.   : 2.89815   Max.   : 2.37405   Max.   : 3.57788
##     ORC6L_s          TSPYL5_s            MELK_s           RUNDC1_s
##  Min.   :-2.7858   Min.   :-2.72517   Min.   :-3.19321   Min.   :-2.50218
##  1st Qu.:-0.6068   1st Qu.:-0.61758   1st Qu.:-0.60447   1st Qu.:-0.73547
##  Median : 0.1021   Median : 0.03182   Median :-0.05112   Median :-0.04086
##  Mean   : 0.0000   Mean   : 0.00000   Mean   : 0.00000   Mean   : 0.00000
##  3rd Qu.: 0.7544   3rd Qu.: 0.68771   3rd Qu.: 0.53324   3rd Qu.: 0.68595
##  Max.   : 2.0877   Max.   : 2.73746   Max.   : 3.74371   Max.   : 2.81010
##    DIAPH3.1_s        C16orf61_s          TGFB3_s           FGF18_s
##  Min.   :-2.66152   Min.   :-2.8602   Min.   :-2.88616   Min.   :-2.9395
##  1st Qu.:-0.75446   1st Qu.:-0.6716   1st Qu.:-0.62970   1st Qu.:-0.5999
##  Median :-0.05369   Median :-0.1757   Median :-0.02116   Median : 0.1261
##  Mean   : 0.00000   Mean   : 0.0000   Mean   : 0.00000   Mean   : 0.0000
##  3rd Qu.: 0.63996   3rd Qu.: 0.6095   3rd Qu.: 0.59423   3rd Qu.: 0.6656
##  Max.   : 2.82725   Max.   : 3.3800   Max.   : 3.08895   Max.   : 2.5853
##    CDC42BPA_s           DTL_s            WISP1_s          DIAPH3.2_s
##  Min.   :-2.41079   Min.   :-2.0065   Min.   :-2.82692   Min.   :-2.76135
##  1st Qu.:-0.72369   1st Qu.:-0.8389   1st Qu.:-0.62776   1st Qu.:-0.74344
##  Median :-0.09906   Median : 0.1071   Median : 0.06794   Median : 0.05976
##  Mean   : 0.00000   Mean   : 0.0000   Mean   : 0.00000   Mean   : 0.00000
##  3rd Qu.: 0.61617   3rd Qu.: 0.7855   3rd Qu.: 0.68072   3rd Qu.: 0.69671
##  Max.   : 2.94497   Max.   : 2.0949   Max.   : 2.25875   Max.   : 2.25636
##     OXCT1_s            ZNF533_s            RFC4_s            KNTC2_s
##  Min.   :-2.43548   Min.   :-1.5654   Min.   :-3.62020   Min.   :-1.9780
##  1st Qu.:-0.58483   1st Qu.:-0.7002   1st Qu.:-0.57305   1st Qu.:-0.7418
##  Median :-0.03602   Median :-0.2730   Median :-0.05709   Median :-0.1287
##  Mean   : 0.00000   Mean   : 0.0000   Mean   : 0.00000   Mean   : 0.0000
##  3rd Qu.: 0.58845   3rd Qu.: 0.3374   3rd Qu.: 0.61096   3rd Qu.: 0.5408
##  Max.   : 3.47227   Max.   : 3.2030   Max.   : 2.98337   Max.   : 3.1695
##     FBXO31_s          GSTM3_l       RP5.860F19.3_l       BBC3_l
##  Min.   :-2.3615   Min.   :0.971   Min.   :0.9462   Min.   :0.6509
##  1st Qu.:-0.6763   1st Qu.:1.049   1st Qu.:1.0622   1st Qu.:0.9808
##  Median :-0.1175   Median :1.092   Median :1.1015   Median :1.0663
##  Mean   : 0.0000   Mean   :1.098   Mean   :1.1017   Mean   :1.0533
##  3rd Qu.: 0.6638   3rd Qu.:1.139   3rd Qu.:1.1324   3rd Qu.:1.1349
##  Max.   : 3.4630   Max.   :1.269   Max.   :1.2792   Max.   :1.2814
##      MMP9_l       Contig35251_RC_l Contig40831_RC_l   ALDH4A1_l
```

```
##  Min.   :0.9186   Min.   :0.7335   Min.   :0.9276   Min.   :0.8029
##  1st Qu.:1.0436   1st Qu.:0.8786   1st Qu.:1.0558   1st Qu.:1.0385
##  Median :1.0826   Median :0.9545   Median :1.1076   Median :1.0972
##  Mean   :1.0837   Mean   :0.9993   Mean   :1.0987   Mean   :1.0865
##  3rd Qu.:1.1275   3rd Qu.:1.1131   3rd Qu.:1.1386   3rd Qu.:1.1435
##  Max.   :1.2575   Max.   :1.3849   Max.   :1.2292   Max.   :1.2818
##     SERF1A_l         SCUBE2_l          MTDH_l           DCK_l
##  Min.   :0.8935   Min.   :0.9102   Min.   :0.8434   Min.   :0.7378
##  1st Qu.:1.0653   1st Qu.:1.0546   1st Qu.:0.9957   1st Qu.:0.9048
##  Median :1.1002   Median :1.0910   Median :1.0704   Median :0.9784
##  Mean   :1.0946   Mean   :1.0891   Mean   :1.0647   Mean   :0.9801
##  3rd Qu.:1.1282   3rd Qu.:1.1233   3rd Qu.:1.1229   3rd Qu.:1.0439
##  Max.   :1.2108   Max.   :1.2346   Max.   :1.2921   Max.   :1.2805
##      FLT1_l          PECI.1_l        QSCN6L1_l         DIAPH3_l
##  Min.   :0.9232   Min.   :0.9425   Min.   :0.9634   Min.   :0.9364
##  1st Qu.:1.0644   1st Qu.:1.0510   1st Qu.:1.0830   1st Qu.:1.0605
##  Median :1.1049   Median :1.0851   Median :1.1012   Median :1.0967
##  Mean   :1.0971   Mean   :1.0859   Mean   :1.1049   Mean   :1.0933
##  3rd Qu.:1.1281   3rd Qu.:1.1180   3rd Qu.:1.1308   3rd Qu.:1.1311
##  Max.   :1.2551   Max.   :1.2564   Max.   :1.2642   Max.   :1.2104
##     SLC2A3_l         GPR180_l        RTN4RL1_l      Contig32125_RC_l
##  Min.   :0.9664   Min.   :0.9726   Min.   :0.8482   Min.   :0.9034
##  1st Qu.:1.0724   1st Qu.:1.0715   1st Qu.:1.0276   1st Qu.:1.0601
##  Median :1.0988   Median :1.0917   Median :1.1001   Median :1.0956
##  Mean   :1.1011   Mean   :1.0931   Mean   :1.0811   Mean   :1.0936
##  3rd Qu.:1.1251   3rd Qu.:1.1184   3rd Qu.:1.1416   3rd Qu.:1.1228
##  Max.   :1.2425   Max.   :1.2031   Max.   :1.2320   Max.   :1.2402
##     STK32B_l          EXT1_l          COL4A2_l         PECI_l
##  Min.   :0.9241   Min.   :0.9251   Min.   :0.876   Min.   :0.9391
##  1st Qu.:1.0498   1st Qu.:1.0411   1st Qu.:1.030   1st Qu.:1.0317
##  Median :1.0908   Median :1.0798   Median :1.081   Median :1.0771
##  Mean   :1.0833   Mean   :1.0795   Mean   :1.076   Mean   :1.0836
##  3rd Qu.:1.1135   3rd Qu.:1.1186   3rd Qu.:1.119   3rd Qu.:1.1303
##  Max.   :1.2407   Max.   :1.2161   Max.   :1.270   Max.   :1.2834
##      GNAZ_l          AYTL2_l       Contig63649_RC_l    RAB6B_l
##  Min.   :0.9868   Min.   :0.8354   Min.   :0.9687   Min.   :0.8882
##  1st Qu.:1.0662   1st Qu.:1.0536   1st Qu.:1.0653   1st Qu.:1.0497
##  Median :1.0931   Median :1.0832   Median :1.0903   Median :1.0811
##  Mean   :1.1007   Mean   :1.0886   Mean   :1.0944   Mean   :1.0906
##  3rd Qu.:1.1260   3rd Qu.:1.1202   3rd Qu.:1.1282   3rd Qu.:1.1280
##  Max.   :1.2327   Max.   :1.2623   Max.   :1.2001   Max.   :1.2512
##  AA555029_RC_l       GPR126_l          ECT2_l          NUSAP1_l
##  Min.   :0.9436   Min.   :0.9633   Min.   :0.9132   Min.   :0.8812
##  1st Qu.:1.0438   1st Qu.:1.0522   1st Qu.:1.0184   1st Qu.:1.0436
##  Median :1.0983   Median :1.0951   Median :1.0712   Median :1.0955
##  Mean   :1.0897   Mean   :1.0916   Mean   :1.0786   Mean   :1.0946
##  3rd Qu.:1.1338   3rd Qu.:1.1307   3rd Qu.:1.1309   3rd Qu.:1.1475
##  Max.   :1.3403   Max.   :1.2353   Max.   :1.3286   Max.   :1.3020
##      GMPS_l          UCHL5_l          ORC6L_l          TSPYL5_l
##  Min.   :0.8790   Min.   :0.9327   Min.   :0.7899   Min.   :0.842
##  1st Qu.:0.9991   1st Qu.:1.0539   1st Qu.:1.0246   1st Qu.:1.037
##  Median :1.0835   Median :1.0857   Median :1.0905   Median :1.090
##  Mean   :1.0744   Mean   :1.0890   Mean   :1.0770   Mean   :1.085
##  3rd Qu.:1.1483   3rd Qu.:1.1288   3rd Qu.:1.1474   3rd Qu.:1.141
```

```
##  Max.   :1.2675   Max.   :1.2700   Max.   :1.2547   Max.   :1.286
##      MELK_l          RUNDC1_l        DIAPH3.1_l       C16orf61_l
##  Min.   :0.7931   Min.   :0.7559   Min.   :0.8028   Min.   :0.8705
##  1st Qu.:1.0334   1st Qu.:0.9819   1st Qu.:1.0093   1st Qu.:1.0336
##  Median :1.0780   Median :1.0584   Median :1.0756   Median :1.0671
##  Mean   :1.0790   Mean   :1.0572   Mean   :1.0764   Mean   :1.0766
##  3rd Qu.:1.1231   3rd Qu.:1.1326   3rd Qu.:1.1371   3rd Qu.:1.1180
##  Max.   :1.3400   Max.   :1.3225   Max.   :1.3097   Max.   :1.2793
##     TGFB3_l          FGF18_l         CDC42BPA_l         DTL_l
##  Min.   :0.9496   Min.   :0.8764   Min.   :0.9383   Min.   :0.5513
##  1st Qu.:1.0673   1st Qu.:1.0507   1st Qu.:1.0467   1st Qu.:0.8542
##  Median :1.0968   Median :1.0991   Median :1.0840   Median :1.0462
##  Mean   :1.0967   Mean   :1.0887   Mean   :1.0881   Mean   :1.0074
##  3rd Qu.:1.1258   3rd Qu.:1.1336   3rd Qu.:1.1251   3rd Qu.:1.1642
##  Max.   :1.2354   Max.   :1.2477   Max.   :1.2482   Max.   :1.3589
##     WISP1_l         DIAPH3.2_l        OXCT1_l          ZNF533_l
##  Min.   :0.9398   Min.   :0.9357   Min.   :0.9447   Min.   :0.9119
##  1st Qu.:1.0690   1st Qu.:1.0571   1st Qu.:1.0680   1st Qu.:1.0075
##  Median :1.1066   Median :1.1016   Median :1.1018   Median :1.0515
##  Mean   :1.1015   Mean   :1.0968   Mean   :1.1022   Mean   :1.0742
##  3rd Qu.:1.1386   3rd Qu.:1.1355   3rd Qu.:1.1389   3rd Qu.:1.1112
##  Max.   :1.2165   Max.   :1.2140   Max.   :1.2945   Max.   :1.3519
##      RFC4_l          KNTC2_l          FBXO31_l
##  Min.   :0.8905   Min.   :0.9435   Min.   :0.9472
##  1st Qu.:1.0707   1st Qu.:1.0353   1st Qu.:1.0512
##  Median :1.0983   Median :1.0779   Median :1.0835
##  Mean   :1.0999   Mean   :1.0844   Mean   :1.0886
##  3rd Qu.:1.1328   3rd Qu.:1.1224   3rd Qu.:1.1269
##  Max.   :1.2468   Max.   :1.2802   Max.   :1.2685
```

```
head(viral34_c)
```

```
##             infection     stime              sind gender              hosp age
## 1 bacterial_infection  7.296372   symptoms_remain female    hospitalization  47
## 2     viral_infection  6.718686   symptoms_remain   male no_hospitalization  47
## 3 bacterial_infection  6.995209 symptoms_finished   male    hospitalization  38
## 4     viral_infection  9.330595   symptoms_remain female    hospitalization  45
## 5     viral_infection  3.438741 symptoms_finished   male no_hospitalization  31
## 6     viral_infection 15.329227   symptoms_remain female no_hospitalization  41
##   ancestry       GSTM3 RP5.860F19.3        BBC3        MMP9 Contig35251_RC
## 1        A  0.14647630  -0.09803689  0.30821656 -0.20635196     -0.5190545
## 2        A -0.03543524  -0.02103562 -0.09643536  0.24147416     -0.5319210
## 3        B -0.26258909   0.01080372  0.08854258 -0.02584139     -0.3400320
## 4        A  0.33787726   0.34173748  0.32773524 -0.34135513     -0.6258146
## 5        A  0.09657176   0.07818674 -0.40614429 -0.17861930     -0.6432336
## 6        A -0.21568976  -0.02222821  0.25438095  0.22220903      0.5438846
##   Contig40831_RC     ALDH4A1      SERF1A       SCUBE2        MTDH        DCK
## 1   -0.115149133  0.16674915 -0.08378990  0.006056118 -0.13776980 -0.3355029
## 2   -0.003368632  0.09845308  0.10674758  0.097569969  0.49423377 -0.5800370
## 3   -0.080972535  0.06714804  0.06265814 -0.125834721  0.07226344  0.1389293
## 4   -0.126344331  0.60304554 -0.52384308  0.085869554 -0.57695111 -0.5250435
## 5    0.084764382  0.15974585 -0.01786414 -0.144122253 -0.51943819 -0.1974320
## 6    0.155627613 -0.11070614 -0.15845941  0.003581842  0.01179057 -0.5446215
##          FLT1      PECI.1     QSCN6L1      DIAPH3      SLC2A3       GPR180
```

```
## 1 -0.05592013 -0.08777178 -0.12461618  0.08084737  0.36857538 -0.045212204
## 2  0.16900027 -0.09680589  0.26558747  0.12485686  0.46424057 -0.187539926
## 3  0.06697535 -0.11336268  0.09573915  0.23404700 -0.07758043 -0.005409022
## 4 -0.03036967  0.10875631 -0.08282874 -0.17743971 -0.22028177  0.115940757
## 5  0.06349948 -0.18639378 -0.08196639  0.10069529 -0.04723501 -0.111992419
## 6 -0.01398411 -0.29762197  0.25657503 -0.03541076 -0.05243454  0.117939467
##       RTN4RL1 Contig32125_RC      STK32B        EXT1      COL4A2        PECI
## 1 -0.16287206    -0.01326674  0.02778352 -0.13445797 -0.02990369  0.17349207
## 2  0.20005740    -0.01326124  0.12973975 -0.19559405 -0.22673707  0.21204866
## 3  0.12189144    -0.07876585 -0.06204672 -0.11337939 -0.20833506  0.04232249
## 4 -0.64421170    -0.02356354  0.01292946  0.02194074  0.10271000  0.47963136
## 5  0.02815844    -0.10416021 -0.12139572 -0.31679313 -0.25780916  0.10048933
## 6  0.38596644    -0.16397529 -0.11000363 -0.12637540 -0.35700047 -0.10518681
##          GNAZ       AYTL2 Contig63649_RC       RAB6B AA555029_RC      GPR126
## 1  0.07047200  0.23928807    0.029619073  0.46141386 -0.04808210 -0.10022007
## 2 -0.03168261  0.01566120    0.004188184  0.01856611  0.15926624  0.28115470
## 3  0.06296012 -0.12755049    0.050402671 -0.14251272  0.11420782  0.05710594
## 4  0.33487679  0.53361497    0.294420521  0.32544391 -0.31064082  0.19116150
## 5 -0.14666447  0.11709402    0.026705563 -0.08731065 -0.22007577 -0.08256859
## 6 -0.20496584 -0.05130563   -0.303730416 -0.27515282  0.05199264 -0.16971181
##          ECT2      NUSAP1        GMPS       UCHL5      ORC6L      TSPYL5
## 1  0.03544526  0.10981625  0.21805322 -0.03809348  0.1728180  0.15589646
## 2 -0.03772432  0.03225047  0.18573594 -0.27078994  0.1017376 -0.05882551
## 3 -0.18130437 -0.04820767  0.04043471 -0.04321627 -0.1501144  0.09089323
## 4  0.03338104  0.65528392  0.23712422 -0.19231373  0.1926975  0.54098979
## 5  0.32874172  0.07952344  0.17836256 -0.14087134  0.1256798  0.04560219
## 6 -0.21921949 -0.27156456 -0.33843161  0.13085033 -0.3269674 -0.21852037
##          MELK       RUNDC1    DIAPH3.1    C16orf61       TGFB3       FGF18
## 1 -0.34581318  0.558363335 -0.44455761  0.05912505  0.08180754 -0.04819787
## 2 -0.01081727 -0.358850367 -0.24259102 -0.05018147  0.11869773  0.27382112
## 3 -0.13659874  0.129923754 -0.05644101 -0.27369996  0.11315389  0.03470079
## 4 -0.33968909 -0.071808178 -0.44559038  0.13548833 -0.02046981  0.10391225
## 5 -0.24841783 -0.007654665 -0.09683179  0.20482626  0.12745233  0.18982466
## 6 -0.13351906 -0.495218907  0.20965383 -0.17109518  0.26659311 -0.19909908
##      CDC42BPA        DTL        WISP1     DIAPH3.2       OXCT1      ZNF533
## 1  0.19203352 -1.0115741 -0.004976858 -0.297776741  0.03135030  0.86482037
## 2  0.12535125 -0.1460709  0.217921322  0.020572007  0.16334775  0.01575178
## 3 -0.22807824  0.4482309  0.071255809 -0.144136303  0.05694880 -0.14760060
## 4  0.03281893 -1.0765161 -0.440424509  0.051227778 -0.20543872 -0.20651322
## 5 -0.07390664 -0.8427116  0.119424302  0.008235576 -0.13898008  0.28849255
## 6 -0.26596006  0.7104907 -0.166179324  0.121529650  0.08704478 -0.15575775
##          RFC4       KNTC2      FBXO31    GSTM3_s RP5.860F19.3_s      BBC3_s
## 1  0.06185603  0.59748058 -0.04140661  0.7138813    -0.57440211  1.26618208
## 2  0.01687964 -0.32724674 -0.13521580 -0.2060675    -0.18510052  0.04967577
## 3 -0.05427720  0.09654722  0.03522958 -1.3548121    -0.02412778  0.60577550
## 4  0.39446170  0.04598343  0.06070769  1.6818192     1.64900082  1.32486112
## 5  0.06241235 -0.09261277  0.26401621  0.4615079     0.31654599 -0.88140315
## 6 -0.19665862 -0.18365899 -0.01361086 -1.1176365    -0.19112999  1.10433590
##        MMP9_s Contig35251_RC_s Contig40831_RC_s  ALDH4A1_s    SERF1A_s
## 1 -0.81209763       -0.6159014      -0.67108894  0.8726701 -0.44912514
## 2  1.33518865       -0.6455369      -0.04954233  0.5661609  0.66594479
## 3  0.05343442       -0.2035604      -0.48105268  0.4256655  0.40792311
## 4 -1.45942582       -0.8618013      -0.73333895  2.8307454 -3.02441939
## 5 -0.67912202       -0.9019223       0.44051430  0.8412396 -0.06331218
```

```
## 6  1.24281408        1.8323579       0.83454368 -0.3725348 -0.88610863
##     SCUBE2_s     MTDH_s       DCK_s      FLT1_s   PECI.1_s  QSCN6L1_s
## 1  0.1942805 -0.1821611 -0.05112308 -0.35139159 -0.3325357 -1.0997776
## 2  0.7808869  2.0740241 -0.93365927  1.07514078 -0.3880154  1.8335965
## 3 -0.6511435  0.5676351  1.66112734  0.42805919 -0.4896928  0.5567539
## 4  0.7058869 -1.7499914 -0.73518462 -0.18934069  0.8743696 -0.7856385
## 5 -0.7683671 -1.5446764  0.44718207  0.40601389 -0.9381866 -0.7791558
## 6  0.1784203  0.3517534 -0.80584276 -0.08541711 -1.6212538  1.7658451
##     DIAPH3_s   SLC2A3_s    GPR180_s  RTN4RL1_s Contig32125_RC_s   STK32B_s
## 1  0.5392279  2.2788093 -0.24330929 -0.4879262      -0.01428243  0.4260294
## 2  0.7979150  2.8891253 -1.34271285  0.9697711      -0.01424700  1.0552938
## 3  1.4397330 -0.5675343  0.06414844  0.6558194      -0.43623247 -0.1283947
## 4 -0.9789797 -1.4779272  1.00150889 -2.4212151      -0.08061514  0.3343515
## 5  0.6558937 -0.3739394 -0.75914981  0.2793431      -0.59982477 -0.4946913
## 6 -0.1441357 -0.4071109  1.01694783  1.7164702      -0.98515784 -0.4243804
##       EXT1_s   COL4A2_s     PECI_s     GNAZ_s    AYTL2_s Contig63649_RC_s
## 1 -0.4813926  0.1461046  0.9860751  0.3995737  1.5729248       0.28173399
## 2 -0.8379918 -0.8211227  1.1664517 -0.2763188  0.2428707       0.09793693
## 3 -0.3584438 -0.7306964  0.3724338  0.3498723 -0.6089022       0.43194361
## 4  0.4308618  0.7977604  2.4182648  2.1489730  3.3234779       2.19553759
## 5 -1.5449311 -0.9738091  0.6445516 -1.0370811  0.8461575       0.26067713
## 6 -0.4342479 -1.4612293 -0.3176484 -1.4228244 -0.1554245      -2.12748792
##      RAB6B_s AA555029_RC_s   GPR126_s     ECT2_s   NUSAP1_s     GMPS_s
## 1  2.3501669    -0.1449979 -0.5035802  0.3578867  0.4808374  1.0797140
## 2  0.1756376     0.9631957  1.7873898  0.0514164  0.1499810  0.9544301
## 3 -0.6153129     0.7223765  0.4414985 -0.5499677 -0.1932127  0.3911430
## 4  1.6825094    -1.5482689  1.2467884  0.3492407  2.8075265  1.1536463
## 5 -0.3442524    -1.0642351 -0.3975453  1.5863570  0.3516236  0.9258458
## 6 -1.2666197     0.3898614 -0.9210265 -0.7087750 -1.1459402 -1.0776023
##       UCHL5_s    ORC6L_s   TSPYL5_s     MELK_s   RUNDC1_s   DIAPH3.1_s
## 1 -0.08519916  0.8392967  0.7914942 -1.2786512  2.1740843 -1.455677597
## 2 -1.50870499  0.5735413 -0.1130234  0.1658503 -0.8279555 -0.703133863
## 3 -0.11653751 -0.3680828  0.5176677 -0.3765191  0.7718018 -0.009523967
## 4 -1.02863177  0.9136225  2.4137021 -1.2522442  0.1115334 -1.459525754
## 5 -0.71393620  0.6630564  0.3268789 -0.8586827  0.3215078 -0.160023276
## 6  0.94830393 -1.0293010 -0.7857389 -0.3632395 -1.2742896  0.981966756
##    C16orf61_s    TGFB3_s    FGF18_s CDC42BPA_s      DTL_s    WISP1_s
## 1  0.61186055  0.5877739 -0.1281211  1.2597867 -1.5254135 -0.1128211
## 2  0.04624479  0.8456102  1.5191381  0.8751973  0.1207287  1.2764815
## 3 -1.11037043  0.8068628  0.2959395 -1.1632042  1.2510601  0.3623297
## 4  1.00700876 -0.1270723  0.6499845  0.3415180 -1.6489300 -2.8269239
## 5  1.36580373  0.9067987  1.0894618 -0.2740209 -1.2042460  0.6625592
## 6 -0.57943337  1.8792943 -0.9000427 -1.3816873  1.7498645 -1.1175804
##    DIAPH3.2_s    OXCT1_s   ZNF533_s      RFC4_s    KNTC2_s    FBXO31_s  GSTM3_l
## 1 -1.82122539  0.0835812  3.2029626  0.34098731  3.1695391 -0.09573504 1.146283
## 2  0.13180112  0.8077063  0.2600142  0.05613357 -1.4582792 -0.65490852 1.086730
## 3 -0.87866209  0.2240121 -0.3061799 -0.39453158  0.6626070  0.36107443 1.007013
## 4  0.31987013 -1.2154201 -0.5103764  2.44751394  0.4095594  0.51294316 1.205335
## 5  0.05611878 -0.8508345  1.2053583  0.34451067 -0.2840483  1.72481525 1.130296
## 6  0.75116261  0.3891157 -0.3344533 -1.29629087 -0.7396911  0.06994858 1.024000
##   RP5.860F19.3_l    BBC3_l    MMP9_l Contig35251_RC_l Contig40831_RC_l
## 1       1.065387 1.1964092 1.0273483        0.9086397         1.059473
## 2       1.091576 1.0659392 1.1760282        0.9034401         1.097489
## 3       1.102207 1.1276993 1.0899612        0.9783141         1.071251
```

```
## 4       1.206491 1.2022920 0.9778165        0.8646544         1.055585
## 5       1.124341 0.9531455 1.0372264        0.8572905         1.126475
## 6       1.091175 1.1800021 1.1700672        1.2652235         1.149187
##   ALDH4A1_l  SERF1A_l SCUBE2_l    MTDH_l     DCK_l   FLT1_l  PECI.1_l QSCN6L1_l
## 1  1.152706 1.0702849 1.100629 1.0516011 0.9800153 1.079796 1.0689185  1.056186
## 2  1.130903 1.1335764 1.130618 1.2511141 0.8837522 1.153416 1.0658115  1.183440
## 3  1.120748 1.1192832 1.055762 1.1224146 1.1438818 1.120692 1.0600923  1.130027
## 4  1.281779 0.9067077 1.126833 0.8850266 0.9062228 1.088437 1.1342227  1.070614
## 5  1.150492 1.0926398 1.049379 0.9084851 1.0305361 1.119558 1.0344670  1.070910
## 6  1.061012 1.0443464 1.099806 1.1025348 0.8982809 1.093940 0.9941321  1.180676
##   DIAPH3_l SLC2A3_l GPR180_l RTN4RL1_l Contig32125_RC_l STK32B_l    EXT1_l
## 1 1.125205 1.214490 1.083427 1.0427923         1.094180 1.107831 1.0527575
## 2 1.139388 1.242493 1.034060 1.1631687         1.094182 1.140950 1.0311917
## 3 1.173734 1.072412 1.096808 1.1384390         1.072006 1.077713 1.0600865
## 4 1.037644 1.022350 1.136531 0.8568754         1.090727 1.102913 1.1058993
## 5 1.131626 1.082742 1.060567 1.1079547         1.063275 1.057306 0.9870127
## 6 1.086738 1.080980 1.137172 1.2196394         1.042403 1.061255 1.0555742
##    COL4A2_l   PECI_l   GNAZ_l  AYTL2_l Contig63649_RC_l  RAB6B_l AA555029_RC_l
## 1 1.0885944 1.154833 1.121831 1.175354        1.1084369 1.241677     1.0824551
## 2 1.0200246 1.166909 1.087995 1.103819        1.1000074 1.104782     1.1503398
## 3 1.0266382 1.112621 1.119382 1.055165        1.1152736 1.049943     1.1359748
## 4 1.1322759 1.246926 1.204436 1.262321        1.1922303 1.201603     0.9893029
## 5 1.0087572 1.131560 1.048489 1.136901        1.1074748 1.069077     1.0224237
## 6 0.9719145 1.062921 1.027844 1.081362        0.9918692 1.002412     1.1157947
##   GPR126_l   ECT2_l NUSAP1_l    GMPS_l  UCHL5_l   ORC6L_l TSPYL5_l    MELK_l
## 1 1.064635 1.110358 1.134564 1.1687766 1.085833 1.1546201 1.149273 0.9761383
## 2 1.188195 1.085958 1.109305 1.1586833 1.004012 1.1319625 1.078809 1.0950000
## 3 1.117469 1.036274 1.082413 1.1120005 1.084102 1.0472789 1.128460 1.0520102
## 4 1.160385 1.109678 1.296174 1.1746853 1.032361 1.1608662 1.264406 0.9784430
## 5 1.070704 1.202594 1.124775 1.1563661 1.050517 1.1396518 1.113699 1.0121761
## 6 1.040379 1.022732 1.003728 0.9789156 1.141305 0.9832136 1.022983 1.0530851
##    RUNDC1_l DIAPH3.1_l C16orf61_l  TGFB3_l  FGF18_l CDC42BPA_l     DTL_l
## 1 1.2693007  0.9382254   1.118129 1.125516 1.082416   1.160658 0.6873433
## 2 0.9712143  1.0142915   1.081744 1.137416 1.185958   1.139547 1.0486967
## 3 1.1410086  1.0796194   1.002945 1.135636 1.110113   1.019541 1.2378613
## 4 1.0743851  0.9378211   1.142785 1.091766 1.132663   1.109493 0.6541381
## 5 1.0960575  1.0658026   1.164658 1.140219 1.159966   1.073668 0.7688520
## 6 0.9182013  1.1661631   1.039890 1.183748 1.029941   1.005780 1.3111641
##     WISP1_l DIAPH3.2_l  OXCT1_l ZNF533_l   RFC4_l   KNTC2_l FBXO31_l
## 1 1.0969520  0.9940749 1.109008 1.351915 1.119021 1.2802338 1.084714
## 2 1.1687356  1.1054462 1.151631 1.103849 1.104223 0.9831091 1.052493
## 3 1.1220865  1.0493743 1.117417 1.048161 1.080354 1.1302877 1.110287
## 4 0.9398414  1.1155441 1.027675 1.027291 1.222145 1.1138238 1.118646
## 5 1.1376485  1.1013537 1.051178 1.190429 1.119203 1.0672548 1.182958
## 6 1.0416259  1.1383232 1.127214 1.045297 1.030812 1.0354385 1.094065
```

```
#PERFORMING UNIVARIATE analysis of the first 6 columns:
#ATTACH/DETACH METHODS NOT USED HERE:

#INFECTION
#Categorical data that Had numerical value 1 or 0 and was later factored

# absolute frequencies
freq.cc<-table(viral34_c$infection)
```

```
freq.cc
```

```
##
## bacterial_infection     viral_infection
##                  69                  71
```

```
#bacterial_infection     viral_infection
#69                   71

# relative frequencies
relfreq.cc<-freq.cc/nrow(viral34_c)
relfreq.cc
```

```
##
## bacterial_infection     viral_infection
##           0.4928571           0.5071429
```

```
#bacterial_infection     viral_infection
#0.4928571           0.5071429

# relative frequencies (ALTERNATIVE METHOD)
relfreq.cc<-prop.table(table(viral34_c$infection))
relfreq.cc
```

```
##
## bacterial_infection     viral_infection
##           0.4928571           0.5071429
```

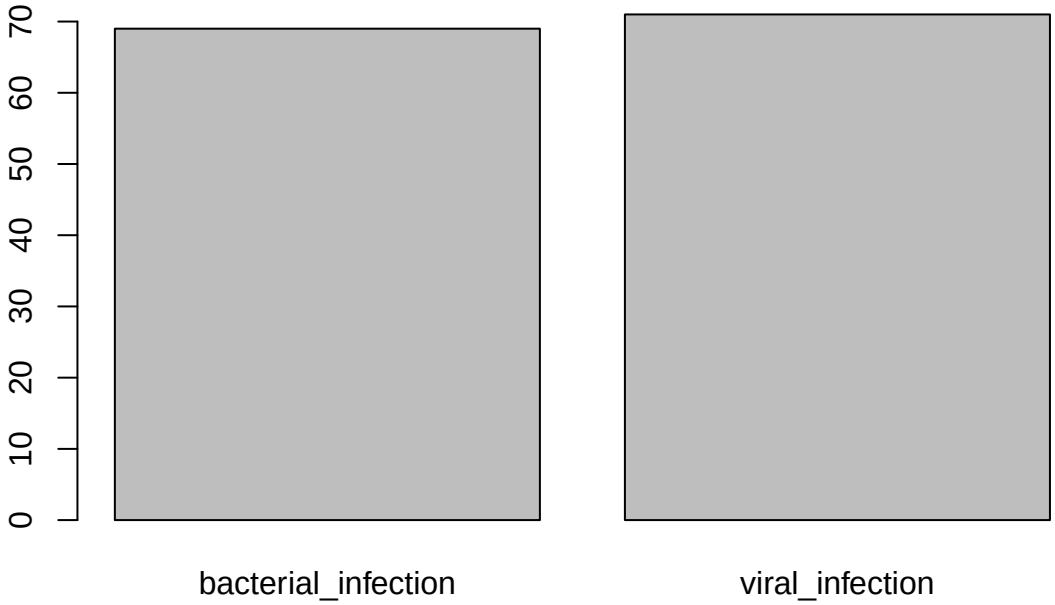
```
# function cbinb() is used to combine two tables
freqtablecc<-cbind(freq.cc, relfreq.cc)
freqtablecc
```

```
##                      freq.cc relfreq.cc
## bacterial_infection      69  0.4928571
## viral_infection          71  0.5071429
```

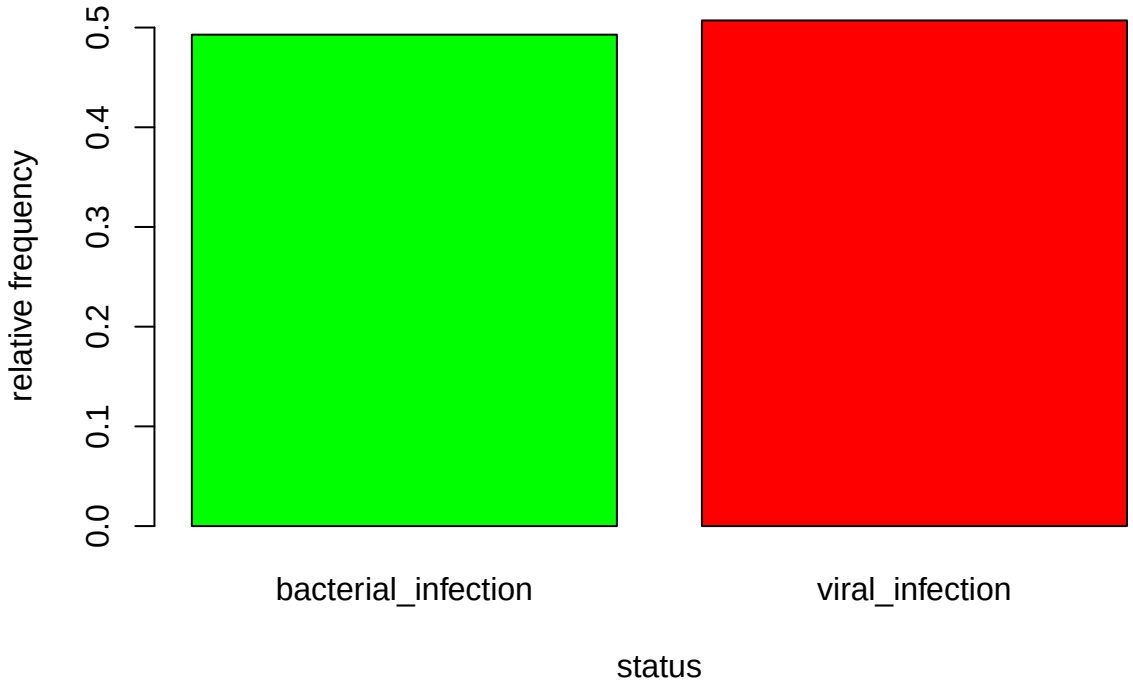
```
options(digits=4)
freqtablecc
```

```
##                      freq.cc relfreq.cc
## bacterial_infection      69     0.4929
## viral_infection          71     0.5071
```

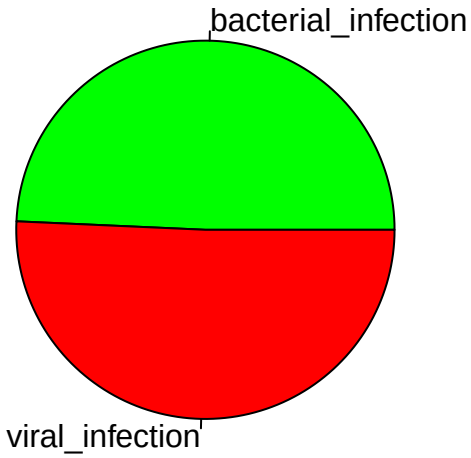
```
#freq.cc relfreq.cc
#bacterial_infection      69     0.4929
#viral_infection          71     0.5071

barplot(freq.cc)
```




```
barplot(relfreq.cc, xlab="status", ylab="relative frequency", names.arg=c("bacterial_infection", "viral_
```




```
pie(relfreq.cc, labels=c("bacterial_infection", "viral_infection"),col=c("green", "red"))
```




```
#AGE
#Age is a continuous, numerical variable
#Summary statistics
# mean or average
mean(viral34_c$age)
```

```
## [1] 44.25
```

```
## [1] 44.25
# median
median(viral34_c$age)
```

```
## [1] 45
```

```
## [1] 45
# range
max(viral34_c$age)-min(viral34_c$age)
```

```
## [1] 27
```

```
## [1] 27
# variance
var(viral34_c$age)
```

```
## [1] 28.81
```

```
## [1] 28.81
# standard deviation
sd(viral34_c$age)
```

```
## [1] 5.367
```

```
## [1] 5.367
# coeficient of variation (in percentage)
100*sd(viral34_c$age)/mean(viral34_c$age)
```

```
## [1] 12.13
```

```
## [1] 12.13
# minimum, first , second and third quartiles, and maximum
quantile(viral34_c$age)
```

```
##   0%  25%  50%  75% 100%
##   26   41   45   49   53
```

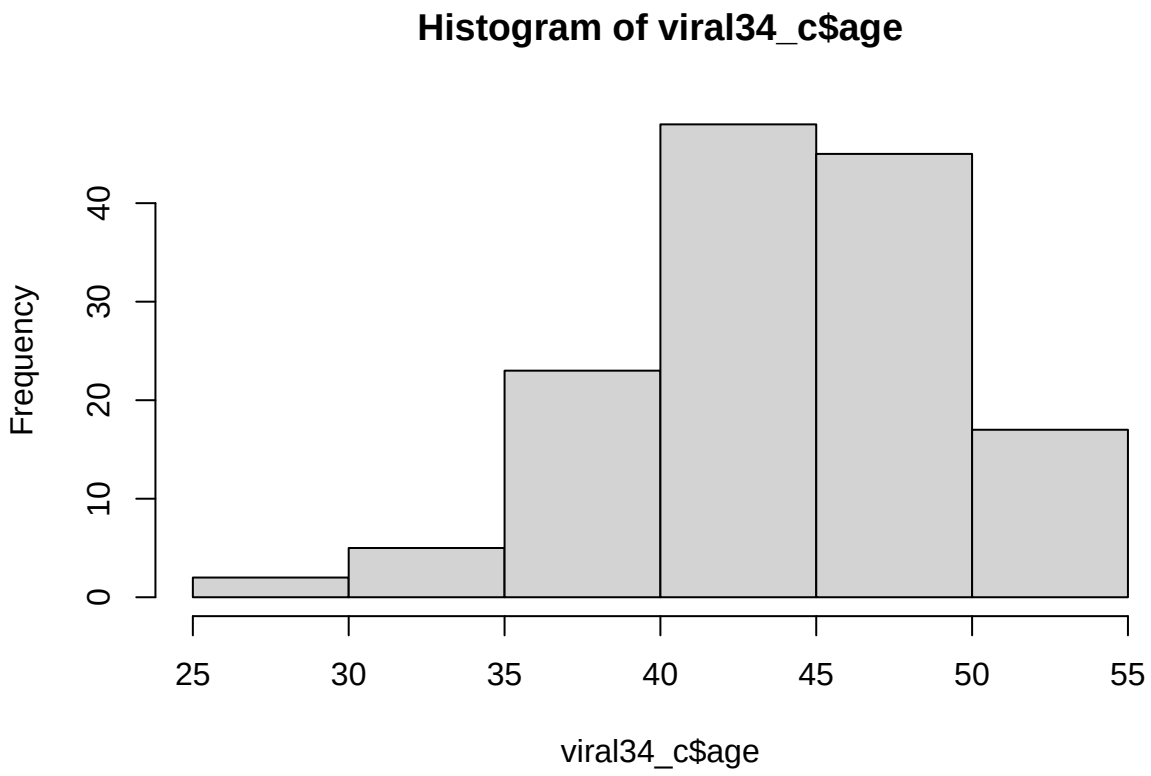
```
#0%  25%  50%  75% 100%
#26   41   45   49   53
# interquartile range
IQR(viral34_c$age)
```

```
## [1] 8
```

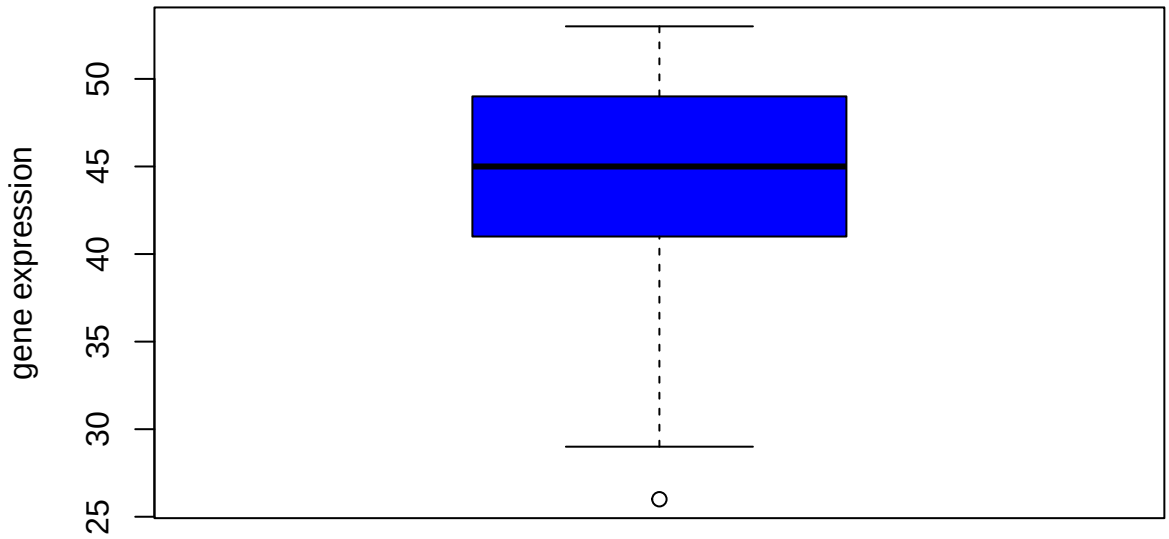
```
## [1]  8

# 35% and 63% quantiles
quantile(viral34_c$age, c(0.35,0.63))
```

```
## 35% 63%
##  42  47
```

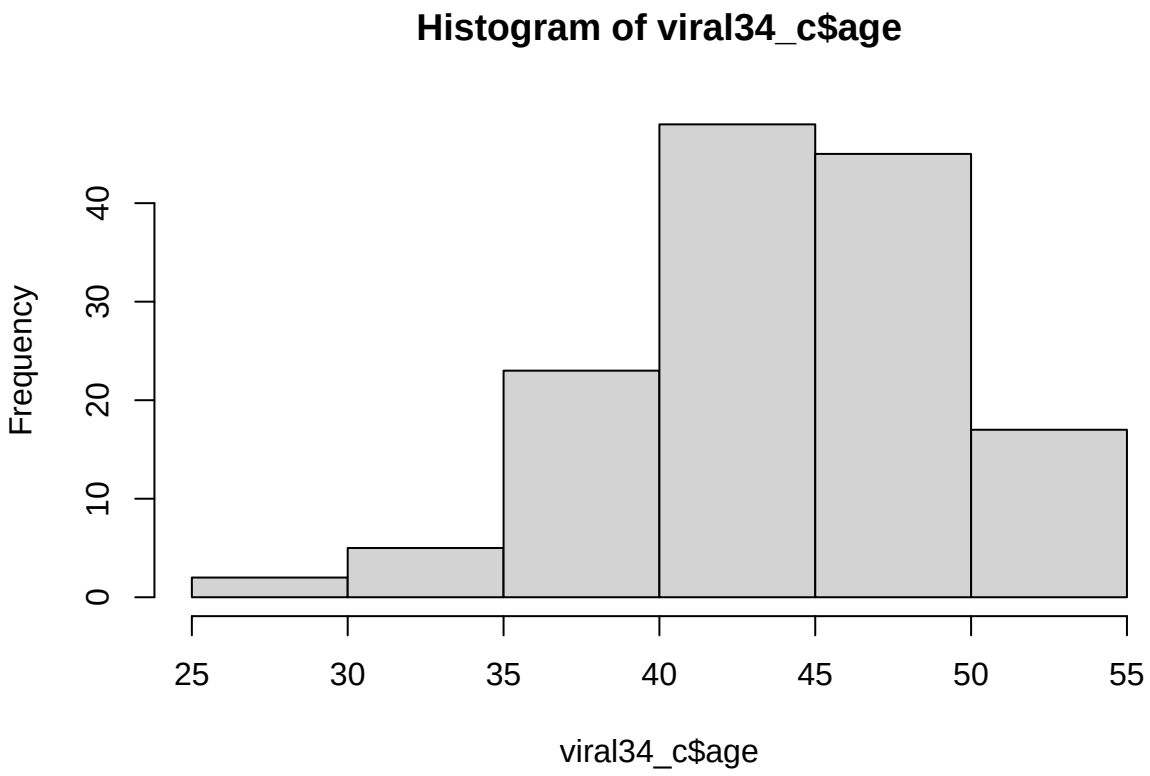
```
#35% 63%
#  42  47

#Histogram
hist(viral34_c$age)
```



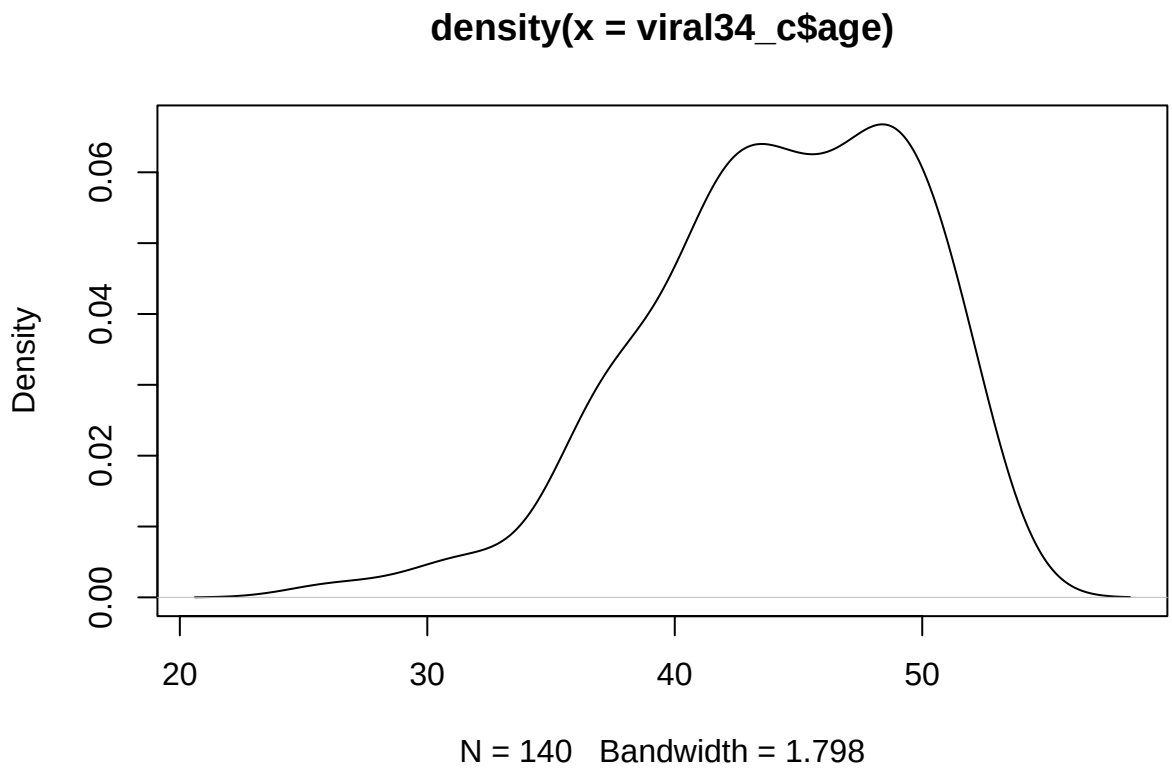

```
#Boxplot:
boxplot(viral34_c$age, ylab="gene expression", col="blue")
```



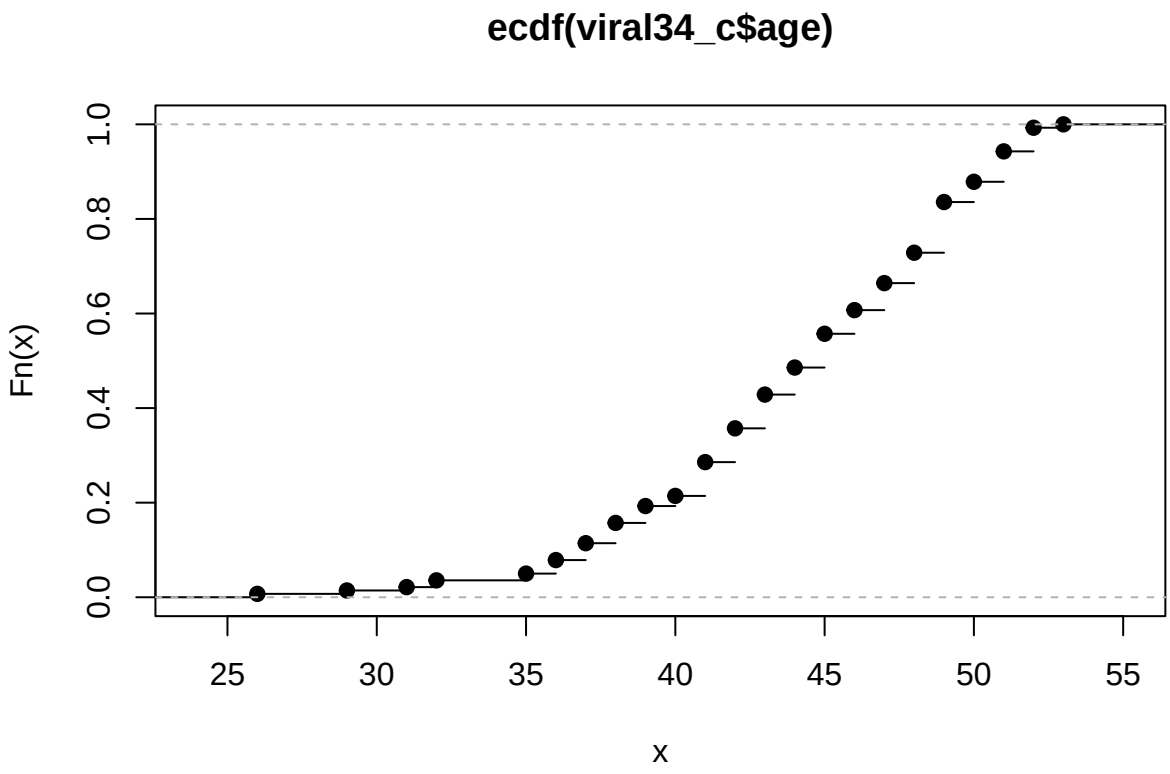

```
#Error in plot.new() : figure margins too large

#Density Function
hist(viral34_c$age)
```




```
density<-density(viral34_c$age)
plot(density)
```




```
#Error in plot.new() : figure margins too large

#Empirical cumulative distribution
f<-ecdf(viral34_c$age)
plot(f)
```




```
#Error in plot.new() : figure margins too large

#Testing for outliers in age:
library(outliers)
```

```
##
## Attaching package: 'outliers'

## The following object is masked from 'package:randomForest':
##
##     outlier
```

```
grubbs.test(viral34_c$age)
```

```
##
##  Grubbs test for one outlier
##
## data:  viral34_c$age
## G = 3.40, U = 0.92, p-value = 0.04
## alternative hypothesis: lowest value 26 is an outlier
```

```
# Grubbs test for one outlier
#
# data:  viral34_c$age
```

```r
# G = 3.40, U = 0.92, p-value = 0.04
# alternative hypothesis: lowest value 26 is an outlier

#Testing for normal distribution of age:
shapiro.test(viral34_c$age)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  viral34_c$age
## W = 0.96, p-value = 3e-04
```

```r
# Shapiro-Wilk normality test
#
# data:  viral34_c$age
# W = 0.96, p-value = 3e-04
#Age is not normally distriuted with p<=0.05


#HOSP
#Hospitalization is Categorical data that Had numerical value 1 or 0 and was earlier factored.
#Indicator of hospitalization (1= hospitalization, 0 = no hospitalization).

# absolute frequencies
freq.cc<-table(viral34_c$hosp)
freq.cc
```

```
##
## no_hospitalization    hospitalization
##                 73                 67
```

```r
# no_hospitalization    hospitalization
# 73                 67

# relative frequencies
relfreq.cc<-freq.cc/nrow(viral34_c)
relfreq.cc
```

```
##
## no_hospitalization    hospitalization
##             0.5214             0.4786
```

```r
# no_hospitalization    hospitalization
# 0.5214             0.4786

# relative frequencies (ALTERNATIVE METHOD)
relfreq.cc<-prop.table(table(viral34_c$hosp))
relfreq.cc
```

```
##
## no_hospitalization    hospitalization
##             0.5214             0.4786
```

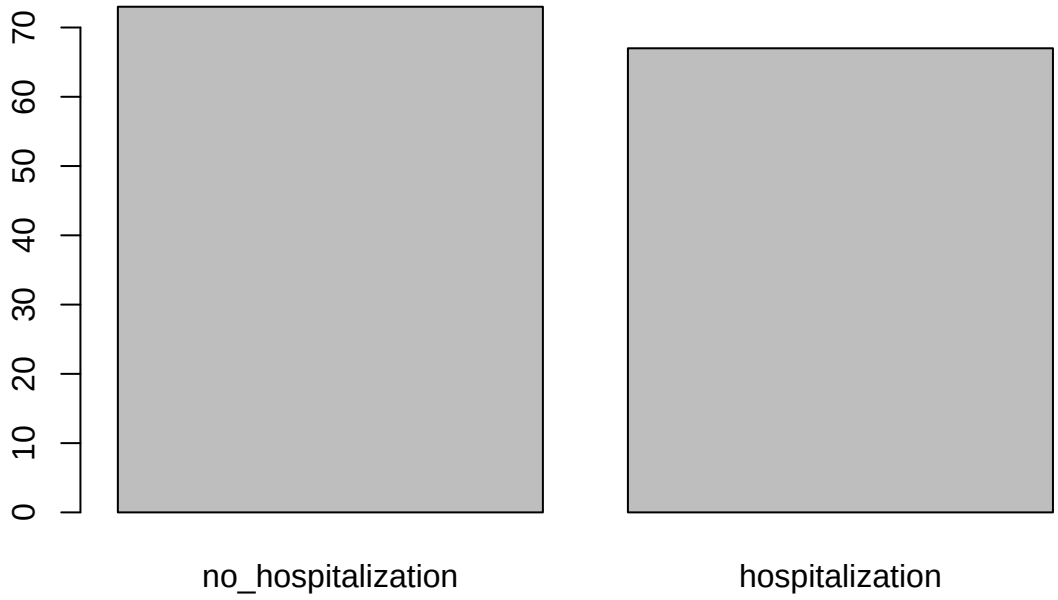
```r
# function cbinb() is used to combine two tables
freqtablecc<-cbind(freq.cc, relfreq.cc)
freqtablecc
```

```
##                    freq.cc relfreq.cc
## no_hospitalization      73     0.5214
## hospitalization         67     0.4786
```

```r
options(digits=4)
freqtablecc
```

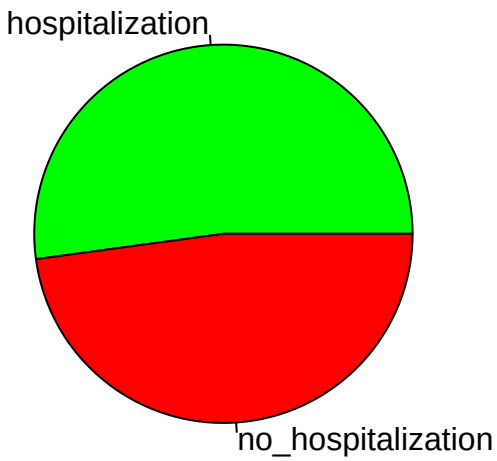
```
##                    freq.cc relfreq.cc
## no_hospitalization      73     0.5214
## hospitalization         67     0.4786
```

```r
# freq.cc relfreq.cc
# no_hospitalization      73     0.5214
# hospitalization         67     0.4786

barplot(freq.cc)
```




```r
barplot(relfreq.cc, xlab="status", ylab="relative frequency", names.arg=c("hospitalization", "no_hospita
```


```
pie(relfreq.cc, labels=c("hospitalization", "no_hospitalization"),col=c("green", "red"))
```




```
#stime
#symptom time is a continuous, numerical variable with units of days
#Summary statistics

# mean or average
mean(viral34_c$stime)
```

```
## [1] 7.356
```

```
# median
## 7.356
median(viral34_c$stime)
```

```
## [1] 6.962
```

```
## 6.962

# range
max(viral34_c$stime)-min(viral34_c$stime)
```

```
## [1] 17.6
```

```r
## 17.6
var(viral34_c$stime)  # variance
```

```
## [1] 16.42
```

```r
## 16.42
sd(viral34_c$stime)  # standard deviation
```

```
## [1] 4.052
```

```r
## 4.052
# coeficient of variation (in percentage)
100*sd(viral34_c$stime)/mean(viral34_c$stime)
```

```
## [1] 55.08
```

```r
## 55.08
# minimum, first , second and third quartiles, and maximum
quantile(viral34_c$stime)
```

```
##       0%      25%      50%      75%     100%
##  0.05476  4.69541  6.96235 10.05681 17.65914
```

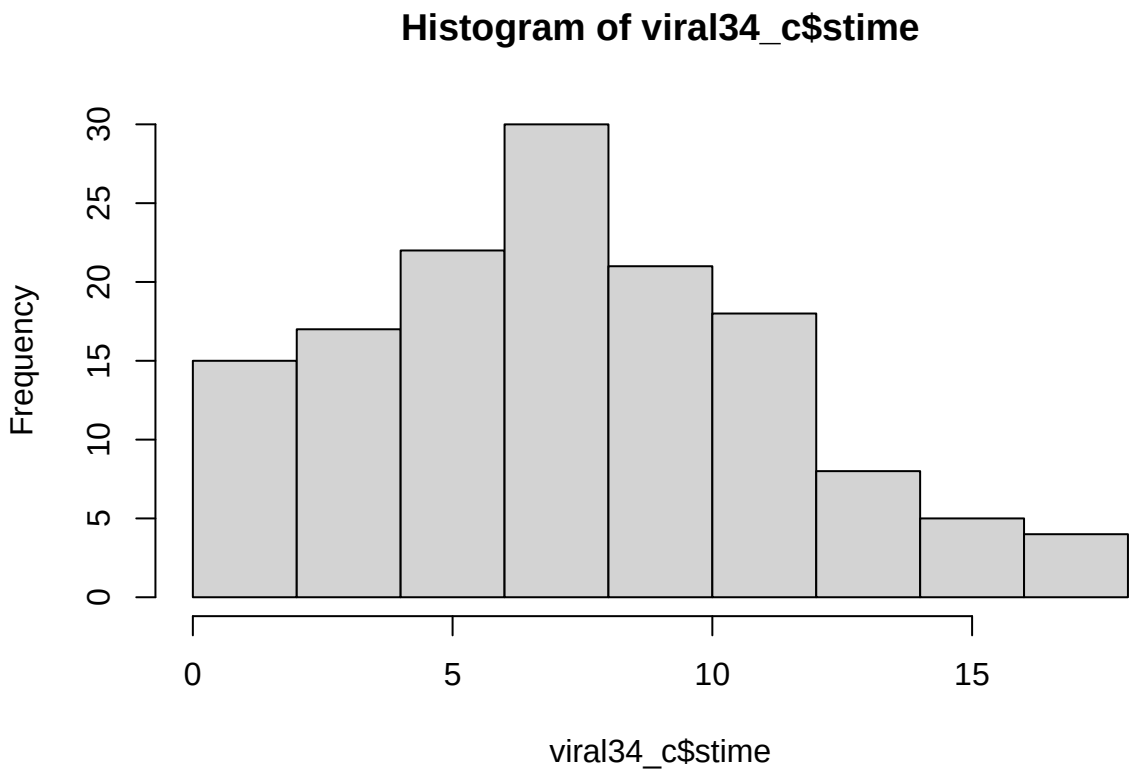
```r
# 0%      25%      50%      75%     100%
# 0.05476  4.69541  6.96235 10.05681 17.65914

# interquartile range
IQR(viral34_c$stime)
```

```
## [1] 5.361
```

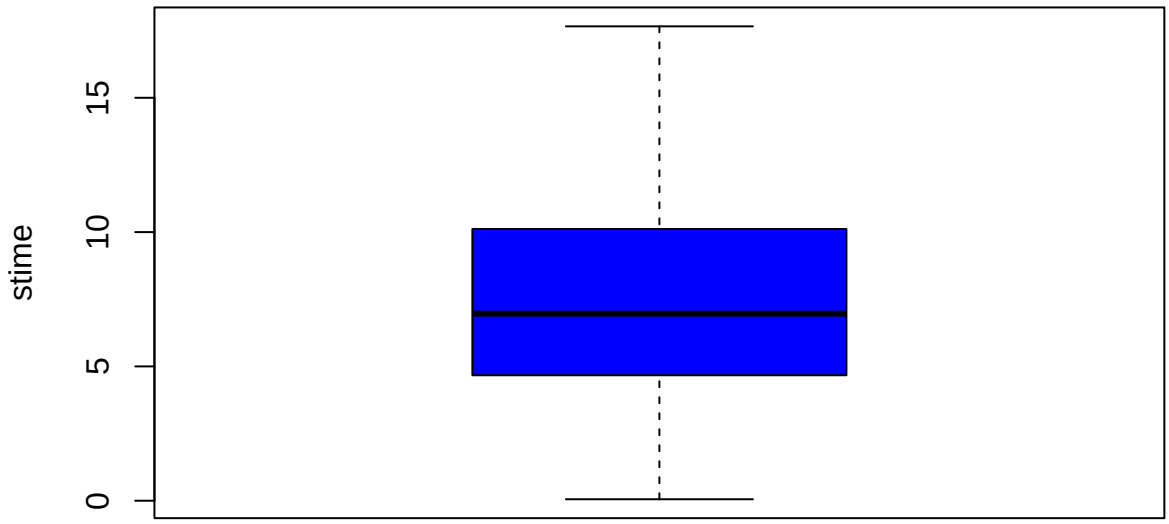
```r
## 5.361
# 35% and 63% quantiles
quantile(viral34_c$stime, c(0.35,0.63))
```

```
##   35%   63%
## 5.566 8.432
```

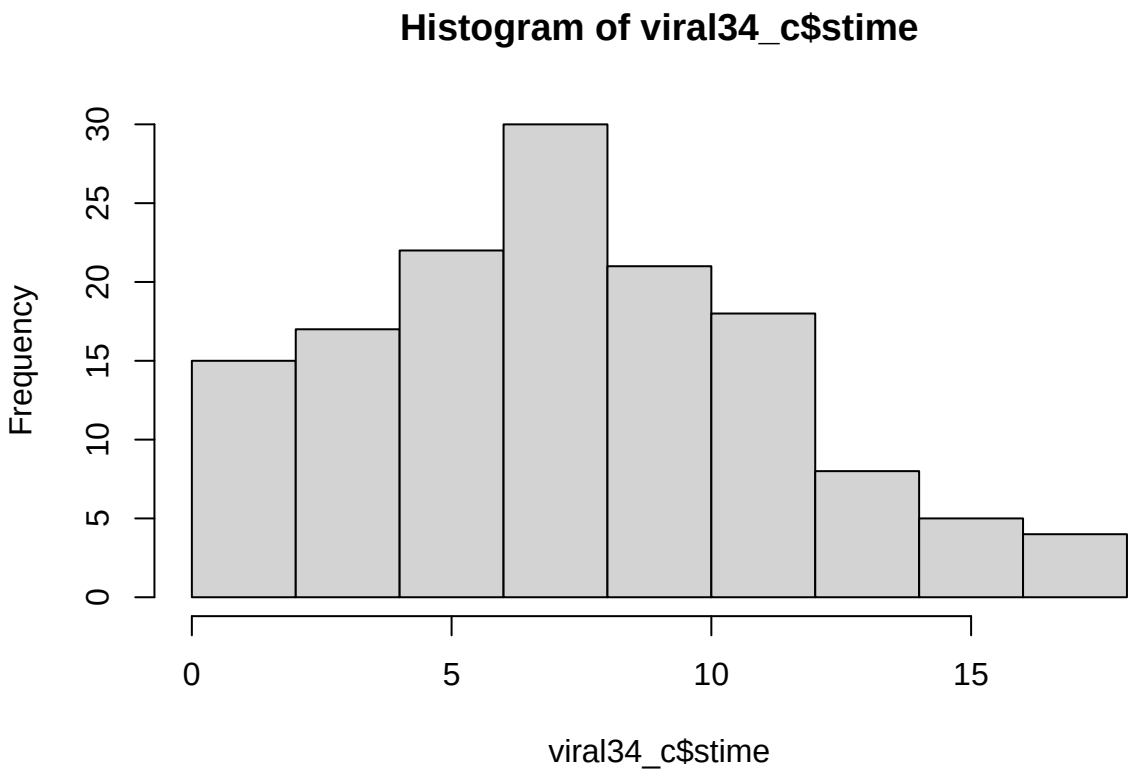
```r
# 35%   63%
#   5.566 8.432
#Histogram
hist(viral34_c$stime)
```



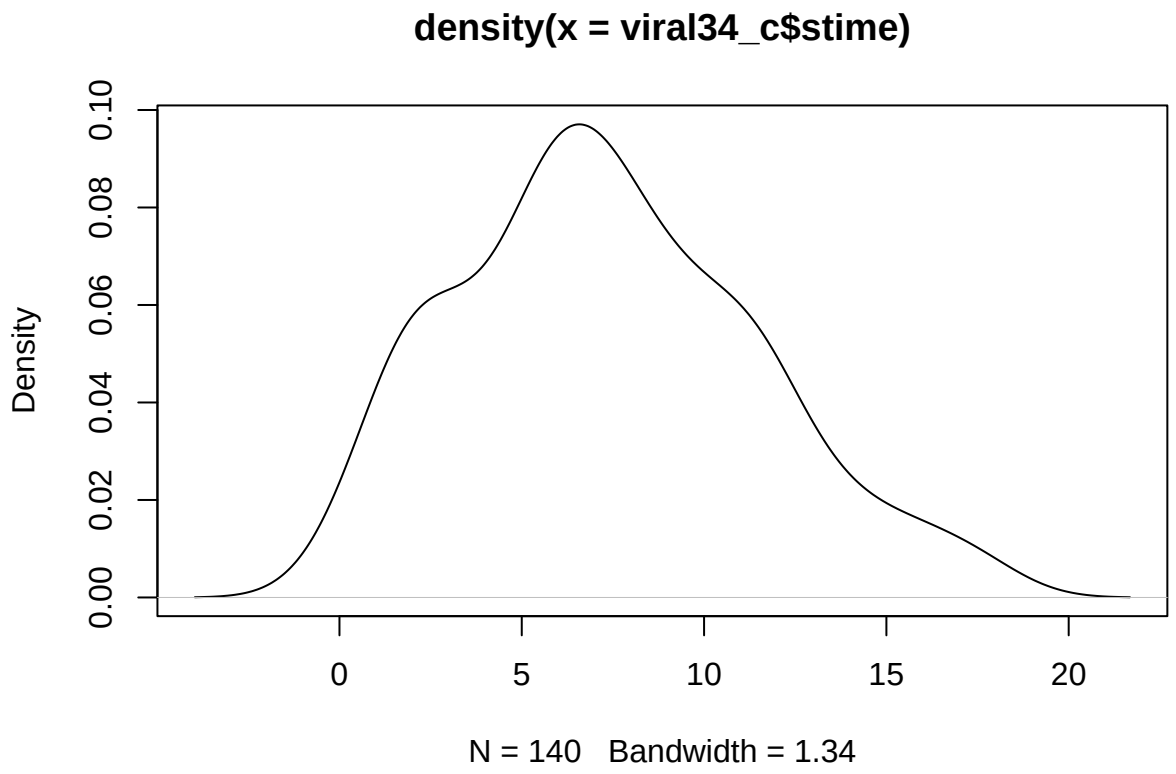

```
#Boxplot:
boxplot(viral34_c$stime, ylab="stime", col="blue")
```



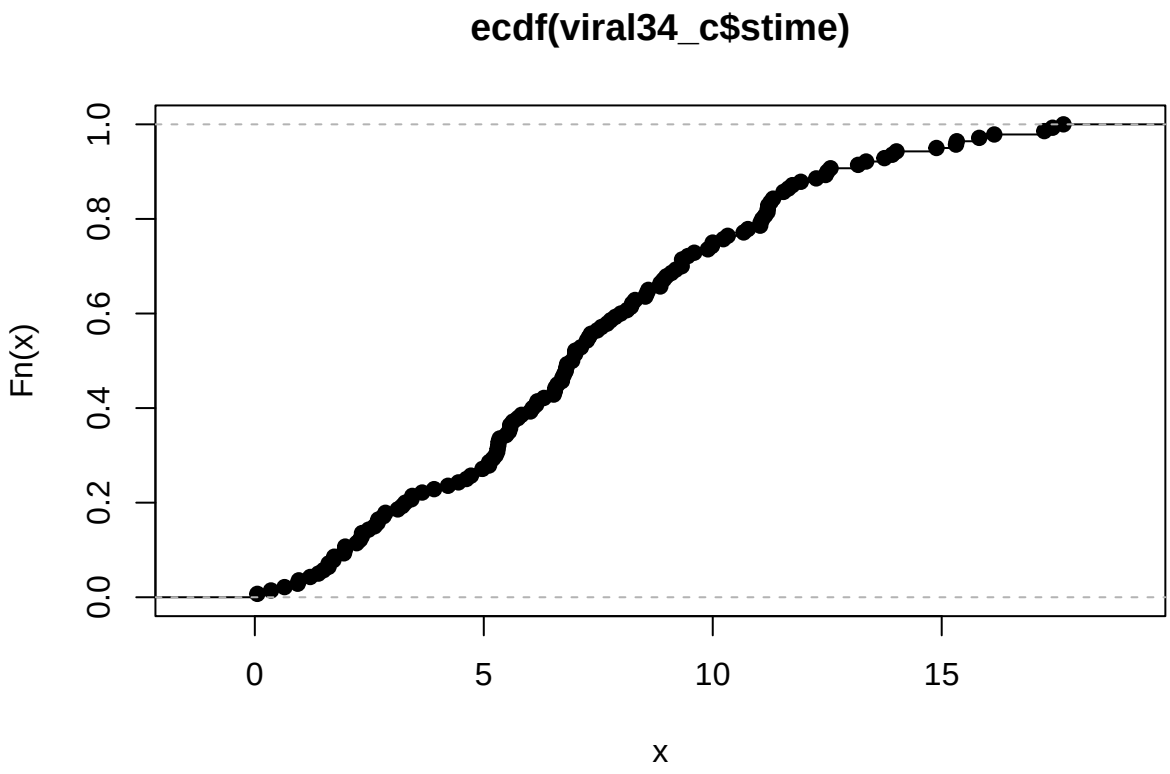

```
#Error in plot.new() : figure margins too large

#Density Function
hist(viral34_c$stime)
```




```
density<-density(viral34_c$stime)
plot(density)
```




```
#Error in plot.new() : figure margins too large

#Empirical cumulative distribution
f<-ecdf(viral34_c$stime)
plot(f)
```




```
#Error in plot.new() : figure margins too large

#Testing for outliers:

grubbs.test(viral34_c$stime)
```

```
##
##  Grubbs test for one outlier
##
## data:  viral34_c$stime
## G = 2.54, U = 0.95, p-value = 0.7
## alternative hypothesis: highest value 17.65913758 is an outlier
```

```
#Testing for normal distribution of age:
shapiro.test(viral34_c$stime)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  viral34_c$stime
## W = 0.98, p-value = 0.02
```

```
#sind
#sind is Categorical data that Had numerical value 1 or 0 and was earlier factored.Indicator of symptom
```

```r
# (1 = symptoms finished; 0 = symptoms remain)

# absolute frequencies
freq.cc<-table(viral34_c$sind)
freq.cc
```

```
##
##   symptoms_remain symptoms_finished
##                93                47
```

```r
# symptoms_remain symptoms_finished
# 93                47

# relative frequencies
relfreq.cc<-freq.cc/nrow(viral34_c)
relfreq.cc
```

```
##
##   symptoms_remain symptoms_finished
##            0.6643            0.3357
```

```r
# symptoms_remain symptoms_finished
# 0.6643            0.3357

# relative frequencies (ALTERNATIVE METHOD)
relfreq.cc<-prop.table(table(viral34_c$sind))
relfreq.cc
```

```
##
##   symptoms_remain symptoms_finished
##            0.6643            0.3357
```

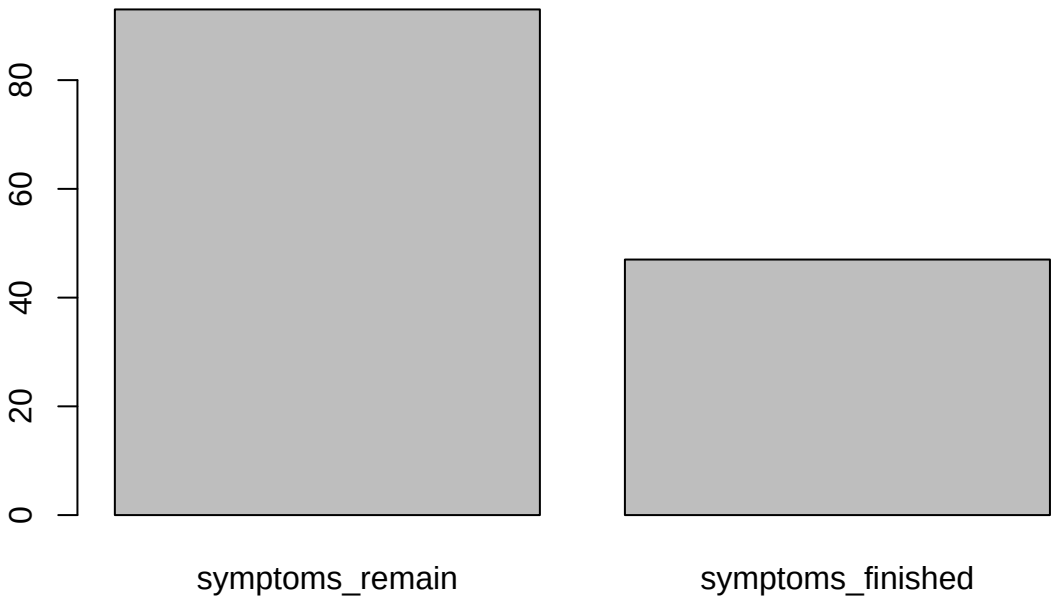
```r
# function cbinb() is used to combine two tables
freqtablecc<-cbind(freq.cc, relfreq.cc)
freqtablecc
```

```
##                   freq.cc relfreq.cc
## symptoms_remain        93     0.6643
## symptoms_finished      47     0.3357
```

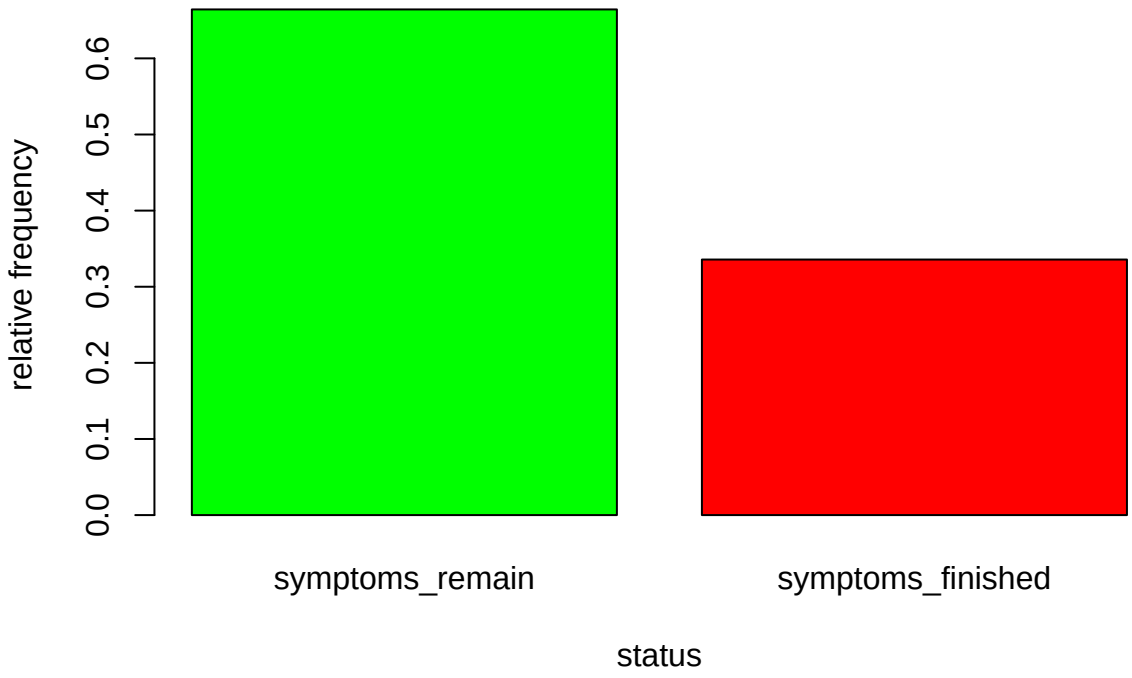
```r
options(digits=4)
freqtablecc
```

```
##                   freq.cc relfreq.cc
## symptoms_remain        93     0.6643
## symptoms_finished      47     0.3357
```

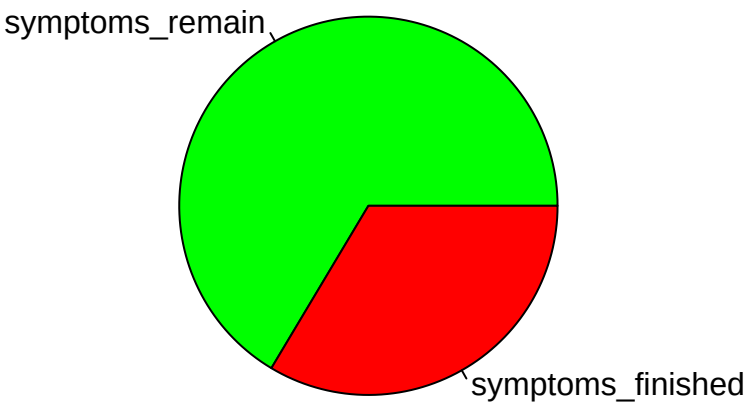
```r
# freq.cc relfreq.cc
# symptoms_remain        93     0.6643
# symptoms_finished      47     0.3357

barplot(freq.cc)
```




```
barplot(relfreq.cc, xlab="status", ylab="relative frequency", names.arg=c("symptoms_remain", "symptoms_
```




```
pie(relfreq.cc, labels=c("symptoms_remain", "symptoms_finished"),col=c("green", "red"))
```




```
#gender
#gender is Categorical data that Had numerical value 1 or 0 and was earlier factored as "female" (base
# absolute frequencies
freq.cc<-table(viral34_c$gender)
freq.cc
```

```
##
## female   male
##     62     78
```

```
# female   male
# 62     78

# relative frequencies
relfreq.cc<-freq.cc/nrow(viral34_c)
relfreq.cc
```

```
##
## female   male
## 0.4429 0.5571
```

```
# female   male
# 0.4429 0.5571
```

```r
# relative frequencies (ALTERNATIVE METHOD)
relfreq.cc<-prop.table(table(viral34_c$gender))
relfreq.cc
```

```
##
## female   male
## 0.4429 0.5571
```

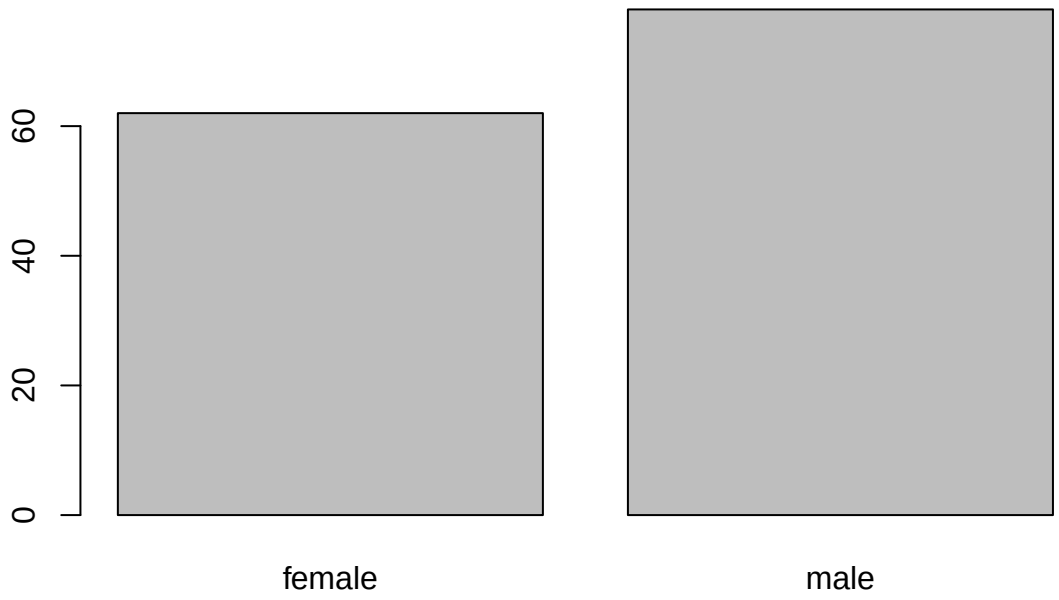
```r
# function cbinb() is used to combine two tables
freqtablecc<-cbind(freq.cc, relfreq.cc)
freqtablecc
```

```
##        freq.cc relfreq.cc
## female      62     0.4429
## male        78     0.5571
```

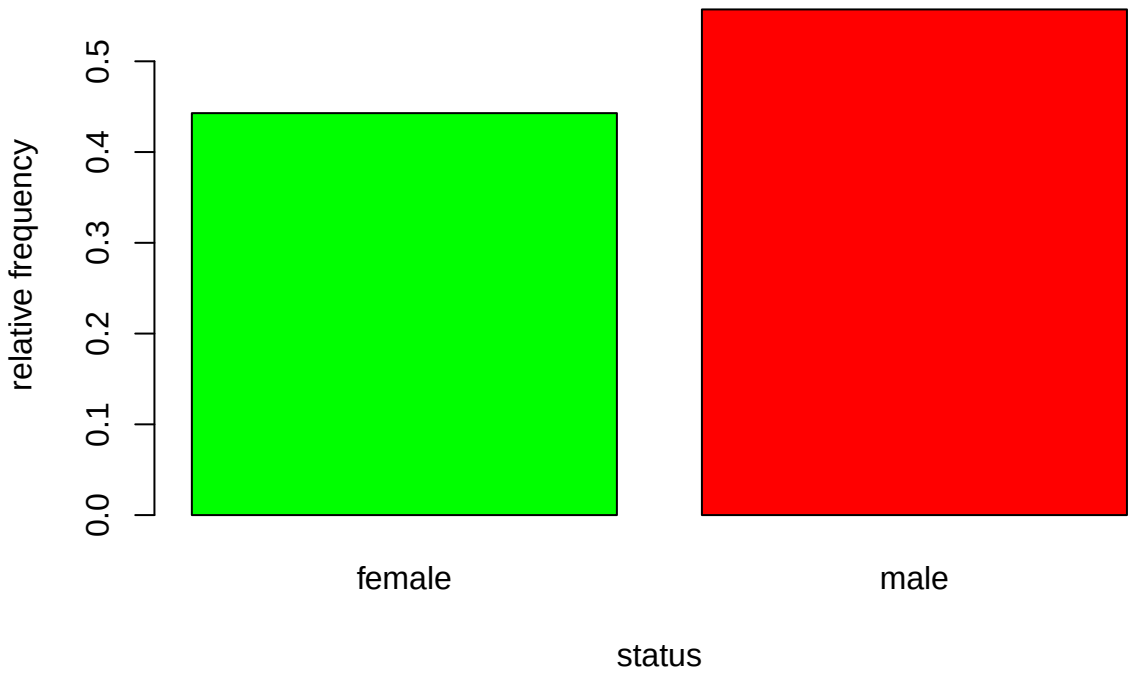
```r
options(digits=4)
freqtablecc
```

```
##        freq.cc relfreq.cc
## female      62     0.4429
## male        78     0.5571
```

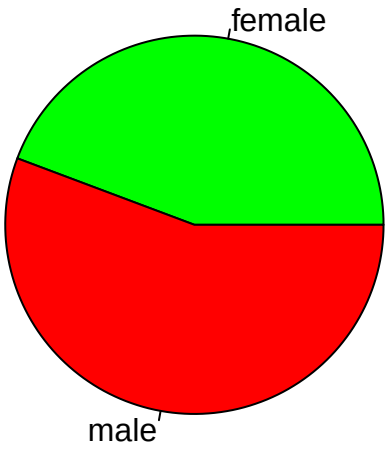
```r
# freq.cc relfreq.cc
# female      62     0.4429
# male        78     0.5571

barplot(freq.cc)
```



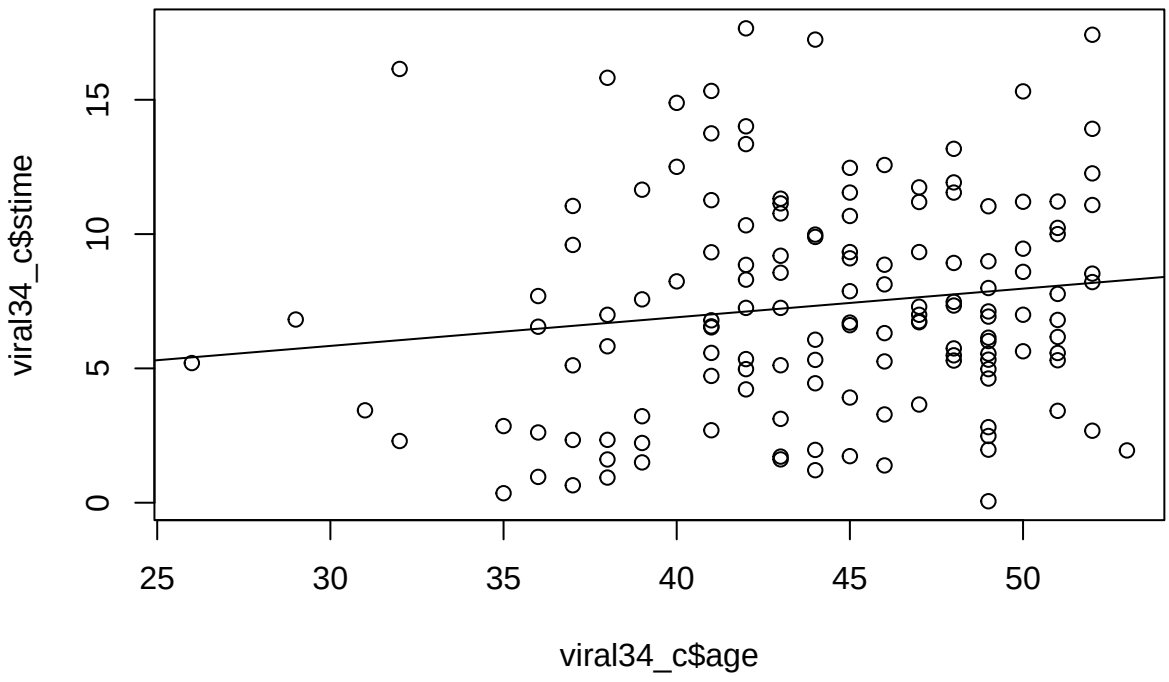

```
barplot(relfreq.cc, xlab="status", ylab="relative frequency", names.arg=c("female", "male"), col=c("gre
```




```
pie(relfreq.cc, labels=c("female", "male"),col=c("green", "red"))
```



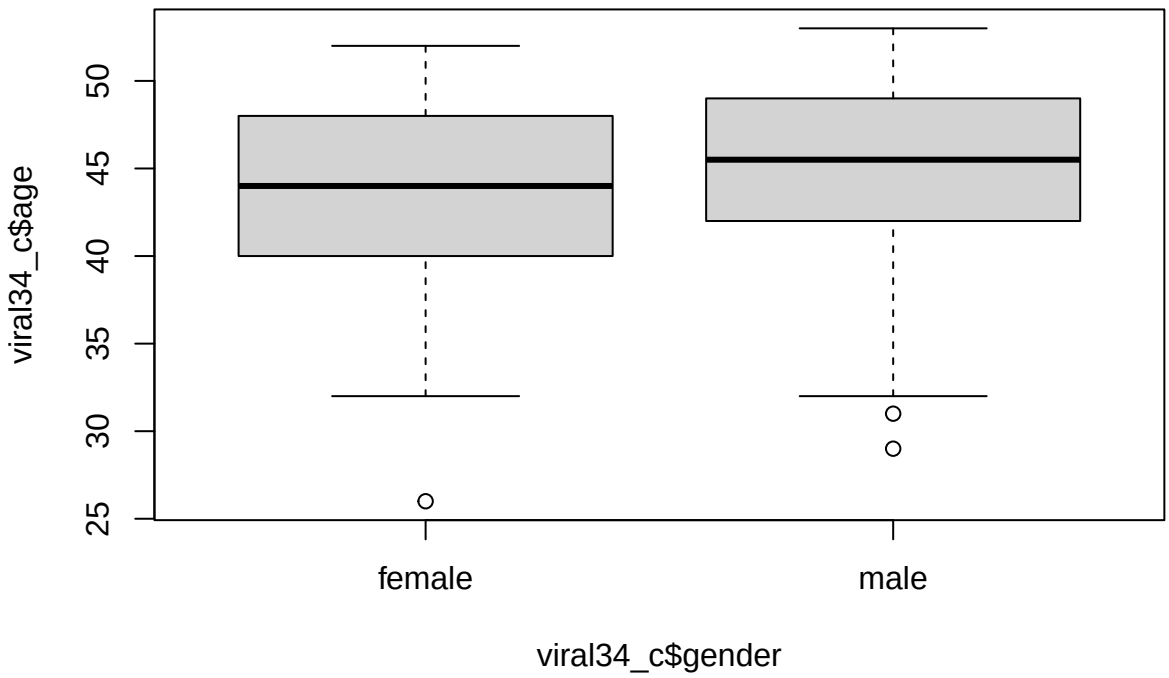

```
#BRIEF BIVARIATE ANALYSIS
#Continuous+Continuous
plot(viral34_c$age, viral34_c$stime)
abline(lm(viral34_c$stime ~ viral34_c$age))
```




```
#Continuous+Categorical
tapply(viral34_c$age, viral34_c$gender, mean)
```

```
## female   male
##  43.56  44.79
```

```
boxplot(viral34_c$age~viral34_c$gender)
```



QUESTION 2 Perform hierarchical clustering of (scaled) gene expression levels and explore possible relationships between genes. How many gene clusters are observed? Performing hierarchical clustering of all GENES (NEED TO TRANSPOSE MATRIX FOR GENE ROWS!) according to all genes using Euclidean distance and average linkage algorithm Note: After standardization of gene expression levels, Euclidean and Correlation Distance are the same. Euclidean and Manhattan distance measure absolute differences between vectors (gene expression levels), but Euclidean is less robust towards outliers

```
class(viral34_c)
```

```
## [1] "data.frame"
```

```
#[1] "data.frame"

str(viral34_c)
```

```
## 'data.frame':    140 obs. of  157 variables:
##  $ infection       : Factor w/ 2 levels "bacterial_infection",..: 1 2 1 2 2 2 1 1 2 1 ...
##  $ stime           : num  7.3 6.72 7 9.33 3.44 ...
##  $ sind            : Factor w/ 2 levels "symptoms_remain",..: 1 1 2 1 2 1 1 2 2 1 ...
##  $ gender          : Factor w/ 2 levels "female","male": 1 2 2 1 2 1 2 2 1 1 ...
##  $ hosp            : Factor w/ 2 levels "no_hospitalization",..: 2 1 2 2 1 1 1 1 2 1 ...
##  $ age             : int  47 47 38 45 31 41 48 47 38 44 ...
##  $ ancestry        : chr  "A" "A" "B" "A" ...
##  $ GSTM3           : num  0.1465 -0.0354 -0.2626 0.3379 0.0966 ...
##  $ RP5.860F19.3    : num  -0.098 -0.021 0.0108 0.3417 0.0782 ...
```

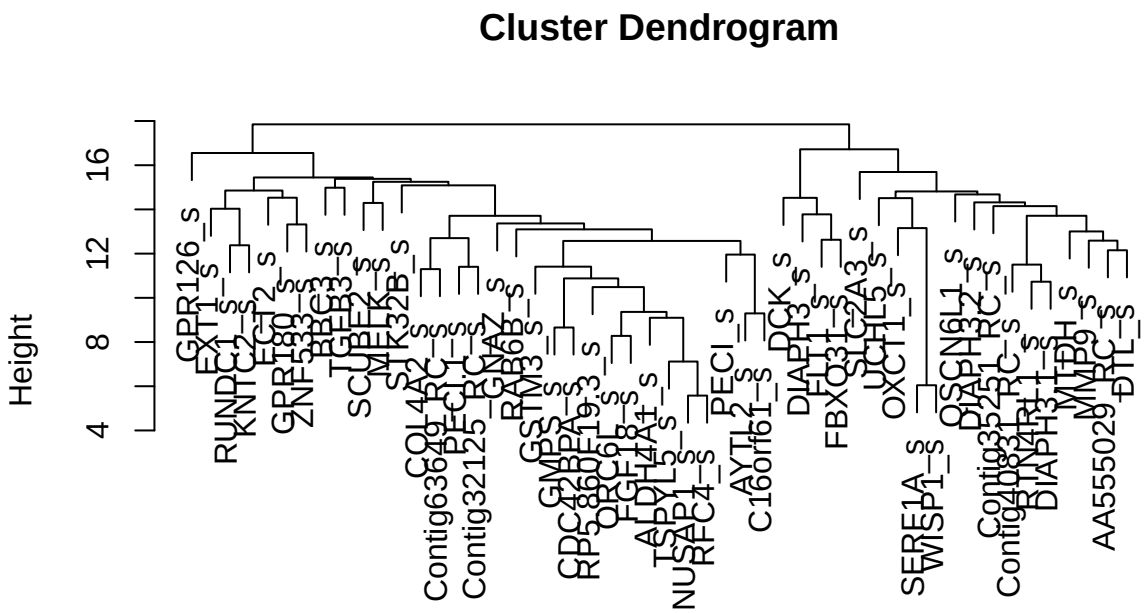
```
##  $ BBC3            : num  0.3082 -0.0964 0.0885 0.3277 -0.4061 ...
##  $ MMP9            : num  -0.2064 0.2415 -0.0258 -0.3414 -0.1786 ...
##  $ Contig35251_RC  : num  -0.519 -0.532 -0.34 -0.626 -0.643 ...
##  $ Contig40831_RC  : num  -0.11515 -0.00337 -0.08097 -0.12634 0.08476 ...
##  $ ALDH4A1         : num  0.1667 0.0985 0.0671 0.603 0.1597 ...
##  $ SERF1A          : num  -0.0838 0.1067 0.0627 -0.5238 -0.0179 ...
##  $ SCUBE2          : num  0.00606 0.09757 -0.12583 0.08587 -0.14412 ...
##  $ MTDH            : num  -0.1378 0.4942 0.0723 -0.577 -0.5194 ...
##  $ DCK             : num  -0.336 -0.58 0.139 -0.525 -0.197 ...
##  $ FLT1            : num  -0.0559 0.169 0.067 -0.0304 0.0635 ...
##  $ PECI.1          : num  -0.0878 -0.0968 -0.1134 0.1088 -0.1864 ...
##  $ QSCN6L1         : num  -0.1246 0.2656 0.0957 -0.0828 -0.082 ...
##  $ DIAPH3          : num  0.0808 0.1249 0.234 -0.1774 0.1007 ...
##  $ SLC2A3          : num  0.3686 0.4642 -0.0776 -0.2203 -0.0472 ...
##  $ GPR180          : num  -0.04521 -0.18754 -0.00541 0.11594 -0.11199 ...
##  $ RTN4RL1         : num  -0.1629 0.2001 0.1219 -0.6442 0.0282 ...
##  $ Contig32125_RC  : num  -0.0133 -0.0133 -0.0788 -0.0236 -0.1042 ...
##  $ STK32B          : num  0.0278 0.1297 -0.062 0.0129 -0.1214 ...
##  $ EXT1            : num  -0.1345 -0.1956 -0.1134 0.0219 -0.3168 ...
##  $ COL4A2          : num  -0.0299 -0.2267 -0.2083 0.1027 -0.2578 ...
##  $ PECI            : num  0.1735 0.212 0.0423 0.4796 0.1005 ...
##  $ GNAZ            : num  0.0705 -0.0317 0.063 0.3349 -0.1467 ...
##  $ AYTL2           : num  0.2393 0.0157 -0.1276 0.5336 0.1171 ...
##  $ Contig63649_RC  : num  0.02962 0.00419 0.0504 0.29442 0.02671 ...
##  $ RAB6B           : num  0.4614 0.0186 -0.1425 0.3254 -0.0873 ...
##  $ AA555029_RC     : num  -0.0481 0.1593 0.1142 -0.3106 -0.2201 ...
##  $ GPR126          : num  -0.1002 0.2812 0.0571 0.1912 -0.0826 ...
##  $ ECT2            : num  0.0354 -0.0377 -0.1813 0.0334 0.3287 ...
##  $ NUSAP1          : num  0.1098 0.0323 -0.0482 0.6553 0.0795 ...
##  $ GMPS            : num  0.2181 0.1857 0.0404 0.2371 0.1784 ...
##  $ UCHL5           : num  -0.0381 -0.2708 -0.0432 -0.1923 -0.1409 ...
##  $ ORC6L           : num  0.173 0.102 -0.15 0.193 0.126 ...
##  $ TSPYL5          : num  0.1559 -0.0588 0.0909 0.541 0.0456 ...
##  $ MELK            : num  -0.3458 -0.0108 -0.1366 -0.3397 -0.2484 ...
##  $ RUNDC1          : num  0.55836 -0.35885 0.12992 -0.07181 -0.00765 ...
##  $ DIAPH3.1        : num  -0.4446 -0.2426 -0.0564 -0.4456 -0.0968 ...
##  $ C16orf61        : num  0.0591 -0.0502 -0.2737 0.1355 0.2048 ...
##  $ TGFB3           : num  0.0818 0.1187 0.1132 -0.0205 0.1275 ...
##  $ FGF18           : num  -0.0482 0.2738 0.0347 0.1039 0.1898 ...
##  $ CDC42BPA        : num  0.192 0.1254 -0.2281 0.0328 -0.0739 ...
##  $ DTL             : num  -1.012 -0.146 0.448 -1.077 -0.843 ...
##  $ WISP1           : num  -0.00498 0.21792 0.07126 -0.44042 0.11942 ...
##  $ DIAPH3.2        : num  -0.29778 0.02057 -0.14414 0.05123 0.00824 ...
##  $ OXCT1           : num  0.0314 0.1633 0.0569 -0.2054 -0.139 ...
##  $ ZNF533          : num  0.8648 0.0158 -0.1476 -0.2065 0.2885 ...
##  $ RFC4            : num  0.0619 0.0169 -0.0543 0.3945 0.0624 ...
##  $ KNTC2           : num  0.5975 -0.3272 0.0965 0.046 -0.0926 ...
##  $ FBXO31          : num  -0.0414 -0.1352 0.0352 0.0607 0.264 ...
##  $ GSTM3_s         : num  0.714 -0.206 -1.355 1.682 0.462 ...
##  $ RP5.860F19.3_s  : num  -0.5744 -0.1851 -0.0241 1.649 0.3165 ...
##  $ BBC3_s          : num  1.2662 0.0497 0.6058 1.3249 -0.8814 ...
##  $ MMP9_s          : num  -0.8121 1.3352 0.0534 -1.4594 -0.6791 ...
##  $ Contig35251_RC_s: num  -0.616 -0.646 -0.204 -0.862 -0.902 ...
##  $ Contig40831_RC_s: num  -0.6711 -0.0495 -0.4811 -0.7333 0.4405 ...
```

```
##  $ ALDH4A1_s       : num  0.873 0.566 0.426 2.831 0.841 ...
##  $ SERF1A_s        : num  -0.4491 0.6659 0.4079 -3.0244 -0.0633 ...
##  $ SCUBE2_s        : num  0.194 0.781 -0.651 0.706 -0.768 ...
##  $ MTDH_s          : num  -0.182 2.074 0.568 -1.75 -1.545 ...
##  $ DCK_s           : num  -0.0511 -0.9337 1.6611 -0.7352 0.4472 ...
##  $ FLT1_s          : num  -0.351 1.075 0.428 -0.189 0.406 ...
##  $ PECI.1_s        : num  -0.333 -0.388 -0.49 0.874 -0.938 ...
##  $ QSCN6L1_s       : num  -1.1 1.834 0.557 -0.786 -0.779 ...
##  $ DIAPH3_s        : num  0.539 0.798 1.44 -0.979 0.656 ...
##  $ SLC2A3_s        : num  2.279 2.889 -0.568 -1.478 -0.374 ...
##  $ GPR180_s        : num  -0.2433 -1.3427 0.0641 1.0015 -0.7591 ...
##  $ RTN4RL1_s       : num  -0.488 0.97 0.656 -2.421 0.279 ...
##  $ Contig32125_RC_s: num  -0.0143 -0.0142 -0.4362 -0.0806 -0.5998 ...
##  $ STK32B_s        : num  0.426 1.055 -0.128 0.334 -0.495 ...
##  $ EXT1_s          : num  -0.481 -0.838 -0.358 0.431 -1.545 ...
##  $ COL4A2_s        : num  0.146 -0.821 -0.731 0.798 -0.974 ...
##  $ PECI_s          : num  0.986 1.166 0.372 2.418 0.645 ...
##  $ GNAZ_s          : num  0.4 -0.276 0.35 2.149 -1.037 ...
##  $ AYTL2_s         : num  1.573 0.243 -0.609 3.323 0.846 ...
##  $ Contig63649_RC_s: num  0.2817 0.0979 0.4319 2.1955 0.2607 ...
##  $ RAB6B_s         : num  2.35 0.176 -0.615 1.683 -0.344 ...
##  $ AA555029_RC_s   : num  -0.145 0.963 0.722 -1.548 -1.064 ...
##  $ GPR126_s        : num  -0.504 1.787 0.441 1.247 -0.398 ...
##  $ ECT2_s          : num  0.3579 0.0514 -0.55 0.3492 1.5864 ...
##  $ NUSAP1_s        : num  0.481 0.15 -0.193 2.808 0.352 ...
##  $ GMPS_s          : num  1.08 0.954 0.391 1.154 0.926 ...
##  $ UCHL5_s         : num  -0.0852 -1.5087 -0.1165 -1.0286 -0.7139 ...
##  $ ORC6L_s         : num  0.839 0.574 -0.368 0.914 0.663 ...
##  $ TSPYL5_s        : num  0.791 -0.113 0.518 2.414 0.327 ...
##  $ MELK_s          : num  -1.279 0.166 -0.377 -1.252 -0.859 ...
##  $ RUNDC1_s        : num  2.174 -0.828 0.772 0.112 0.322 ...
##  $ DIAPH3.1_s      : num  -1.45568 -0.70313 -0.00952 -1.45953 -0.16002 ...
##  $ C16orf61_s      : num  0.6119 0.0462 -1.1104 1.007 1.3658 ...
##  $ TGFB3_s         : num  0.588 0.846 0.807 -0.127 0.907 ...
##  $ FGF18_s         : num  -0.128 1.519 0.296 0.65 1.089 ...
##  $ CDC42BPA_s      : num  1.26 0.875 -1.163 0.342 -0.274 ...
##   [list output truncated]
```

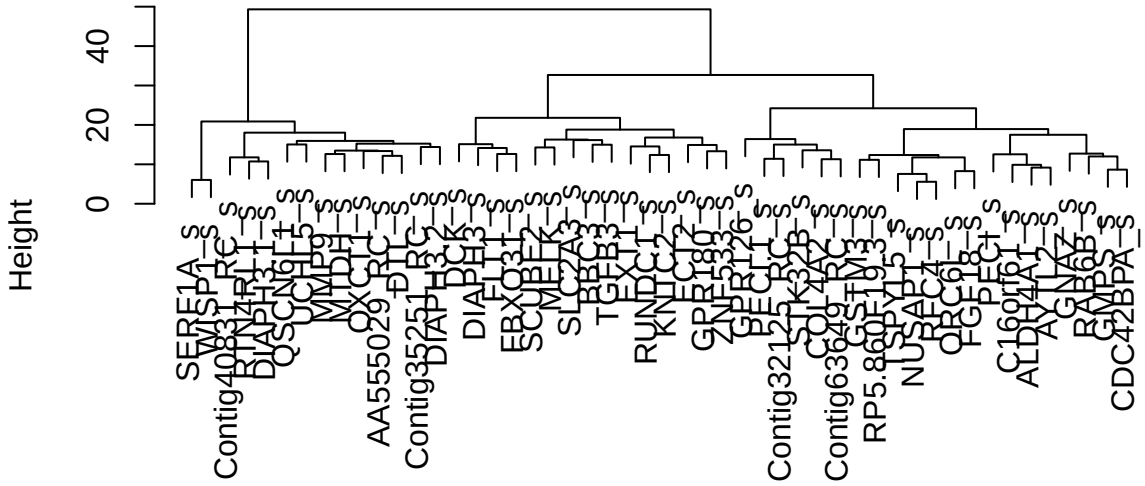
```
dim(viral34_c)
```

```
## [1] 140 157
```

```
hc_genes1<-hclust(dist(t(viral34_c[58:107]),method="euclidean"),method="average")
plot(hc_genes1)
```

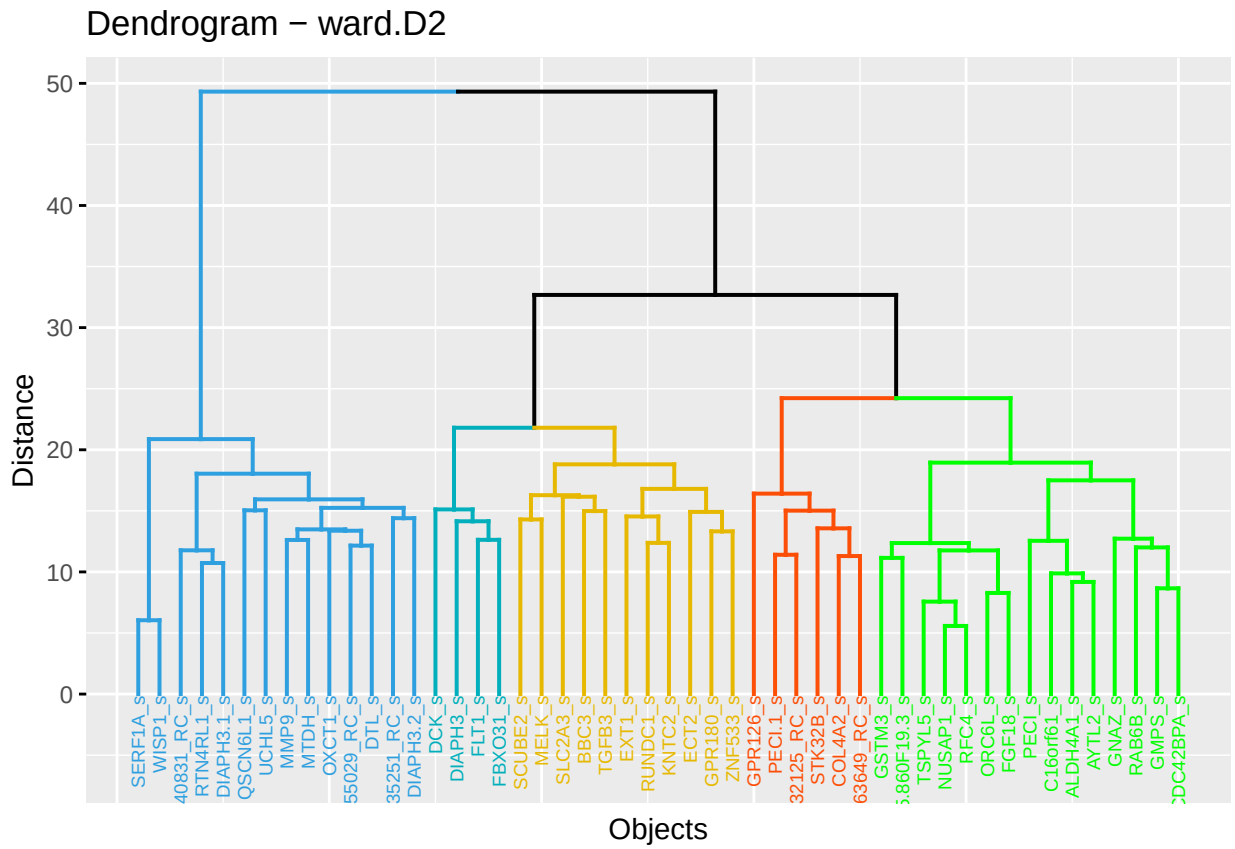



```
#Performing hierarchical clustering of all genes according to all genes using Euclidean distance and (s
hc_genes2<-hclust(dist(t(viral34_c[58:107]),method="euclidean"),method="ward.D2")
plot(hc_genes2)
```



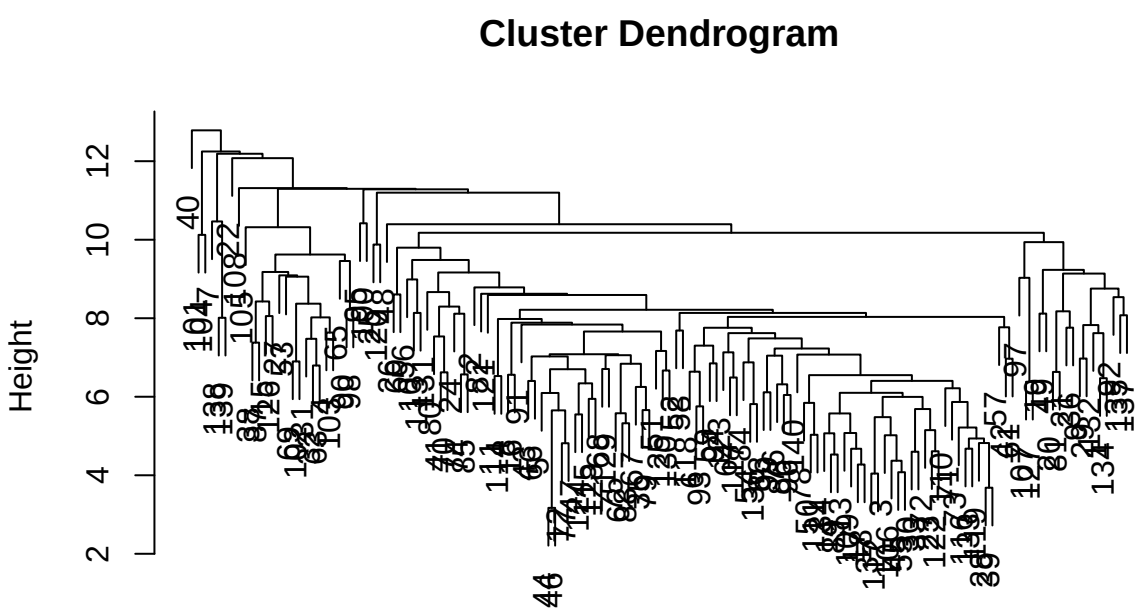

```
# Cut in five groups
# label size
fviz_dend(hc_genes2, k = 5,
          cex = 0.5,
          k_colors = c("#2E9FDF", "#00AFBB", "#E7B800", "#FC4E07", "green"),
          color_labels_by_k = TRUE,  # color labels by groups
          ggtheme = theme_gray(),
          main = "Dendrogram - ward.D2",
          xlab = "Objects", ylab = "Distance", sub = ""
          #Change theme
)
```

```
## Warning: The ‘<scale>‘ argument of ‘guides()‘ cannot be ‘FALSE‘. Use "none" instead as
## of ggplot2 3.3.4.
## i The deprecated feature was likely used in the factoextra package.
##   Please report the issue at <https://github.com/kassambara/factoextra/issues>.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```



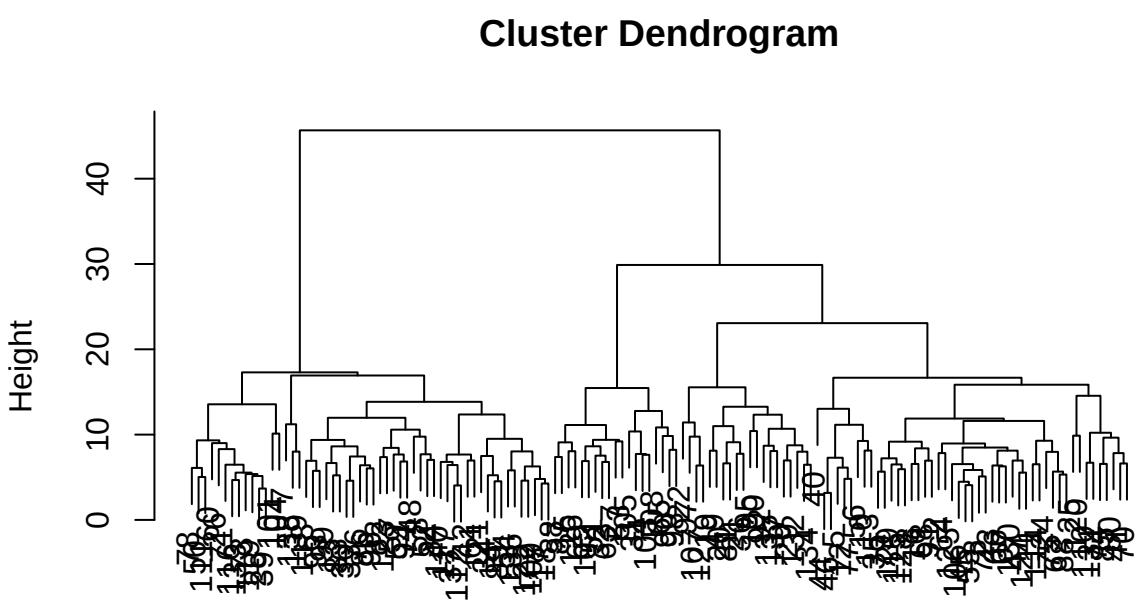
QUESTION 3 Perform hierarchical clustering of INDIVIDUALS according to their (scaled) gene expression levels and explore possible relationships between them. How many clusters of individuals are observed? Check visually whether the clustering is related to infection, gender, hospitalization or ancestry.

```
#Performing hierarchical clustering of all individuals according to all scaled genes using Euclidean di
hc_individuals1<-hclust(dist((viral34_c[1:140,58:107]),method="euclidean"),method="average")
plot(hc_individuals1)
```



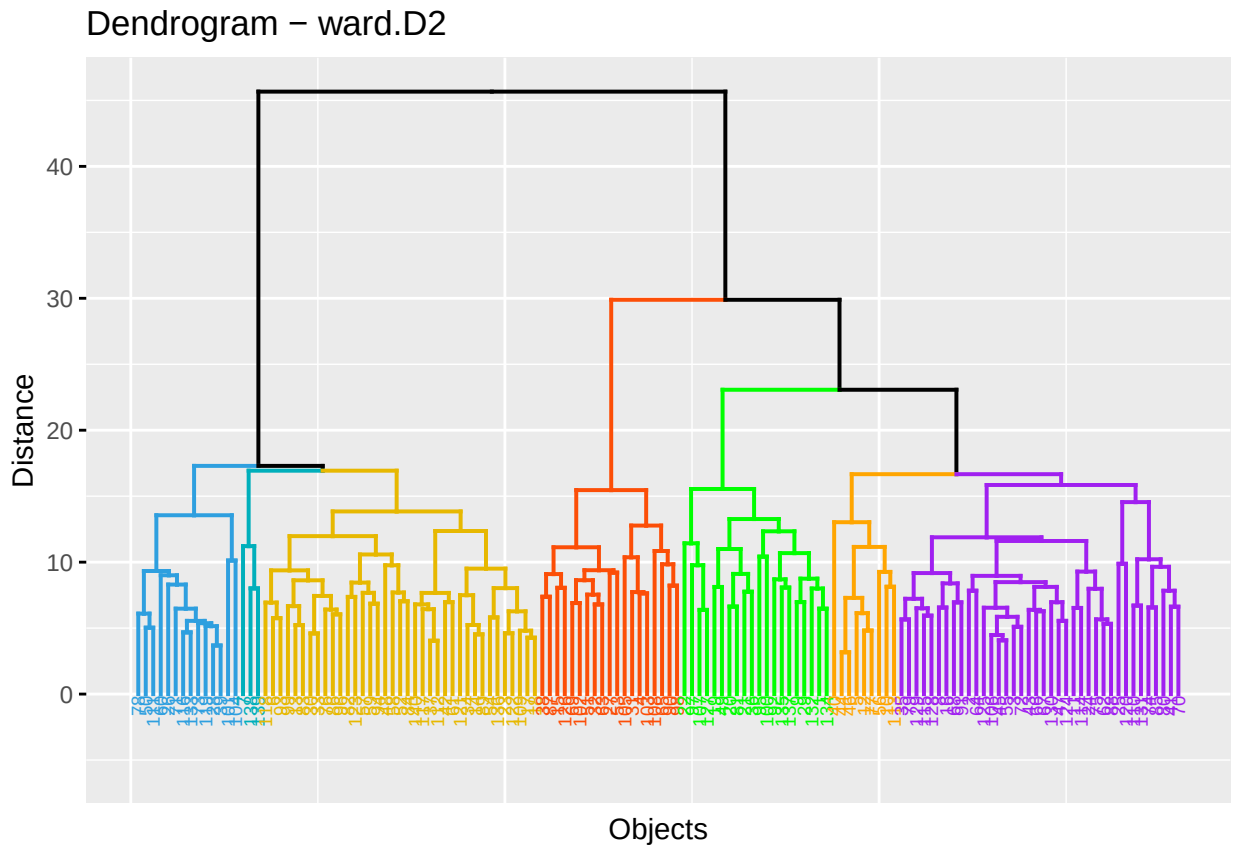
dist((viral34_c[1:140, 58:107]), method = "euclidean")
hclust (*, "average")

```
#Performing hierarchical clustering of all individuals according to all scaled genes using Euclidean di
hc_individuals2<-hclust(dist((viral34_c[1:140,58:107]),method="euclidean"),method="ward.D2")
plot(hc_individuals2)
```

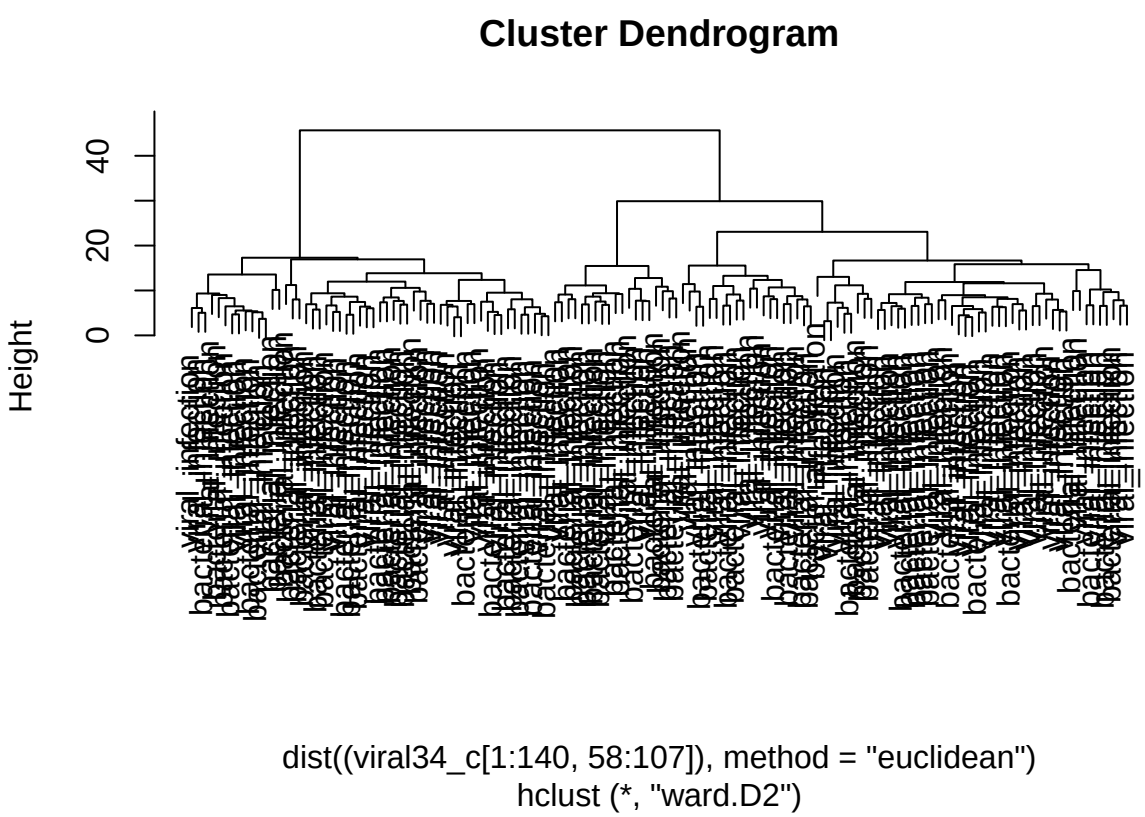


dist((viral34_c[1:140, 58:107]), method = "euclidean")
hclust (*, "ward.D2")

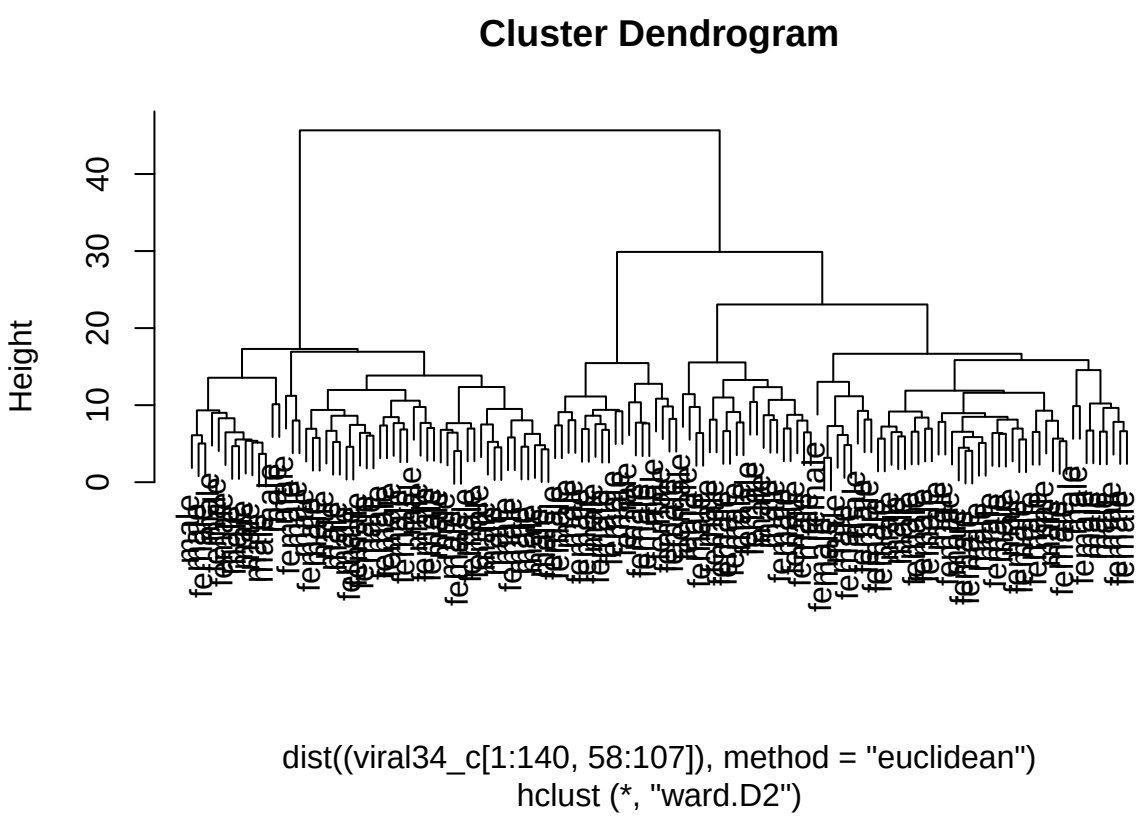
```
# Cut in seven groups
# label size
fviz_dend(hc_individuals2, k = 7,
          cex = 0.5,
          k_colors = c("#2E9FDF", "#00AFBB", "#E7B800", "#FC4E07", "green", "orange", "purple"),
          color_labels_by_k = TRUE,  # color labels by groups
          ggtheme = theme_gray(),
          main = "Dendrogram - ward.D2",
          xlab = "Objects", ylab = "Distance", sub = ""
          # Change theme
)
```



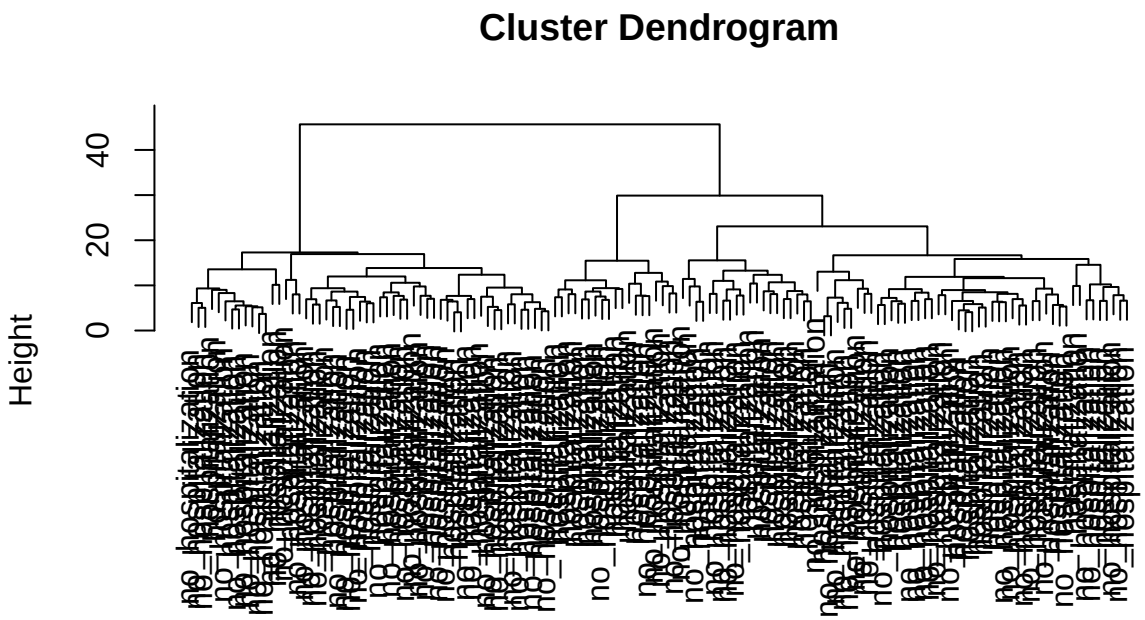

```
#Creating labels to visually check whether the clustering is related to infection, gender, hospitalizat
#x<-viral34_c$infection
#x
# x<-t(viral34_c$infection)
# x
# #SIMILAR TO LECTURE NOTE: x<-factor(golub.cl, labels=c("*","ALL"))
#plot(hc_individuals2, labels=x)



plot(hc_individuals2, labels=viral34_c$infection)#NOT RELATED
```



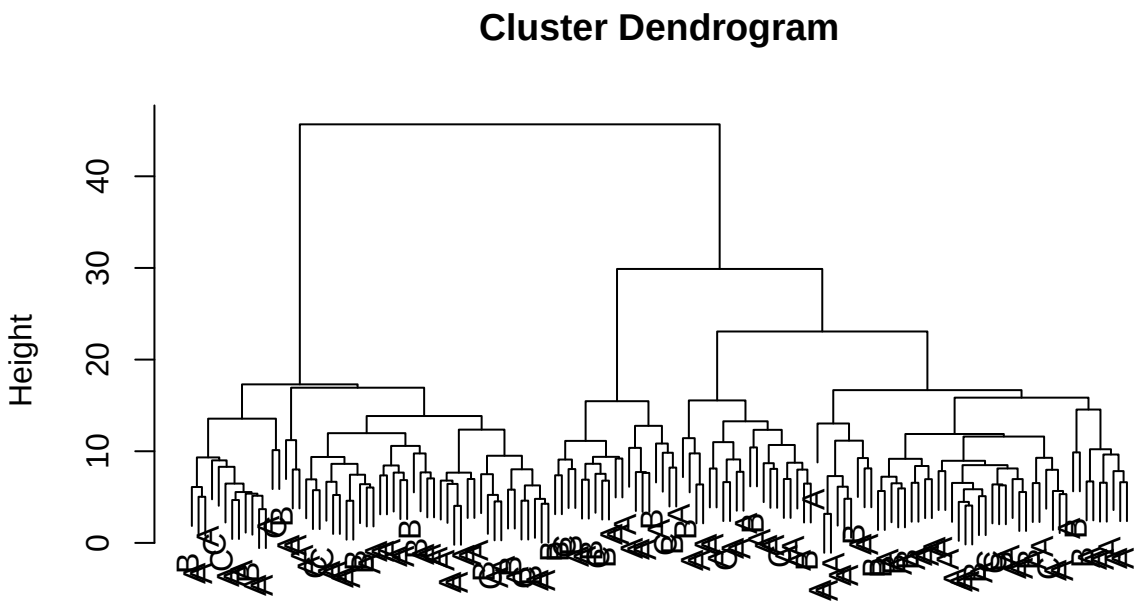

```
plot(hc_individuals2, labels=viral34_c$gender)#NOT RELATED
```




```
plot(hc_individuals2, labels=viral34_c$hosp)#NOT RELATED
```




```
plot(hc_individuals2, labels=viral34_c$ancestry)#NOT RELATED
```




```
#Counted 6 Individual Gene clusters infection, gender, hospitalization or ancestry.
#I observed no relationships between clusters and
```

QUESTION 4 Perform K-means clustering with k=2 and test whether the clustering is associated to (a) the kind of infection and (b) the risk of hospitalization. Interpret the results.

```
#Applying kmeans to continuous variables (same dataframe as before: X = [all rows, columns correspondin
#Since problem statement doesn't specify, I still used here the transposed continuous scaled gene expre

#Clustering individuals according to gene expression
kdata<-(viral34_c[, 58:107])
kcluster<-kmeans(kdata, 2, nstart=10)
kcluster
```

```
## K-means clustering with 2 clusters of sizes 76, 64
##
## Cluster means:
##   GSTM3_s RP5.860F19.3_s  BBC3_s  MMP9_s Contig35251_RC_s Contig40831_RC_s
## 1 -0.4969        -0.5156 -0.2802  0.3547           0.3021           0.2450
## 2  0.5900         0.6123  0.3327 -0.4213          -0.3588          -0.2909
##   ALDH4A1_s SERF1A_s SCUBE2_s  MTDH_s   DCK_s   FLT1_s PECI.1_s QSCN6L1_s
## 1   -0.6094   0.1704  -0.2399  0.3346 -0.1333  0.08164  -0.3778    0.3524
## 2    0.7237  -0.2024   0.2849 -0.3974  0.1583 -0.09694   0.4486   -0.4184
##   DIAPH3_s SLC2A3_s GPR180_s RTN4RL1_s Contig32125_RC_s STK32B_s  EXT1_s
## 1   0.2072  0.02026 -0.08575    0.5555          -0.3284  -0.1506 -0.1916
```

```
## 2  -0.2461 -0.02405  0.10183   -0.6597           0.3900   0.1789  0.2275
##   COL4A2_s  PECI_s  GNAZ_s AYTL2_s Contig63649_RC_s RAB6B_s AA555029_RC_s
## 1  -0.4103 -0.3649 -0.4150 -0.5491          -0.4382 -0.4366        0.4566
## 2   0.4872  0.4333  0.4928  0.6521           0.5203  0.5185       -0.5422
##   GPR126_s  ECT2_s NUSAP1_s  GMPS_s UCHL5_s ORC6L_s TSPYL5_s   MELK_s RUNDC1_s
## 1  0.03420 -0.1940  -0.5214 -0.6688  0.3794 -0.5673  -0.5147 -0.06751  -0.2653
## 2 -0.04061  0.2304   0.6191  0.7942 -0.4506  0.6737   0.6112  0.08017   0.3151
##   DIAPH3.1_s C16orf61_s TGFB3_s FGF18_s CDC42BPA_s   DTL_s WISP1_s DIAPH3.2_s
## 1     0.5321    -0.5266 -0.1713 -0.4996    -0.5677  0.4757  0.1151     0.3390
## 2    -0.6318     0.6253  0.2035  0.5933     0.6742 -0.5649 -0.1367    -0.4026
##   OXCT1_s ZNF533_s  RFC4_s KNTC2_s  FBXO31_s
## 1  0.3097  -0.2738 -0.5429 -0.3487 -0.004618
## 2 -0.3678   0.3251  0.6447  0.4141  0.005484
##
## Clustering vector:
##   [1] 2 1 1 2 2 1 1 1 2 2 2 2 1 1 2 1 1 1 2 2 2 2 1 2 2 1 2 1 2 1 2 2 1 1 1 2 1
##  [38] 2 1 1 1 1 1 2 1 2 1 1 2 1 2 1 2 1 1 1 1 1 1 1 2 1 2 2 2 1 1 1 2 1 1 1 1 2
##  [75] 2 1 2 1 1 2 2 1 1 1 2 1 2 2 1 2 2 2 1 1 2 1 2 2 1 2 1 2 2 1 2 1 2 2 1 1 1
## [112] 1 2 2 2 1 1 1 1 2 1 1 1 2 1 2 2 2 1 2 2 2 1 2 2 1 2 1 1 1
##
## Within cluster sum of squares by cluster:
## [1] 2626 3112
##  (between_SS / total_SS =  17.4 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

```
# Results:
# Within cluster sum of squares by cluster:
#   [1] 117.4392 156.1455
# (between_SS / total_SS =  20.3 %)

#Based on results, K-means clustering with 2 clusters of sizes 76, 64


kmeans(kdata,2)$cluster
```

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

```
#Running summary of dataframe provides proportions of categorical variables that may coincide with thes
summary(viral34_c)
```

```
##                infection      stime                       sind       gender
##  bacterial_infection:69   Min.   : 0.055   symptoms_remain  :93   female:62
##  viral_infection    :71   1st Qu.: 4.695   symptoms_finished:47   male  :78
##                           Median : 6.962
##                           Mean   : 7.356
##                           3rd Qu.:10.057
```

```
##                          Max.   :17.659
##                  hosp          age          ancestry              GSTM3
##  no_hospitalization:73   Min.   :26.0   Length:140         Min.   :-0.3594
##  hospitalization   :67   1st Qu.:41.0   Class :character   1st Qu.:-0.1455
##                          Median :45.0   Mode  :character   Median :-0.0203
##                          Mean   :44.2                      Mean   : 0.0053
##                          3rd Qu.:49.0                      3rd Qu.: 0.1233
##                          Max.   :53.0                      Max.   : 0.5561
##   RP5.860F19.3          BBC3              MMP9          Contig35251_RC
##  Min.   :-0.4242   Min.   :-1.0828   Min.   :-0.4943   Min.   :-0.9177
##  1st Qu.:-0.1072   1st Qu.:-0.3333   1st Qu.:-0.1605   1st Qu.:-0.5925
##  Median : 0.0087   Median :-0.0953   Median :-0.0476   Median :-0.4027
##  Mean   : 0.0156   Mean   :-0.1130   Mean   :-0.0370   Mean   :-0.2517
##  3rd Qu.: 0.1031   3rd Qu.: 0.1110   3rd Qu.: 0.0880   3rd Qu.: 0.0437
##  Max.   : 0.5938   Max.   : 0.6018   Max.   : 0.5168   Max.   : 0.9944
##  Contig40831_RC       ALDH4A1            SERF1A            SCUBE2
##  Min.   :-0.4715   Min.   :-0.7679   Min.   :-0.5563   Min.   :-0.5152
##  1st Qu.:-0.1256   1st Qu.:-0.1749   1st Qu.:-0.0984   1st Qu.:-0.1291
##  Median : 0.0270   Median :-0.0041   Median : 0.0049   Median :-0.0226
##  Mean   : 0.0055   Mean   :-0.0277   Mean   :-0.0070   Mean   :-0.0243
##  3rd Qu.: 0.1225   3rd Qu.: 0.1378   3rd Qu.: 0.0900   3rd Qu.: 0.0749
##  Max.   : 0.4185   Max.   : 0.6030   Max.   : 0.3561   Max.   : 0.4372
##       MTDH              DCK              FLT1             PECI.1
##  Min.   :-0.6756   Min.   :-0.909   Min.   :-0.4826   Min.   :-0.4336
##  1st Qu.:-0.2933   1st Qu.:-0.529   1st Qu.:-0.1008   1st Qu.:-0.1396
##  Median :-0.0834   Median :-0.340   Median : 0.0189   Median :-0.0403
##  Mean   :-0.0867   Mean   :-0.321   Mean   :-0.0005   Mean   :-0.0336
##  3rd Qu.: 0.0738   3rd Qu.:-0.160   3rd Qu.: 0.0897   3rd Qu.: 0.0588
##  Max.   : 0.6406   Max.   : 0.599   Max.   : 0.5083   Max.   : 0.5128
##     QSCN6L1            DIAPH3            SLC2A3            GPR180
##  Min.   :-0.3794   Min.   :-0.4493   Min.   :-0.3716   Min.   :-0.3552
##  1st Qu.:-0.0466   1st Qu.:-0.1120   1st Qu.:-0.0777   1st Qu.:-0.0803
##  Median : 0.0078   Median :-0.0058   Median : 0.0005   Median :-0.0206
##  Mean   : 0.0217   Mean   :-0.0109   Mean   : 0.0114   Mean   :-0.0137
##  3rd Qu.: 0.0981   3rd Qu.: 0.0992   3rd Qu.: 0.0806   3rd Qu.: 0.0598
##  Max.   : 0.5401   Max.   : 0.3549   Max.   : 0.4642   Max.   : 0.3306
##     RTN4RL1        Contig32125_RC        STK32B             EXT1
##  Min.   :-0.6646   Min.   :-0.5321   Min.   :-0.4804   Min.   :-0.4778
##  1st Qu.:-0.2055   1st Qu.:-0.1135   1st Qu.:-0.1429   1st Qu.:-0.1675
##  Median : 0.0046   Median :-0.0090   Median :-0.0235   Median :-0.0558
##  Mean   :-0.0414   Mean   :-0.0110   Mean   :-0.0412   Mean   :-0.0519
##  3rd Qu.: 0.1318   3rd Qu.: 0.0734   3rd Qu.: 0.0449   3rd Qu.: 0.0605
##  Max.   : 0.4281   Max.   : 0.4563   Max.   : 0.4580   Max.   : 0.3741
##      COL4A2             PECI              GNAZ             AYTL2
##  Min.   :-0.5987   Min.   :-0.4423   Min.   :-0.3175   Min.   :-0.6943
##  1st Qu.:-0.1979   1st Qu.:-0.1942   1st Qu.:-0.0956   1st Qu.:-0.1319
##  Median :-0.0528   Median :-0.0637   Median :-0.0164   Median :-0.0460
##  Mean   :-0.0596   Mean   :-0.0373   Mean   : 0.0101   Mean   :-0.0252
##  3rd Qu.: 0.0627   3rd Qu.: 0.0966   3rd Qu.: 0.0834   3rd Qu.: 0.0654
##  Max.   : 0.5602   Max.   : 0.6090   Max.   : 0.4306   Max.   : 0.5336
##  Contig63649_RC        RAB6B          AA555029_RC         GPR126
##  Min.   :-0.3654   Min.   :-0.5692   Min.   :-0.431   Min.   :-0.3797
##  1st Qu.:-0.0984   1st Qu.:-0.1431   1st Qu.:-0.160   1st Qu.:-0.1361
##  Median :-0.0249   Median :-0.0522   Median :-0.001   Median :-0.0105
```

```
##  Mean   :-0.0094   Mean   :-0.0172   Mean   :-0.021   Mean   :-0.0164
##  3rd Qu.: 0.0900   3rd Qu.: 0.0896   3rd Qu.: 0.107   3rd Qu.: 0.0978
##  Max.   : 0.3205   Max.   : 0.4946   Max.   : 0.820   Max.   : 0.4393
##       ECT2             NUSAP1             GMPS             UCHL5
##  Min.   :-0.5077   Min.   :-0.5863   Min.   :-0.5915   Min.   :-0.4585
##  1st Qu.:-0.2311   1st Qu.:-0.1607   1st Qu.:-0.2841   1st Qu.:-0.1311
##  Median :-0.0813   Median :-0.0093   Median :-0.0451   Median :-0.0386
##  Mean   :-0.0500   Mean   :-0.0029   Mean   :-0.0605   Mean   :-0.0242
##  3rd Qu.: 0.0984   3rd Qu.: 0.1504   3rd Qu.: 0.1528   3rd Qu.: 0.0921
##  Max.   : 0.7757   Max.   : 0.6765   Max.   : 0.5519   Max.   : 0.5607
##      ORC6L             TSPYL5             MELK             RUNDC1
##  Min.   :-0.7968   Min.   :-0.6789   Min.   :-0.7898   Min.   :-0.870
##  1st Qu.:-0.2140   1st Qu.:-0.1786   1st Qu.:-0.1895   1st Qu.:-0.331
##  Median :-0.0244   Median :-0.0244   Median :-0.0611   Median :-0.118
##  Mean   :-0.0517   Mean   :-0.0320   Mean   :-0.0493   Mean   :-0.106
##  3rd Qu.: 0.1501   3rd Qu.: 0.1313   3rd Qu.: 0.0744   3rd Qu.: 0.104
##  Max.   : 0.5067   Max.   : 0.6178   Max.   : 0.8189   Max.   : 0.753
##     DIAPH3.1          C16orf61           TGFB3             FGF18
##  Min.   :-0.7682   Min.   :-0.6119   Min.   :-0.4152   Min.   :-0.5978
##  1st Qu.:-0.2564   1st Qu.:-0.1889   1st Qu.:-0.0924   1st Qu.:-0.1404
##  Median :-0.0683   Median :-0.0931   Median :-0.0053   Median : 0.0015
##  Mean   :-0.0539   Mean   :-0.0591   Mean   :-0.0023   Mean   :-0.0232
##  3rd Qu.: 0.1179   3rd Qu.: 0.0587   3rd Qu.: 0.0827   3rd Qu.: 0.1070
##  Max.   : 0.7049   Max.   : 0.5941   Max.   : 0.4397   Max.   : 0.4822
##     CDC42BPA            DTL              WISP1            DIAPH3.2
##  Min.   :-0.4444   Min.   :-1.264   Min.   :-0.4404   Min.   :-0.4510
##  1st Qu.:-0.1519   1st Qu.:-0.651   1st Qu.:-0.0876   1st Qu.:-0.1221
##  Median :-0.0436   Median :-0.153   Median : 0.0240   Median : 0.0088
##  Mean   :-0.0264   Mean   :-0.209   Mean   : 0.0131   Mean   :-0.0009
##  3rd Qu.: 0.0804   3rd Qu.: 0.203   3rd Qu.: 0.1223   3rd Qu.: 0.1127
##  Max.   : 0.4842   Max.   : 0.892   Max.   : 0.3755   Max.   : 0.3669
##      OXCT1             ZNF533             RFC4             KNTC2
##  Min.   :-0.4278   Min.   :-0.5109   Min.   :-0.5636   Min.   :-0.4311
##  1st Qu.:-0.0905   1st Qu.:-0.2613   1st Qu.:-0.0825   1st Qu.:-0.1841
##  Median : 0.0095   Median :-0.1380   Median :-0.0010   Median :-0.0616
##  Mean   : 0.0161   Mean   :-0.0593   Mean   : 0.0080   Mean   :-0.0359
##  3rd Qu.: 0.1234   3rd Qu.: 0.0381   3rd Qu.: 0.1045   3rd Qu.: 0.0722
##  Max.   : 0.6491   Max.   : 0.8648   Max.   : 0.4791   Max.   : 0.5975
##      FBXO31           GSTM3_s       RP5.860F19.3_s        BBC3_s
##  Min.   :-0.4215   Min.   :-1.845   Min.   :-2.2232   Min.   :-2.9155
##  1st Qu.:-0.1388   1st Qu.:-0.763   1st Qu.:-0.6210   1st Qu.:-0.6625
##  Median :-0.0451   Median :-0.130   Median :-0.0348   Median : 0.0531
##  Mean   :-0.0253   Mean   : 0.000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 0.0860   3rd Qu.: 0.597   3rd Qu.: 0.4423   3rd Qu.: 0.6732
##  Max.   : 0.5556   Max.   : 2.786   Max.   : 2.9235   Max.   : 2.1487
##      MMP9_s        Contig35251_RC_s Contig40831_RC_s   ALDH4A1_s
##  Min.   :-2.1926   Min.   :-1.534   Min.   :-2.653   Min.   :-3.322
##  1st Qu.:-0.5924   1st Qu.:-0.785   1st Qu.:-0.729   1st Qu.:-0.661
##  Median :-0.0509   Median :-0.348   Median : 0.120   Median : 0.106
##  Mean   : 0.0000   Mean   : 0.000   Mean   : 0.000   Mean   : 0.000
##  3rd Qu.: 0.5992   3rd Qu.: 0.680   3rd Qu.: 0.651   3rd Qu.: 0.743
##  Max.   : 2.6553   Max.   : 2.870   Max.   : 2.296   Max.   : 2.831
##     SERF1A_s          SCUBE2_s           MTDH_s            DCK_s
##  Min.   :-3.214   Min.   :-3.1470   Min.   :-2.1023   Min.   :-2.120
```

```
##  1st Qu.:-0.534   1st Qu.:-0.6724   1st Qu.:-0.7373   1st Qu.:-0.748
##  Median : 0.070   Median : 0.0104   Median : 0.0118   Median :-0.066
##  Mean   : 0.000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.000
##  3rd Qu.: 0.568   3rd Qu.: 0.6356   3rd Qu.: 0.5733   3rd Qu.: 0.584
##  Max.   : 2.125   Max.   : 2.9577   Max.   : 2.5964   Max.   : 3.320
##      FLT1_s          PECI.1_s        QSCN6L1_s         DIAPH3_s
##  Min.   :-3.057   Min.   :-2.456   Min.   :-3.015   Min.   :-2.5771
##  1st Qu.:-0.636   1st Qu.:-0.651   1st Qu.:-0.513   1st Qu.:-0.5946
##  Median : 0.123   Median :-0.041   Median :-0.105   Median : 0.0302
##  Mean   : 0.000   Mean   : 0.000   Mean   : 0.000   Mean   : 0.0000
##  3rd Qu.: 0.572   3rd Qu.: 0.568   3rd Qu.: 0.575   3rd Qu.: 0.6471
##  Max.   : 3.227   Max.   : 3.356   Max.   : 3.897   Max.   : 2.1500
##     SLC2A3_s         GPR180_s        RTN4RL1_s      Contig32125_RC_s
##  Min.   :-2.4430   Min.   :-2.638   Min.   :-2.503   Min.   :-3.357
##  1st Qu.:-0.5685   1st Qu.:-0.515   1st Qu.:-0.659   1st Qu.:-0.660
##  Median :-0.0693   Median :-0.053   Median : 0.185   Median : 0.013
##  Mean   : 0.0000   Mean   : 0.000   Mean   : 0.000   Mean   : 0.000
##  3rd Qu.: 0.4415   3rd Qu.: 0.568   3rd Qu.: 0.696   3rd Qu.: 0.544
##  Max.   : 2.8891   Max.   : 2.659   Max.   : 1.886   Max.   : 3.011
##     STK32B_s          EXT1_s          COL4A2_s          PECI_s
##  Min.   :-2.711   Min.   :-2.4843   Min.   :-2.6489   Min.   :-1.895
##  1st Qu.:-0.627   1st Qu.:-0.6743   1st Qu.:-0.6795   1st Qu.:-0.734
##  Median : 0.110   Median :-0.0225   Median : 0.0334   Median :-0.124
##  Mean   : 0.000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.000
##  3rd Qu.: 0.531   3rd Qu.: 0.6559   3rd Qu.: 0.6012   3rd Qu.: 0.626
##  Max.   : 3.082   Max.   : 2.4850   Max.   : 3.0457   Max.   : 3.023
##      GNAZ_s          AYTL2_s       Contig63649_RC_s    RAB6B_s
##  Min.   :-2.167   Min.   :-3.980   Min.   :-2.573   Min.   :-2.710
##  1st Qu.:-0.700   1st Qu.:-0.635   1st Qu.:-0.643   1st Qu.:-0.618
##  Median :-0.175   Median :-0.124   Median :-0.112   Median :-0.172
##  Mean   : 0.000   Mean   : 0.000   Mean   : 0.000   Mean   : 0.000
##  3rd Qu.: 0.485   3rd Qu.: 0.539   3rd Qu.: 0.718   3rd Qu.: 0.524
##  Max.   : 2.782   Max.   : 3.323   Max.   : 2.384   Max.   : 2.513
##  AA555029_RC_s       GPR126_s          ECT2_s          NUSAP1_s
##  Min.   :-2.190   Min.   :-2.1825   Min.   :-1.917   Min.   :-2.4885
##  1st Qu.:-0.743   1st Qu.:-0.7189   1st Qu.:-0.759   1st Qu.:-0.6731
##  Median : 0.106   Median : 0.0356   Median :-0.131   Median :-0.0273
##  Mean   : 0.000   Mean   : 0.0000   Mean   : 0.000   Mean   : 0.0000
##  3rd Qu.: 0.687   3rd Qu.: 0.6862   3rd Qu.: 0.621   3rd Qu.: 0.6540
##  Max.   : 4.495   Max.   : 2.7371   Max.   : 3.458   Max.   : 2.8981
##      GMPS_s          UCHL5_s          ORC6L_s          TSPYL5_s
##  Min.   :-2.0588   Min.   :-2.657   Min.   :-2.786   Min.   :-2.7252
##  1st Qu.:-0.8669   1st Qu.:-0.654   1st Qu.:-0.607   1st Qu.:-0.6176
##  Median : 0.0594   Median :-0.088   Median : 0.102   Median : 0.0318
##  Mean   : 0.0000   Mean   : 0.000   Mean   : 0.000   Mean   : 0.0000
##  3rd Qu.: 0.8269   3rd Qu.: 0.711   3rd Qu.: 0.754   3rd Qu.: 0.6877
##  Max.   : 2.3741   Max.   : 3.578   Max.   : 2.088   Max.   : 2.7375
##      MELK_s          RUNDC1_s        DIAPH3.1_s        C16orf61_s
##  Min.   :-3.193   Min.   :-2.5022   Min.   :-2.6615   Min.   :-2.860
##  1st Qu.:-0.604   1st Qu.:-0.7355   1st Qu.:-0.7545   1st Qu.:-0.672
##  Median :-0.051   Median :-0.0409   Median :-0.0537   Median :-0.176
##  Mean   : 0.000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.000
##  3rd Qu.: 0.533   3rd Qu.: 0.6859   3rd Qu.: 0.6400   3rd Qu.: 0.609
##  Max.   : 3.744   Max.   : 2.8101   Max.   : 2.8273   Max.   : 3.380
```

```
##     TGFB3_s           FGF18_s          CDC42BPA_s          DTL_s
##  Min.   :-2.8862   Min.   :-2.939   Min.   :-2.4108   Min.   :-2.006
##  1st Qu.:-0.6297   1st Qu.:-0.600   1st Qu.:-0.7237   1st Qu.:-0.839
##  Median :-0.0212   Median : 0.126   Median :-0.0991   Median : 0.107
##  Mean   : 0.0000   Mean   : 0.000   Mean   : 0.0000   Mean   : 0.000
##  3rd Qu.: 0.5942   3rd Qu.: 0.666   3rd Qu.: 0.6162   3rd Qu.: 0.785
##  Max.   : 3.0889   Max.   : 2.585   Max.   : 2.9450   Max.   : 2.095
##     WISP1_s          DIAPH3.2_s         OXCT1_s          ZNF533_s
##  Min.   :-2.8269   Min.   :-2.7614   Min.   :-2.435   Min.   :-1.565
##  1st Qu.:-0.6278   1st Qu.:-0.7434   1st Qu.:-0.585   1st Qu.:-0.700
##  Median : 0.0679   Median : 0.0598   Median :-0.036   Median :-0.273
##  Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.000   Mean   : 0.000
##  3rd Qu.: 0.6807   3rd Qu.: 0.6967   3rd Qu.: 0.588   3rd Qu.: 0.337
##  Max.   : 2.2588   Max.   : 2.2564   Max.   : 3.472   Max.   : 3.203
##      RFC4_s          KNTC2_s          FBXO31_s         GSTM3_1
##  Min.   :-3.620   Min.   :-1.978   Min.   :-2.362   Min.   :0.971
##  1st Qu.:-0.573   1st Qu.:-0.742   1st Qu.:-0.676   1st Qu.:1.049
##  Median :-0.057   Median :-0.129   Median :-0.117   Median :1.092
##  Mean   : 0.000   Mean   : 0.000   Mean   : 0.000   Mean   :1.098
##  3rd Qu.: 0.611   3rd Qu.: 0.541   3rd Qu.: 0.664   3rd Qu.:1.139
##  Max.   : 2.983   Max.   : 3.170   Max.   : 3.463   Max.   :1.269
##  RP5.860F19.3_1      BBC3_1          MMP9_1      Contig35251_RC_1
##  Min.   :0.946   Min.   :0.651   Min.   :0.919   Min.   :0.734
##  1st Qu.:1.062   1st Qu.:0.981   1st Qu.:1.044   1st Qu.:0.879
##  Median :1.101   Median :1.066   Median :1.083   Median :0.955
##  Mean   :1.102   Mean   :1.053   Mean   :1.084   Mean   :0.999
##  3rd Qu.:1.132   3rd Qu.:1.135   3rd Qu.:1.127   3rd Qu.:1.113
##  Max.   :1.279   Max.   :1.281   Max.   :1.258   Max.   :1.385
##  Contig40831_RC_1   ALDH4A1_1        SERF1A_1        SCUBE2_1
##  Min.   :0.928    Min.   :0.803   Min.   :0.893   Min.   :0.91
##  1st Qu.:1.056    1st Qu.:1.038   1st Qu.:1.065   1st Qu.:1.05
##  Median :1.108    Median :1.097   Median :1.100   Median :1.09
##  Mean   :1.099    Mean   :1.087   Mean   :1.095   Mean   :1.09
##  3rd Qu.:1.139    3rd Qu.:1.143   3rd Qu.:1.128   3rd Qu.:1.12
##  Max.   :1.229    Max.   :1.282   Max.   :1.211   Max.   :1.23
##     MTDH_1          DCK_1           FLT1_1         PECI.1_1
##  Min.   :0.843   Min.   :0.738   Min.   :0.923   Min.   :0.942
##  1st Qu.:0.996   1st Qu.:0.905   1st Qu.:1.064   1st Qu.:1.051
##  Median :1.070   Median :0.978   Median :1.105   Median :1.085
##  Mean   :1.065   Mean   :0.980   Mean   :1.097   Mean   :1.086
##  3rd Qu.:1.123   3rd Qu.:1.044   3rd Qu.:1.128   3rd Qu.:1.118
##  Max.   :1.292   Max.   :1.280   Max.   :1.255   Max.   :1.256
##    QSCN6L1_1        DIAPH3_1        SLC2A3_1        GPR180_1
##  Min.   :0.963   Min.   :0.936   Min.   :0.966   Min.   :0.973
##  1st Qu.:1.083   1st Qu.:1.060   1st Qu.:1.072   1st Qu.:1.071
##  Median :1.101   Median :1.097   Median :1.099   Median :1.092
##  Mean   :1.105   Mean   :1.093   Mean   :1.101   Mean   :1.093
##  3rd Qu.:1.131   3rd Qu.:1.131   3rd Qu.:1.125   3rd Qu.:1.118
##  Max.   :1.264   Max.   :1.210   Max.   :1.242   Max.   :1.203
##    RTN4RL1_1     Contig32125_RC_1    STK32B_1         EXT1_1
##  Min.   :0.848   Min.   :0.903    Min.   :0.924   Min.   :0.925
##  1st Qu.:1.028   1st Qu.:1.060    1st Qu.:1.050   1st Qu.:1.041
##  Median :1.100   Median :1.096    Median :1.091   Median :1.080
##  Mean   :1.081   Mean   :1.094    Mean   :1.083   Mean   :1.079
```

```
##  3rd Qu.:1.142   3rd Qu.:1.123    3rd Qu.:1.113   3rd Qu.:1.119
##  Max.   :1.232   Max.   :1.240    Max.   :1.241   Max.   :1.216
##     COL4A2_l         PECI_l          GNAZ_l         AYTL2_l
##  Min.   :0.876   Min.   :0.939   Min.   :0.987   Min.   :0.835
##  1st Qu.:1.030   1st Qu.:1.032   1st Qu.:1.066   1st Qu.:1.054
##  Median :1.081   Median :1.077   Median :1.093   Median :1.083
##  Mean   :1.076   Mean   :1.084   Mean   :1.101   Mean   :1.089
##  3rd Qu.:1.119   3rd Qu.:1.130   3rd Qu.:1.126   3rd Qu.:1.120
##  Max.   :1.270   Max.   :1.283   Max.   :1.233   Max.   :1.262
##  Contig63649_RC_l    RAB6B_l      AA555029_RC_l      GPR126_l
##  Min.   :0.969    Min.   :0.888   Min.   :0.944   Min.   :0.963
##  1st Qu.:1.065    1st Qu.:1.050   1st Qu.:1.044   1st Qu.:1.052
##  Median :1.090    Median :1.081   Median :1.098   Median :1.095
##  Mean   :1.094    Mean   :1.091   Mean   :1.090   Mean   :1.092
##  3rd Qu.:1.128    3rd Qu.:1.128   3rd Qu.:1.134   3rd Qu.:1.131
##  Max.   :1.200    Max.   :1.251   Max.   :1.340   Max.   :1.235
##      ECT2_l         NUSAP1_l         GMPS_l         UCHL5_l         ORC6L_l
##  Min.   :0.913   Min.   :0.881   Min.   :0.879   Min.   :0.933   Min.   :0.79
##  1st Qu.:1.018   1st Qu.:1.044   1st Qu.:0.999   1st Qu.:1.054   1st Qu.:1.02
##  Median :1.071   Median :1.095   Median :1.083   Median :1.086   Median :1.09
##  Mean   :1.079   Mean   :1.095   Mean   :1.074   Mean   :1.089   Mean   :1.08
##  3rd Qu.:1.131   3rd Qu.:1.147   3rd Qu.:1.148   3rd Qu.:1.129   3rd Qu.:1.15
##  Max.   :1.329   Max.   :1.302   Max.   :1.268   Max.   :1.270   Max.   :1.25
##     TSPYL5_l         MELK_l         RUNDC1_l       DIAPH3.1_l
##  Min.   :0.842   Min.   :0.793   Min.   :0.756   Min.   :0.803
##  1st Qu.:1.037   1st Qu.:1.033   1st Qu.:0.982   1st Qu.:1.009
##  Median :1.090   Median :1.078   Median :1.058   Median :1.076
##  Mean   :1.085   Mean   :1.079   Mean   :1.057   Mean   :1.076
##  3rd Qu.:1.141   3rd Qu.:1.123   3rd Qu.:1.133   3rd Qu.:1.137
##  Max.   :1.286   Max.   :1.340   Max.   :1.323   Max.   :1.310
##    C16orf61_l       TGFB3_l        FGF18_l        CDC42BPA_l        DTL_l
##  Min.   :0.871   Min.   :0.95   Min.   :0.876   Min.   :0.938   Min.   :0.551
##  1st Qu.:1.034   1st Qu.:1.07   1st Qu.:1.051   1st Qu.:1.047   1st Qu.:0.854
##  Median :1.067   Median :1.10   Median :1.099   Median :1.084   Median :1.046
##  Mean   :1.077   Mean   :1.10   Mean   :1.089   Mean   :1.088   Mean   :1.007
##  3rd Qu.:1.118   3rd Qu.:1.13   3rd Qu.:1.134   3rd Qu.:1.125   3rd Qu.:1.164
##  Max.   :1.279   Max.   :1.24   Max.   :1.248   Max.   :1.248   Max.   :1.359
##     WISP1_l       DIAPH3.2_l       OXCT1_l         ZNF533_l         RFC4_l
##  Min.   :0.94   Min.   :0.936   Min.   :0.945   Min.   :0.912   Min.   :0.89
##  1st Qu.:1.07   1st Qu.:1.057   1st Qu.:1.068   1st Qu.:1.008   1st Qu.:1.07
##  Median :1.11   Median :1.102   Median :1.102   Median :1.052   Median :1.10
##  Mean   :1.10   Mean   :1.097   Mean   :1.102   Mean   :1.074   Mean   :1.10
##  3rd Qu.:1.14   3rd Qu.:1.135   3rd Qu.:1.139   3rd Qu.:1.111   3rd Qu.:1.13
##  Max.   :1.22   Max.   :1.214   Max.   :1.294   Max.   :1.352   Max.   :1.25
##     KNTC2_l         FBXO31_l
##  Min.   :0.944   Min.   :0.947
##  1st Qu.:1.035   1st Qu.:1.051
##  Median :1.078   Median :1.083
##  Mean   :1.084   Mean   :1.089
##  3rd Qu.:1.122   3rd Qu.:1.127
##  Max.   :1.280   Max.   :1.268
```

```
#Based on summary results, categorical proportions are as follows:
#infection:
```

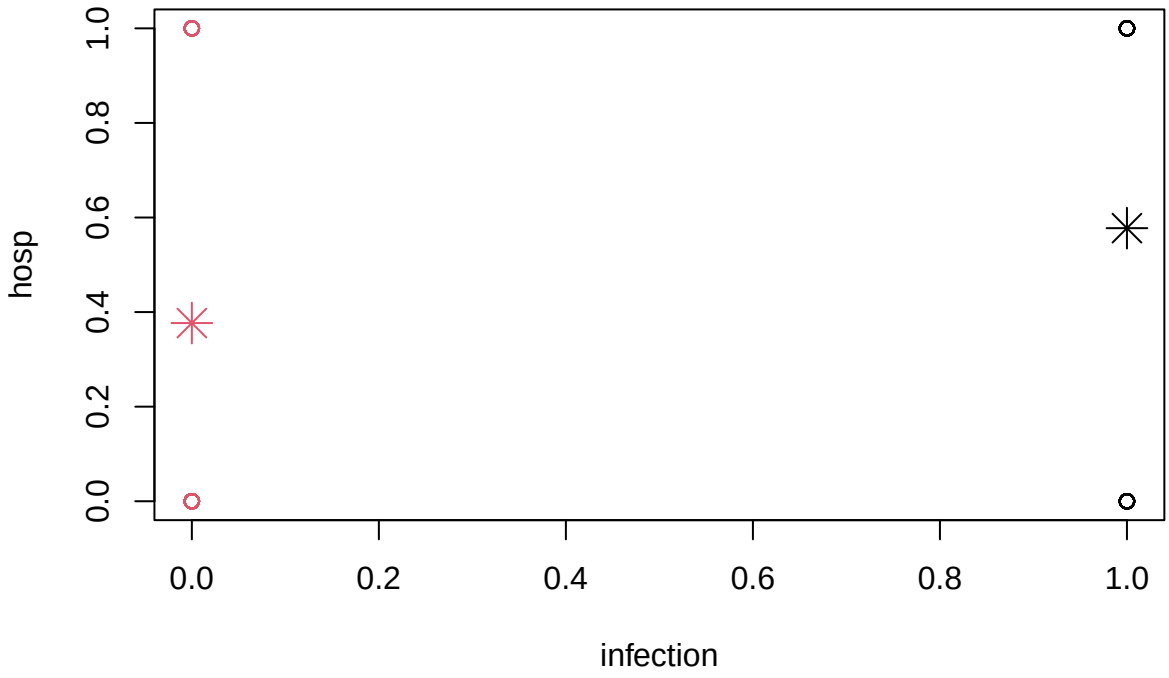
```
#bacterial_infection:69
#viral_infection    :71

#sind
#symptoms_remain  :93
#symptoms_finished:47

#gender
#female:62
#male  :78

#no_hospitalization:73
#hospitalization   :67

#Based on these results, it appears that clustering is more associated with sind (symptoms remain vs. f
#to type of infection or risk of hospitalization

#Clustering is graphically depicted as follows:

#plot(kdata, col=kcluster$cluster)
#points(kcluster$centers, col=1:2, pch=8, cex=2)

#Clustering individuals according to the kind of infection (nned to factor)
#k_infection<-as.factor(viral34_c$infection)
#kcluster_infection<-kmeans(k_infection, 2, nstart=10)
#kcluster_infection
#Clustering individuals according to the risk of hospitalization (nned to factor)
#k_hosp<-as.factor(viral34_c$hosp)
#kcluster_hosp<-kmeans(k_hosp, 2, nstart=10)
#kcluster_hosp

#If the factored column variables of infection and hospitalization are used for 2 kmeans clustering,
# the following error message is generated: Error in kmeans(kdata, 2, nstart = 10) : more cluster cente
#In addition: Warning message:In storage.mode(x) <- "double" : NAs introduced by coercion
#This occurs because kmeans uses the mean of data points for clustering and our dataset is made of plai
#(i.e not numbers). The 2 raw unfactored data columns can be used, or the data can be preprocessed via
#(e.g. "one hot encoding" method that transforms the 2 category column into 4 multiple columns that eac
#belongs to the relevant category (i.e column with 3 ancestries will get 3 new binary (1 or 0) columns.
# Other methods include ROCK algorithm (kaggle notebook)and "Kmode" which is similar to kmeans for cate

#https://www.kaggle.com/code/vijjikiran/clustering-of-categorical-data/report
# k-means is the classical unsupervised clustering algorithm for numerical data. But computing the eucl
#So instead, I will therefore run categorical data through the following algorithms for clustering -
# (1)By applying one-hot encoding, the data will be converted to numeric data and then it will be run t
# (2)The data will be run through k-modes algorithm that uses modes of categorical attributes instead o
# (3)The data will be run through the Rock(Robust clustering using links) algorithm that is designed fo
# I will evaluate how the purity of the clusters, a simple evaluative measure, is different for the eac
#Purity of clustering is a simple measure of the accuracy, which is between 0 and 1. 0 indicates poor c

#Demonstrating 2 alternative methods:
#Method#1: Extract 2 unfactored categorical columns (infection and hospitalization) from original viral
kdata1<-viral34[,c(1,5)]
head(kdata1)
```

```
##   infection hosp
## 1         0    1
## 2         1    0
## 3         0    1
## 4         1    1
## 5         1    0
## 6         1    0
```

```
kcluster1<-kmeans(kdata1, 2, nstart=10)
kcluster1
```

```
## K-means clustering with 2 clusters of sizes 71, 69
##
## Cluster means:
##   infection   hosp
## 1         1 0.5775
## 2         0 0.3768
##
## Clustering vector:
##   [1] 2 1 2 1 1 1 2 2 1 2 1 1 1 2 2 2 1 2 2 1 2 1 2 1 2 2 2 2 2 2 2 2 2 1 1 1 1
##  [38] 1 1 2 1 2 1 1 1 1 1 2 1 1 2 1 2 2 1 2 2 2 2 1 1 1 1 1 2 1 1 2 2 1 2 1 1 1
##  [75] 1 2 2 1 1 2 2 2 1 1 2 1 1 1 1 2 2 1 2 2 1 1 1 2 2 1 2 2 1 2 1 1 1 2 2 1 2
## [112] 1 2 1 1 2 1 2 2 1 1 1 2 1 2 2 2 2 1 2 2 1 1 2 1 2 2 2 2 1
##
## Within cluster sum of squares by cluster:
## [1] 17.32 16.20
##  (between_SS / total_SS =  52.1 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

```
#Results: Within cluster sum of squares by cluster:
#   [1] 17.32394 16.20290
# (between_SS / total_SS =  52.1 %)
plot(kdata1, col=kcluster1$cluster)
points(kcluster1$centers, col=1:2, pch=8, cex=2)
```




```
#YES! The clustering is visually related to infection risk and hostpitalization status##############

#Method#2: "Hot-start"
library(caret)

# Dummify data
dmy <- dummyVars(" ~ .", viral34_c)
trsf <- data.frame(predict(dmy, newdata = viral34_c))
#Show what dataframe looks like
kdata2<-trsf[,c(1,2,8,9)]
head(kdata2)
```

```
##   infection.bacterial_infection infection.viral_infection
## 1                             1                         0
## 2                             0                         1
## 3                             1                         0
## 4                             0                         1
## 5                             0                         1
## 6                             0                         1
##   hosp.no_hospitalization hosp.hospitalization
## 1                       0                    1
## 2                       1                    0
## 3                       0                    1
## 4                       0                    1
## 5                       1                    0
## 6                       1                    0
```

```r
#Since 4 columns of categorical data were transformed, using 4 kmeans
kcluster2<-kmeans(kdata2, 4, nstart=10)
kcluster2
```

```
## K-means clustering with 4 clusters of sizes 26, 30, 41, 43
##
## Cluster means:
##   infection.bacterial_infection infection.viral_infection
## 1                             1                         0
## 2                             0                         1
## 3                             0                         1
## 4                             1                         0
##   hosp.no_hospitalization hosp.hospitalization
## 1                       0                    1
## 2                       1                    0
## 3                       0                    1
## 4                       1                    0
##
## Clustering vector:
##   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19  20
##   1   2   1   3   2   2   4   4   3   4   2   2   3   4   1   4   3   4   1   2
##  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36  37  38  39  40
##   4   2   4   2   1   1   1   4   1   4   1   4   4   2   3   3   3   3   2   4
##  41  42  43  44  45  46  47  48  49  50  51  52  53  54  55  56  57  58  59  60
##   2   4   3   3   3   2   3   4   2   2   1   2   1   1   3   4   4   4   1   3
##  61  62  63  64  65  66  67  68  69  70  71  72  73  74  75  76  77  78  79  80
##   2   3   3   3   4   3   2   4   1   3   4   3   2   3   3   4   4   2   2   4
##  81  82  83  84  85  86  87  88  89  90  91  92  93  94  95  96  97  98  99 100
##   4   4   2   3   1   3   3   2   2   1   4   3   4   4   3   3   3   4   1   3
## 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
##   4   1   2   4   3   3   2   4   4   2   4   3   4   3   3   1   2   4   4   3
## 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140
##   3   3   1   2   4   1   1   1   3   4   4   3   2   1   3   4   1   1   4   2
##
## Within cluster sum of squares by cluster:
## [1] 0 0 0 0
##  (between_SS / total_SS = 100.0 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

```r
#Results: K-means clustering with 4 clusters of sizes 43, 30, 41, 26

#I will later demonstrate two additional clustering approaches using the following libraries tools:
#For kmode
library(klaR)
#For ROCK
library(cba)
```

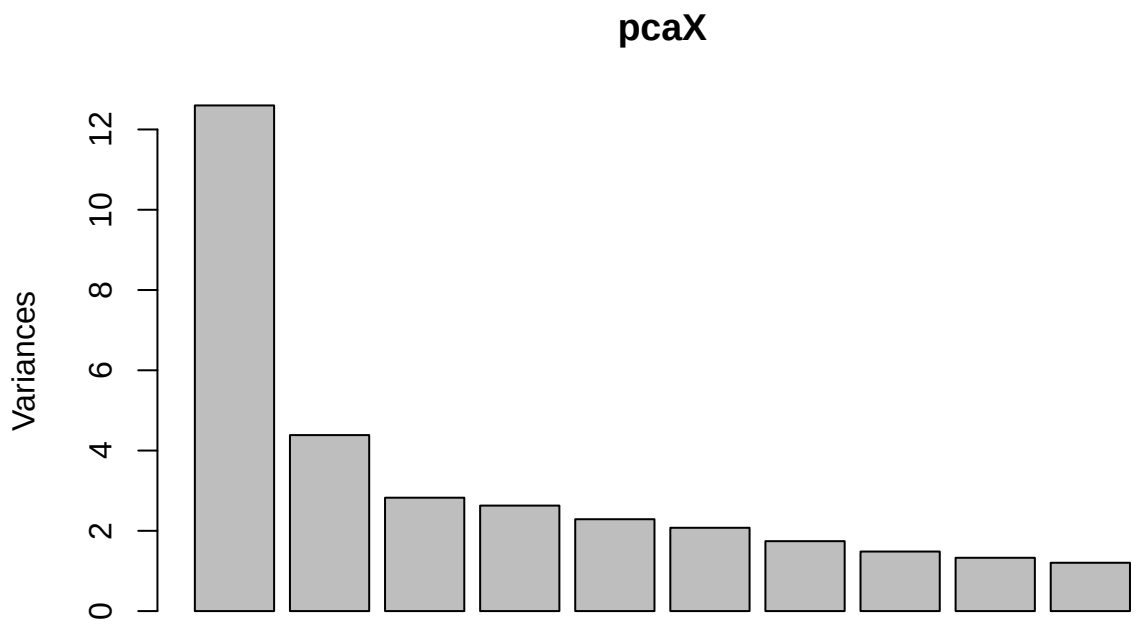
QUESTION 5 Perform PCA for exploring possible relationships BETWEEN INDIVIDUALS according to their (scaled) GENE EXPRESSION LEVELS. Provide the variance explained plot. How much variability

is explained by the first two principal components? Which is the eigen-value of PC1 and how can be interpreted? Check, using concentration ellipses, whether PCA projections of individuals are associated to infection, gender, hospitalization or ancestry. Which are the 10 genes that most contribute to PC1 and PC2? (follow similar steps as in section 1.5.8 in "Solutions Exercises section 2"). Discuss the results.

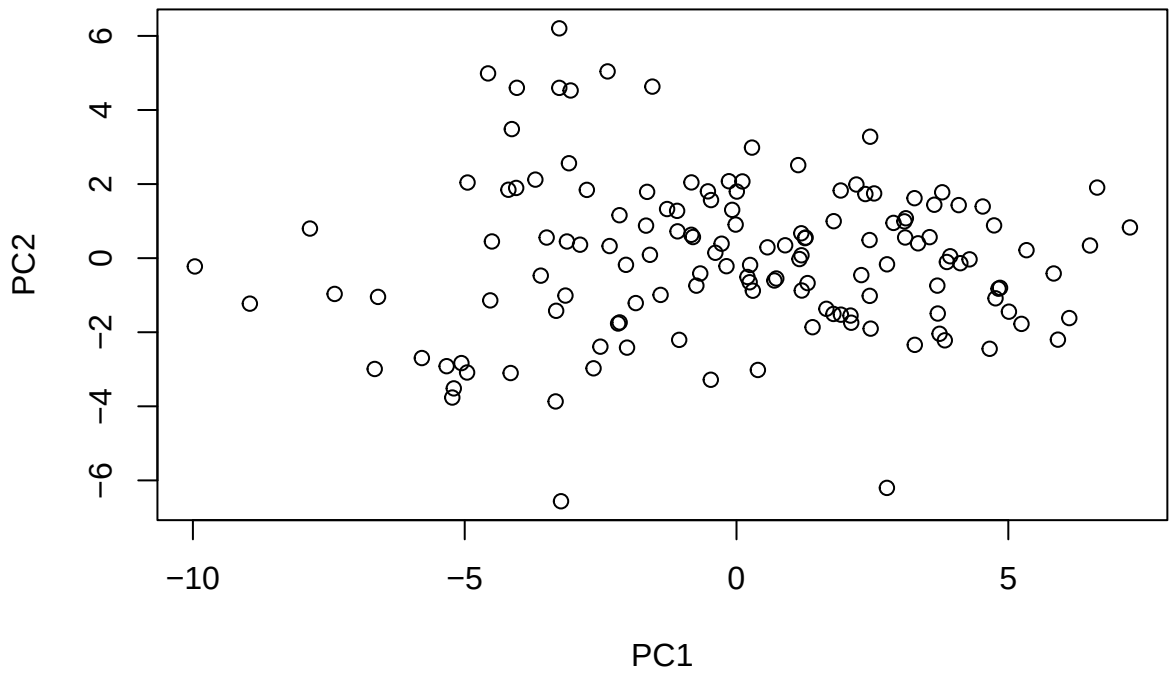
```
X<-viral34_c[,58:107]
pcaX <-prcomp(X, scale =TRUE)
summary(pcaX)
```

```
## Importance of components:
##                           PC1    PC2    PC3    PC4    PC5    PC6    PC7    PC8
## Standard deviation     3.549 2.0941 1.6807 1.6212 1.5128 1.4407 1.3199 1.2180
## Proportion of Variance 0.252 0.0877 0.0565 0.0526 0.0458 0.0415 0.0348 0.0297
## Cumulative Proportion  0.252 0.3397 0.3962 0.4487 0.4945 0.5360 0.5709 0.6005
##                           PC9   PC10   PC11   PC12   PC13   PC14   PC15   PC16
## Standard deviation     1.1524 1.0973 1.0868 1.0422 1.0277 0.9919 0.9643 0.9246
## Proportion of Variance 0.0266 0.0241 0.0236 0.0217 0.0211 0.0197 0.0186 0.0171
## Cumulative Proportion  0.6271 0.6512 0.6748 0.6965 0.7177 0.7373 0.7559 0.7730
##                           PC17   PC18   PC19   PC20   PC21   PC22   PC23   PC24
## Standard deviation     0.8979 0.8855 0.8715 0.8598 0.8316 0.7959 0.7584 0.7216
## Proportion of Variance 0.0161 0.0157 0.0152 0.0148 0.0138 0.0127 0.0115 0.0104
## Cumulative Proportion  0.7892 0.8048 0.8200 0.8348 0.8486 0.8613 0.8728 0.8832
##                           PC25    PC26    PC27    PC28    PC29    PC30    PC31
## Standard deviation     0.70540 0.69312 0.65886 0.62757 0.60388 0.57832 0.57115
## Proportion of Variance 0.00995 0.00961 0.00868 0.00788 0.00729 0.00669 0.00652
## Cumulative Proportion  0.89318 0.90279 0.91147 0.91935 0.92664 0.93333 0.93985
##                           PC32    PC33    PC34    PC35    PC36    PC37    PC38
## Standard deviation     0.55491 0.54517 0.52600 0.49787 0.48352 0.47144 0.44215
## Proportion of Variance 0.00616 0.00594 0.00553 0.00496 0.00468 0.00445 0.00391
## Cumulative Proportion  0.94601 0.95196 0.95749 0.96245 0.96712 0.97157 0.97548
##                           PC39    PC40    PC41    PC42    PC43    PC44    PC45
## Standard deviation     0.41424 0.40245 0.38027 0.35457 0.34884 0.32447 0.29799
## Proportion of Variance 0.00343 0.00324 0.00289 0.00251 0.00243 0.00211 0.00178
## Cumulative Proportion  0.97891 0.98215 0.98504 0.98756 0.98999 0.99210 0.99387
##                           PC46    PC47    PC48    PC49    PC50
## Standard deviation     0.28162 0.27687 0.24151 0.23198 0.19576
## Proportion of Variance 0.00159 0.00153 0.00117 0.00108 0.00077
## Cumulative Proportion  0.99546 0.99699 0.99816 0.99923 1.00000
```

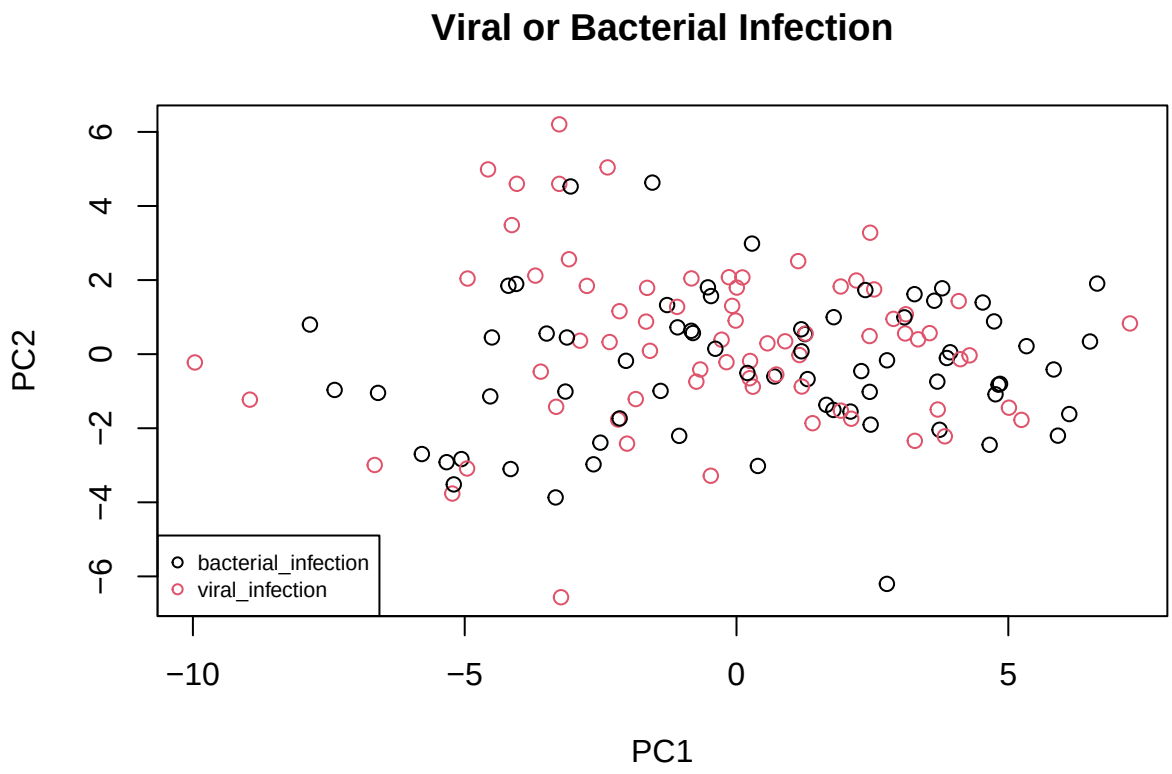
```
#PC1 accounts for 0.252, PC2 accounts for 0.08771. PC3 accounts for 0.0565 for a total of 0.3962 of var
plot(pcaX)
```



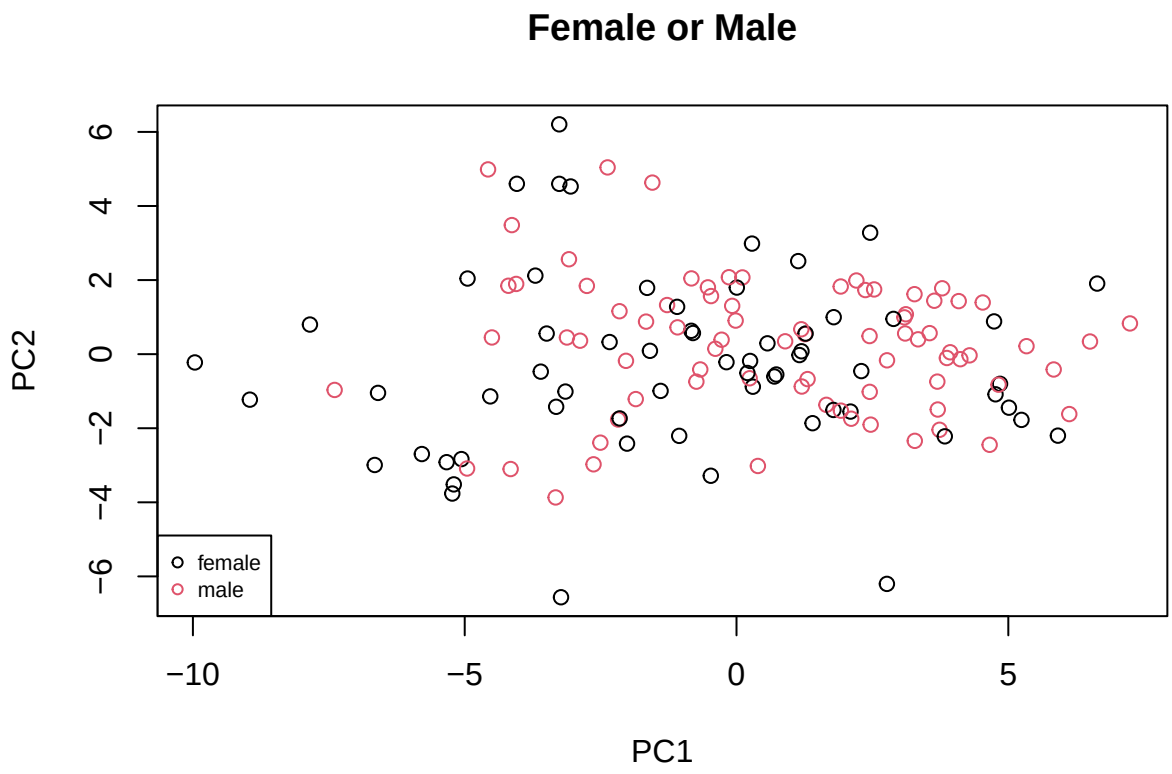

```
#Plotting the data on the first two principal components
PC1 <- pcaX$x[,1]
PC2 <- pcaX$x[,2]
plot(PC1,PC2)
```



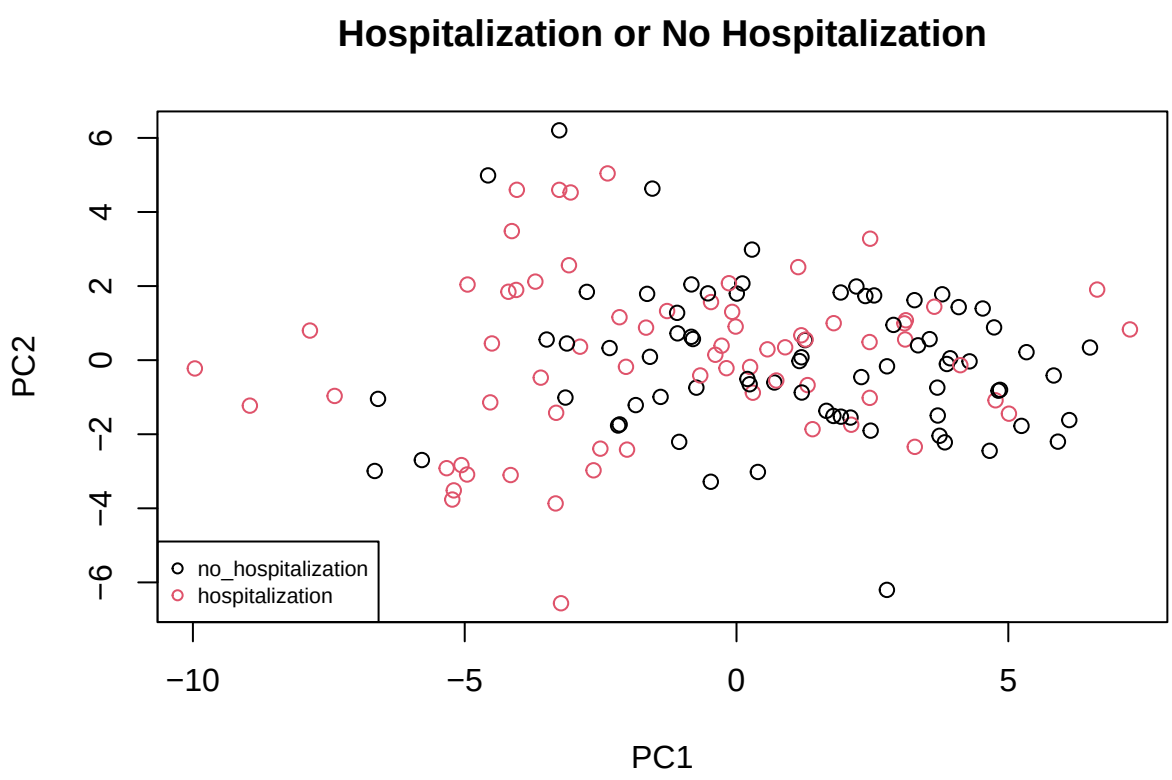

```
#Plotting the data on the first two principal components and color the points according to "infection"
plot(PC1,PC2, col=viral34_c$infection, main = "Viral or Bacterial Infection")
legend("bottomleft", col=1:2, legend=levels(viral34_c$infection), pch=1, cex=0.7)
```



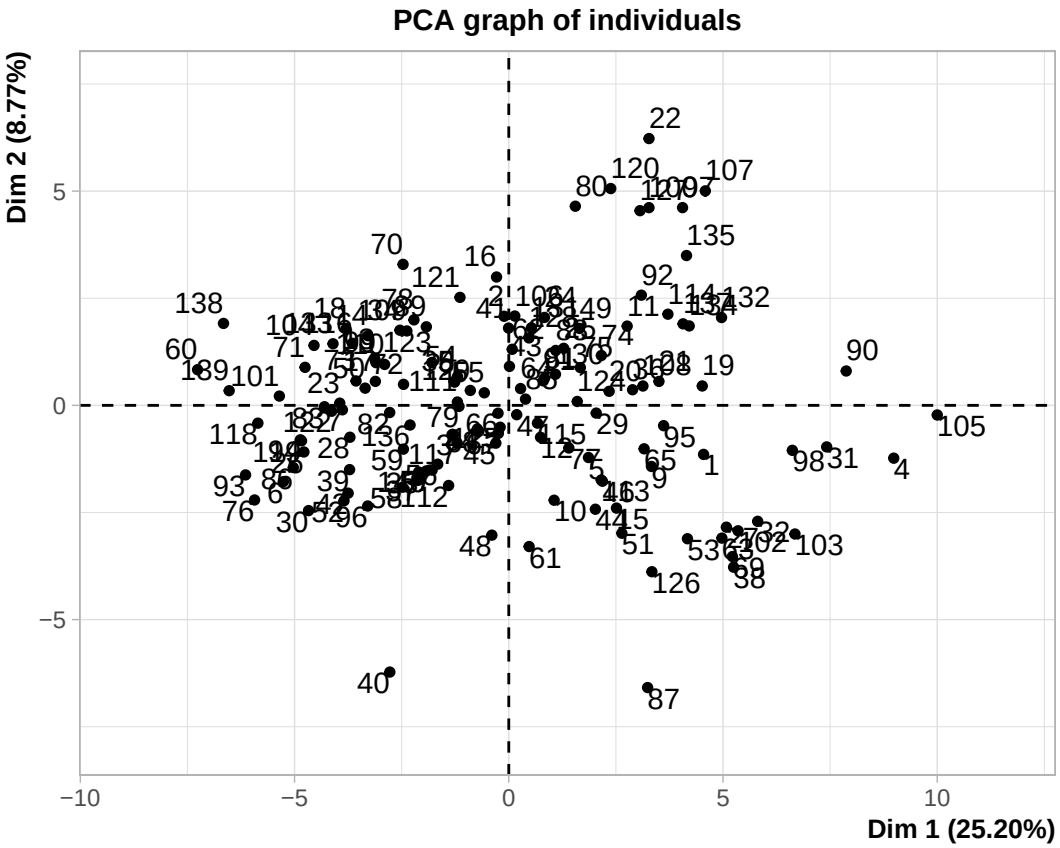

```
#There is no clear association between PCA projections and "infection"

#Plotting the data on the first two principal components and color the points according to "gender"
plot(PC1,PC2, col=viral34_c$gender, main = "Female or Male")
legend("bottomleft", col=1:2, legend=levels(viral34_c$gender), pch=1, cex=0.7)
```

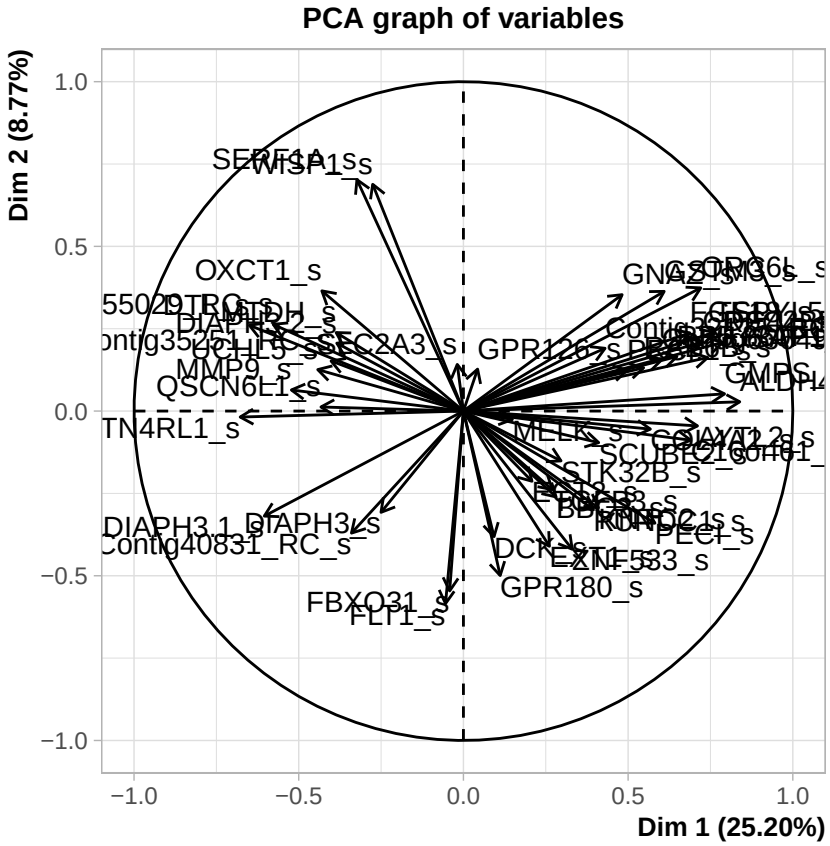



```
#There is no clear association between PCA projections and "gender"

#Plotting the data on the first two principal components and color the points according to "hosp"
plot(PC1,PC2, col=viral34_c$hosp, main = "Hospitalization or No Hospitalization")
legend("bottomleft", col=1:2, legend=levels(viral34_c$hosp), pch=1, cex=0.7)
```




```
#There is no clear association between PCA projections and "hosp"

#Plotting the data on the first two principal components and color the points according to "ancestry"
#plot(PC1,PC2, col=viral34_c$ancestry, main = "A, B, or C")
#legend("bottomleft", col=factor(viral34_c$ancestry), legend=levels(viral34_c$ancestry), pch=1, cex=0.7
#There is no clear association between PCA projections and "ancestry"

#To make the visualization of results easier we use the function 'PCA' from FactoMineR package that all
library(FactoMineR)
pcaX<-PCA(X, scale.unit = TRUE, ncp = 3) # We  use the first 3 PC
```

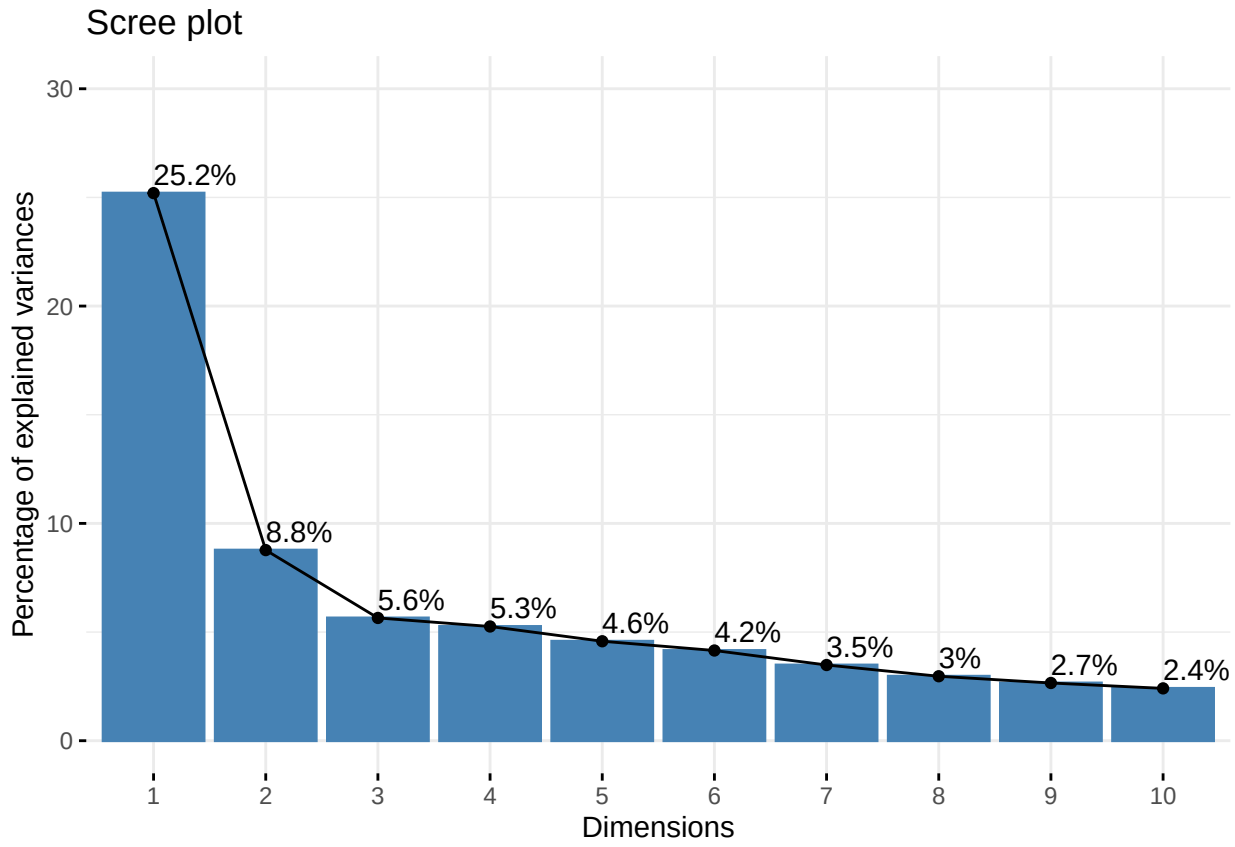

PCA graph of individuals
Dim 2 (8.77%)
Dim 1 (25.20%)
5
0
−5
−10
−5
0
5
10



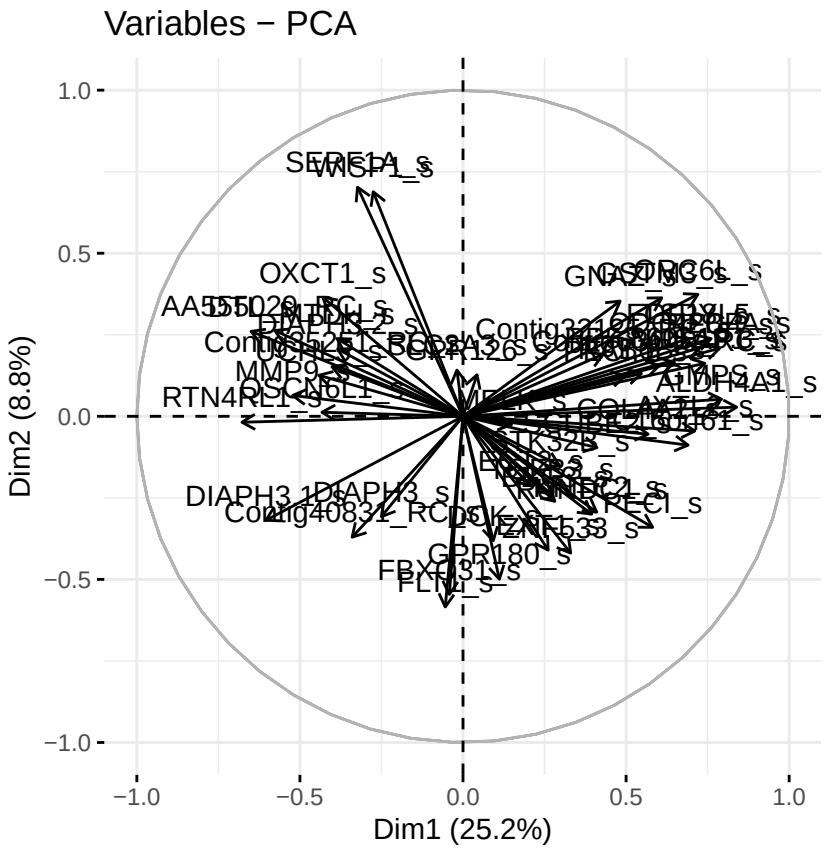

```
library("factoextra")
#PCA relies on eigenvalue decomposition of the covariance matrix. The following function provides the e
eig.val<- get_eigenvalue(pcaX)
head(eig.val)
```

```
##       eigenvalue variance.percent cumulative.variance.percent
## Dim.1     12.599           25.198                       25.20
## Dim.2      4.385            8.771                       33.97
## Dim.3      2.825            5.650                       39.62
## Dim.4      2.628            5.256                       44.87
## Dim.5      2.289            4.577                       49.45
## Dim.6      2.076            4.151                       53.60
```

```
#Results:


# Interpretation:
# The total variance of a matrix is the sum of the variances of the variables. When variables are scale
# of variables (57 in our case) since the variance of each (scaled) original variable is 1.The eigen va
# Thus, an eigen value larger than 1 indicates that PC accounts for more variance than the (scaled) ori
# The proportion of variance explained is obtained by dividing the variance (eigen value) by the total
# % variance = 100·eigen_value/total_variance. In our case: % variance explained by PC1 = 100·12.598858

#Variance explained plot:

fviz_eig(pcaX, addlabels = TRUE, ylim = c(0, 30))
```



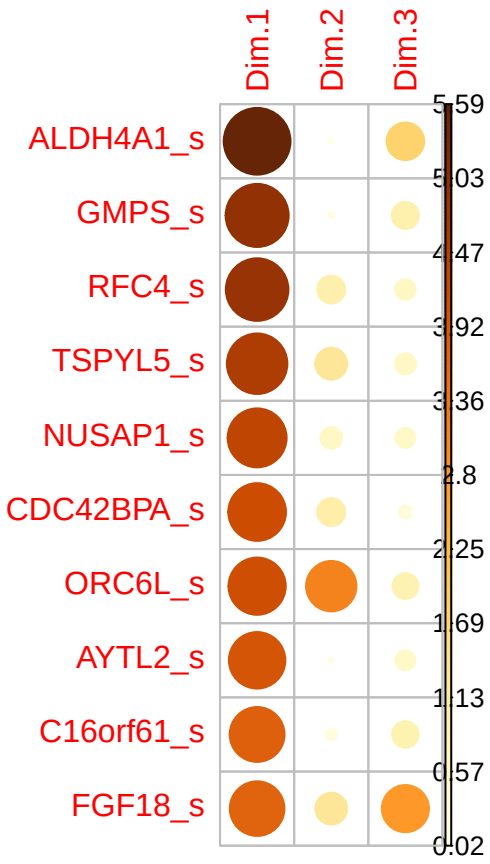

```
#Variables plot:
fviz_pca_var(pcaX, col.var = "black")
```



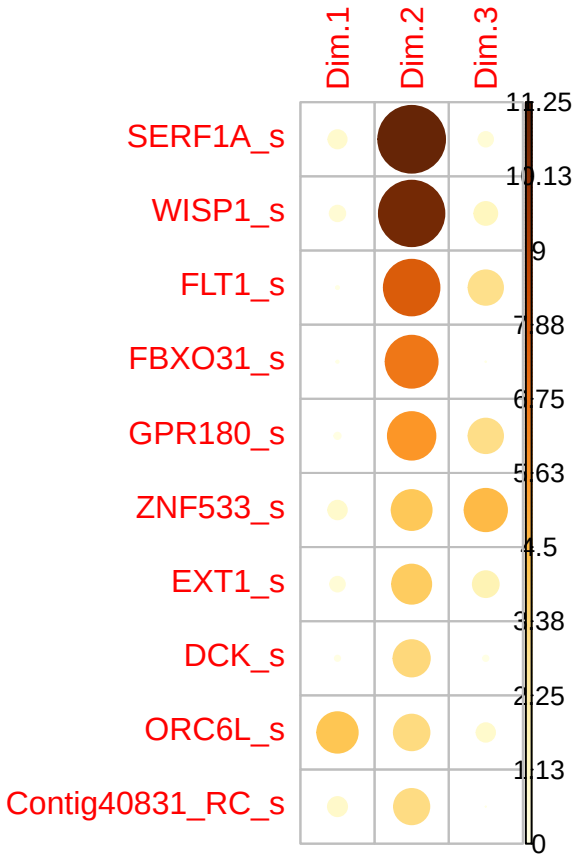

```
#The following function provides a list of matrices containing all the relevant information in a PCA, l
var <- get_pca_var(pcaX)
var
```

```
## Principal Component Analysis Results for variables
##  ===================================================
##   Name       Description
## 1 "$coord"   "Coordinates for the variables"
## 2 "$cor"     "Correlations between variables and dimensions"
## 3 "$cos2"    "Cos2 for the variables"
## 4 "$contrib" "contributions of the variables"
```

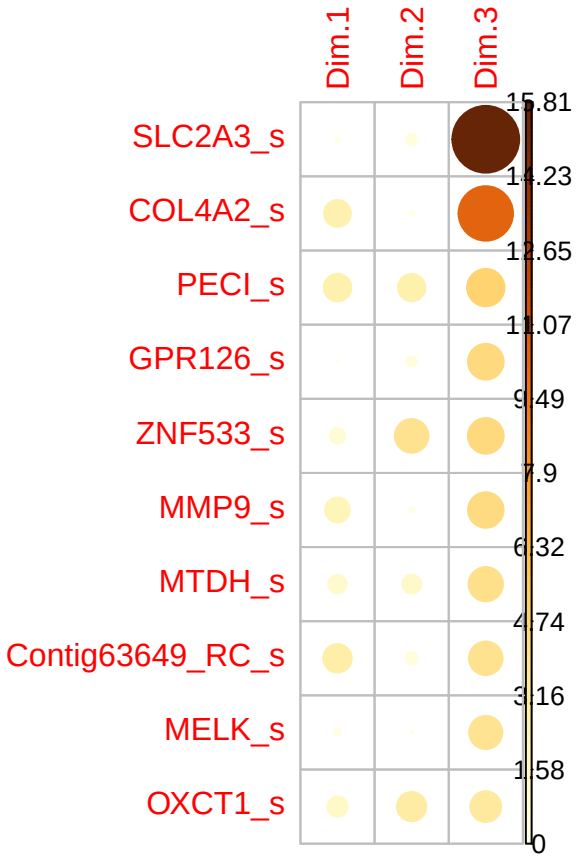
```
library(corrplot)
#We order the results according to the contribution value and restrict the 10 most important variables
corrplot(var$contrib[order(var$contrib[,1],decreasing = T)[1:10],], is.corr=FALSE)
```



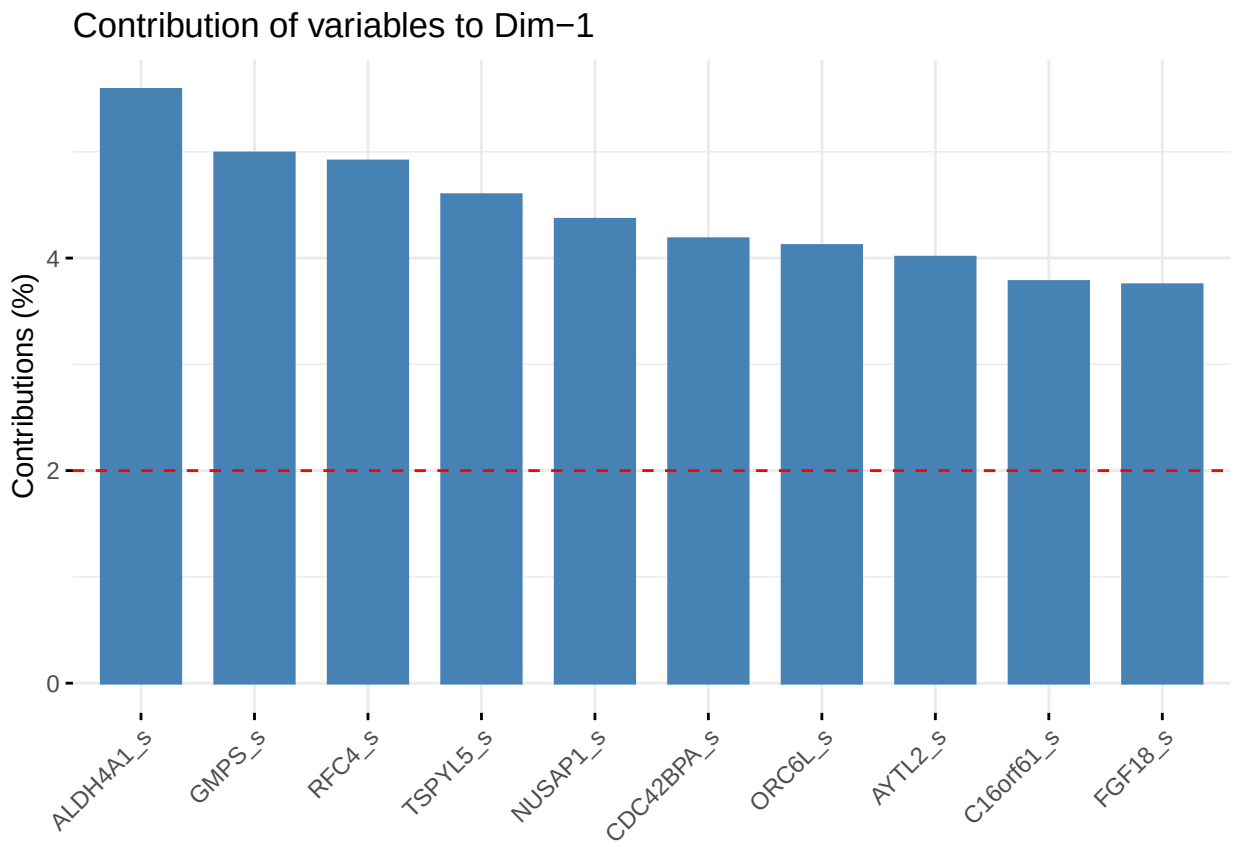

```
#Results: The 10 most important genes for PC1=ALDH4A1, GMPS, RFC4, TSPYL5, NUSAP1, CDC42BPA, ORC6L, AYT

#We order the results according to the contribution value and restrict the 10 most important variables
corrplot(var$contrib[order(var$contrib[,2],decreasing = T)[1:10],], is.corr=FALSE)
```



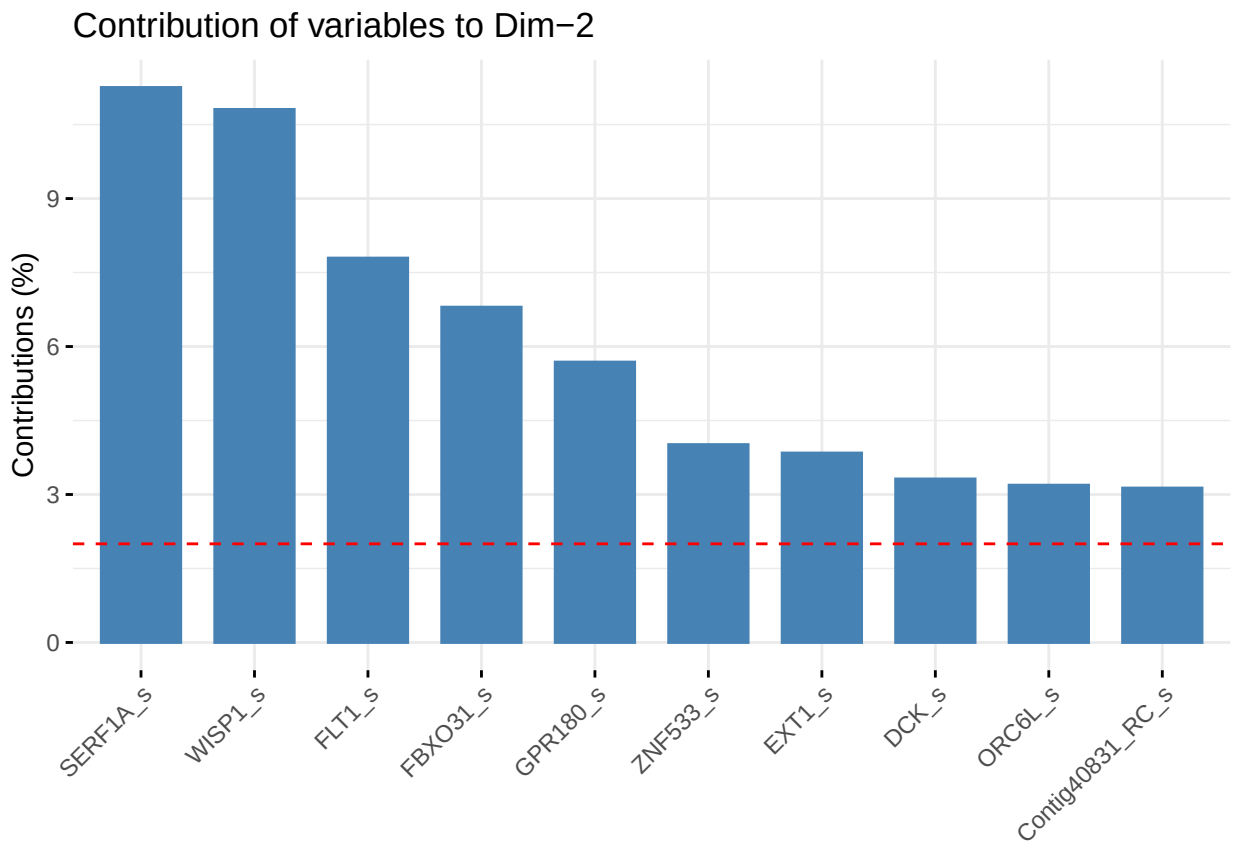

```
#The 10 most important genes for PC2= SERF1A, WISP1,FLT1,FBXo31,GPR180,ZNF533,EXT1,DCK,ORC6L, Contig408

#We order the results according to the contribution value and restrict the 10 most important variables
corrplot(var$contrib[order(var$contrib[,3],decreasing = T)[1:10],], is.corr=FALSE)
```



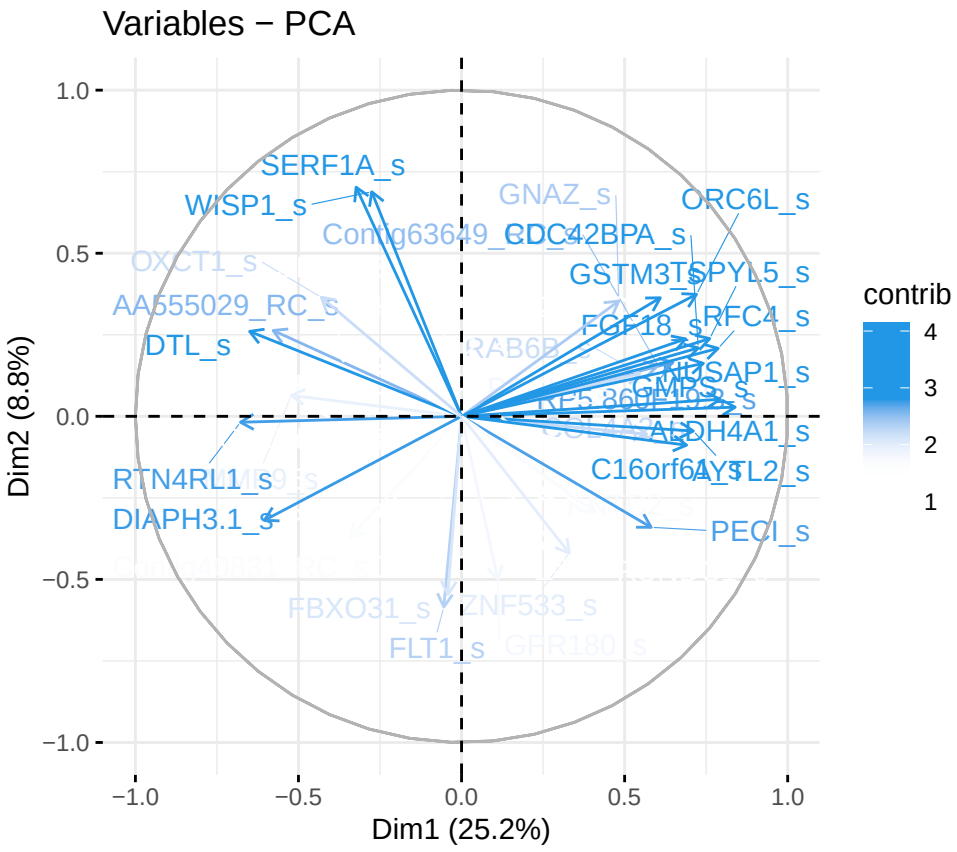

```
#The 10 most important genes for PC3= SLC2A3,COL4A2,PECI,GPR126,ZNF533,MMP9,MTDH,Contig63649_RC,MELK, O

#We can also plot the most important variables to PC1 as follows:
fviz_contrib(pcaX, choice = "var", axes = 1, top = 10)
```




```
#Contributions of variables to PC2
fviz_contrib(pcaX, choice = "var", axes = 2, top = 10)
```



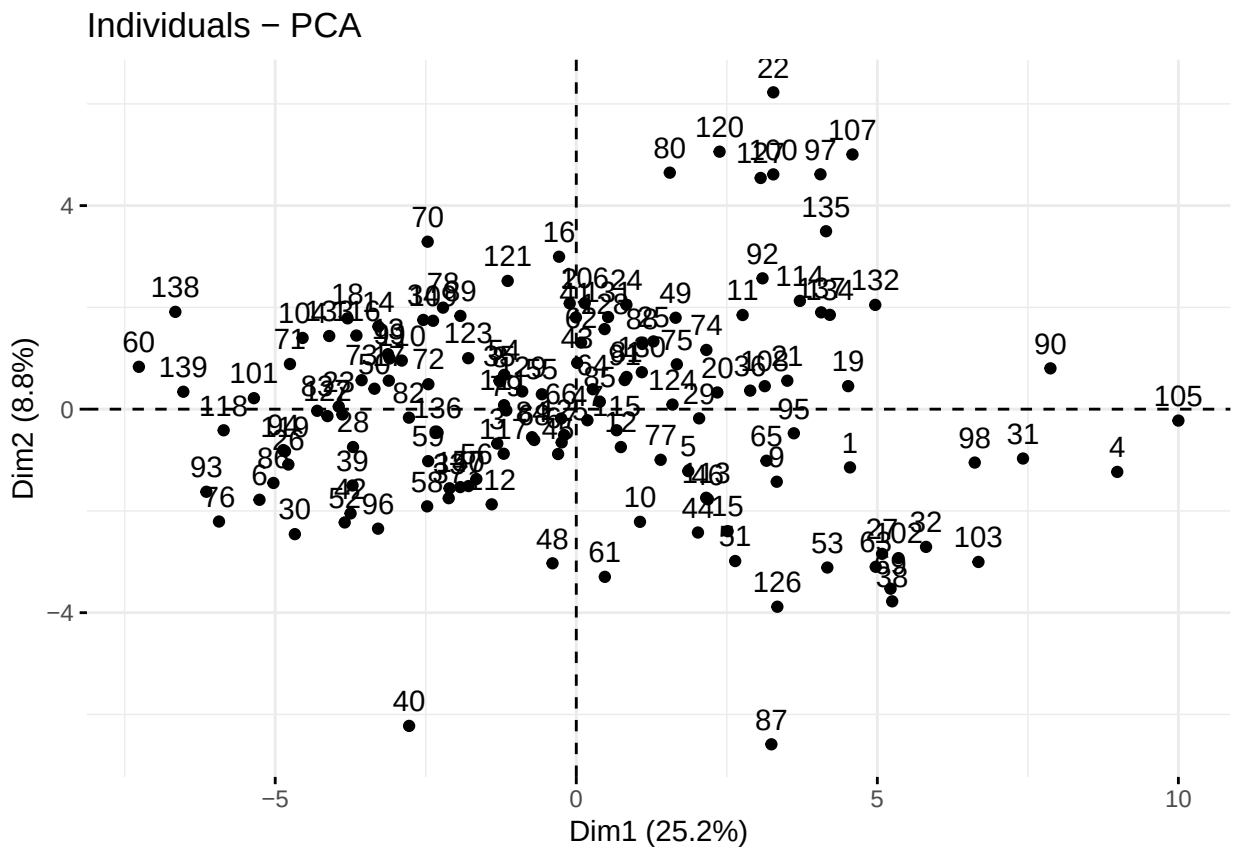

```
#Plot of the variables according to contrib values:
fviz_pca_var(pcaX, col.var = "contrib",
             gradient.cols = c(0,0,4,4),#c("#00AFBB", "#E7B800", "#FC4E07"),
             repel = TRUE # Avoid text overlapping
)
```



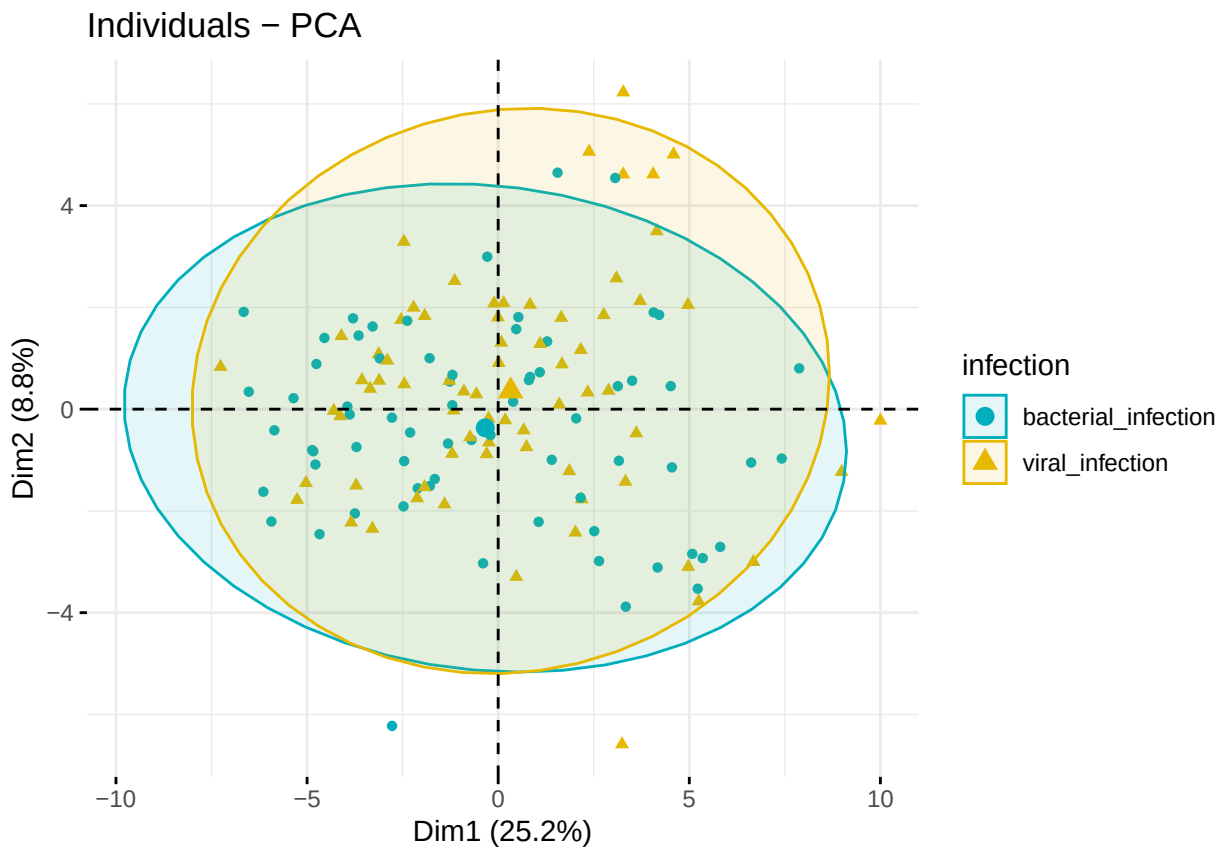

```
#Graphical representation of individuals
ind <- get_pca_ind(pcaX)
ind
```

```
## Principal Component Analysis Results for individuals
##  ===================================================
##   Name       Description
## 1 "$coord"   "Coordinates for the individuals"
## 2 "$cos2"    "Cos2 for the individuals"
## 3 "$contrib" "contributions of the individuals"
```

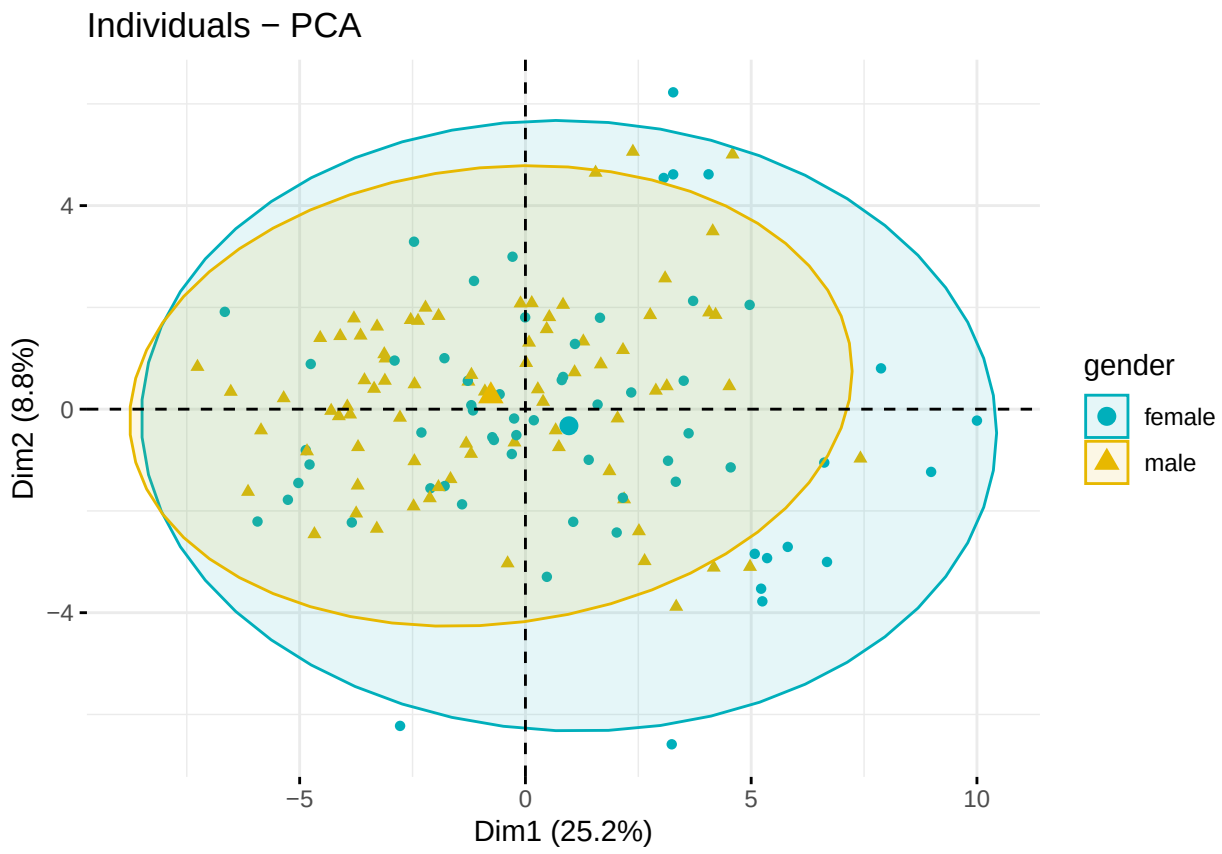
```
fviz_pca_ind(pcaX)
```



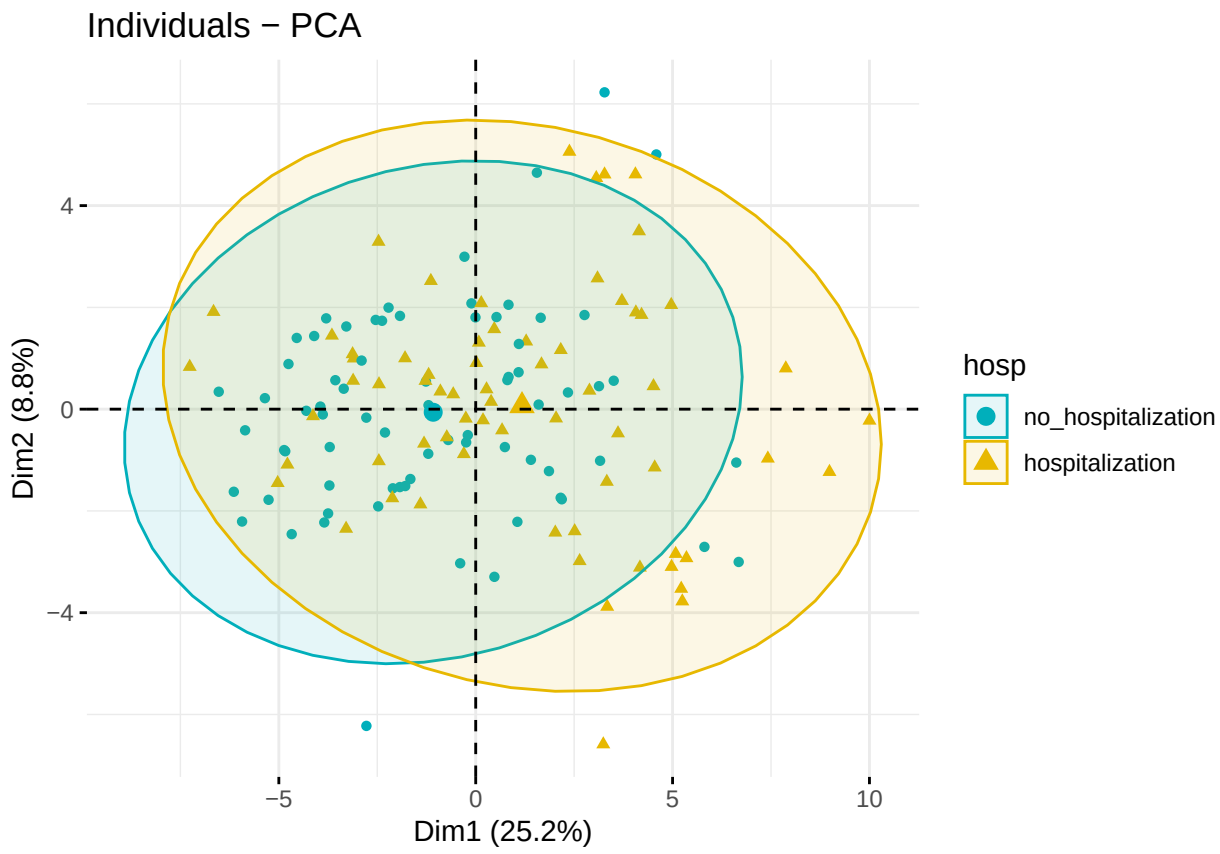

```
#We check whether the PCA projection of individuals is related to the "infection" variable by adding co
#We see that the two ellipses overlap which means no association between infection type and PCA project


# show points only (nbut not "text")
# color by groups NOT AS FACTOR!!!!!!!!!
# Concentration ellipses
fviz_pca_ind(pcaX,
             geom.ind = "point",
             col.ind = viral34_c$infection,
             palette = c("#00AFBB", "#E7B800", "#FC4E07"),
             addEllipses = TRUE,
             legend.title = "infection"
)
```



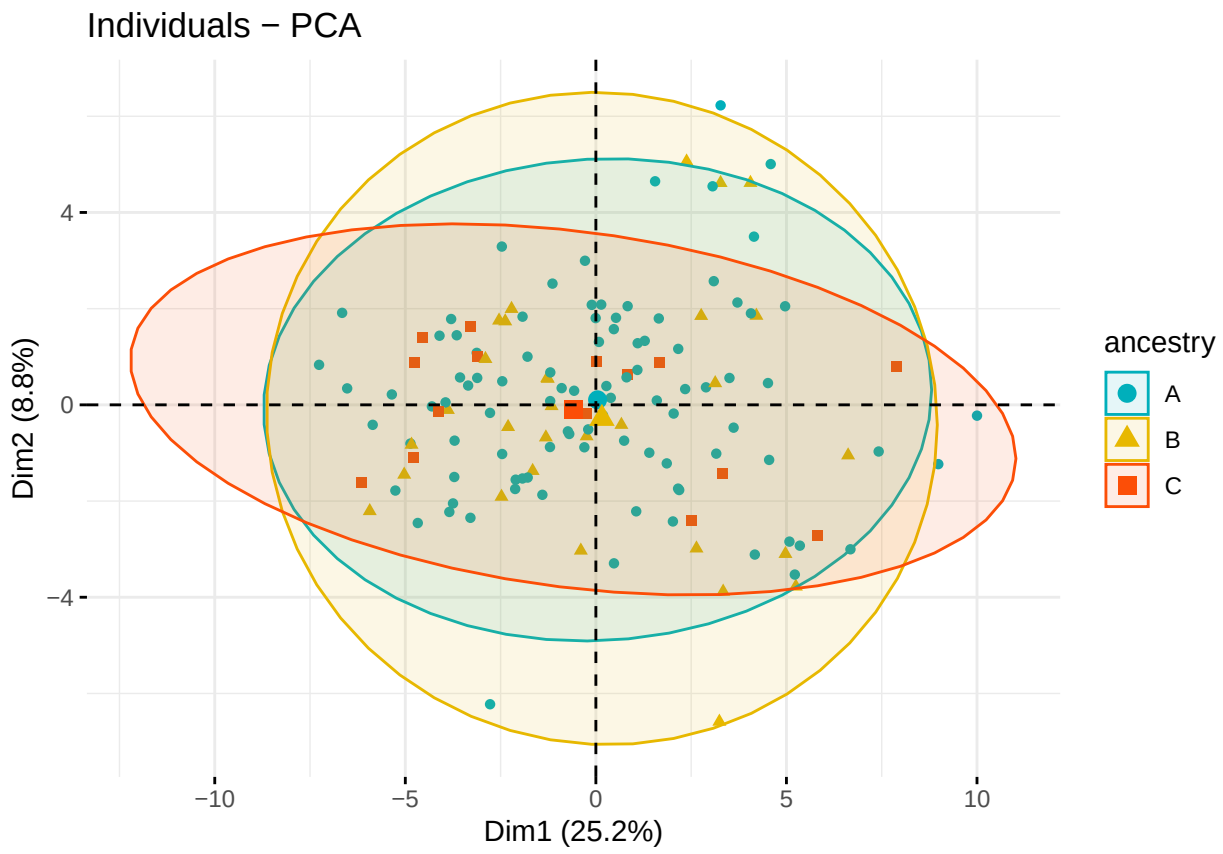

```
#Now we check whether the PCA projection of individuals is related to the "gender" variable by adding c
#by gender indicator. We see that the two ellipses are separated which implies different gene expressio
#positive and negative individuals. Large values of PC1 and small values of PC2 are related to gender n

# Concentration ellipses
# color by groups
# show points only (nbut not "text")
fviz_pca_ind(pcaX,
             geom.ind = "point",
             col.ind = viral34_c$gender,
             palette = c("#00AFBB", "#E7B800", "#FC4E07"),
             addEllipses = TRUE,
             legend.title = "gender"
)
```




```
#Now we check whether the PCA projection of individuals is related to the "hospitalization": Some separ
fviz_pca_ind(pcaX,
             geom.ind = "point", # show points only (nbut not "text")
             col.ind = viral34_c$hosp, # color by groups
             palette = c("#00AFBB", "#E7B800", "#FC4E07"),
             addEllipses = TRUE, # Concentration ellipses
             legend.title = "hosp"
)
```

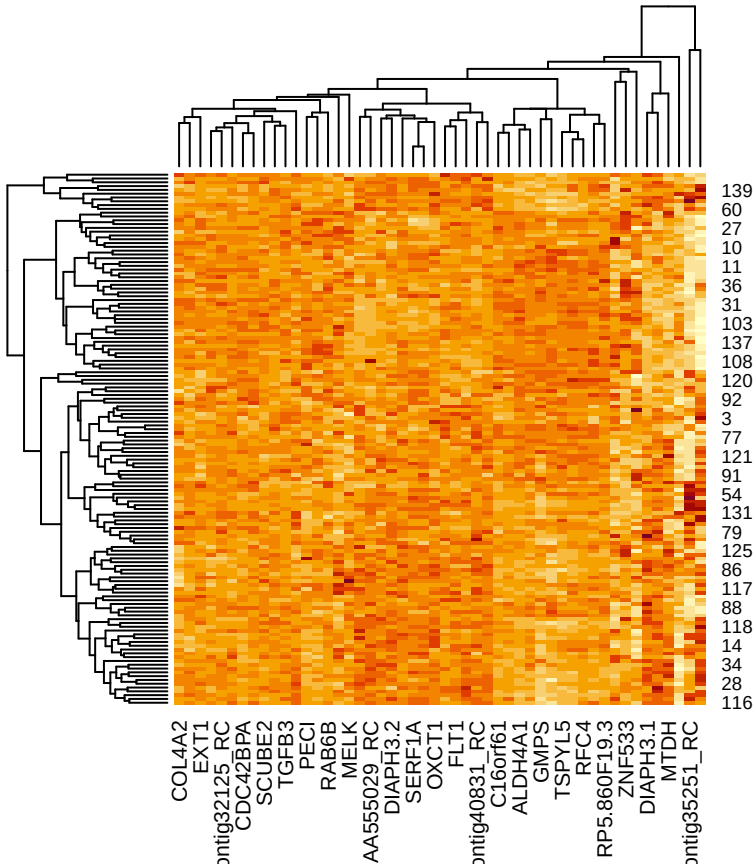



```
#Now we check whether the PCA projection of individuals is related to the "ancestry": Some separation
fviz_pca_ind(pcaX,
             geom.ind = "point", # show points only (nbut not "text")
             col.ind = viral34_c$ancestry, # color by groups
             palette = c("#00AFBB", "#E7B800", "#FC4E07"),
             addEllipses = TRUE, # Concentration ellipses
             legend.title = "ancestry"
)
```



QUESTION 6 Perform a nice heatmap with dendrograms for genes and individuals, individuals divided in two groups according to k-means (k=2), and annotations for infection and hospitalization (similar to the one proposed in section 1.4 in "Solutions Exercises section 2").

```
heatmap(as.matrix(viral34_c[,8:57])) # Using NON-SCALED gene expression levels
```

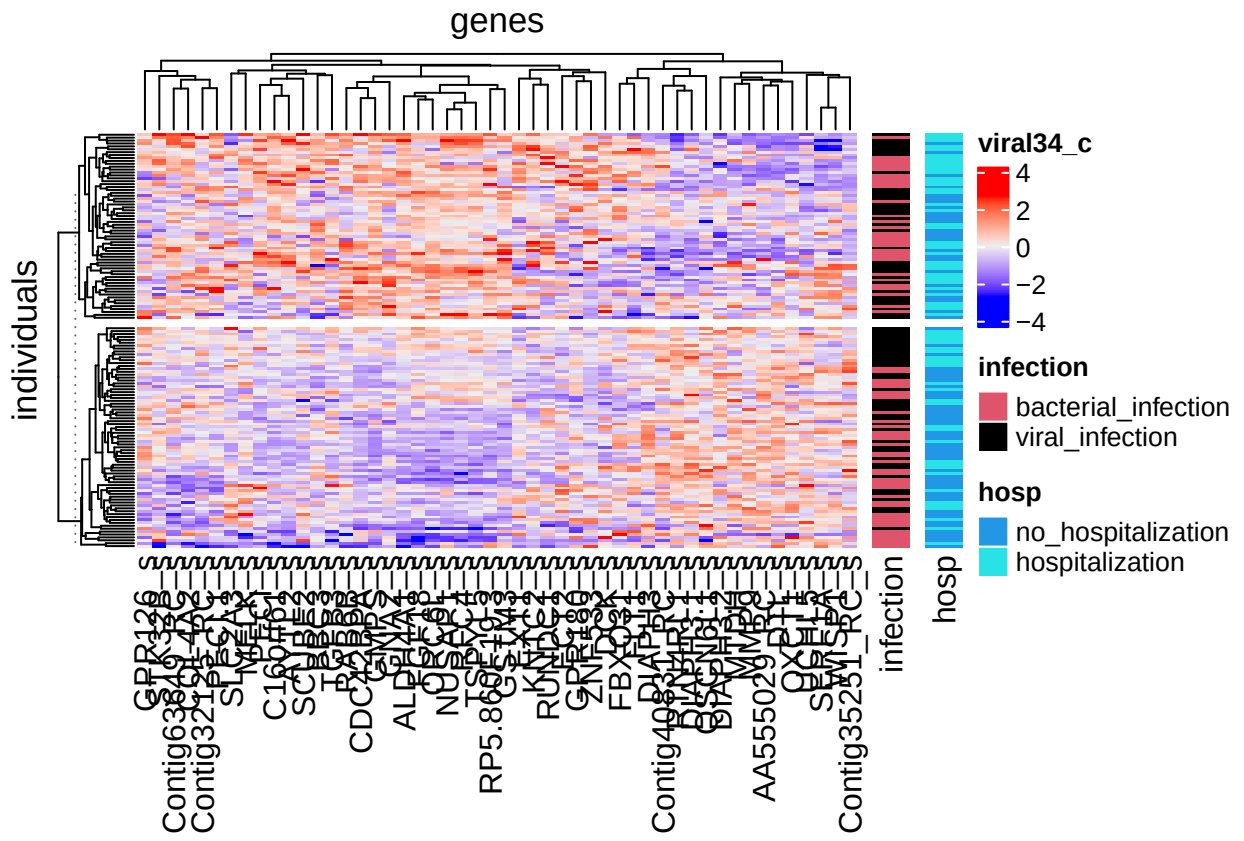



```
#Using instead 'Heatmap()' function from ComplexHeatmap package with examples at following website:
#https://www.datanovia.com/en/lessons/heatmap-in-r-static-and-interactive-visualization/

library(ComplexHeatmap)
```

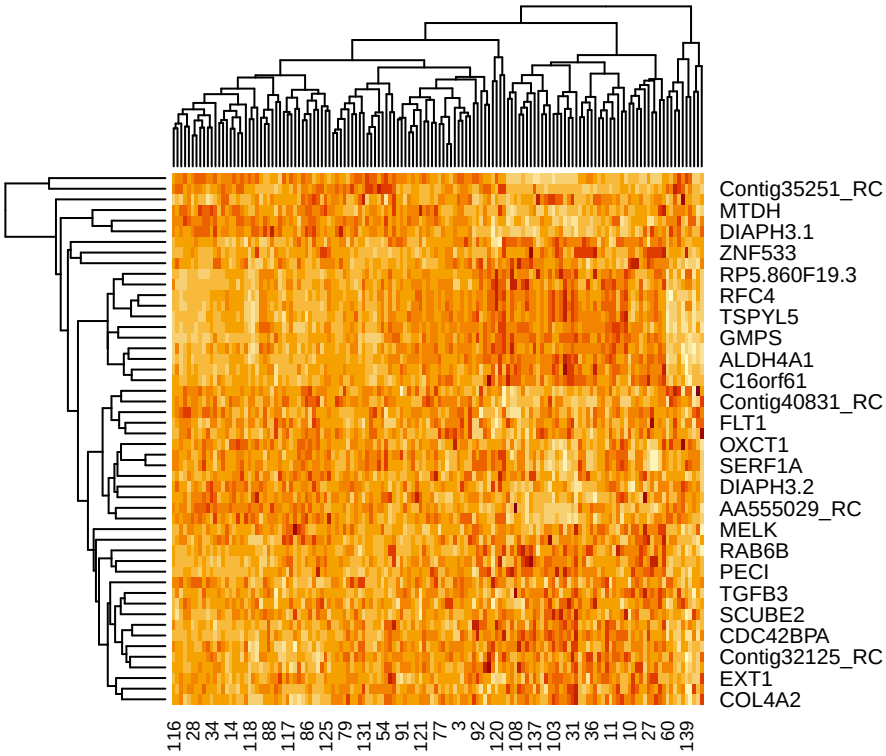
```
## ========================================
## ComplexHeatmap version 2.20.0
## Bioconductor page: http://bioconductor.org/packages/ComplexHeatmap/
## Github page: https://github.com/jokergoo/ComplexHeatmap
## Documentation: http://jokergoo.github.io/ComplexHeatmap-reference
##
## If you use it in published research, please cite either one:
## - Gu, Z. Complex Heatmap Visualization. iMeta 2022.
## - Gu, Z. Complex heatmaps reveal patterns and correlations in multidimensional
##     genomic data. Bioinformatics 2016.
##
##
## The new InteractiveComplexHeatmap package can directly export static
## complex heatmaps into an interactive Shiny app with zero effort. Have a try!
##
## This message can be suppressed by:
##   suppressPackageStartupMessages(library(ComplexHeatmap))
## ========================================
```

```
#Using already-scaled gene expression levels
# Text size for row names
#title of legend
# individuals are divided into 4 groups using Kmeans clustering
set.seed(1234)
Heatmap(viral34_c[,58:107],
        name = "viral34_c",
        column_title = "genes", row_title = "individuals",
        row_names_gp = gpar(fontsize = 7),
        km=2,
        show_row_names = FALSE, show_column_names = T
)+Heatmap(viral34_c$infection, name = "infection", width = unit(5, "mm"), col=c(2,1))+  Heatmap(viral34_
```

```
## Warning: The input is a data frame-like object, convert it to a matrix.
```




```
#Now repeating with transposition:
heatmap(as.matrix(t(viral34_c[,8:57]))) # Using NON-SCALED gene expression levels
```




```
#set.seed(1234)
#Heatmap(t(viral34_c[,58:107]),
#        name = "viral34_c",
#        column_title = "individuals", row_title = "genes",
#        row_names_gp = gpar(fontsize = 7),
#        km=2,
#        show_row_names = FALSE, show_column_names = T
#)+Heatmap(viral34_c$infection, name = "infection", width = unit(5, "mm"), col=c(2,1))+  Heatmap(viral3

#OBSERVATIONS FROM HEATMAP:
#I observed 4 gene clusters and 2 individual clusters with following observations:
#(LEGEND: GC=GENE CLUSTER, IC=INDIVIDUAL CLUSTER, Over=Over-expressed(red), Under=Underexpressed(blue))
#GC#1(Left of Dendogram): IC#1 (Left of Dendogram)=Over, IC#2=Under
#GC#2: IC#1=Under, IC#2=Over
#GC#3: IC#1=Over, IC#2=Under
#GC#4:For IC#1: DTL=Under, Contig35251_RC=Under, RUNDC1=Over, BBC3=Under, ECT2=Over, MTDH=Under, ZNF533
#GC#4: For IC#2: DTL=Over, (right CONTIG35251_RC=Over, Rest=Under), DCK=Under, MTDH=Over, BBC3=Under,ME
#IC#1: Left=Viral, Hospitalization; Middle=Bacterial, Right=Viral
#IC#2: (Right=Hospitalization, Viral Infection), Left= Bacterial, Middle: No Hospitalization

# You can see that the last cluster is strongly associated to negative ER and is characterized by low es
# on the right and high expression of the rest. The other three cluster mainly correspond to ER positiv
# We can observe that in the first cluster, that is characterized by high expression of the genes on th
# low expression of the rest, is the cluster with few events, i.e. this gene expression pattern is asso

library(glmnet)
```

```r
library(penalized)
```

```
## Welcome to penalized. For extended examples, see vignette("penalized").
```

```r
#EXAMINING BRIEFLY REALTIONSHIPS BETWEEN 10 genes:


# PerformanceAnalytics::chart.Correlation() #Argument R missing?
# corrr::network_plot() #Argument rdf missing?
# psych::pairs.panels() #Argument x missing?
# corrplot::corrplot.mixed() #Argument corr missing?
# GGally::ggpairs() #Argument data missing?
# ggcorrplot::ggcorrplot() #Argument corr missing?

library("PerformanceAnalytics")
```

```
## Loading required package: xts
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
##
## ############################ Warning from 'xts' package ##########################
## #                                                                                #
## # The dplyr lag() function breaks how base R's lag() function is supposed to  #
## # work, which breaks lag(my_xts). Calls to lag(my_xts) that you type or       #
## # source() into this session won't work correctly.                             #
## #                                                                                #
## # Use stats::lag() to make sure you're not using dplyr::lag(), or you can add #
## # conflictRules('dplyr', exclude = 'lag') to your .Rprofile to stop           #
## # dplyr from breaking base R's lag() function.                                 #
## #                                                                                #
## # Code in packages is not affected. It's protected by R's namespace mechanism #
## # Set `options(xts.warn_dplyr_breaks_lag = FALSE)` to suppress this warning.  #
## #                                                                                #
## ##################################################################################
```

```
##
## Attaching package: 'xts'
```

```
## The following objects are masked from 'package:dplyr':
##
##     first, last
```

```
##
## Attaching package: 'PerformanceAnalytics'
```

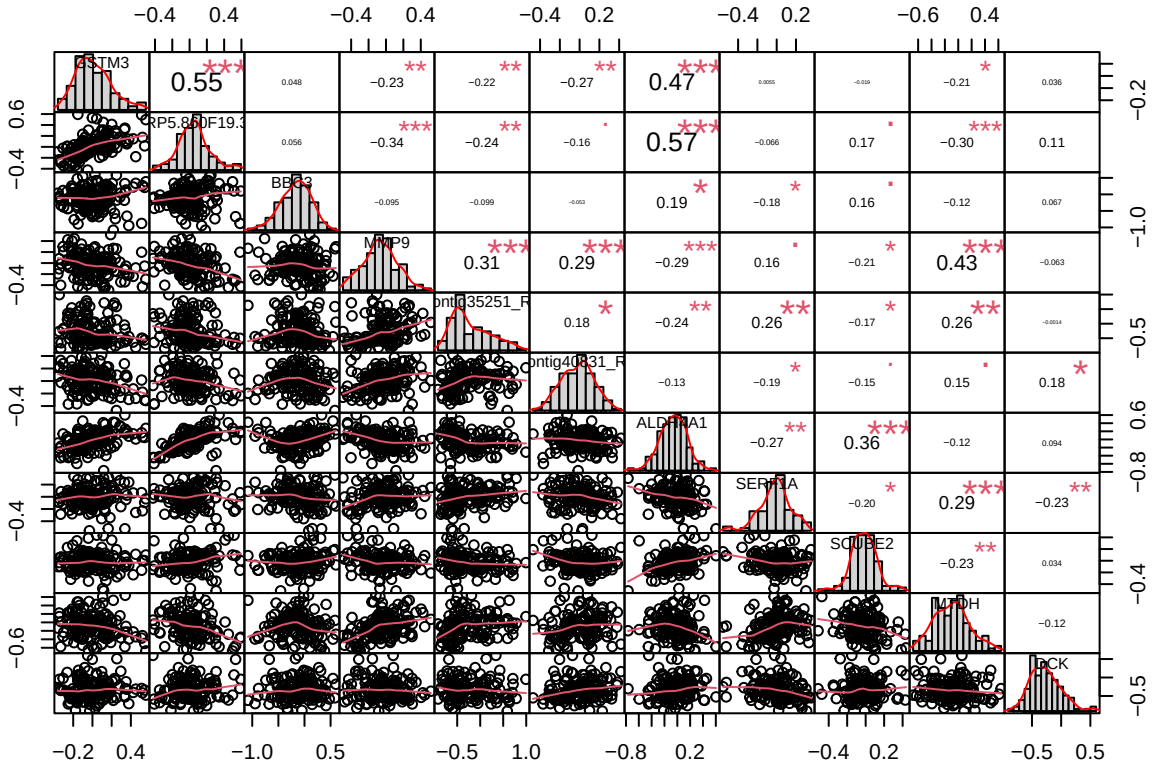
```
## The following object is masked from 'package:graphics':
##
##     legend
```

```r
chart.Correlation(viral34_c[,8:18], histogram=TRUE, pch=19)
```

```
## Warning in par(usr): argument 1 does not name a graphical parameter
```

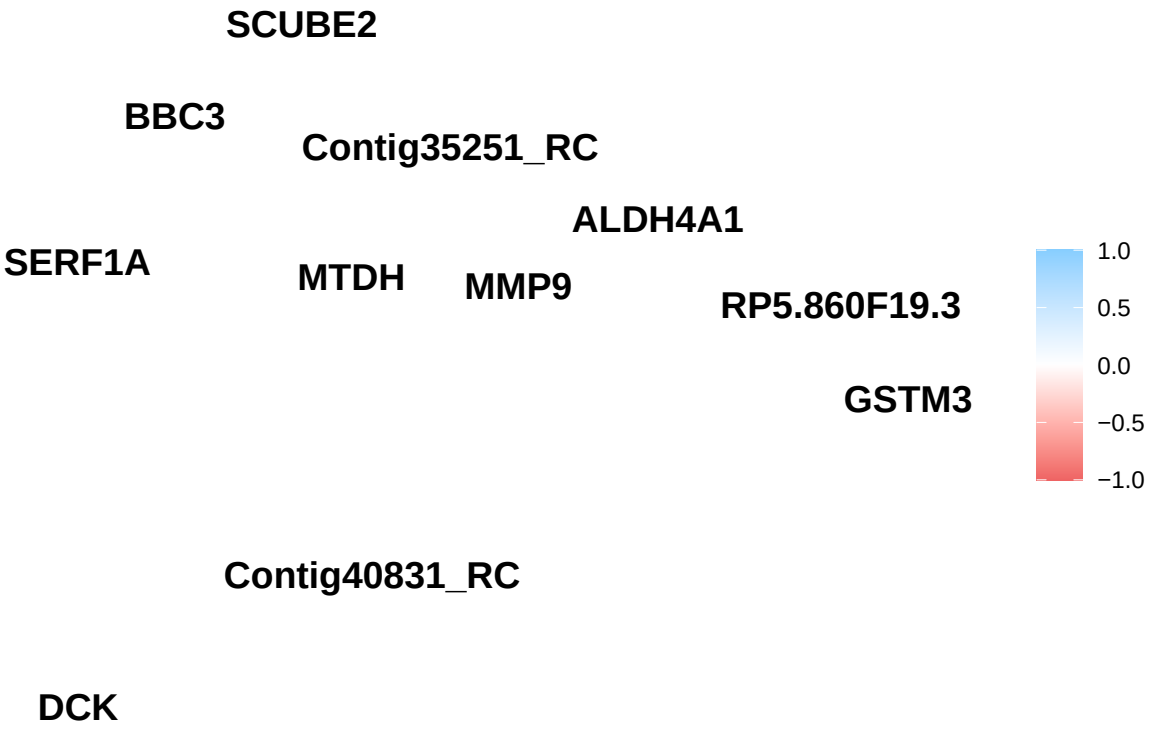
```
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
```

```
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
## Warning in par(usr): argument 1 does not name a graphical parameter
```



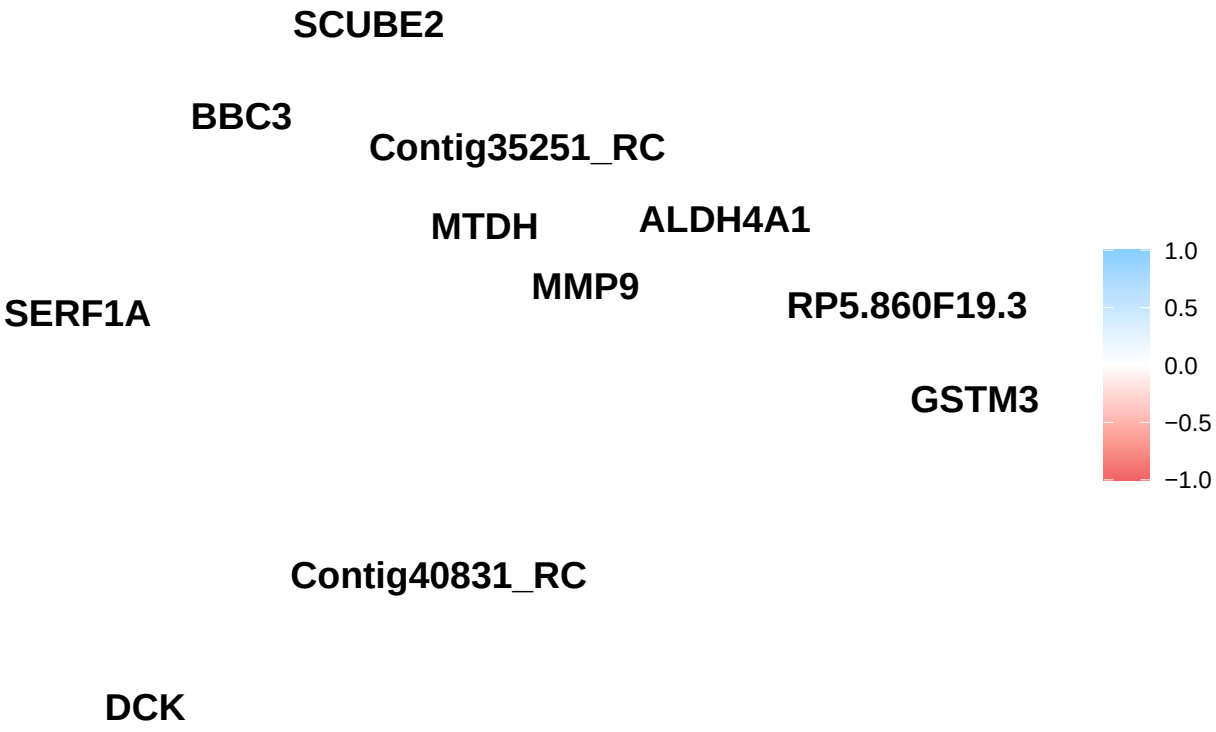

```
library(corrr)
network_plot(correlate(viral34_c[,8:18]), min_cor=0.6)
```

```
## Correlation computed with
## * Method: 'pearson'
## * Missing treated using: 'pairwise.complete.obs'
```




```
library(dplyr)
viral34_c[8:18] %>% correlate() %>% network_plot(min_cor=0.6)
```

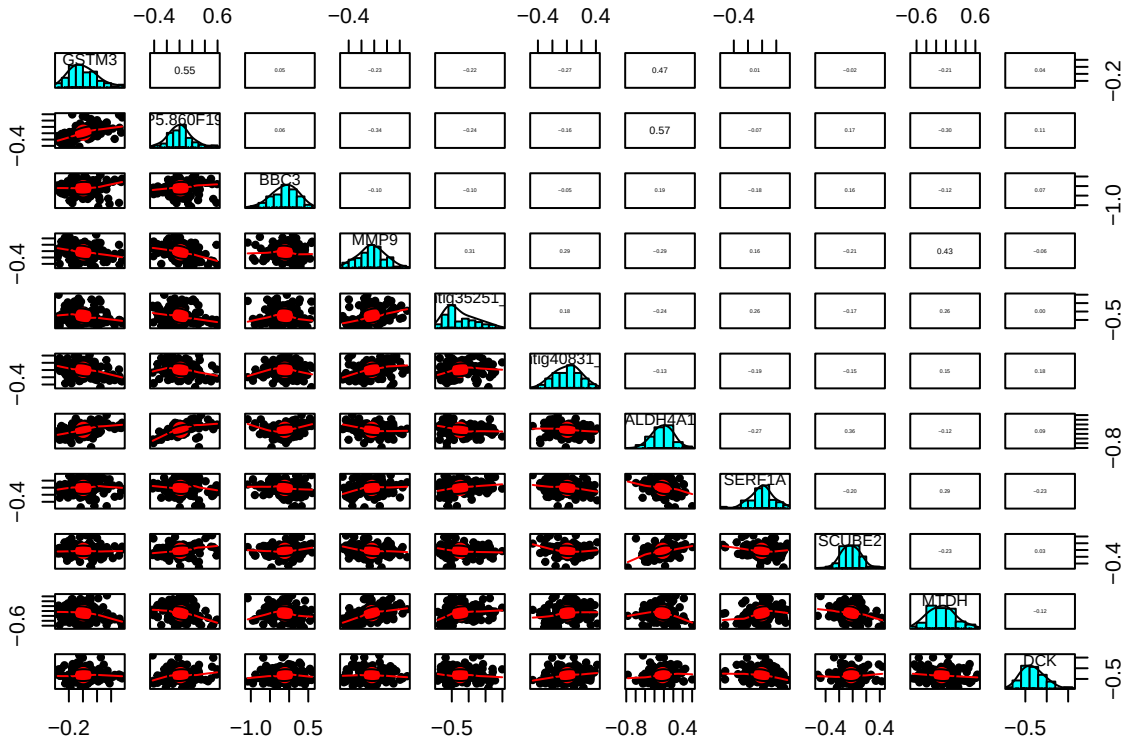
```
## Correlation computed with
## * Method: 'pearson'
## * Missing treated using: 'pairwise.complete.obs'
```



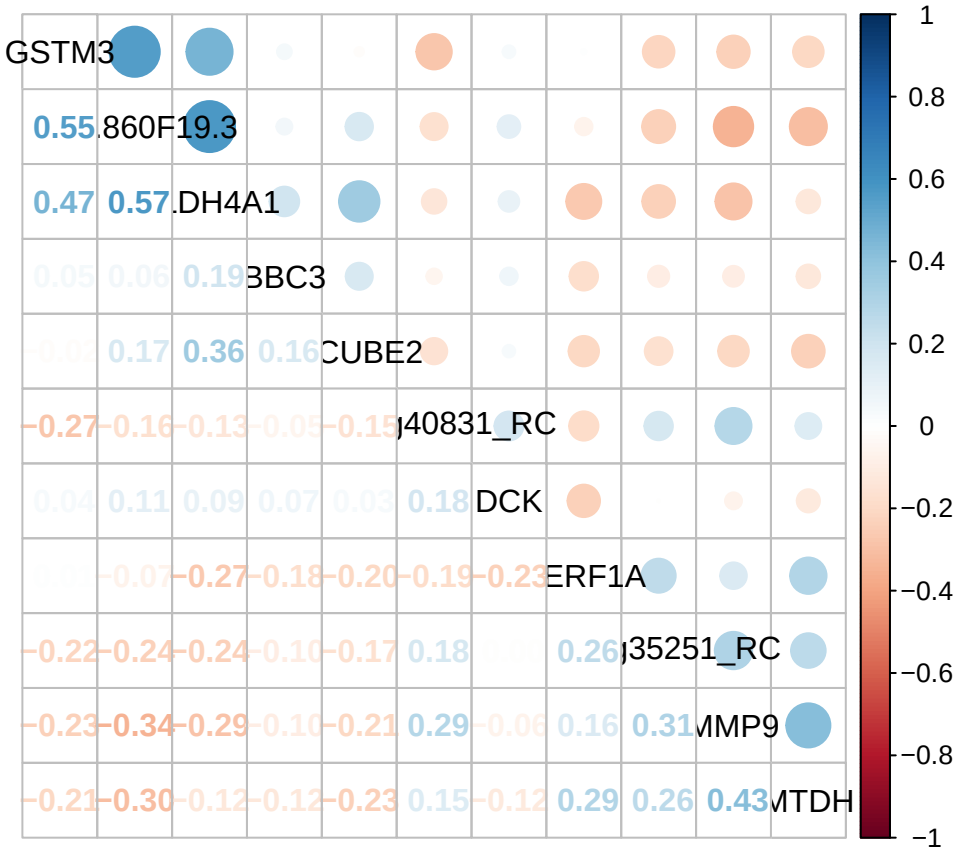

```
library(psych)
```

```
##
## Attaching package: 'psych'

## The following object is masked from 'package:outliers':
##
##     outlier

## The following object is masked from 'package:randomForest':
##
##     outlier

## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```
pairs.panels(viral34_c[8:18], scale=TRUE)
```



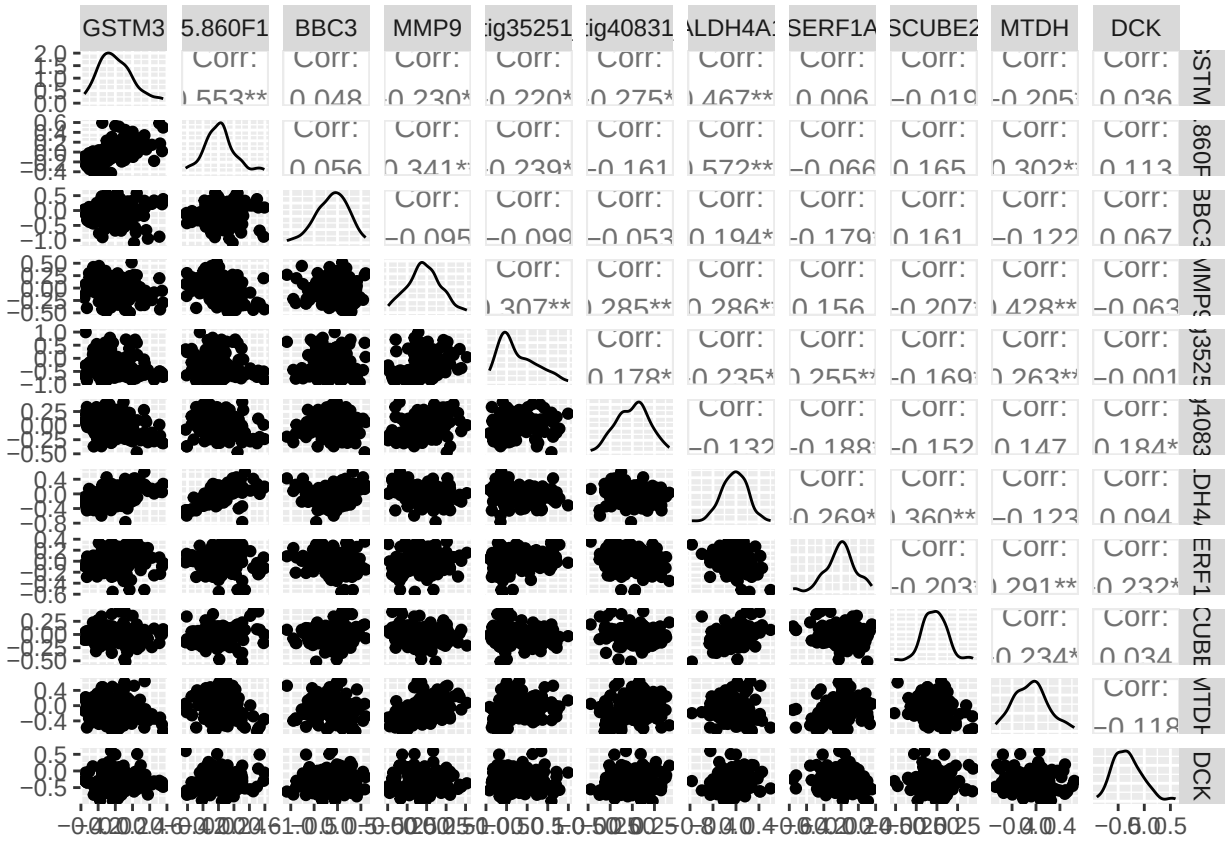

```
library(corrplot)
corrplot.mixed(cor(viral34_c[8:18]), order="hclust", tl.col="black")
```



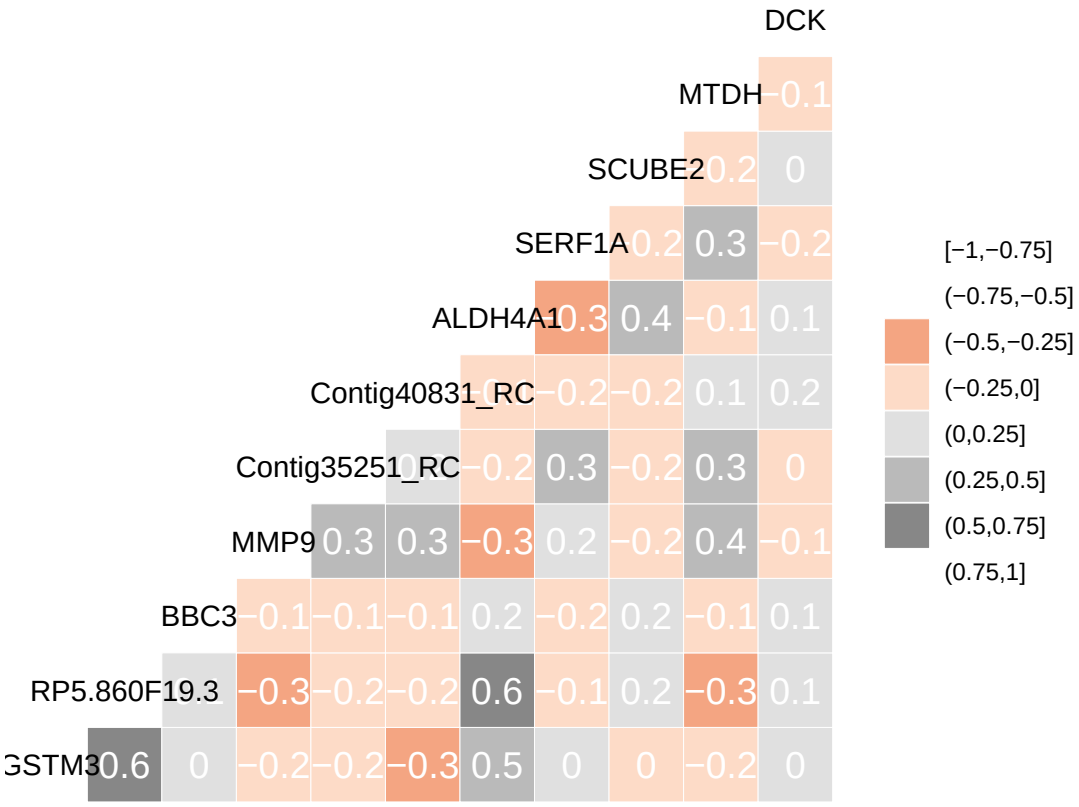

```
library(GGally)
```

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

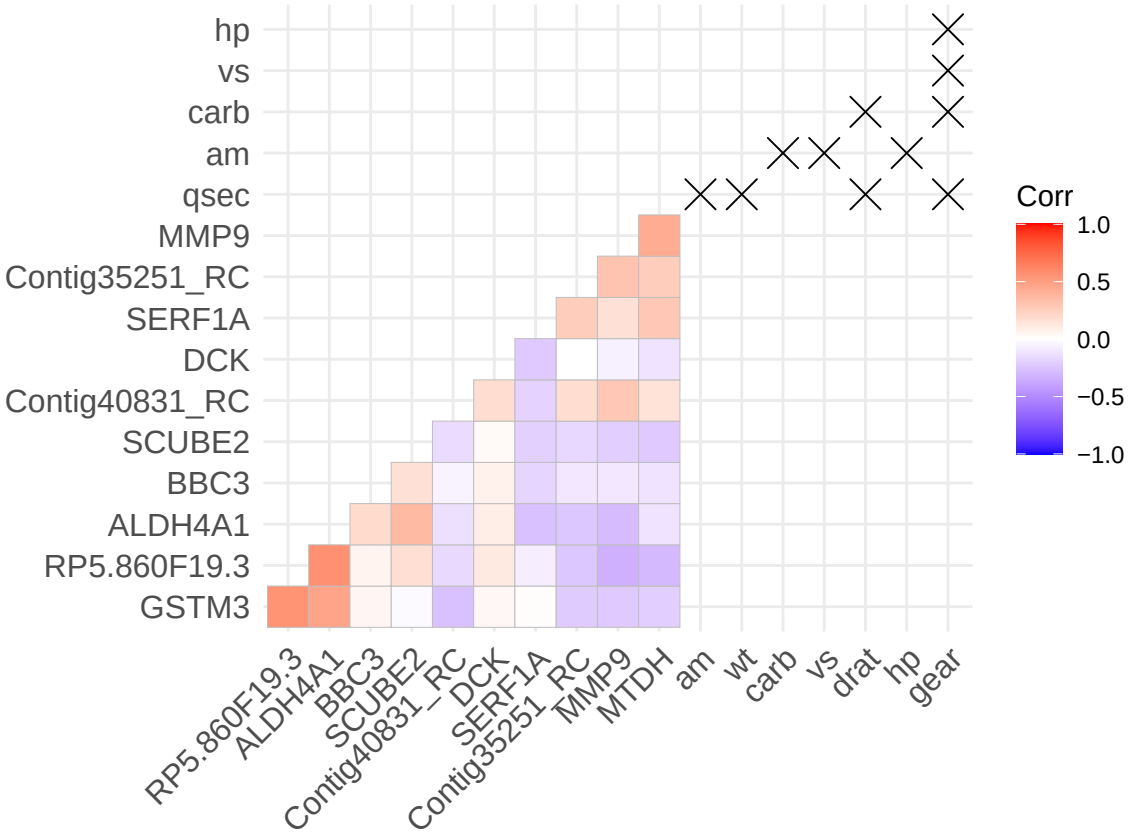
```
ggpairs(viral34_c[8:18])
```




```
ggcorr(viral34_c[8:18], nbreaks=8, palette='RdGy', label=TRUE, label_size=5, label_color='white')
```




```
library(ggcorrplot)
ggcorrplot(cor(viral34_c[8:18]), p.mat = cor_pmat(mtcars), hc.order=TRUE, type='lower')
```



QUESTION 7 Test if the mean expression levels of the first gene are different between viral and bacterial infections. An alpha=0.05 is assumed:

```
#Testing the following hypothesis:
#HO=Null Hypothesis: HO: u1=u2
#H1=Alternative Hypothesis: H1: u1!=u2

#Determine if gene GSTM3 (column#8) is a continuous, numerical variable, for which a mean and sd can be
#In general, scale variables for PCA but NOT for t-test, anova, etc, as these methods take advantage of
#of the data to implement the analysis

is.numeric(viral34_c$GSTM3)
```

```
## [1] TRUE
```

```
#Results show this is numeric (TRUE)
#Determine if non-scaled (raw) expression levels of first gene follows a normal distribution via shapir
#HO=Null Hypothesis: HO: X~N(u, sig)
#H1=Alternative Hypothesis: H1: X !~N(u, sig)

shapiro.test(viral34_c$GSTM3)
```

```
##
##  Shapiro-Wilk normality test
##
```

```
## data:  viral34_c$GSTM3
## W = 0.97, p-value = 0.004
```

```
##Because the p-value=0.00387<0.05, the non-scaled (raw) expression levels of first gene is not normall

#Determine if Z-normalized/standardized/scaled expression levels of first gene follows a normal distrib
#H0=Null Hypothesis: Ho: X~Nu, sig)
#H1=Alternative Hypothesis: H1: X !~N(u, sig)

shapiro.test(viral34_c$GSTM3_s)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  viral34_c$GSTM3_s
## W = 0.97, p-value = 0.004
```

```
#Because the p-value=0.00387<0.05, the Z-normalized/standardized/scaled expression levels of first gene

#Determine if log-transformed expression levels of first gene follows a normal distribution via shapiro

shapiro.test(viral34_c$GSTM3_l)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  viral34_c$GSTM3_l
## W = 0.98, p-value = 0.05
```

```
#Because the p-value=0.0463<=0.05 (JUST BARELY!), the log-transformed expression levels of first gene a

#Since neither GSTM, GSTM3_s or GSTM3_l are normally distributed, we apply a Wilcoxon test:
wilcox.test(viral34_c$GSTM3 ~ viral34_c$infection)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  viral34_c$GSTM3 by viral34_c$infection
## W = 2181, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
```

```
#Result: p-value=0.264>0.05

wilcox.test(viral34_c$GSTM3_s ~ viral34_c$infection)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  viral34_c$GSTM3_s by viral34_c$infection
## W = 2181, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
```

```r
#Result: p-value=0.264>0.05

wilcox.test(viral34_c$GSTM3_l ~ viral34_c$infection)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  viral34_c$GSTM3_l by viral34_c$infection
## W = 2181, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
```

```r
#Result: p-value=0.264>0.05

# We therefore DO NOT REJECT null hypothesis that means are equal. We conclude that there is NO statist
#suggest that the means of first gene are different between viral and bacterial infections are differen
```

QUESTION 8 Test if the mean expression levels of the first gene are different among ancestry groups. An alpha=0.05 is assumed:

```r
#Testing the following hypothesis:
#H0=Null Hypothesis: H0: u1=u2=u3 (equality of mean GSTM3 expression levels for all 3 ancestry types )
#H1=Alternative Hypothesis: H1: u1!=u2!=!u3 (at least 1 of  3 GSTM3 expression levels for all 3 ancestr
#Since GSTM3 gene expression levels  was found to not be normally distributed we apply a Kruskal-Wallis

kruskal.test(viral34_c$GSTM3 ~ viral34_c$ancestry)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  viral34_c$GSTM3 by viral34_c$ancestry
## Kruskal-Wallis chi-squared = 0.46, df = 2, p-value = 0.8
```

```r
#Results: p-value=0.7959>0.05
#Therefore, there is no statistically significant evidence that the mean GSTM3 expression levels are di
#WE do not reject the null hypothesis that these means are equal.

# However, ONLY for demonstration and to be comprehensive here, let's ASSUME that the (log-transformed)
#distributed (as the p-value from shapiro.test was almost > 0.05). Therefore, in this case, we would pe
#1-WAY ANOVA test as follows:One-factor ANOVAis as test for association between a continuous variable Y
#from the decomposition of the total variability in two components: the variability between groups and
#distribution.

#First we test whether the variances are equal (homoscedasticity)

library(lmtest)
bptest(lm(viral34_c$GSTM3_l ~ viral34_c$ancestry),studentize = F)
```

```
##
##  Breusch-Pagan test
##
## data:  lm(viral34_c$GSTM3_l ~ viral34_c$ancestry)
## BP = 4.4, df = 2, p-value = 0.1
```

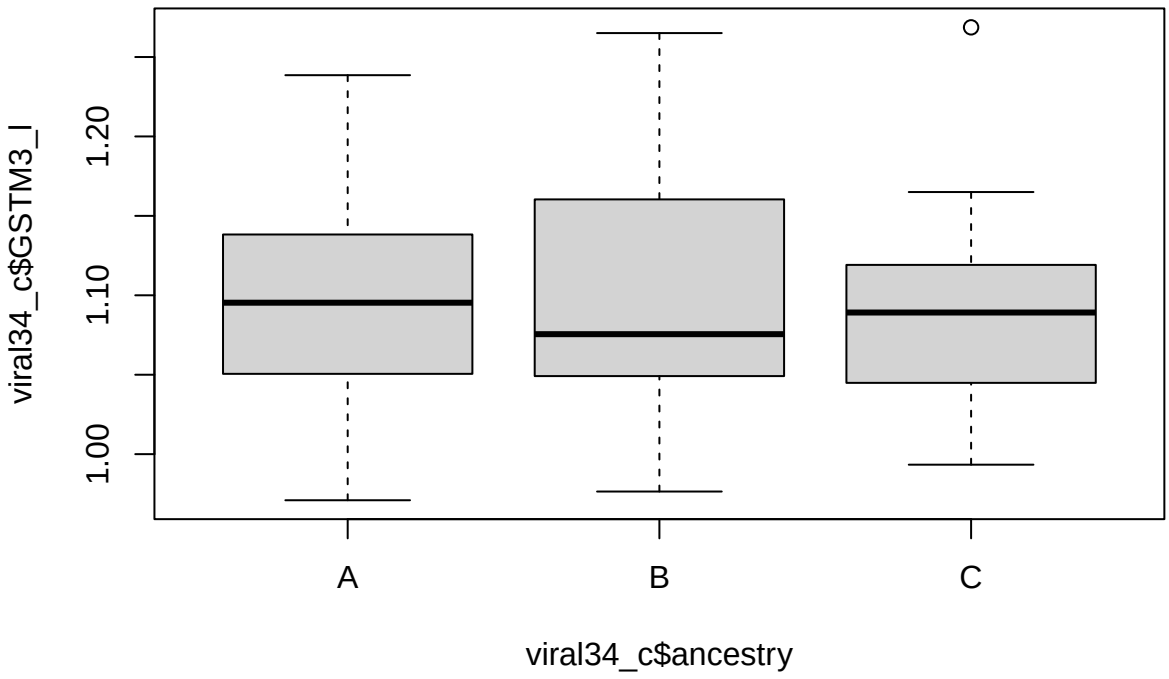
```r
#Results: p-value=0.1116>0.05. Therefore, homoscesaticity is fulfilled.

#Then, we perform one-factor ANOVA:

summary(aov(viral34_c$GSTM3_l ~ viral34_c$ancestry))
```

```
##                      Df Sum Sq Mean Sq F value Pr(>F)
## viral34_c$ancestry    2  0.001 0.00027    0.06   0.94
## Residuals           137  0.584 0.00426
```

```r
#Results: Because Pr(>F)=0.938>0.05, we conclude that there is statistically no significant
#difference in population mean GSTM3 gene expression levels between 3 ancestral groups A, B, C

#The higher the F-value, the lower the corresponding p-value. With p-value > threshold (e.g. alpha= . 0
#we cannot reject the null hypothesis of the ANOVA and cannot conclude that there is a statistically si
#difference between ancestry group means.

#We can follow-up by
tapply(viral34_c$GSTM3_l,viral34_c$ancestry, mean)
```

```
##     A     B     C
## 1.099 1.100 1.093
```

```r
#Also:
boxplot(viral34_c$GSTM3_l~viral34_c$ancestry)
```



```
#These confirm that the expression means among ancestries are similar

#If we had rejected and found means to be different, then TukeyHSD(anova1factor) can be used:
```

QUESTION 9 Test whether mean expression levels of the first and second genes are equal for viral infections.An alpha=0.05 is assumed:

```
##This is paired data since there is a pair of values (gene 1 and gene2) for each individual. Because e
#shown that Gene 1 expression values are not normally distributed, I will use the wilcox test for testi
#means for paired, non-normally-distributed data as alternative to to the t-test for equality of means
#Hypothesis:
#H0: distribution A = Distribution B
#H1: distribution A = Distribution B

#Preliminarily, I tested to see if non-scaled Gene #2 expression values were also normally distributed:

shapiro.test(viral34_c$RP5.860F19.3)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  viral34_c$RP5.860F19.3
## W = 0.98, p-value = 0.02
```

```
#Results: Because, p-value=0.01577<=0.01577, we conclude that Gene#2 Expression values are also not nor

#Furthermore, I tested for correlation between the 2 raw (non-scaled) gene extression levels
#Hypothesis:
#Ho: rho = 0 Gene1 and Gene2 are uncorrelated
#H1: rho not equal 0 Gene1 and Gene2 are correlated

#Using Spearman correlation for non-normally distributed observations

cor(viral34_c$GSTM3,viral34_c$RP5.860F19.3, method="spearman")
```

```
## [1] 0.6124
```

```
cor.test(viral34_c$GSTM3,viral34_c$RP5.860F19.3,method="spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  viral34_c$GSTM3 and viral34_c$RP5.860F19.3
## S = 177236, p-value <2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##    rho
## 0.6124
```

```
#Results indicated a p-value=2.2E-16<=0.05 and also a high correlation coefficient of 0.612, suggesting
#both genes are correlated and allowing us to reject null hypothesis that they are uncorrelated

#Finally, I tested for equality of Gene1 mean expression values to Gene mean expression vales for viral
#Because I am only evaluating mean gene 1 and 2 expression values for viral infection, I first created
#contain these 2 columns: #slice(), select(1:3)

library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v stringr   1.5.1
## v lubridate 1.9.3     v tibble    3.2.1
## v purrr     1.0.2     v tidyr     1.3.1
## v readr     2.1.5
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x psych::%+%()             masks ggplot2::%+%()
## x psych::alpha()           masks ggplot2::alpha()
## x randomForest::combine()  masks Biobase::combine(), BiocGenerics::combine(), dplyr::combine()
## x tidyr::expand()          masks Matrix::expand()
## x dplyr::filter()          masks stats::filter()
## x xts::first()             masks dplyr::first()
## x dplyr::lag()             masks stats::lag()
## x xts::last()              masks dplyr::last()
## x purrr::lift()            masks caret::lift()
## x randomForest::margin()   masks ggplot2::margin()
## x tidyr::pack()            masks Matrix::pack()
## x BiocGenerics::Position() masks ggplot2::Position(), base::Position()
## x MASS::select()           masks dplyr::select()
## x tidyr::unpack()          masks Matrix::unpack()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
subviral34 <- viral34_c %>% filter(infection == "viral_infection")

#Now, I test whether non-scaled mean expression levels of the first and second genes are equal for viral
#Wilcoxon rank test for the equality of two means for paired data:

wilcox.test(subviral34$GSTM3,subviral34$RP5.860F19.3,paired=T)
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  subviral34$GSTM3 and subviral34$RP5.860F19.3
## V = 1148, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
```

```
#Results show that because p-value=0.4581>0.05, we cannot reject the null hypothesis that the
#population mean expression levels of the first and second genes are equal. There is not enough statist
#evidence to suggest that means are different.
```

QUESTION 10 Perform a nonparametric test for association of the kind of infection (viral or bacterial) and the risk of hospitalization.Provide the OR of the risk of hospitalization for viral vs bacterial infections.

```r
#Using females as reference group based on default alphabetical order and to yield an OR>1:

library("epitools")
```

```
##
## Attaching package: 'epitools'
```

```
## The following object is masked from 'package:survival':
##
##     ratetable
```

```r
# var1=risk factor, var2=hospitalization_status
table<-table(viral34_c$infection, viral34_c$hosp)
table
```

```
##
##                       no_hospitalization hospitalization
##   bacterial_infection                 43              26
##   viral_infection                     30              41
```

```r
# OR of having the disease(Y=1) for the category in the
oddsratio(table)
```

```
## $data
##
##                       no_hospitalization hospitalization Total
##   bacterial_infection                 43              26    69
##   viral_infection                     30              41    71
##   Total                               73              67   140
##
## $measure
##                      odds ratio with 95% C.I.
##                       estimate lower upper
##   bacterial_infection    1.000    NA    NA
##   viral_infection        2.242 1.142 4.473
##
## $p.value
##                      two-sided
##                       midp.exact fisher.exact chi.square
##   bacterial_infection         NA           NA         NA
##   viral_infection         0.0188      0.01904     0.0175
##
## $correction
## [1] FALSE
##
## attr(,"method")
## [1] "median-unbiased estimate & mid-p exact CI"
```

```r
# 2nd row with respect to reference group (1rst row)

#Based on results, the Odds Ratio (OR) = 2.242376 suggests that viral infections are 2.24X higher risk
#to bacterial infection.
```

QUESTION 11 Test the normality of expression levels of the 50 genes (use function apply). How many genes are not normally distributed and which are their names?

```
pvector<-apply(viral34_c[,8:57], 2, function(x) shapiro.test(x)$p.value)
pvector
```

```
##          GSTM3   RP5.860F19.3           BBC3           MMP9 Contig35251_RC
##      3.870e-03      1.577e-02      3.992e-01      6.833e-01      8.627e-07
## Contig40831_RC        ALDH4A1         SERF1A         SCUBE2           MTDH
##      6.319e-01      7.705e-01      2.491e-02      2.754e-02      3.443e-01
##            DCK           FLT1         PECI.1        QSCN6L1         DIAPH3
##      7.808e-03      4.557e-01      1.592e-01      2.166e-03      8.208e-02
##         SLC2A3         GPR180        RTN4RL1 Contig32125_RC         STK32B
##      5.022e-02      8.169e-02      2.064e-03      8.139e-01      3.931e-01
##           EXT1         COL4A2           PECI           GNAZ          AYTL2
##      8.380e-01      9.376e-01      3.960e-03      7.872e-04      6.778e-04
## Contig63649_RC          RAB6B    AA555029_RC         GPR126           ECT2
##      3.678e-01      5.207e-04      1.692e-03      5.015e-01      1.502e-02
##         NUSAP1           GMPS          UCHL5          ORC6L         TSPYL5
##      9.286e-01      1.914e-02      5.756e-01      1.224e-01      9.740e-01
##           MELK         RUNDC1       DIAPH3.1       C16orf61          TGFB3
##      1.190e-01      1.782e-01      8.669e-01      4.056e-02      5.729e-01
##          FGF18       CDC42BPA            DTL          WISP1       DIAPH3.2
##      1.450e-01      1.975e-01      2.683e-03      3.204e-01      1.583e-01
##          OXCT1         ZNF533           RFC4          KNTC2         FBXO31
##      3.130e-01      7.075e-09      9.092e-02      7.851e-05      2.450e-01
```

```
#Counting the genes that are not normally distributed:
length(which(pvector<=0.05))
```

```
## [1] 19
```

```
#Based on results, there are 19 genes that are not normally distributed

#Getting names of genes that are not normally distributed:
non_normal_genes<-(names(which(pvector<=0.05)))
non_normal_genes
```

```
##  [1] "GSTM3"          "RP5.860F19.3"   "Contig35251_RC" "SERF1A"
##  [5] "SCUBE2"         "DCK"            "QSCN6L1"        "RTN4RL1"
##  [9] "PECI"           "GNAZ"           "AYTL2"          "RAB6B"
## [13] "AA555029_RC"    "ECT2"           "GMPS"           "C16orf61"
## [17] "DTL"            "ZNF533"         "KNTC2"
```

QUESTION 12 Identify those genes that are differentially expressed between viral and bacterial infections (use function apply). Create a function that checks whether the gene expression levels are normally distributed or not and, accordingly, applies the most appropriate test for comparing gene expression levels between viral and bacterial infections. Adjust the p-values for multiple testing according to an fdr threshold equal to 0.1. Interpret the results.

```r
#Function will ultimately perform a statistical test for the equality of two means (gene expression lev
# I tried this alternave with "apply" but was not successful:
#pvector<-apply(viral34_c[,8:57], 2, function(x) shapiro.test(x)$p.value)


diffex<-function(df, alpha){
  #Initialize empty variables that will contain the pvalues of the k different t-tests
  pfinal<-NULL
  pnormal<-vector()
  pvar<-c()
  pval<-vector()
  #pval<-numeric()
  for(i in 8:57){
    pnormal<-shapiro.test(df[,i])$p.value
    if(pnormal<=alpha){
      pval<-wilcox.test(df[,i] ~ df$infection)$p.value
      names(pval) <- colnames(df)[i]
    }
    else {
      pvar<-var.test(df[,i]~df$infection)$p.value
      if(pvar<=alpha){
        pval<-t.test(df[,i] ~ df$infection, var.equal=F)$p.value
        names(pval) <- colnames(df)[i]
      }
      else{
        pval<-t.test(df[,i] ~ df$infection, var.equal=T)$p.value
        names(pval) <- colnames(df)[i]
      }
    }
    pfinal<-c(pval,pfinal) # Add new pval to the pfinal vector
  }
  #close loop
  #Adjusting p-values using conservative Bonferroni multiple testing correction
  #pfinal_bonferroni<-p.adjust(pfinal, method = "bonferroni", n = length(pfinal))
  #Adjusting p-values using using Benjamini & Hochberg multiple testing correction
  pfinal_fdr<-p.adjust(pfinal, method = "fdr", n = length(pfinal))
  #q=0.1 for FDR()????????????????????????????????
  names(pfinal_fdr) <- names(pfinal)
  return(pfinal_fdr)
}#close function

#Note, could not find FDR function allowing specificatio of fdr=0.1 cutoff thershold=
#Getting all calculated, adjusted p-values from the statistical tests:
pvalues<-(diffex(viral34_c, 0.05))
pvalues
```

```
##         FBXO31          KNTC2           RFC4         ZNF533          OXCT1
##      0.9341046      0.6264431      0.1048154      0.8478197      0.4160092
##       DIAPH3.2          WISP1            DTL       CDC42BPA          FGF18
##      0.0008148      0.8345801      0.4097221      0.3086543      0.0615507
##          TGFB3       C16orf61       DIAPH3.1         RUNDC1           MELK
##      0.8290377      0.5750973      0.7770385      0.9341046      0.9341046
##         TSPYL5          ORC6L          UCHL5           GMPS         NUSAP1
```

```
##      0.0373311      0.1674404      0.9063010      0.4097221      0.1048154
##           ECT2         GPR126    AA555029_RC          RAB6B Contig63649_RC
##      0.1100606      0.0277589      0.5882743      0.4160092      0.1674404
##          AYTL2           GNAZ           PECI         COL4A2           EXT1
##      0.3086543      0.5882743      0.9341046      0.4160092      0.8339875
##         STK32B Contig32125_RC        RTN4RL1         GPR180         SLC2A3
##      0.7355522      0.4097221      0.4160092      0.5512460      0.5882743
##         DIAPH3        QSCN6L1         PECI.1           FLT1            DCK
##      0.8339875      0.9063010      0.3086543      0.5201829      0.5882743
##           MTDH         SCUBE2         SERF1A        ALDH4A1 Contig40831_RC
##      0.2463937      0.8663983      0.8663983      0.0146467      0.3086543
## Contig35251_RC           MMP9           BBC3   RP5.860F19.3          GSTM3
##      0.3086543      0.0027674      0.9063010      0.4632253      0.5076624
```

```
#Getting number of significant results obtained after Benjamini & Hochberg correction:
num_sig<-sum((diffex(viral34_c, 0.05)<0.05))
num_sig
```

```
## [1] 5
```

```
#Getting names of genes whose non-scaled expression levels are significantly different among infection
diff_genes_names<-(names(which(diffex(viral34_c, 0.05)<=0.05)))
diff_genes_names
```

```
## [1] "DIAPH3.2" "TSPYL5"   "GPR126"   "ALDH4A1"  "MMP9"
```

```
#Based on results, there are 5 such nd genes: "DIAPH3.2", "TSPYL5",   "GPR126",   "ALDH4A1",  "MMP9"
```

QUESTION 13 Consider a regression model for the kind of infection as a function of gender, age and ancestry and the first 10 genes (scaled). Use stepwise variable selection and denote the selected model as "best.model". Interpret the obtained model.

```
#BACKGROUND: The logistic regression is used with dichotomous dependent variables.A generalized regress
#probabilistic outcome (Y=0/Y=1) where the probability is bound by an interval of [0,1], necessitating

#The FULL fitted model will be obtained first before step-wise variable selection:

library(glmnet)

#In general, it is recommended to center the age predictor by subtracting the mean:
m<-mean(viral34_c$age)
m
```

```
## [1] 44.25
```

```
##Results: [1] 44.25

c.age<-(viral34_c$age)-m
mean(c.age)
```

```
## [1] 0
```

```r
#This corrected age column is added to the viral34_c dataframe:
viral34_ca<-cbind(viral34_c,c.age)
summary(viral34_ca)
```

```
##                infection      stime                          sind       gender
##  bacterial_infection:69   Min.   : 0.055   symptoms_remain  :93   female:62
##  viral_infection    :71   1st Qu.: 4.695   symptoms_finished:47   male  :78
##                           Median : 6.962
##                           Mean   : 7.356
##                           3rd Qu.:10.057
##                           Max.   :17.659
##                   hosp          age         ancestry             GSTM3
##  no_hospitalization:73   Min.   :26.0   Length:140         Min.   :-0.3594
##  hospitalization   :67   1st Qu.:41.0   Class :character   1st Qu.:-0.1455
##                          Median :45.0   Mode  :character   Median :-0.0203
##                          Mean   :44.2                      Mean   : 0.0053
##                          3rd Qu.:49.0                      3rd Qu.: 0.1233
##                          Max.   :53.0                      Max.   : 0.5561
##   RP5.860F19.3          BBC3              MMP9          Contig35251_RC
##  Min.   :-0.4242   Min.   :-1.0828   Min.   :-0.4943   Min.   :-0.9177
##  1st Qu.:-0.1072   1st Qu.:-0.3333   1st Qu.:-0.1605   1st Qu.:-0.5925
##  Median : 0.0087   Median :-0.0953   Median :-0.0476   Median :-0.4027
##  Mean   : 0.0156   Mean   :-0.1130   Mean   :-0.0370   Mean   :-0.2517
##  3rd Qu.: 0.1031   3rd Qu.: 0.1110   3rd Qu.: 0.0880   3rd Qu.: 0.0437
##  Max.   : 0.5938   Max.   : 0.6018   Max.   : 0.5168   Max.   : 0.9944
##  Contig40831_RC       ALDH4A1            SERF1A            SCUBE2
##  Min.   :-0.4715   Min.   :-0.7679   Min.   :-0.5563   Min.   :-0.5152
##  1st Qu.:-0.1256   1st Qu.:-0.1749   1st Qu.:-0.0984   1st Qu.:-0.1291
##  Median : 0.0270   Median :-0.0041   Median : 0.0049   Median :-0.0226
##  Mean   : 0.0055   Mean   :-0.0277   Mean   :-0.0070   Mean   :-0.0243
##  3rd Qu.: 0.1225   3rd Qu.: 0.1378   3rd Qu.: 0.0900   3rd Qu.: 0.0749
##  Max.   : 0.4185   Max.   : 0.6030   Max.   : 0.3561   Max.   : 0.4372
##       MTDH              DCK              FLT1             PECI.1
##  Min.   :-0.6756   Min.   :-0.909   Min.   :-0.4826   Min.   :-0.4336
##  1st Qu.:-0.2933   1st Qu.:-0.529   1st Qu.:-0.1008   1st Qu.:-0.1396
##  Median :-0.0834   Median :-0.340   Median : 0.0189   Median :-0.0403
##  Mean   :-0.0867   Mean   :-0.321   Mean   :-0.0005   Mean   :-0.0336
##  3rd Qu.: 0.0738   3rd Qu.:-0.160   3rd Qu.: 0.0897   3rd Qu.: 0.0588
##  Max.   : 0.6406   Max.   : 0.599   Max.   : 0.5083   Max.   : 0.5128
##     QSCN6L1            DIAPH3            SLC2A3            GPR180
##  Min.   :-0.3794   Min.   :-0.4493   Min.   :-0.3716   Min.   :-0.3552
##  1st Qu.:-0.0466   1st Qu.:-0.1120   1st Qu.:-0.0777   1st Qu.:-0.0803
##  Median : 0.0078   Median :-0.0058   Median : 0.0005   Median :-0.0206
##  Mean   : 0.0217   Mean   :-0.0109   Mean   : 0.0114   Mean   :-0.0137
##  3rd Qu.: 0.0981   3rd Qu.: 0.0992   3rd Qu.: 0.0806   3rd Qu.: 0.0598
##  Max.   : 0.5401   Max.   : 0.3549   Max.   : 0.4642   Max.   : 0.3306
##     RTN4RL1        Contig32125_RC        STK32B             EXT1
##  Min.   :-0.6646   Min.   :-0.5321   Min.   :-0.4804   Min.   :-0.4778
##  1st Qu.:-0.2055   1st Qu.:-0.1135   1st Qu.:-0.1429   1st Qu.:-0.1675
##  Median : 0.0046   Median :-0.0090   Median :-0.0235   Median :-0.0558
##  Mean   :-0.0414   Mean   :-0.0110   Mean   :-0.0412   Mean   :-0.0519
##  3rd Qu.: 0.1318   3rd Qu.: 0.0734   3rd Qu.: 0.0449   3rd Qu.: 0.0605
##  Max.   : 0.4281   Max.   : 0.4563   Max.   : 0.4580   Max.   : 0.3741
```

```
##      COL4A2             PECI              GNAZ             AYTL2        
##  Min.   :-0.5987   Min.   :-0.4423   Min.   :-0.3175   Min.   :-0.6943  
##  1st Qu.:-0.1979   1st Qu.:-0.1942   1st Qu.:-0.0956   1st Qu.:-0.1319  
##  Median :-0.0528   Median :-0.0637   Median :-0.0164   Median :-0.0460  
##  Mean   :-0.0596   Mean   :-0.0373   Mean   : 0.0101   Mean   :-0.0252  
##  3rd Qu.: 0.0627   3rd Qu.: 0.0966   3rd Qu.: 0.0834   3rd Qu.: 0.0654  
##  Max.   : 0.5602   Max.   : 0.6090   Max.   : 0.4306   Max.   : 0.5336  
##  Contig63649_RC       RAB6B          AA555029_RC         GPR126       
##  Min.   :-0.3654   Min.   :-0.5692   Min.   :-0.431   Min.   :-0.3797  
##  1st Qu.:-0.0984   1st Qu.:-0.1431   1st Qu.:-0.160   1st Qu.:-0.1361  
##  Median :-0.0249   Median :-0.0522   Median :-0.001   Median :-0.0105  
##  Mean   :-0.0094   Mean   :-0.0172   Mean   :-0.021   Mean   :-0.0164  
##  3rd Qu.: 0.0900   3rd Qu.: 0.0896   3rd Qu.: 0.107   3rd Qu.: 0.0978  
##  Max.   : 0.3205   Max.   : 0.4946   Max.   : 0.820   Max.   : 0.4393  
##       ECT2             NUSAP1             GMPS             UCHL5        
##  Min.   :-0.5077   Min.   :-0.5863   Min.   :-0.5915   Min.   :-0.4585  
##  1st Qu.:-0.2311   1st Qu.:-0.1607   1st Qu.:-0.2841   1st Qu.:-0.1311  
##  Median :-0.0813   Median :-0.0093   Median :-0.0451   Median :-0.0386  
##  Mean   :-0.0500   Mean   :-0.0029   Mean   :-0.0605   Mean   :-0.0242  
##  3rd Qu.: 0.0984   3rd Qu.: 0.1504   3rd Qu.: 0.1528   3rd Qu.: 0.0921  
##  Max.   : 0.7757   Max.   : 0.6765   Max.   : 0.5519   Max.   : 0.5607  
##      ORC6L              TSPYL5             MELK             RUNDC1       
##  Min.   :-0.7968   Min.   :-0.6789   Min.   :-0.7898   Min.   :-0.870   
##  1st Qu.:-0.2140   1st Qu.:-0.1786   1st Qu.:-0.1895   1st Qu.:-0.331   
##  Median :-0.0244   Median :-0.0244   Median :-0.0611   Median :-0.118   
##  Mean   :-0.0517   Mean   :-0.0320   Mean   :-0.0493   Mean   :-0.106   
##  3rd Qu.: 0.1501   3rd Qu.: 0.1313   3rd Qu.: 0.0744   3rd Qu.: 0.104   
##  Max.   : 0.5067   Max.   : 0.6178   Max.   : 0.8189   Max.   : 0.753   
##     DIAPH3.1          C16orf61           TGFB3            FGF18        
##  Min.   :-0.7682   Min.   :-0.6119   Min.   :-0.4152   Min.   :-0.5978  
##  1st Qu.:-0.2564   1st Qu.:-0.1889   1st Qu.:-0.0924   1st Qu.:-0.1404  
##  Median :-0.0683   Median :-0.0931   Median :-0.0053   Median : 0.0015  
##  Mean   :-0.0539   Mean   :-0.0591   Mean   :-0.0023   Mean   :-0.0232  
##  3rd Qu.: 0.1179   3rd Qu.: 0.0587   3rd Qu.: 0.0827   3rd Qu.: 0.1070  
##  Max.   : 0.7049   Max.   : 0.5941   Max.   : 0.4397   Max.   : 0.4822  
##     CDC42BPA            DTL              WISP1            DIAPH3.2       
##  Min.   :-0.4444   Min.   :-1.264   Min.   :-0.4404   Min.   :-0.4510  
##  1st Qu.:-0.1519   1st Qu.:-0.651   1st Qu.:-0.0876   1st Qu.:-0.1221  
##  Median :-0.0436   Median :-0.153   Median : 0.0240   Median : 0.0088  
##  Mean   :-0.0264   Mean   :-0.209   Mean   : 0.0131   Mean   :-0.0009  
##  3rd Qu.: 0.0804   3rd Qu.: 0.203   3rd Qu.: 0.1223   3rd Qu.: 0.1127  
##  Max.   : 0.4842   Max.   : 0.892   Max.   : 0.3755   Max.   : 0.3669  
##      OXCT1              ZNF533             RFC4             KNTC2        
##  Min.   :-0.4278   Min.   :-0.5109   Min.   :-0.5636   Min.   :-0.4311  
##  1st Qu.:-0.0905   1st Qu.:-0.2613   1st Qu.:-0.0825   1st Qu.:-0.1841  
##  Median : 0.0095   Median :-0.1380   Median :-0.0010   Median :-0.0616  
##  Mean   : 0.0161   Mean   :-0.0593   Mean   : 0.0080   Mean   :-0.0359  
##  3rd Qu.: 0.1234   3rd Qu.: 0.0381   3rd Qu.: 0.1045   3rd Qu.: 0.0722  
##  Max.   : 0.6491   Max.   : 0.8648   Max.   : 0.4791   Max.   : 0.5975  
##      FBXO31           GSTM3_s        RP5.860F19.3_s        BBC3_s       
##  Min.   :-0.4215   Min.   :-1.845   Min.   :-2.2232   Min.   :-2.9155  
##  1st Qu.:-0.1388   1st Qu.:-0.763   1st Qu.:-0.6210   1st Qu.:-0.6625  
##  Median :-0.0451   Median :-0.130   Median :-0.0348   Median : 0.0531  
##  Mean   :-0.0253   Mean   : 0.000   Mean   : 0.0000   Mean   : 0.0000
```

```
##  3rd Qu.: 0.0860   3rd Qu.: 0.597   3rd Qu.: 0.4423   3rd Qu.: 0.6732
##  Max.   : 0.5556   Max.   : 2.786   Max.   : 2.9235   Max.   : 2.1487
##      MMP9_s        Contig35251_RC_s Contig40831_RC_s   ALDH4A1_s
##  Min.   :-2.1926   Min.   :-1.534   Min.   :-2.653   Min.   :-3.322
##  1st Qu.:-0.5924   1st Qu.:-0.785   1st Qu.:-0.729   1st Qu.:-0.661
##  Median :-0.0509   Median :-0.348   Median : 0.120   Median : 0.106
##  Mean   : 0.0000   Mean   : 0.000   Mean   : 0.000   Mean   : 0.000
##  3rd Qu.: 0.5992   3rd Qu.: 0.680   3rd Qu.: 0.651   3rd Qu.: 0.743
##  Max.   : 2.6553   Max.   : 2.870   Max.   : 2.296   Max.   : 2.831
##     SERF1A_s         SCUBE2_s           MTDH_s            DCK_s
##  Min.   :-3.214   Min.   :-3.1470   Min.   :-2.1023   Min.   :-2.120
##  1st Qu.:-0.534   1st Qu.:-0.6724   1st Qu.:-0.7373   1st Qu.:-0.748
##  Median : 0.070   Median : 0.0104   Median : 0.0118   Median :-0.066
##  Mean   : 0.000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.000
##  3rd Qu.: 0.568   3rd Qu.: 0.6356   3rd Qu.: 0.5733   3rd Qu.: 0.584
##  Max.   : 2.125   Max.   : 2.9577   Max.   : 2.5964   Max.   : 3.320
##      FLT1_s          PECI.1_s        QSCN6L1_s         DIAPH3_s
##  Min.   :-3.057   Min.   :-2.456   Min.   :-3.015   Min.   :-2.5771
##  1st Qu.:-0.636   1st Qu.:-0.651   1st Qu.:-0.513   1st Qu.:-0.5946
##  Median : 0.123   Median :-0.041   Median :-0.105   Median : 0.0302
##  Mean   : 0.000   Mean   : 0.000   Mean   : 0.000   Mean   : 0.0000
##  3rd Qu.: 0.572   3rd Qu.: 0.568   3rd Qu.: 0.575   3rd Qu.: 0.6471
##  Max.   : 3.227   Max.   : 3.356   Max.   : 3.897   Max.   : 2.1500
##     SLC2A3_s          GPR180_s        RTN4RL1_s      Contig32125_RC_s
##  Min.   :-2.4430   Min.   :-2.638   Min.   :-2.503   Min.   :-3.357
##  1st Qu.:-0.5685   1st Qu.:-0.515   1st Qu.:-0.659   1st Qu.:-0.660
##  Median :-0.0693   Median :-0.053   Median : 0.185   Median : 0.013
##  Mean   : 0.0000   Mean   : 0.000   Mean   : 0.000   Mean   : 0.000
##  3rd Qu.: 0.4415   3rd Qu.: 0.568   3rd Qu.: 0.696   3rd Qu.: 0.544
##  Max.   : 2.8891   Max.   : 2.659   Max.   : 1.886   Max.   : 3.011
##     STK32B_s          EXT1_s           COL4A2_s           PECI_s
##  Min.   :-2.711   Min.   :-2.4843   Min.   :-2.6489   Min.   :-1.895
##  1st Qu.:-0.627   1st Qu.:-0.6743   1st Qu.:-0.6795   1st Qu.:-0.734
##  Median : 0.110   Median :-0.0225   Median : 0.0334   Median :-0.124
##  Mean   : 0.000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.000
##  3rd Qu.: 0.531   3rd Qu.: 0.6559   3rd Qu.: 0.6012   3rd Qu.: 0.626
##  Max.   : 3.082   Max.   : 2.4850   Max.   : 3.0457   Max.   : 3.023
##      GNAZ_s          AYTL2_s       Contig63649_RC_s    RAB6B_s
##  Min.   :-2.167   Min.   :-3.980   Min.   :-2.573   Min.   :-2.710
##  1st Qu.:-0.700   1st Qu.:-0.635   1st Qu.:-0.643   1st Qu.:-0.618
##  Median :-0.175   Median :-0.124   Median :-0.112   Median :-0.172
##  Mean   : 0.000   Mean   : 0.000   Mean   : 0.000   Mean   : 0.000
##  3rd Qu.: 0.485   3rd Qu.: 0.539   3rd Qu.: 0.718   3rd Qu.: 0.524
##  Max.   : 2.782   Max.   : 3.323   Max.   : 2.384   Max.   : 2.513
##  AA555029_RC_s       GPR126_s          ECT2_s          NUSAP1_s
##  Min.   :-2.190   Min.   :-2.1825   Min.   :-1.917   Min.   :-2.4885
##  1st Qu.:-0.743   1st Qu.:-0.7189   1st Qu.:-0.759   1st Qu.:-0.6731
##  Median : 0.106   Median : 0.0356   Median :-0.131   Median :-0.0273
##  Mean   : 0.000   Mean   : 0.0000   Mean   : 0.000   Mean   : 0.0000
##  3rd Qu.: 0.687   3rd Qu.: 0.6862   3rd Qu.: 0.621   3rd Qu.: 0.6540
##  Max.   : 4.495   Max.   : 2.7371   Max.   : 3.458   Max.   : 2.8981
##      GMPS_s          UCHL5_s          ORC6L_s          TSPYL5_s
##  Min.   :-2.0588   Min.   :-2.657   Min.   :-2.786   Min.   :-2.7252
##  1st Qu.:-0.8669   1st Qu.:-0.654   1st Qu.:-0.607   1st Qu.:-0.6176
```

```
##  Median : 0.0594   Median :-0.088   Median : 0.102   Median : 0.0318
##  Mean   : 0.0000   Mean   : 0.000   Mean   : 0.000   Mean   : 0.0000
##  3rd Qu.: 0.8269   3rd Qu.: 0.711   3rd Qu.: 0.754   3rd Qu.: 0.6877
##  Max.   : 2.3741   Max.   : 3.578   Max.   : 2.088   Max.   : 2.7375
##      MELK_s          RUNDC1_s         DIAPH3.1_s        C16orf61_s
##  Min.   :-3.193   Min.   :-2.5022   Min.   :-2.6615   Min.   :-2.860
##  1st Qu.:-0.604   1st Qu.:-0.7355   1st Qu.:-0.7545   1st Qu.:-0.672
##  Median :-0.051   Median :-0.0409   Median :-0.0537   Median :-0.176
##  Mean   : 0.000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.000
##  3rd Qu.: 0.533   3rd Qu.: 0.6859   3rd Qu.: 0.6400   3rd Qu.: 0.609
##  Max.   : 3.744   Max.   : 2.8101   Max.   : 2.8273   Max.   : 3.380
##     TGFB3_s           FGF18_s         CDC42BPA_s          DTL_s
##  Min.   :-2.8862   Min.   :-2.939   Min.   :-2.4108   Min.   :-2.006
##  1st Qu.:-0.6297   1st Qu.:-0.600   1st Qu.:-0.7237   1st Qu.:-0.839
##  Median :-0.0212   Median : 0.126   Median :-0.0991   Median : 0.107
##  Mean   : 0.0000   Mean   : 0.000   Mean   : 0.0000   Mean   : 0.000
##  3rd Qu.: 0.5942   3rd Qu.: 0.666   3rd Qu.: 0.6162   3rd Qu.: 0.785
##  Max.   : 3.0889   Max.   : 2.585   Max.   : 2.9450   Max.   : 2.095
##     WISP1_s          DIAPH3.2_s         OXCT1_s          ZNF533_s
##  Min.   :-2.8269   Min.   :-2.7614   Min.   :-2.435   Min.   :-1.565
##  1st Qu.:-0.6278   1st Qu.:-0.7434   1st Qu.:-0.585   1st Qu.:-0.700
##  Median : 0.0679   Median : 0.0598   Median :-0.036   Median :-0.273
##  Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.000   Mean   : 0.000
##  3rd Qu.: 0.6807   3rd Qu.: 0.6967   3rd Qu.: 0.588   3rd Qu.: 0.337
##  Max.   : 2.2588   Max.   : 2.2564   Max.   : 3.472   Max.   : 3.203
##      RFC4_s          KNTC2_s          FBXO31_s         GSTM3_l
##  Min.   :-3.620   Min.   :-1.978   Min.   :-2.362   Min.   :0.971
##  1st Qu.:-0.573   1st Qu.:-0.742   1st Qu.:-0.676   1st Qu.:1.049
##  Median :-0.057   Median :-0.129   Median :-0.117   Median :1.092
##  Mean   : 0.000   Mean   : 0.000   Mean   : 0.000   Mean   :1.098
##  3rd Qu.: 0.611   3rd Qu.: 0.541   3rd Qu.: 0.664   3rd Qu.:1.139
##  Max.   : 2.983   Max.   : 3.170   Max.   : 3.463   Max.   :1.269
##  RP5.860F19.3_l      BBC3_l          MMP9_l      Contig35251_RC_l
##  Min.   :0.946   Min.   :0.651   Min.   :0.919   Min.   :0.734
##  1st Qu.:1.062   1st Qu.:0.981   1st Qu.:1.044   1st Qu.:0.879
##  Median :1.101   Median :1.066   Median :1.083   Median :0.955
##  Mean   :1.102   Mean   :1.053   Mean   :1.084   Mean   :0.999
##  3rd Qu.:1.132   3rd Qu.:1.135   3rd Qu.:1.127   3rd Qu.:1.113
##  Max.   :1.279   Max.   :1.281   Max.   :1.258   Max.   :1.385
##  Contig40831_RC_l   ALDH4A1_l        SERF1A_l        SCUBE2_l
##  Min.   :0.928    Min.   :0.803   Min.   :0.893   Min.   :0.91
##  1st Qu.:1.056    1st Qu.:1.038   1st Qu.:1.065   1st Qu.:1.05
##  Median :1.108    Median :1.097   Median :1.100   Median :1.09
##  Mean   :1.099    Mean   :1.087   Mean   :1.095   Mean   :1.09
##  3rd Qu.:1.139    3rd Qu.:1.143   3rd Qu.:1.128   3rd Qu.:1.12
##  Max.   :1.229    Max.   :1.282   Max.   :1.211   Max.   :1.23
##      MTDH_l          DCK_l           FLT1_l         PECI.1_l
##  Min.   :0.843   Min.   :0.738   Min.   :0.923   Min.   :0.942
##  1st Qu.:0.996   1st Qu.:0.905   1st Qu.:1.064   1st Qu.:1.051
##  Median :1.070   Median :0.978   Median :1.105   Median :1.085
##  Mean   :1.065   Mean   :0.980   Mean   :1.097   Mean   :1.086
##  3rd Qu.:1.123   3rd Qu.:1.044   3rd Qu.:1.128   3rd Qu.:1.118
##  Max.   :1.292   Max.   :1.280   Max.   :1.255   Max.   :1.256
##    QSCN6L1_l        DIAPH3_l        SLC2A3_l        GPR180_l
```

```
##  Min.   :0.963   Min.   :0.936   Min.   :0.966   Min.   :0.973
##  1st Qu.:1.083   1st Qu.:1.060   1st Qu.:1.072   1st Qu.:1.071
##  Median :1.101   Median :1.097   Median :1.099   Median :1.092
##  Mean   :1.105   Mean   :1.093   Mean   :1.101   Mean   :1.093
##  3rd Qu.:1.131   3rd Qu.:1.131   3rd Qu.:1.125   3rd Qu.:1.118
##  Max.   :1.264   Max.   :1.210   Max.   :1.242   Max.   :1.203
##    RTN4RL1_l     Contig32125_RC_l    STK32B_l         EXT1_l
##  Min.   :0.848   Min.   :0.903    Min.   :0.924   Min.   :0.925
##  1st Qu.:1.028   1st Qu.:1.060    1st Qu.:1.050   1st Qu.:1.041
##  Median :1.100   Median :1.096    Median :1.091   Median :1.080
##  Mean   :1.081   Mean   :1.094    Mean   :1.083   Mean   :1.079
##  3rd Qu.:1.142   3rd Qu.:1.123    3rd Qu.:1.113   3rd Qu.:1.119
##  Max.   :1.232   Max.   :1.240    Max.   :1.241   Max.   :1.216
##     COL4A2_l         PECI_l          GNAZ_l         AYTL2_l
##  Min.   :0.876   Min.   :0.939   Min.   :0.987   Min.   :0.835
##  1st Qu.:1.030   1st Qu.:1.032   1st Qu.:1.066   1st Qu.:1.054
##  Median :1.081   Median :1.077   Median :1.093   Median :1.083
##  Mean   :1.076   Mean   :1.084   Mean   :1.101   Mean   :1.089
##  3rd Qu.:1.119   3rd Qu.:1.130   3rd Qu.:1.126   3rd Qu.:1.120
##  Max.   :1.270   Max.   :1.283   Max.   :1.233   Max.   :1.262
##  Contig63649_RC_l    RAB6B_l      AA555029_RC_l      GPR126_l
##  Min.   :0.969    Min.   :0.888   Min.   :0.944   Min.   :0.963
##  1st Qu.:1.065    1st Qu.:1.050   1st Qu.:1.044   1st Qu.:1.052
##  Median :1.090    Median :1.081   Median :1.098   Median :1.095
##  Mean   :1.094    Mean   :1.091   Mean   :1.090   Mean   :1.092
##  3rd Qu.:1.128    3rd Qu.:1.128   3rd Qu.:1.134   3rd Qu.:1.131
##  Max.   :1.200    Max.   :1.251   Max.   :1.340   Max.   :1.235
##      ECT2_l         NUSAP1_l         GMPS_l         UCHL5_l         ORC6L_l
##  Min.   :0.913   Min.   :0.881   Min.   :0.879   Min.   :0.933   Min.   :0.79
##  1st Qu.:1.018   1st Qu.:1.044   1st Qu.:0.999   1st Qu.:1.054   1st Qu.:1.02
##  Median :1.071   Median :1.095   Median :1.083   Median :1.086   Median :1.09
##  Mean   :1.079   Mean   :1.095   Mean   :1.074   Mean   :1.089   Mean   :1.08
##  3rd Qu.:1.131   3rd Qu.:1.147   3rd Qu.:1.148   3rd Qu.:1.129   3rd Qu.:1.15
##  Max.   :1.329   Max.   :1.302   Max.   :1.268   Max.   :1.270   Max.   :1.25
##     TSPYL5_l         MELK_l         RUNDC1_l       DIAPH3.1_l
##  Min.   :0.842   Min.   :0.793   Min.   :0.756   Min.   :0.803
##  1st Qu.:1.037   1st Qu.:1.033   1st Qu.:0.982   1st Qu.:1.009
##  Median :1.090   Median :1.078   Median :1.058   Median :1.076
##  Mean   :1.085   Mean   :1.079   Mean   :1.057   Mean   :1.076
##  3rd Qu.:1.141   3rd Qu.:1.123   3rd Qu.:1.133   3rd Qu.:1.137
##  Max.   :1.286   Max.   :1.340   Max.   :1.323   Max.   :1.310
##    C16orf61_l       TGFB3_l        FGF18_l        CDC42BPA_l        DTL_l
##  Min.   :0.871   Min.   :0.95   Min.   :0.876   Min.   :0.938   Min.   :0.551
##  1st Qu.:1.034   1st Qu.:1.07   1st Qu.:1.051   1st Qu.:1.047   1st Qu.:0.854
##  Median :1.067   Median :1.10   Median :1.099   Median :1.084   Median :1.046
##  Mean   :1.077   Mean   :1.10   Mean   :1.089   Mean   :1.088   Mean   :1.007
##  3rd Qu.:1.118   3rd Qu.:1.13   3rd Qu.:1.134   3rd Qu.:1.125   3rd Qu.:1.164
##  Max.   :1.279   Max.   :1.24   Max.   :1.248   Max.   :1.248   Max.   :1.359
##     WISP1_l       DIAPH3.2_l       OXCT1_l         ZNF533_l         RFC4_l
##  Min.   :0.94   Min.   :0.936   Min.   :0.945   Min.   :0.912   Min.   :0.89
##  1st Qu.:1.07   1st Qu.:1.057   1st Qu.:1.068   1st Qu.:1.008   1st Qu.:1.07
##  Median :1.11   Median :1.102   Median :1.102   Median :1.052   Median :1.10
##  Mean   :1.10   Mean   :1.097   Mean   :1.102   Mean   :1.074   Mean   :1.10
##  3rd Qu.:1.14   3rd Qu.:1.135   3rd Qu.:1.139   3rd Qu.:1.111   3rd Qu.:1.13
```

```
##  Max.   :1.22   Max.   :1.214   Max.   :1.294   Max.   :1.352   Max.   :1.25
##     KNTC2_l         FBXO31_l         c.age
##  Min.   :0.944   Min.   :0.947   Min.   :-18.25
##  1st Qu.:1.035   1st Qu.:1.051   1st Qu.: -3.25
##  Median :1.078   Median :1.083   Median :  0.75
##  Mean   :1.084   Mean   :1.089   Mean   :  0.00
##  3rd Qu.:1.122   3rd Qu.:1.127   3rd Qu.:  4.75
##  Max.   :1.280   Max.   :1.268   Max.   :  8.75
```

```
dim(viral34_ca)
```

```
## [1] 140 158
```

```
#Assigning the dependent, factored categorical variable "infection type) to a variable Y
Y<-viral34_c$infection
Y
```

```
##   [1] bacterial_infection viral_infection     bacterial_infection
##   [4] viral_infection     viral_infection     viral_infection
##   [7] bacterial_infection bacterial_infection viral_infection
##  [10] bacterial_infection viral_infection     viral_infection
##  [13] viral_infection     bacterial_infection bacterial_infection
##  [16] bacterial_infection viral_infection     bacterial_infection
##  [19] bacterial_infection viral_infection     bacterial_infection
##  [22] viral_infection     bacterial_infection viral_infection
##  [25] bacterial_infection bacterial_infection bacterial_infection
##  [28] bacterial_infection bacterial_infection bacterial_infection
##  [31] bacterial_infection bacterial_infection bacterial_infection
##  [34] viral_infection     viral_infection     viral_infection
##  [37] viral_infection     viral_infection     viral_infection
##  [40] bacterial_infection viral_infection     bacterial_infection
##  [43] viral_infection     viral_infection     viral_infection
##  [46] viral_infection     viral_infection     bacterial_infection
##  [49] viral_infection     viral_infection     bacterial_infection
##  [52] viral_infection     bacterial_infection bacterial_infection
##  [55] viral_infection     bacterial_infection bacterial_infection
##  [58] bacterial_infection bacterial_infection viral_infection
##  [61] viral_infection     viral_infection     viral_infection
##  [64] viral_infection     bacterial_infection viral_infection
##  [67] viral_infection     bacterial_infection bacterial_infection
##  [70] viral_infection     bacterial_infection viral_infection
##  [73] viral_infection     viral_infection     viral_infection
##  [76] bacterial_infection bacterial_infection viral_infection
##  [79] viral_infection     bacterial_infection bacterial_infection
##  [82] bacterial_infection viral_infection     viral_infection
##  [85] bacterial_infection viral_infection     viral_infection
##  [88] viral_infection     viral_infection     bacterial_infection
##  [91] bacterial_infection viral_infection     bacterial_infection
##  [94] bacterial_infection viral_infection     viral_infection
##  [97] viral_infection     bacterial_infection bacterial_infection
## [100] viral_infection     bacterial_infection bacterial_infection
## [103] viral_infection     bacterial_infection viral_infection
## [106] viral_infection     viral_infection     bacterial_infection
```

```
## [109] bacterial_infection viral_infection     bacterial_infection
## [112] viral_infection     bacterial_infection viral_infection
## [115] viral_infection     bacterial_infection viral_infection
## [118] bacterial_infection bacterial_infection viral_infection
## [121] viral_infection     viral_infection     bacterial_infection
## [124] viral_infection     bacterial_infection bacterial_infection
## [127] bacterial_infection bacterial_infection viral_infection
## [130] bacterial_infection bacterial_infection viral_infection
## [133] viral_infection     bacterial_infection viral_infection
## [136] bacterial_infection bacterial_infection bacterial_infection
## [139] bacterial_infection viral_infection
## Levels: bacterial_infection viral_infection
```

```
#Index Directory of Co-Variate Columns in dataframe viral34_ca
# gender=4
# age=6
# ancestry=7,
# First scaled 10 genes=58-67
# c.age=148

#Obtaining first the FULL logistic model with "infection" as dependent variable and variables (includin
model1<-glm(Y~., data=viral34_c[,c(4,6,7,58:67)],family="binomial")
model1
```

```
##
## Call:  glm(formula = Y ~ ., family = "binomial", data = viral34_c[,
##     c(4, 6, 7, 58:67)])
##
## Coefficients:
##      (Intercept)        gendermale               age         ancestryB
##         0.776969         -0.706281         -0.000755         -0.661874
##        ancestryC           GSTM3_s    RP5.860F19.3_s           BBC3_s
##        -1.374640         -0.180613          0.111857         -0.207958
##           MMP9_s  Contig35251_RC_s  Contig40831_RC_s         ALDH4A1_s
##         1.424142          0.229409          0.142125          1.601389
##         SERF1A_s          SCUBE2_s            MTDH_s
##         0.116305         -0.281441         -0.009742
##
## Degrees of Freedom: 139 Total (i.e. Null);  125 Residual
## Null Deviance:       194
## Residual Deviance: 140   AIC: 170
```

```
summary(model1)
```

```
##
## Call:
## glm(formula = Y ~ ., family = "binomial", data = viral34_c[,
##     c(4, 6, 7, 58:67)])
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       0.776969   1.782134    0.44    0.663
## gendermale       -0.706281   0.455474   -1.55    0.121
```

```
## age              -0.000755   0.040859   -0.02    0.985
## ancestryB        -0.661874   0.570617   -1.16    0.246
## ancestryC        -1.374640   0.723780   -1.90    0.058 .
## GSTM3_s          -0.180613   0.290497   -0.62    0.534
## RP5.860F19.3_s    0.111857   0.296742    0.38    0.706
## BBC3_s           -0.207958   0.226570   -0.92    0.359
## MMP9_s            1.424142   0.332779    4.28  1.9e-05 ***
## Contig35251_RC_s  0.229409   0.243580    0.94    0.346
## Contig40831_RC_s  0.142125   0.243031    0.58    0.559
## ALDH4A1_s         1.601389   0.385158    4.16  3.2e-05 ***
## SERF1A_s          0.116305   0.265279    0.44    0.661
## SCUBE2_s         -0.281441   0.251780   -1.12    0.264
## MTDH_s           -0.009742   0.264942   -0.04    0.971
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 194.05  on 139  degrees of freedom
## Residual deviance: 140.11  on 125  degrees of freedom
## AIC: 170.1
##
## Number of Fisher Scoring iterations: 5
```

```
# Null deviance: 194.05  on 139  degrees of freedom
# Residual deviance: 140.11  on 125  degrees of freedom
# AIC: 170.11

#Obtaining first the FULL logistic model with "infection" as dependent variable and variables (includin
model1a<-glm(Y~., data=viral34_ca[,c(4,148,7,58:67)],family="binomial")
model1a
```

```
##
## Call:  glm(formula = Y ~ ., family = "binomial", data = viral34_ca[,
##     c(4, 148, 7, 58:67)])
##
## Coefficients:
##      (Intercept)        gendermale           FGF18_l         ancestryB
##          -6.1156           -0.8018            6.3427           -0.6907
##        ancestryC           GSTM3_s    RP5.860F19.3_s           BBC3_s
##          -1.3728           -0.2650            0.0970           -0.1399
##           MMP9_s  Contig35251_RC_s  Contig40831_RC_s         ALDH4A1_s
##           1.4432            0.2498            0.1848            1.4022
##         SERF1A_s          SCUBE2_s            MTDH_s
##           0.1195           -0.3329           -0.0127
##
## Degrees of Freedom: 139 Total (i.e. Null);  125 Residual
## Null Deviance:       194
## Residual Deviance: 138   AIC: 168
```

```
summary(model1a)
```

```
##
```

```
## Call:
## glm(formula = Y ~ ., family = "binomial", data = viral34_ca[,
##     c(4, 148, 7, 58:67)])
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -6.1156     5.3325   -1.15  0.25144
## gendermale        -0.8018     0.4616   -1.74  0.08238 .
## FGF18_l            6.3427     4.9228    1.29  0.19759
## ancestryB         -0.6907     0.5761   -1.20  0.23057
## ancestryC         -1.3728     0.7322   -1.87  0.06081 .
## GSTM3_s           -0.2650     0.2997   -0.88  0.37668
## RP5.860F19.3_s     0.0970     0.3036    0.32  0.74938
## BBC3_s            -0.1399     0.2343   -0.60  0.55035
## MMP9_s             1.4432     0.3360    4.29  1.7e-05 ***
## Contig35251_RC_s   0.2498     0.2414    1.03  0.30075
## Contig40831_RC_s   0.1848     0.2436    0.76  0.44810
## ALDH4A1_s          1.4022     0.4116    3.41  0.00066 ***
## SERF1A_s           0.1195     0.2560    0.47  0.64068
## SCUBE2_s          -0.3329     0.2549   -1.31  0.19147
## MTDH_s            -0.0127     0.2652   -0.05  0.96190
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 194.05  on 139  degrees of freedom
## Residual deviance: 138.41  on 125  degrees of freedom
## AIC: 168.4
##
## Number of Fisher Scoring iterations: 5
```

```
# Null deviance: 194.05  on 139  degrees of freedom
# Residual deviance: 138.41  on 125  degrees of freedom
# AIC: 168.41

#Comparing these two FULL models, it is evident that the deviance and AIC was lower after "correcting"
#used for subsequent variable selection:

#When the number of variables is large, stepwise selection of variables is used to remove those variabl
#Given a multivariate FULL, age-adjusted logistic model, Forward selection, Backward elimination, and S
#will be used to remove variables that are unimportant and unassociated with the infection-type respons
#Comparing the fit of different models will require adjusting for the number k of covariates in the mod
#more complex model always provides the smallest R2. Other adjusted measures include:
#AIC (Akaike information criterion): Deviance+2(k+1)
#BIC (Bayesian information criterion): Deviance+ ln(n)*(k+1)
#The model with the smallest AIC or BIC will be chosen as best.model for Problem#14 Validation


#Variable Selection:

#Performing forward selection:
forwardmodel1<-step(model1,direction="forward")
```

```
## Start:  AIC=170.1
## Y ~ gender + age + ancestry + GSTM3_s + RP5.860F19.3_s + BBC3_s +
##     MMP9_s + Contig35251_RC_s + Contig40831_RC_s + ALDH4A1_s +
##     SERF1A_s + SCUBE2_s + MTDH_s
```

```
#Results show an AIC=170.11 after forward selection:
# Start:  AIC=170.11
#
# Y ~ gender + age + ancestry + GSTM3_s + RP5.860F19.3_s + BBC3_s +
#   MMP9_s + Contig35251_RC_s + Contig40831_RC_s + ALDH4A1_s +
#   SERF1A_s + SCUBE2_s + MTDH_s

summary(forwardmodel1) # best forward model
```

```
##
## Call:
## glm(formula = Y ~ gender + age + ancestry + GSTM3_s + RP5.860F19.3_s +
##     BBC3_s + MMP9_s + Contig35251_RC_s + Contig40831_RC_s + ALDH4A1_s +
##     SERF1A_s + SCUBE2_s + MTDH_s, family = "binomial", data = viral34_c[,
##     c(4, 6, 7, 58:67)])
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       0.776969   1.782134    0.44    0.663
## gendermale       -0.706281   0.455474   -1.55    0.121
## age              -0.000755   0.040859   -0.02    0.985
## ancestryB        -0.661874   0.570617   -1.16    0.246
## ancestryC        -1.374640   0.723780   -1.90    0.058 .
## GSTM3_s          -0.180613   0.290497   -0.62    0.534
## RP5.860F19.3_s    0.111857   0.296742    0.38    0.706
## BBC3_s           -0.207958   0.226570   -0.92    0.359
## MMP9_s            1.424142   0.332779    4.28  1.9e-05 ***
## Contig35251_RC_s  0.229409   0.243580    0.94    0.346
## Contig40831_RC_s  0.142125   0.243031    0.58    0.559
## ALDH4A1_s         1.601389   0.385158    4.16  3.2e-05 ***
## SERF1A_s          0.116305   0.265279    0.44    0.661
## SCUBE2_s         -0.281441   0.251780   -1.12    0.264
## MTDH_s           -0.009742   0.264942   -0.04    0.971
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 194.05  on 139  degrees of freedom
## Residual deviance: 140.11  on 125  degrees of freedom
## AIC: 170.1
##
## Number of Fisher Scoring iterations: 5
```

```
#Performing backward selection:
backwardmodel1<-step(model1,direction="backward")
```

```
## Start:  AIC=170.1
```

```
## Y ~ gender + age + ancestry + GSTM3_s + RP5.860F19.3_s + BBC3_s +
##     MMP9_s + Contig35251_RC_s + Contig40831_RC_s + ALDH4A1_s +
##     SERF1A_s + SCUBE2_s + MTDH_s
##
##                     Df Deviance AIC
## - age                1      140 168
## - MTDH_s             1      140 168
## - RP5.860F19.3_s     1      140 168
## - SERF1A_s           1      140 168
## - Contig40831_RC_s   1      140 168
## - GSTM3_s            1      140 168
## - BBC3_s             1      141 169
## - Contig35251_RC_s   1      141 169
## - SCUBE2_s           1      141 169
## <none>                      140 170
## - gender             1      143 171
## - ancestry           2      145 171
## - ALDH4A1_s          1      163 191
## - MMP9_s             1      166 194
##
## Step:  AIC=168.1
## Y ~ gender + ancestry + GSTM3_s + RP5.860F19.3_s + BBC3_s + MMP9_s +
##     Contig35251_RC_s + Contig40831_RC_s + ALDH4A1_s + SERF1A_s +
##     SCUBE2_s + MTDH_s
##
##                     Df Deviance AIC
## - MTDH_s             1      140 166
## - RP5.860F19.3_s     1      140 166
## - SERF1A_s           1      140 166
## - Contig40831_RC_s   1      140 166
## - GSTM3_s            1      140 166
## - BBC3_s             1      141 167
## - Contig35251_RC_s   1      141 167
## - SCUBE2_s           1      141 167
## <none>                      140 168
## - gender             1      143 169
## - ancestry           2      145 169
## - ALDH4A1_s          1      163 189
## - MMP9_s             1      166 192
##
## Step:  AIC=166.1
## Y ~ gender + ancestry + GSTM3_s + RP5.860F19.3_s + BBC3_s + MMP9_s +
##     Contig35251_RC_s + Contig40831_RC_s + ALDH4A1_s + SERF1A_s +
##     SCUBE2_s
##
##                     Df Deviance AIC
## - RP5.860F19.3_s     1      140 164
## - SERF1A_s           1      140 164
## - Contig40831_RC_s   1      140 164
## - GSTM3_s            1      140 164
## - BBC3_s             1      141 165
## - Contig35251_RC_s   1      141 165
## - SCUBE2_s           1      141 165
## <none>                      140 166
```

```
## - gender            1      143 167
## - ancestry          2      145 167
## - ALDH4A1_s         1      165 189
## - MMP9_s            1      168 192
##
## Step:  AIC=164.3
## Y ~ gender + ancestry + GSTM3_s + BBC3_s + MMP9_s + Contig35251_RC_s +
##     Contig40831_RC_s + ALDH4A1_s + SERF1A_s + SCUBE2_s
##
##                    Df Deviance AIC
## - SERF1A_s          1      140 162
## - GSTM3_s           1      141 163
## - Contig40831_RC_s  1      141 163
## - BBC3_s            1      141 163
## - Contig35251_RC_s  1      141 163
## - SCUBE2_s          1      142 164
## <none>                     140 164
## - gender            1      143 165
## - ancestry          2      145 165
## - ALDH4A1_s         1      168 190
## - MMP9_s            1      168 190
##
## Step:  AIC=162.5
## Y ~ gender + ancestry + GSTM3_s + BBC3_s + MMP9_s + Contig35251_RC_s +
##     Contig40831_RC_s + ALDH4A1_s + SCUBE2_s
##
##                    Df Deviance AIC
## - GSTM3_s           1      141 161
## - Contig40831_RC_s  1      141 161
## - BBC3_s            1      142 162
## - Contig35251_RC_s  1      142 162
## - SCUBE2_s          1      142 162
## <none>                     140 162
## - gender            1      143 163
## - ancestry          2      145 163
## - ALDH4A1_s         1      168 188
## - MMP9_s            1      169 189
##
## Step:  AIC=160.8
## Y ~ gender + ancestry + BBC3_s + MMP9_s + Contig35251_RC_s +
##     Contig40831_RC_s + ALDH4A1_s + SCUBE2_s
##
##                    Df Deviance AIC
## - Contig40831_RC_s  1      141 159
## - BBC3_s            1      142 160
## - SCUBE2_s          1      142 160
## - Contig35251_RC_s  1      142 160
## <none>                     141 161
## - gender            1      143 161
## - ancestry          2      146 162
## - MMP9_s            1      170 188
## - ALDH4A1_s         1      172 190
##
## Step:  AIC=159.2
```

```
## Y ~ gender + ancestry + BBC3_s + MMP9_s + Contig35251_RC_s +
##     ALDH4A1_s + SCUBE2_s
##
##                    Df Deviance AIC
## - BBC3_s            1      142 158
## - SCUBE2_s          1      142 158
## - Contig35251_RC_s  1      143 159
## <none>                     141 159
## - gender            1      144 160
## - ancestry          2      146 160
## - ALDH4A1_s         1      173 189
## - MMP9_s            1      174 190
##
## Step:  AIC=158.2
## Y ~ gender + ancestry + MMP9_s + Contig35251_RC_s + ALDH4A1_s +
##     SCUBE2_s
##
##                    Df Deviance AIC
## - SCUBE2_s          1      144 158
## - Contig35251_RC_s  1      144 158
## <none>                     142 158
## - gender            1      144 158
## - ancestry          2      147 159
## - ALDH4A1_s         1      173 187
## - MMP9_s            1      174 188
##
## Step:  AIC=157.5
## Y ~ gender + ancestry + MMP9_s + Contig35251_RC_s + ALDH4A1_s
##
##                    Df Deviance AIC
## - gender            1      145 157
## - Contig35251_RC_s  1      146 158
## <none>                     144 158
## - ancestry          2      148 158
## - ALDH4A1_s         1      173 185
## - MMP9_s            1      176 188
##
## Step:  AIC=157.3
## Y ~ ancestry + MMP9_s + Contig35251_RC_s + ALDH4A1_s
##
##                    Df Deviance AIC
## - Contig35251_RC_s  1      147 157
## <none>                     145 157
## - ancestry          2      150 158
## - ALDH4A1_s         1      175 185
## - MMP9_s            1      176 186
##
## Step:  AIC=156.8
## Y ~ ancestry + MMP9_s + ALDH4A1_s
##
##             Df Deviance AIC
## <none>              147 157
## - ancestry   2      152 158
## - ALDH4A1_s  1      175 183
```

```
## - MMP9_s     1      181 189
```

```
#Results show reduced AIC=156.84 and reduction of covariates to only ancestry, scaled Gene expression M
#ALDH4A1_s (Alcohol Dehydrogenase) after forward selection:

# Step:  AIC=156.84
# Y ~ ancestry + MMP9_s + ALDH4A1_s
#
# Df Deviance    AIC
# <none>           146.84 156.84
# - ancestry   2   151.94 157.94
# - ALDH4A1_s  1   175.43 183.43
# - MMP9_s     1   180.59 188.59

summary(backwardmodel1)
```

```
##
## Call:
## glm(formula = Y ~ ancestry + MMP9_s + ALDH4A1_s, family = "binomial",
##     data = viral34_c[, c(4, 6, 7, 58:67)])
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.351      0.243    1.45    0.148
## ancestryB     -0.776      0.538   -1.44    0.149
## ancestryC     -1.308      0.694   -1.89    0.059 .
## MMP9_s         1.375      0.288    4.78  1.8e-06 ***
## ALDH4A1_s      1.267      0.282    4.49  7.0e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 194.05  on 139  degrees of freedom
## Residual deviance: 146.84  on 135  degrees of freedom
## AIC: 156.8
##
## Number of Fisher Scoring iterations: 5
```

```
#Performing both selection:
#backward-forward (both) selection
bothmodel1<-step(model1,direction="both")
```

```
## Start:  AIC=170.1
## Y ~ gender + age + ancestry + GSTM3_s + RP5.860F19.3_s + BBC3_s +
##     MMP9_s + Contig35251_RC_s + Contig40831_RC_s + ALDH4A1_s +
##     SERF1A_s + SCUBE2_s + MTDH_s
##
##                    Df Deviance AIC
## - age               1      140 168
## - MTDH_s            1      140 168
## - RP5.860F19.3_s    1      140 168
## - SERF1A_s          1      140 168
```

```
## - Contig40831_RC_s  1      140 168
## - GSTM3_s           1      140 168
## - BBC3_s            1      141 169
## - Contig35251_RC_s  1      141 169
## - SCUBE2_s          1      141 169
## <none>                     140 170
## - gender            1      143 171
## - ancestry          2      145 171
## - ALDH4A1_s         1      163 191
## - MMP9_s            1      166 194
##
## Step:  AIC=168.1
## Y ~ gender + ancestry + GSTM3_s + RP5.860F19.3_s + BBC3_s + MMP9_s +
##     Contig35251_RC_s + Contig40831_RC_s + ALDH4A1_s + SERF1A_s +
##     SCUBE2_s + MTDH_s
##
##                    Df Deviance AIC
## - MTDH_s            1      140 166
## - RP5.860F19.3_s    1      140 166
## - SERF1A_s          1      140 166
## - Contig40831_RC_s  1      140 166
## - GSTM3_s           1      140 166
## - BBC3_s            1      141 167
## - Contig35251_RC_s  1      141 167
## - SCUBE2_s          1      141 167
## <none>                     140 168
## - gender            1      143 169
## - ancestry          2      145 169
## + age               1      140 170
## - ALDH4A1_s         1      163 189
## - MMP9_s            1      166 192
##
## Step:  AIC=166.1
## Y ~ gender + ancestry + GSTM3_s + RP5.860F19.3_s + BBC3_s + MMP9_s +
##     Contig35251_RC_s + Contig40831_RC_s + ALDH4A1_s + SERF1A_s +
##     SCUBE2_s
##
##                    Df Deviance AIC
## - RP5.860F19.3_s    1      140 164
## - SERF1A_s          1      140 164
## - Contig40831_RC_s  1      140 164
## - GSTM3_s           1      140 164
## - BBC3_s            1      141 165
## - Contig35251_RC_s  1      141 165
## - SCUBE2_s          1      141 165
## <none>                     140 166
## - gender            1      143 167
## - ancestry          2      145 167
## + MTDH_s            1      140 168
## + age               1      140 168
## - ALDH4A1_s         1      165 189
## - MMP9_s            1      168 192
##
## Step:  AIC=164.3
```

```
## Y ~ gender + ancestry + GSTM3_s + BBC3_s + MMP9_s + Contig35251_RC_s +
##     Contig40831_RC_s + ALDH4A1_s + SERF1A_s + SCUBE2_s
##
##                    Df Deviance AIC
## - SERF1A_s          1      140 162
## - GSTM3_s           1      141 163
## - Contig40831_RC_s  1      141 163
## - BBC3_s            1      141 163
## - Contig35251_RC_s  1      141 163
## - SCUBE2_s          1      142 164
## <none>                     140 164
## - gender            1      143 165
## - ancestry          2      145 165
## + RP5.860F19.3_s    1      140 166
## + MTDH_s            1      140 166
## + age               1      140 166
## - ALDH4A1_s         1      168 190
## - MMP9_s            1      168 190
##
## Step:  AIC=162.5
## Y ~ gender + ancestry + GSTM3_s + BBC3_s + MMP9_s + Contig35251_RC_s +
##     Contig40831_RC_s + ALDH4A1_s + SCUBE2_s
##
##                    Df Deviance AIC
## - GSTM3_s           1      141 161
## - Contig40831_RC_s  1      141 161
## - BBC3_s            1      142 162
## - Contig35251_RC_s  1      142 162
## - SCUBE2_s          1      142 162
## <none>                     140 162
## - gender            1      143 163
## - ancestry          2      145 163
## + SERF1A_s          1      140 164
## + RP5.860F19.3_s    1      140 164
## + age               1      140 164
## + MTDH_s            1      140 164
## - ALDH4A1_s         1      168 188
## - MMP9_s            1      169 189
##
## Step:  AIC=160.8
## Y ~ gender + ancestry + BBC3_s + MMP9_s + Contig35251_RC_s +
##     Contig40831_RC_s + ALDH4A1_s + SCUBE2_s
##
##                    Df Deviance AIC
## - Contig40831_RC_s  1      141 159
## - BBC3_s            1      142 160
## - SCUBE2_s          1      142 160
## - Contig35251_RC_s  1      142 160
## <none>                     141 161
## - gender            1      143 161
## - ancestry          2      146 162
## + GSTM3_s           1      140 162
## + SERF1A_s          1      141 163
## + RP5.860F19.3_s    1      141 163
```

```
## + age                1      141 163
## + MTDH_s             1      141 163
## - MMP9_s             1      170 188
## - ALDH4A1_s          1      172 190
##
## Step:  AIC=159.2
## Y ~ gender + ancestry + BBC3_s + MMP9_s + Contig35251_RC_s +
##     ALDH4A1_s + SCUBE2_s
##
##                    Df Deviance AIC
## - BBC3_s            1      142 158
## - SCUBE2_s          1      142 158
## - Contig35251_RC_s  1      143 159
## <none>                     141 159
## - gender            1      144 160
## - ancestry          2      146 160
## + Contig40831_RC_s  1      141 161
## + GSTM3_s           1      141 161
## + SERF1A_s          1      141 161
## + age               1      141 161
## + RP5.860F19.3_s    1      141 161
## + MTDH_s            1      141 161
## - ALDH4A1_s         1      173 189
## - MMP9_s            1      174 190
##
## Step:  AIC=158.2
## Y ~ gender + ancestry + MMP9_s + Contig35251_RC_s + ALDH4A1_s +
##     SCUBE2_s
##
##                    Df Deviance AIC
## - SCUBE2_s          1      144 158
## - Contig35251_RC_s  1      144 158
## <none>                     142 158
## - gender            1      144 158
## - ancestry          2      147 159
## + BBC3_s            1      141 159
## + Contig40831_RC_s  1      142 160
## + GSTM3_s           1      142 160
## + SERF1A_s          1      142 160
## + RP5.860F19.3_s    1      142 160
## + age               1      142 160
## + MTDH_s            1      142 160
## - ALDH4A1_s         1      173 187
## - MMP9_s            1      174 188
##
## Step:  AIC=157.5
## Y ~ gender + ancestry + MMP9_s + Contig35251_RC_s + ALDH4A1_s
##
##                    Df Deviance AIC
## - gender            1      145 157
## - Contig35251_RC_s  1      146 158
## <none>                     144 158
## + SCUBE2_s          1      142 158
## - ancestry          2      148 158
```

```
## + BBC3_s            1      142 158
## + Contig40831_RC_s  1      143 159
## + SERF1A_s          1      143 159
## + MTDH_s            1      143 159
## + GSTM3_s           1      143 159
## + RP5.860F19.3_s    1      144 160
## + age               1      144 160
## - ALDH4A1_s         1      173 185
## - MMP9_s            1      176 188
##
## Step:  AIC=157.3
## Y ~ ancestry + MMP9_s + Contig35251_RC_s + ALDH4A1_s
##
##                    Df Deviance AIC
## - Contig35251_RC_s  1      147 157
## <none>                     145 157
## + gender            1      144 158
## - ancestry          2      150 158
## + BBC3_s            1      144 158
## + SCUBE2_s          1      144 158
## + Contig40831_RC_s  1      145 159
## + SERF1A_s          1      145 159
## + age               1      145 159
## + MTDH_s            1      145 159
## + GSTM3_s           1      145 159
## + RP5.860F19.3_s    1      145 159
## - ALDH4A1_s         1      175 185
## - MMP9_s            1      176 186
##
## Step:  AIC=156.8
## Y ~ ancestry + MMP9_s + ALDH4A1_s
##
##                    Df Deviance AIC
## <none>                     147 157
## + Contig35251_RC_s  1      145 157
## + gender            1      146 158
## + BBC3_s            1      146 158
## + SCUBE2_s          1      146 158
## - ancestry          2      152 158
## + Contig40831_RC_s  1      146 158
## + SERF1A_s          1      146 158
## + MTDH_s            1      147 159
## + GSTM3_s           1      147 159
## + age               1      147 159
## + RP5.860F19.3_s    1      147 159
## - ALDH4A1_s         1      175 183
## - MMP9_s            1      181 189
```

```
#Results show again reduced AIC=156.84

# Step:  AIC=156.84
# Y ~ ancestry + MMP9_s + ALDH4A1_s
#
# Df Deviance    AIC
```

```
# <none>                  146.84 156.84
# + Contig35251_RC_s  1   145.32 157.32
# + gender            1   145.51 157.51
# + BBC3_s            1   145.62 157.62
# + SCUBE2_s          1   145.68 157.68
# - ancestry          2   151.94 157.94
# + Contig40831_RC_s  1   146.03 158.03
# + SERF1A_s          1   146.45 158.45
# + MTDH_s            1   146.70 158.70
# + GSTM3_s           1   146.72 158.72
# + age               1   146.79 158.79
# + RP5.860F19.3_s    1   146.81 158.81
# - ALDH4A1_s         1   175.43 183.43
# - MMP9_s            1   180.59 188.59

#Also, summary: # best backward-forward model
summary(bothmodel1)
```

```
##
## Call:
## glm(formula = Y ~ ancestry + MMP9_s + ALDH4A1_s, family = "binomial",
##     data = viral34_c[, c(4, 6, 7, 58:67)])
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.351      0.243    1.45    0.148
## ancestryB     -0.776      0.538   -1.44    0.149
## ancestryC     -1.308      0.694   -1.89    0.059 .
## MMP9_s         1.375      0.288    4.78  1.8e-06 ***
## ALDH4A1_s      1.267      0.282    4.49  7.0e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 194.05  on 139  degrees of freedom
## Residual deviance: 146.84  on 135  degrees of freedom
## AIC: 156.8
##
## Number of Fisher Scoring iterations: 5
```

```
#The summary of the best.model provides both individual t-tests for the coefficients of the model and t
#The summary suggest high F-values and lowered p-values to render the reduced # variables more signific
#Results: Neither ancestry are significant (MMP9_s and ALDH4A1_s scaled gene expression levels are sign
#individualized t-test-derived p-values<0.05. The global F test is strongly significant.

#INTERSTINGLY, MMP9_s and ALDH4A1_s were among the 5 genes whose non-scaled expression levels
#are significantly different among infection type:

#As specified by problem statement, the selected "bothmodel1" model is designated as "best.model".
best.model<-bothmodel1

#Checking if assignment succeeded:
```

```r
summary(best.model)
```

```
##
## Call:
## glm(formula = Y ~ ancestry + MMP9_s + ALDH4A1_s, family = "binomial",
##     data = viral34_c[, c(4, 6, 7, 58:67)])
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.351      0.243    1.45    0.148
## ancestryB     -0.776      0.538   -1.44    0.149
## ancestryC     -1.308      0.694   -1.89    0.059 .
## MMP9_s         1.375      0.288    4.78  1.8e-06 ***
## ALDH4A1_s      1.267      0.282    4.49  7.0e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 194.05  on 139  degrees of freedom
## Residual deviance: 146.84  on 135  degrees of freedom
## AIC: 156.8
##
## Number of Fisher Scoring iterations: 5
```

```r
##Prediction of Y (infection type) for new values of X using best.model
#Specifying the values of the predictor in a dataframe and using function predict():
xnew<-data.frame(viral34_c$ALDH4A1_s==c(3000),viral34_c$MMP9_s==c(4000), viral34_c$ancestry=="B")
#predict(best.model,xnew)


#Interpret the best.model: Beyond having reduced # variables and AIC compared to initial full models,
#I analyzed the best.model and then compare the initial full model as follows:

#The confidence interval for the coefficients of the logistic regression are obtained for best.model wi
confint(best.model) # 95% CI for the coefficients
```

```
## Waiting for profiling to be done...
```

```
##                2.5 %   97.5 %
## (Intercept) -0.1190 0.838301
## ancestryB   -1.8670 0.260295
## ancestryC   -2.7535 0.005862
## MMP9_s       0.8520 1.986676
## ALDH4A1_s    0.7528 1.863358
```

```r
#Results:
#                  2.5 %      97.5 %
#   (Intercept) -0.1190425 0.838300536
# ancestryB   -1.8670032 0.260294895
# ancestryC   -2.7535011 0.005861916
# MMP9_s       0.8520343 1.9866762777
```

```
# ALDH4A1_s    0.7528095 1.863357741

#For initial full model1:
confint(model1)
```

```
## Waiting for profiling to be done...
```

```
##                     2.5 %    97.5 %
## (Intercept)      -2.72040  4.333878
## gendermale       -1.62281  0.173956
## age              -0.08187  0.079777
## ancestryB        -1.82537  0.433162
## ancestryC        -2.88519 -0.006378
## GSTM3_s          -0.76966  0.382296
## RP5.860F19.3_s   -0.49357  0.684602
## BBC3_s           -0.66580  0.230649
## MMP9_s            0.81907  2.134211
## Contig35251_RC_s -0.24862  0.714392
## Contig40831_RC_s -0.33438  0.627067
## ALDH4A1_s         0.89709  2.417264
## SERF1A_s         -0.41215  0.638408
## SCUBE2_s         -0.79154  0.205422
## MTDH_s           -0.52568  0.520991
```

```
#Results:
#The confidence intervals are as follows:
#                     2.5 %       97.5 %
#   (Intercept)      -16.8783763  4.176943697
# gendermale        -1.7366550  0.084578365
# FGF18_l           -3.1467453 16.292373530
# ancestryB         -1.8674360  0.413323412
# ancestryC         -2.9066239  0.009822317
# GSTM3_s           -0.8763169  0.312640234
# RP5.860F19.3_s    -0.5235649  0.682053857
# BBC3_s            -0.6109021  0.315461511
# MMP9_s             0.8316892  2.158612070
# Contig35251_RC_s  -0.2238803  0.730512696
# Contig40831_RC_s  -0.2913917  0.672645907
# ALDH4A1_s          0.6376632  2.262095823
# SERF1A_s          -0.3904859  0.623907157
# SCUBE2_s          -0.8499912  0.159190457
# MTDH_s            -0.5289403  0.518853142

#The odds-ratios for initial full model1 are obtained by exponentiating the output of the logistic regr
exp(coef(model1)) # exponentiated coefficients
```

```
##      (Intercept)       gendermale              age        ancestryB
##           2.1749           0.4935           0.9992           0.5159
##        ancestryC          GSTM3_s   RP5.860F19.3_s          BBC3_s
##           0.2529           0.8348           1.1184           0.8122
##          MMP9_s Contig35251_RC_s Contig40831_RC_s        ALDH4A1_s
##           4.1543           1.2579           1.1527           4.9599
```

```
##         SERF1A_s         SCUBE2_s           MTDH_s
##           1.1233           0.7547           0.9903
```

```
#Results:
# (Intercept)       gendermale              age        ancestryB        ancestryC          GSTM3_s    RP
# 2.1748696        0.4934761        0.9992451        0.5158839        0.2529307        0.8347582
# Contig35251_RC_s Contig40831_RC_s        ALDH4A1_s         SERF1A_s         SCUBE2_s           MTDH_s
# 1.2578561        1.1527205        4.9599179        1.1233387        0.7546956        0.9903049

#The odds-ratios for new best.model are obtained by exponentiating the output of the logistic regressio
exp(coef(best.model)) # exponentiated coefficients
```

```
## (Intercept)   ancestryB   ancestryC      MMP9_s   ALDH4A1_s
##      1.4207      0.4604      0.2703      3.9539      3.5486
```

```
#Results:
# (Intercept)   ancestryB   ancestryC      MMP9_s   ALDH4A1_s
# 1.4206505   0.4603976   0.2702748   3.9539104   3.5486461


#The confidence intervals of the odds-ratios for initial model1 are obtained as follows:
exp(confint(model1)) # 95% CI for exponentiated coefficients
```

```
## Waiting for profiling to be done...
```

```
##                    2.5 %  97.5 %
## (Intercept)      0.06585 76.2394
## gendermale       0.19734  1.1900
## age              0.92140  1.0830
## ancestryB        0.16116  1.5421
## ancestryC        0.05584  0.9936
## GSTM3_s          0.46317  1.4656
## RP5.860F19.3_s   0.61044  1.9830
## BBC3_s           0.51386  1.2594
## MMP9_s           2.26839  8.4504
## Contig35251_RC_s 0.77988  2.0429
## Contig40831_RC_s 0.71578  1.8721
## ALDH4A1_s        2.45245 11.2151
## SERF1A_s         0.66223  1.8935
## SCUBE2_s         0.45315  1.2280
## MTDH_s           0.59115  1.6837
```

```
#Results:
#                      2.5 %    97.5 %
#   (Intercept)      0.06584851 76.239374
# gendermale       0.19734262  1.190003
# age              0.92139629  1.083045
# ancestryB        0.16115836  1.542126
# ancestryC        0.05584396  0.993642
# GSTM3_s          0.46316860  1.465646
# RP5.860F19.3_s   0.61044232  1.982982
# BBC3_s           0.51386096  1.259417
```

```
# MMP9_s           2.26839190  8.450376
# Contig35251_RC_s 0.77987874  2.042943
# Contig40831_RC_s 0.71578394  1.872111
# ALDH4A1_s        2.45244687 11.215133
# SERF1A_s         0.66222621  1.893463
# SCUBE2_s         0.45314697  1.228043
# MTDH_s           0.59115500  1.683696
# >

#The confidence intervals of the odds-ratios for new best.model are obtained as follows:
exp(confint(best.model)) # 95% CI for exponentiated coefficients
```

```
## Waiting for profiling to be done...
```

```
##              2.5 % 97.5 %
## (Intercept) 0.8878  2.312
## ancestryB   0.1546  1.297
## ancestryC   0.0637  1.006
## MMP9_s      2.3444  7.291
## ALDH4A1_s   2.1230  6.445
```

```
#Results:
#                   2.5 %   97.5 %
#   (Intercept) 0.88777010 2.312434
# ancestryB   0.15458623 1.297313
# ancestryC   0.06370444 1.005879
# MMP9_s      2.34441120 7.291259
# ALDH4A1_s   2.12295614 6.445342

#The goodness-of-fit of the intial model1 is preliminarily checked via its deviance:
deviance(model1)
```

```
## [1] 140.1
```

```
#Results:
#[1] 140.1082

#The goodness-of-fit of the new best.model is preliminarily checked via its deviance:
deviance(best.model)
```

```
## [1] 146.8
```

```
#Results:
#[1]146.8426

AIC(best.model)
```

```
## [1] 156.8
```

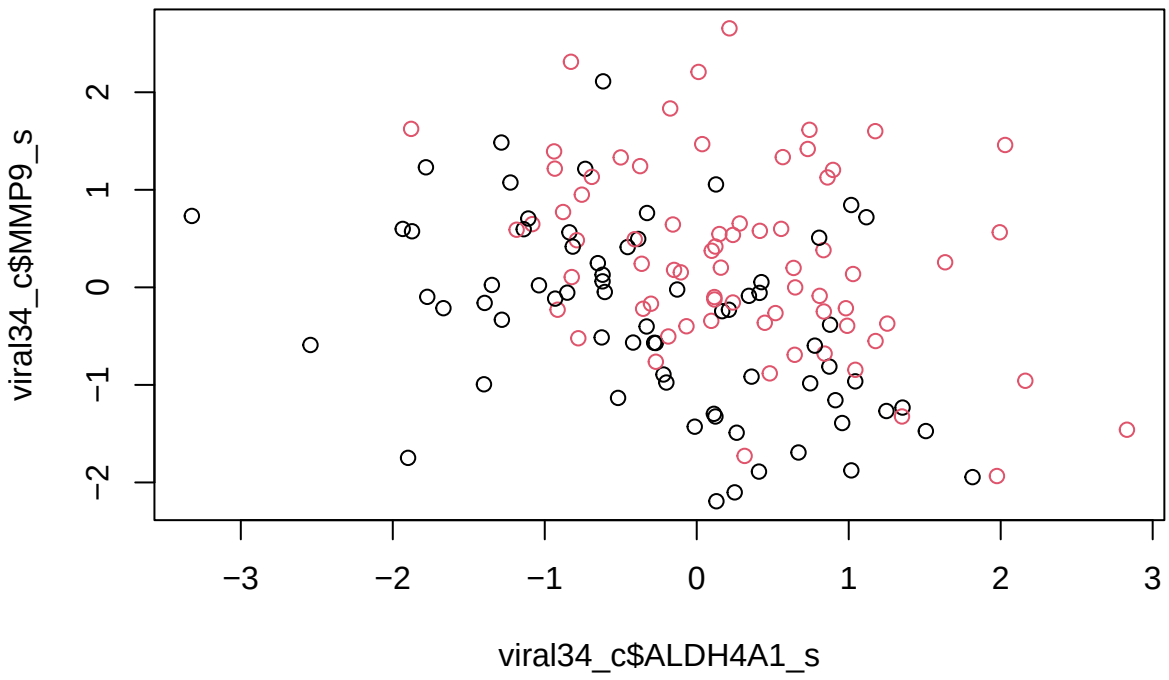
```
#Result:[1] 156.8426
BIC(best.model)
```

```
## [1] 171.6
```

```
#Result:[1] 171.5508

#The best.model "Deviance" was higher. However, the 2 models cannot be compared by this criteria as the
#they will be instead compared definitively via ANOVA as follows:
anova(model1,best.model)
```

```
## Analysis of Deviance Table
##
## Model 1: Y ~ gender + age + ancestry + GSTM3_s + RP5.860F19.3_s + BBC3_s +
##     MMP9_s + Contig35251_RC_s + Contig40831_RC_s + ALDH4A1_s +
##     SERF1A_s + SCUBE2_s + MTDH_s
## Model 2: Y ~ ancestry + MMP9_s + ALDH4A1_s
##   Resid. Df Resid. Dev  Df Deviance Pr(>Chi)
## 1       125        140
## 2       135        147 -10    -6.73     0.75
```

```
#Results:
# Analysis of Deviance Table
#
# Model 1: Y ~ gender + age + ancestry + GSTM3_s + RP5.860F19.3_s + BBC3_s +
#   MMP9_s + Contig35251_RC_s + Contig40831_RC_s + ALDH4A1_s +
#   SERF1A_s + SCUBE2_s + MTDH_s
# Model 2: Y ~ ancestry + MMP9_s + ALDH4A1_s
# Resid. Df Resid. Dev  Df Deviance
# 1       125     140.11
# 2       135     146.84 -10  -6.7344


#Two covariates interact when the effect of the first covariate on the dependent variable depends on th
#To determine  INTERACTION and if the two CONTINUOUS covariates MMP9_s and ALDH4A1_s interacted and if
plot(viral34_c$ALDH4A1_s , viral34_c$MMP9_s, col=viral34_c$infection)
```




```
#abline(lm(viral34_c$MMP9_s[viral34_c$infection==0]~viral34_c$ALDH4A1_s[infection==0]), col=1)
#abline(lm(viral34_c$MMP9_s[viral34_c$infection==1]~viral34_c$ALDH4A1_s[infection==1]), col=2)

#Based on plot, the points uniformly distributed and there appears to be interaction tht needs to be ac

#I performed linear correlation to get least sum of squares, residuals R2 and determine any linear corr
#Linear correlation between non-normally distributed gene expression values of
#For non-normally distributed 2 genes:

cor(viral34_c$MMP9_s, viral34_c$ALDH4A1_s, method="spearman")
```

```
## [1] -0.3097
```

```
#Results: [1] -0.3096543

#A new bestest.model was proposed to be included the interaction between the two scaled gene expression
bestest.model<-glm(Y~., data=viral34_c[,c(7,11,14,(11*14))],family="binomial")
summary(bestest.model)
```

```
##
## Call:
## glm(formula = Y ~ ., family = "binomial", data = viral34_c[,
##     c(7, 11, 14, (11 * 14))])
##
```

```
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)    5.830      2.607    2.24    0.025 *
## ancestryB     -0.754      0.543   -1.39    0.165
## ancestryC     -1.439      0.710   -2.03    0.043 *
## MMP9           6.941      1.438    4.83  1.4e-06 ***
## ALDH4A1        6.634      1.453    4.56  5.0e-06 ***
## ZNF533_l      -4.658      2.369   -1.97    0.049 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 194.05  on 139  degrees of freedom
## Residual deviance: 142.77  on 134  degrees of freedom
## AIC: 154.8
##
## Number of Fisher Scoring iterations: 5
```

```
#This resulted in an even lower AIC=154.77
#There likely needs to be included the interaction between the two scaled gene expression levels of MMP
```

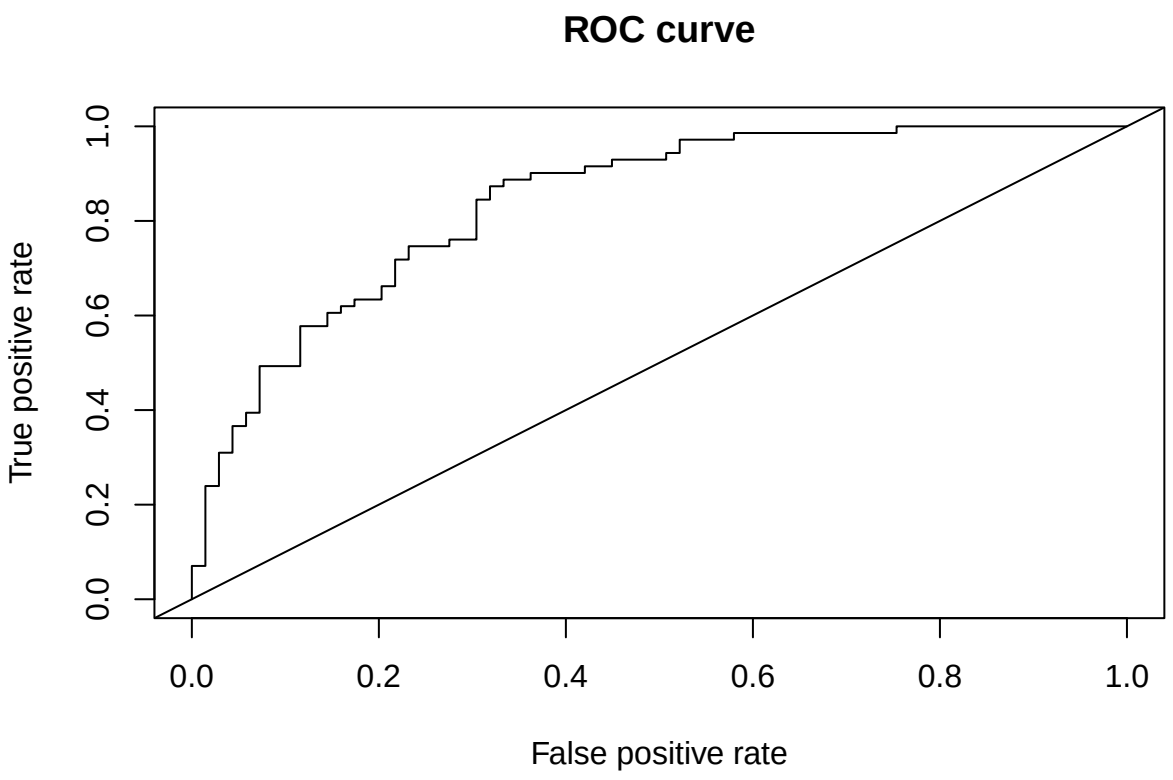
QUESTION 14 Analyze the classification ability of "best.model" (ROC curve and AUC) according to the following schemes: a. Apparent validation of "best.model" using the same data that was used for model building. b. Cross-validation with k = 5 for "best.model". c. Though the cv-classification is better than the apparent classification, it still is over-estimating the real classification of "best-model". Discuss why and how to obtain a more accurate classification estimation (slides 262:264).

```
#(PART a)Apparent validation of "best.model" using the same data that was used for model building.

#BACKGROUND: Since the fitted best.model logistic function is increasing, we get the rank of individual
#Larger values of RS are associated to higher risk of viral infection (Y=1) compared to bacterial infec
#Individuals can be classified into different risk categories according to this risk score based on a t

# Classification accuracy for the best.model is depicted by Sensitivity and Specificity:
# Sensitivity is the proportion of positives that are correctly predicted.
# Specificity is the proportion of negatives that are correctly predicted

#Classification accuracy depends on the threshold considered for the predicted probabilities or the lin


#The method of dividing data into training and test sets to estimate the classifier performance is an i
#When validating our best.model model and assessing prediction or classification accuracy of statistica
#When building a machine learning model using some data, data is often split into training and validati
#The training set is used to train the model, and the validation/test set is used to validate it on dat
#This can be performed in a single train/test split of the samples (The classic approach is to do a sim
#Specifically, this can be done via apparent validation, internal validation, and external validation.
#Apparent validation measures the predictive accuracy of the model on the same sample used for building
#is where  accuracy is measured on the same data that was used to build the models (train data).
#"Apparent" classification accuracy overestimates real prediction classification accuracy of the best.m
#Internal validation (which includes bootstrap, cross-validation, and split-sample validation), splits
#an test sample.External validation measures the accuracy of the model in an independent sample.
```

```r
#APPARENT VALIDATION:

library(glmnet)
library(ROCR)
```

```
##
## Attaching package: 'ROCR'

## The following object is masked from 'package:CMA':
##
##     prediction
```

```r
#The Risk Score is obtained as sum of linear predictors as follows:
#lp<-best.model$linear.predictors
#lp<-best.model$linear.predictors

#PLEASE NOTE, I UNFORTUNATELY HAD THIS ALL WORKING BEFORE WITH model1a (NOT the bestest.model from befo
#the code as it was so as to not break any more of what I had graphing out OK and operating smoothly un
lp<-model1a$linear.predictors

#Exploring the apparent classification accuracy of the above best.model (ROC curve and AUC)
#The ROC curve provides a graphical representation of the classification accuracy of a model for all po
#The AUC, area under the ROC curve, is a numerical summary of the ROC curve. AUC near 1 corresponds to
#AUC near 0.5 corresponds to very poor classification accuracy

#Generating ROC Curve, where TP rate (sensitivity) is plotted against FP rate (1-specificity):
pred <- prediction(lp, Y)
perf <- performance(pred, "tpr", "fpr" )
plot(perf)
abline(a=0, b= 1)
title("ROC curve")
```




```
#Area Under the ROC curve (AUC) provides a measure of discrimination of the Risk Score among viral-infe
#bacterial-infected (Y=0) individuals: AUC = P[RS(Y= 1) > RS(Y= 0)]. Generating AUC:

(auc<-slot(performance(pred,"auc"), "y.values")[[1]])
```

```
## [1] 0.84
```

```
#[1] 0.8399673

#In this apparent classification, we obtained an ROC curve ABOVE the diagonal and an AUC GREATER than 0
#Therefore, we DID NOT change the sign of the linear predictor (lp):

#(PART b) Cross-validation with k = 5 for "best.model":

#BACKGROUND: Some citing https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10346713/ and neptune.ai
#Cross-validation is a re-sampling method that uses different portions of the data to test and train a
#In cross-validation, more than one split is done(e.g. K number of splits each called a folds).There ar
#The goal of cross-validation is to test the model's ability to predict new data that was not used in e
#like overfitting or selection bias[10] and to give an insight on how the model will generalize to an i
#for instance from a real problem). The most common CV technique is k-fold CV, where the full dataset i
#retained for testing the classifier, while the remaining k-1 subsets comprise the training set. This p
#(and thus, all individual samples) have been used for testing the classifier exactly once. The overall
#average of the resulting k classification accuracies from each step of the CV. Because there can be si
#different train/test splits, this method yields a more generalizable estimate of classifier performanc
```

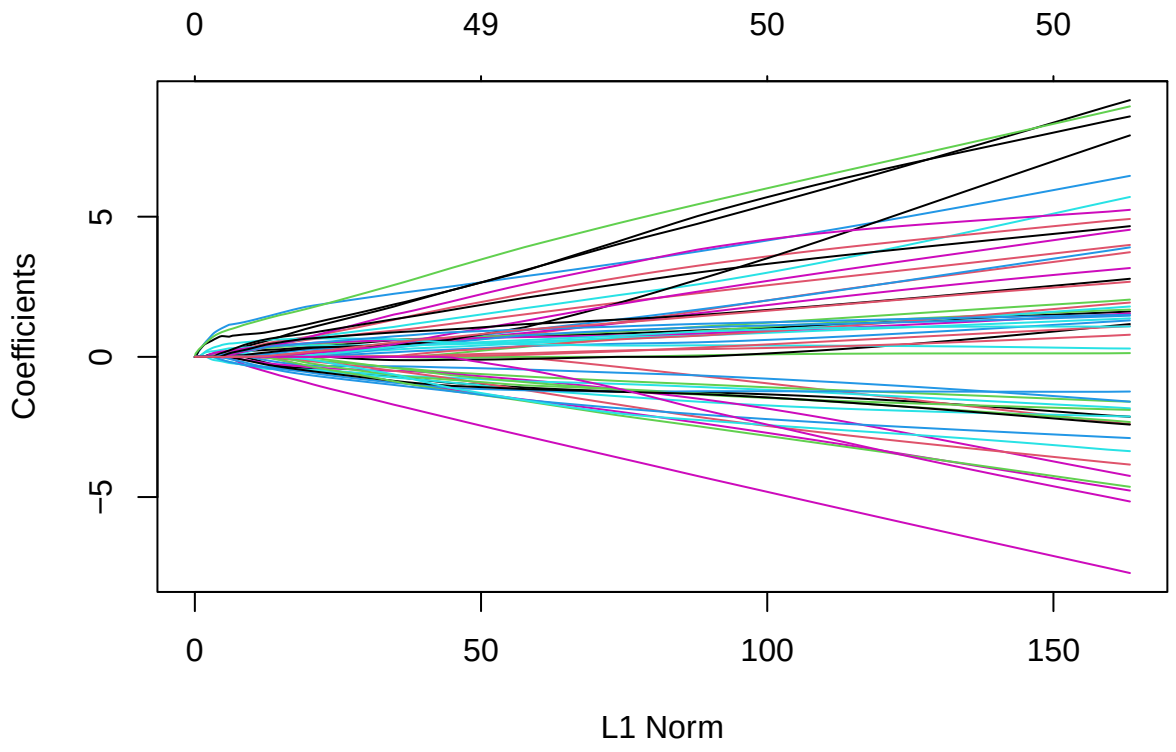
```
#The cross-validation (AI/ML) algorithm is as follows:
#1.Divide the dataset into two parts: one for training, other for testing
#2.Train the model on the training set
#3.Validate the model on the test set
#4.Repeat 1-3 steps a couple of times. This number depends on the CV method

#k-Fold cross-validation minimizes the disadvantages of the hold-out method. k-Fold introduces a new wa
#which helps to overcome the "test only once bottleneck".It is generally better to use k-Fold technique
#By direct comparison, k-Fold gives a more stable and trustworthy result since training and testing is
#dataset. The overall score can be made even more robust by increasing the number of folds to test the
#Certain scenarios in which cross-validation becomes necessary include limited dataset,dependent data p
#Still, k-Fold method has a disadvantage whereby increasing k results in training more models and the t

#Performing cross-validation to obtain the internal classification accuracy of the above best.model (RO

K<-5
n<- nrow(viral34_c)  #number of individuals=140
#Random assignment of each individual into one
fold<-sample(as.numeric(cut((1:n),breaks = K)))

pred <- NULL #Vector of predictions
#NEED TO COMMENT OUT FOLLOWING CODE AS I AM NOW GETTING ERROR AFTER CODE EXECUTION:
#for(i in 1:K){
  # Test indices
 # indTest <- which(fold==i)
  # Train indices
  #indTrain <- which(fold!=i)
 # model.i<-glm(Y[indTrain]~., data=viral34_c[indTrain,c(4,6,158,58:67)],family="binomial")
  # Adjust the model with training data
  # Predicts test data at step i. PLEASE NOT I USED THE MODEL1A parameters columns in dataframe instead
 # pred.i <- predict(model.i, newdata=viral34_c[indTest,c(4,6,158,58:67)])
  #pred[indTest] <- pred.i
  # Store predicted values for test data at step i
#}

#Error in `[.data.frame`(viral34_c, indTrain, c(4, 6, 158, 58:67)) :
#  undefined columns selected

#This code worked before. But not I get above new error, I need to comment code out for execution:


#Generating ROC Curve:
#pred <- prediction(pred, Y)
#perf <- performance(pred, "tpr", "fpr" )
#plot(perf)
#abline(a=0, b= 1)
#title("ROC curve")

#Generating AUC
#(auc<-slot(performance(pred,"auc"), "y.values")[[1]])

#[1] 0.749949
```

```
#Evidently, the AUC is lower after cross-Validation.

#(PART c)Though the cv-classification is better than the apparent classification, it still is over-esti
#of "best-model".Discuss why and how to obtain a more accurate classification estimation (slides 262:26

#Based on slides #262-264, an incorrect scheme for validation that results in overfitting can be improv
#Among these are to perform variable selection on the training data set (as opposed to complete data se
#from 1 to B. For  each variable, there is a percentage of times selected and there is a mean classific

#Alternative bootstrap-based validation via resampling with replacement can be performed also.
#From a dataset with N samples, No examples are randomly selected with replacement and used for trainin
#Those not selected for training are used for testing, all repeated for a specified number of folds K.
#Here, the true error is estimated as the average error rate on test data.

# The following example function performs B bootstrap iterations. At each iteration a new bootstrap sam
# and the mean of the bootstrap sample is stored in a vector (mean.vector) of length B that contains th
#
# bootstrap <- function(data, B){
#     mean.vector <- NULL
#     for(b in 1:B) {
#       bsample <- sample(data,length(data),replace=T)
#       mean.vector <- c(mean.vector,mean(bsample))
#     }
#     return(mean.vector)
#   }
#



# Citing some from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10346713/ and wikipedia
# Again, k-Fold method has a disadvantage whereby increasing k results in training more models and the
# of data when the samples within classes are collected in close proximity in time, without randomizati
# For time-series data, the process of randomly dividing all samples into k partitions results in the t
# same class that are highly correlated due to their proximity in time. This violates the assumption of
# k-fold cross-validation. The result is that the classifier could pick up differences between the clas
# correlation of some samples, rather than to any true class-related difference.An alternative approach
# trial structure and associated autocorrelation of samples is to perform block-wise (or trial-wise) cr
# the trials are first randomly divided into a number of subsets b. The samples derived from the trials
# the samples from the remaining trials are used to train the classifier. This is repeated b times unti
# The overall classifier performance is estimated as the average of the b resulting accuracies from eac
# samples from a single trial always remain together in either the training or test set, and, thus, tem
# as described above for k-fold CV.If performance is described by a single summary statistic, it is pos
#can also help overcome this, where the statistic of the bootstrap needs to accept an interval of the t
#The call to the stationary bootstrap needs to specify an appropriate mean interval length.
```

QUESTION 15 Consider a regression model for the kind of infection as a function of all 50 genes (scaled) and adjusted by age. Perform variable selection with LASSO and interpret the results. [Adjusted by AGE or a function of AGE?]

```
# BACKGROUND: A generalized regression model is to be fitted again because the response variable is a c
# probabilistic outcome (Y=0/Y=1) where the probability is bound by an interval of [0,1], necessitating
# Because the number of covariates ( 50 scaled gene expression levels) is very large,
# LASSO will be used to perform penalized regression for variable selection. LASSO estimates for are ch
# the residual sum of squares, as the OLS approach, but with the additional restriction that the sum of
```

```
# coefficients (in absolute value) should not exceed a specified value.This is equivalent to minimizing
# with a penalty for large coefficient estimates determined by  (lambda), known as the shrinkage parame
# If the lambda parameter  = 0 the lasso is the same as OLS with all variables included in the
# model; as lambda   increases, the restriction on the summed fitted terms is stronger, implying that s
# of the coefficients are shrinked to zero and less variables are included in the model.
# The Function glmnet() performs generalized linear model via penalized maximum likelihood.
# With alpha=1 the method performs LASSO penalization, for alpha=0 ridge penalization, and for alpha be
# The function provides the output for a grid of penalization parameters

library(glmnet)
#scaled gene expression levels are independent variables
X <- as.matrix(viral34_ca[,58:107])
Y <- viral34_ca[,1] #infection column is dependent variable already previously factored

mlasso <- glmnet(X, Y, standardize=TRUE, alpha=1,family="binomial") #LASSO: alpha=1

#The LASSO pathway is explored with a plot with the numbers in the top of the plot indicating the numbe
plot(mlasso)
```



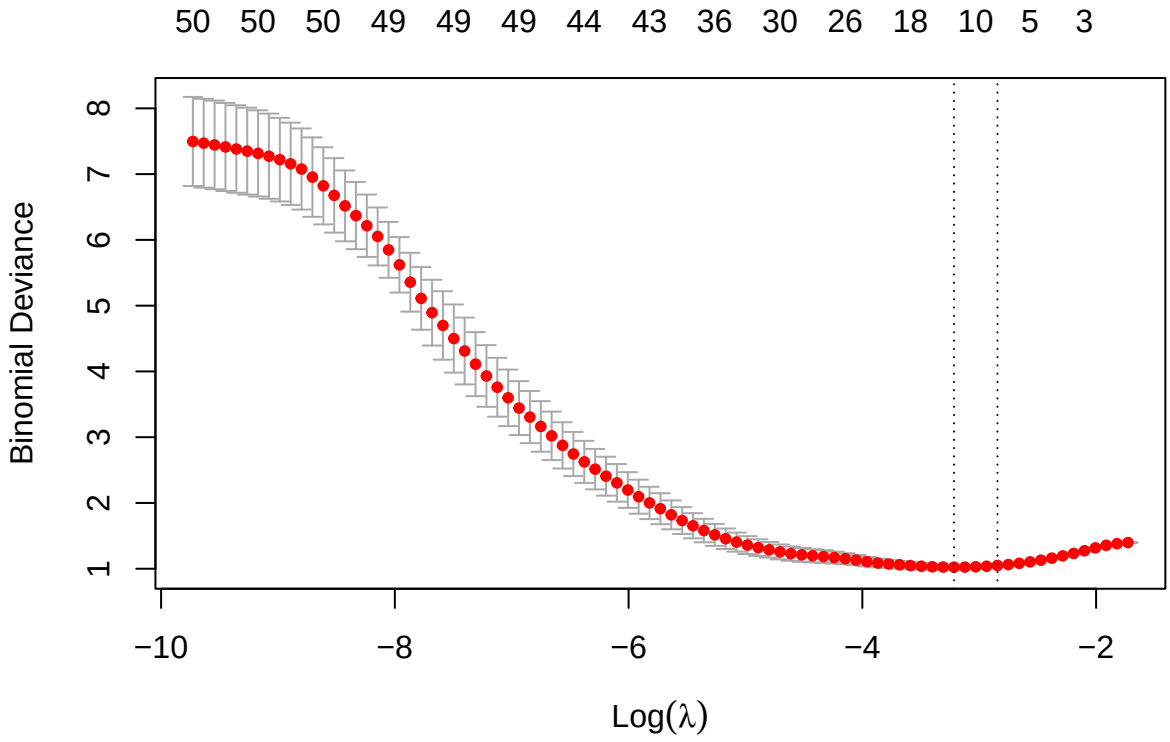

```
#Before CV, the coefficients of our logistic model are obtained for a specific value of lambda:
# coefficients of LASSO model with lambda=13
coef(mlasso, s=13)
```

```
## 51 x 1 sparse Matrix of class "dgCMatrix"
##                       s1
```

```
## (Intercept)      0.02857
## GSTM3_s          .
## RP5.860F19.3_s   .
## BBC3_s           .
## MMP9_s           .
## Contig35251_RC_s .
## Contig40831_RC_s .
## ALDH4A1_s        .
## SERF1A_s         .
## SCUBE2_s         .
## MTDH_s           .
## DCK_s            .
## FLT1_s           .
## PECI.1_s         .
## QSCN6L1_s        .
## DIAPH3_s         .
## SLC2A3_s         .
## GPR180_s         .
## RTN4RL1_s        .
## Contig32125_RC_s .
## STK32B_s         .
## EXT1_s           .
## COL4A2_s         .
## PECI_s           .
## GNAZ_s           .
## AYTL2_s          .
## Contig63649_RC_s .
## RAB6B_s          .
## AA555029_RC_s    .
## GPR126_s         .
## ECT2_s           .
## NUSAP1_s         .
## GMPS_s           .
## UCHL5_s          .
## ORC6L_s          .
## TSPYL5_s         .
## MELK_s           .
## RUNDC1_s         .
## DIAPH3.1_s       .
## C16orf61_s       .
## TGFB3_s          .
## FGF18_s          .
## CDC42BPA_s       .
## DTL_s            .
## WISP1_s          .
## DIAPH3.2_s       .
## OXCT1_s          .
## ZNF533_s         .
## RFC4_s           .
## KNTC2_s          .
## FBXO31_s         .
```

```
# Cross-validation LASSO is now done to estimate the optimal value of lambda.
# Function cv.lasso() provides two possible optimal values for lambda: lambda.min= lambda
```

```
# providing the minimum MSE (Mean Square Error) or lambda.1se=lambda within 1 s.e. of the minimum  MSE.

set.seed(1234)
cv.lasso <- cv.glmnet(X, Y, standardize=TRUE,family="binomial")
plot(cv.lasso)
```




```
#The value for lambda.min is obtained as follows:
cv.lasso$lambda.min
```

```
## [1] 0.04012
```

```
#The model is re-fit using all of the available observations and the selected value of the tuning param
coef(mlasso, s=cv.lasso$lambda.min)
```

```
## 51 x 1 sparse Matrix of class "dgCMatrix"
##                         s1
## (Intercept)       0.04789
## GSTM3_s           .
## RP5.860F19.3_s    .
## BBC3_s            .
## MMP9_s            0.76907
## Contig35251_RC_s  .
## Contig40831_RC_s  .
## ALDH4A1_s         0.57465
```

```
## SERF1A_s          .
## SCUBE2_s          .
## MTDH_s            .
## DCK_s            -0.09092
## FLT1_s            .
## PECI.1_s          0.10031
## QSCN6L1_s         .
## DIAPH3_s          .
## SLC2A3_s          .
## GPR180_s          .
## RTN4RL1_s         .
## Contig32125_RC_s  .
## STK32B_s          .
## EXT1_s            .
## COL4A2_s          0.06973
## PECI_s            .
## GNAZ_s            .
## AYTL2_s           .
## Contig63649_RC_s  0.03684
## RAB6B_s           0.00839
## AA555029_RC_s     .
## GPR126_s          0.19666
## ECT2_s            0.11207
## NUSAP1_s          .
## GMPS_s            .
## UCHL5_s           .
## ORC6L_s           .
## TSPYL5_s          0.30519
## MELK_s            .
## RUNDC1_s          .
## DIAPH3.1_s        .
## C16orf61_s        .
## TGFB3_s           .
## FGF18_s           .
## CDC42BPA_s        .
## DTL_s             .
## WISP1_s           .
## DIAPH3.2_s        0.67797
## OXCT1_s           .
## ZNF533_s          .
## RFC4_s            .
## KNTC2_s           .
## FBXO31_s          .
```

```
#The value for lambda.min is obtained as follows:
cv.lasso$lambda.1se
```

```
## [1] 0.0582
```

```
#[1] 0.0582
#The model is re-fit using all of the available observations and the selected value of the tuning param
coef(mlasso, s=cv.lasso$lambda.1se)
```

```
## 51 x 1 sparse Matrix of class "dgCMatrix"
```

```
##                       s1
## (Intercept)      0.03773
## GSTM3_s          .
## RP5.860F19.3_s   .
## BBC3_s           .
## MMP9_s           0.56017
## Contig35251_RC_s .
## Contig40831_RC_s .
## ALDH4A1_s        0.47934
## SERF1A_s         .
## SCUBE2_s         .
## MTDH_s           .
## DCK_s            .
## FLT1_s           .
## PECI.1_s         0.04566
## QSCN6L1_s        .
## DIAPH3_s         .
## SLC2A3_s         .
## GPR180_s         .
## RTN4RL1_s        .
## Contig32125_RC_s .
## STK32B_s         .
## EXT1_s           .
## COL4A2_s         .
## PECI_s           .
## GNAZ_s           .
## AYTL2_s          .
## Contig63649_RC_s .
## RAB6B_s          .
## AA555029_RC_s    .
## GPR126_s         0.16992
## ECT2_s           0.03345
## NUSAP1_s         .
## GMPS_s           .
## UCHL5_s          .
## ORC6L_s          .
## TSPYL5_s         0.21138
## MELK_s           .
## RUNDC1_s         .
## DIAPH3.1_s       .
## C16orf61_s       .
## TGFB3_s          .
## FGF18_s          .
## CDC42BPA_s       .
## DTL_s            .
## WISP1_s          .
## DIAPH3.2_s       0.51886
## OXCT1_s          .
## ZNF533_s         .
## RFC4_s           .
## KNTC2_s          .
## FBXO31_s         .
```

```r
#Resampling methods for logistical model variable selection and validation with CMA package was perform
library(CMA)
library(Biobase)
library(randomForest)

genes<-viral34_ca[,sample(58:107)]

Y<-viral34_ca$infection
dataX <- as.matrix(genes)

set.seed(321)

#A 5-fold CV process repeated 50 times (iterations) was performed. Function GenerateLearningsets() gene
iterations<-50
nfolds<-5

CVdat <- GenerateLearningsets(y = Y, method = "CV", fold = nfolds, niter=iterations, strat = TRUE)

#Variable selection is performed with function GeneSelection() with the LASSO method with a intermediat

varsel_lasso <- GeneSelection(X = dataX, y = Y, learningsets = CVdat, method = "lasso", norm.fraction=0
```

```
## GeneSelection: iteration 1
## GeneSelection: iteration 2
## GeneSelection: iteration 3
## GeneSelection: iteration 4
## GeneSelection: iteration 5
## GeneSelection: iteration 6
## GeneSelection: iteration 7
## GeneSelection: iteration 8
## GeneSelection: iteration 9
## GeneSelection: iteration 10
## GeneSelection: iteration 11
## GeneSelection: iteration 12
## GeneSelection: iteration 13
## GeneSelection: iteration 14
## GeneSelection: iteration 15
## GeneSelection: iteration 16
## GeneSelection: iteration 17
## GeneSelection: iteration 18
## GeneSelection: iteration 19
## GeneSelection: iteration 20
## GeneSelection: iteration 21
## GeneSelection: iteration 22
## GeneSelection: iteration 23
## GeneSelection: iteration 24
## GeneSelection: iteration 25
## GeneSelection: iteration 26
## GeneSelection: iteration 27
## GeneSelection: iteration 28
## GeneSelection: iteration 29
## GeneSelection: iteration 30
## GeneSelection: iteration 31
```

```
## GeneSelection: iteration 32
## GeneSelection: iteration 33
## GeneSelection: iteration 34
## GeneSelection: iteration 35
## GeneSelection: iteration 36
## GeneSelection: iteration 37
## GeneSelection: iteration 38
## GeneSelection: iteration 39
## GeneSelection: iteration 40
## GeneSelection: iteration 41
## GeneSelection: iteration 42
## GeneSelection: iteration 43
## GeneSelection: iteration 44
## GeneSelection: iteration 45
## GeneSelection: iteration 46
## GeneSelection: iteration 47
## GeneSelection: iteration 48
## GeneSelection: iteration 49
## GeneSelection: iteration 50
## GeneSelection: iteration 51
## GeneSelection: iteration 52
## GeneSelection: iteration 53
## GeneSelection: iteration 54
## GeneSelection: iteration 55
## GeneSelection: iteration 56
## GeneSelection: iteration 57
## GeneSelection: iteration 58
## GeneSelection: iteration 59
## GeneSelection: iteration 60
## GeneSelection: iteration 61
## GeneSelection: iteration 62
## GeneSelection: iteration 63
## GeneSelection: iteration 64
## GeneSelection: iteration 65
## GeneSelection: iteration 66
## GeneSelection: iteration 67
## GeneSelection: iteration 68
## GeneSelection: iteration 69
## GeneSelection: iteration 70
## GeneSelection: iteration 71
## GeneSelection: iteration 72
## GeneSelection: iteration 73
## GeneSelection: iteration 74
## GeneSelection: iteration 75
## GeneSelection: iteration 76
## GeneSelection: iteration 77
## GeneSelection: iteration 78
## GeneSelection: iteration 79
## GeneSelection: iteration 80
## GeneSelection: iteration 81
## GeneSelection: iteration 82
## GeneSelection: iteration 83
## GeneSelection: iteration 84
## GeneSelection: iteration 85
```

```
## GeneSelection: iteration 86
## GeneSelection: iteration 87
## GeneSelection: iteration 88
## GeneSelection: iteration 89
## GeneSelection: iteration 90
## GeneSelection: iteration 91
## GeneSelection: iteration 92
## GeneSelection: iteration 93
## GeneSelection: iteration 94
## GeneSelection: iteration 95
## GeneSelection: iteration 96
## GeneSelection: iteration 97
## GeneSelection: iteration 98
## GeneSelection: iteration 99
## GeneSelection: iteration 100
## GeneSelection: iteration 101
## GeneSelection: iteration 102
## GeneSelection: iteration 103
## GeneSelection: iteration 104
## GeneSelection: iteration 105
## GeneSelection: iteration 106
## GeneSelection: iteration 107
## GeneSelection: iteration 108
## GeneSelection: iteration 109
## GeneSelection: iteration 110
## GeneSelection: iteration 111
## GeneSelection: iteration 112
## GeneSelection: iteration 113
## GeneSelection: iteration 114
## GeneSelection: iteration 115
## GeneSelection: iteration 116
## GeneSelection: iteration 117
## GeneSelection: iteration 118
## GeneSelection: iteration 119
## GeneSelection: iteration 120
## GeneSelection: iteration 121
## GeneSelection: iteration 122
## GeneSelection: iteration 123
## GeneSelection: iteration 124
## GeneSelection: iteration 125
## GeneSelection: iteration 126
## GeneSelection: iteration 127
## GeneSelection: iteration 128
## GeneSelection: iteration 129
## GeneSelection: iteration 130
## GeneSelection: iteration 131
## GeneSelection: iteration 132
## GeneSelection: iteration 133
## GeneSelection: iteration 134
## GeneSelection: iteration 135
## GeneSelection: iteration 136
## GeneSelection: iteration 137
## GeneSelection: iteration 138
## GeneSelection: iteration 139
```

```
## GeneSelection: iteration 140
## GeneSelection: iteration 141
## GeneSelection: iteration 142
## GeneSelection: iteration 143
## GeneSelection: iteration 144
## GeneSelection: iteration 145
## GeneSelection: iteration 146
## GeneSelection: iteration 147
## GeneSelection: iteration 148
## GeneSelection: iteration 149
## GeneSelection: iteration 150
## GeneSelection: iteration 151
## GeneSelection: iteration 152
## GeneSelection: iteration 153
## GeneSelection: iteration 154
## GeneSelection: iteration 155
## GeneSelection: iteration 156
## GeneSelection: iteration 157
## GeneSelection: iteration 158
## GeneSelection: iteration 159
## GeneSelection: iteration 160
## GeneSelection: iteration 161
## GeneSelection: iteration 162
## GeneSelection: iteration 163
## GeneSelection: iteration 164
## GeneSelection: iteration 165
## GeneSelection: iteration 166
## GeneSelection: iteration 167
## GeneSelection: iteration 168
## GeneSelection: iteration 169
## GeneSelection: iteration 170
## GeneSelection: iteration 171
## GeneSelection: iteration 172
## GeneSelection: iteration 173
## GeneSelection: iteration 174
## GeneSelection: iteration 175
## GeneSelection: iteration 176
## GeneSelection: iteration 177
## GeneSelection: iteration 178
## GeneSelection: iteration 179
## GeneSelection: iteration 180
## GeneSelection: iteration 181
## GeneSelection: iteration 182
## GeneSelection: iteration 183
## GeneSelection: iteration 184
## GeneSelection: iteration 185
## GeneSelection: iteration 186
## GeneSelection: iteration 187
## GeneSelection: iteration 188
## GeneSelection: iteration 189
## GeneSelection: iteration 190
## GeneSelection: iteration 191
## GeneSelection: iteration 192
## GeneSelection: iteration 193
```

```
## GeneSelection: iteration 194
## GeneSelection: iteration 195
## GeneSelection: iteration 196
## GeneSelection: iteration 197
## GeneSelection: iteration 198
## GeneSelection: iteration 199
## GeneSelection: iteration 200
## GeneSelection: iteration 201
## GeneSelection: iteration 202
## GeneSelection: iteration 203
## GeneSelection: iteration 204
## GeneSelection: iteration 205
## GeneSelection: iteration 206
## GeneSelection: iteration 207
## GeneSelection: iteration 208
## GeneSelection: iteration 209
## GeneSelection: iteration 210
## GeneSelection: iteration 211
## GeneSelection: iteration 212
## GeneSelection: iteration 213
## GeneSelection: iteration 214
## GeneSelection: iteration 215
## GeneSelection: iteration 216
## GeneSelection: iteration 217
## GeneSelection: iteration 218
## GeneSelection: iteration 219
## GeneSelection: iteration 220
## GeneSelection: iteration 221
## GeneSelection: iteration 222
## GeneSelection: iteration 223
## GeneSelection: iteration 224
## GeneSelection: iteration 225
## GeneSelection: iteration 226
## GeneSelection: iteration 227
## GeneSelection: iteration 228
## GeneSelection: iteration 229
## GeneSelection: iteration 230
## GeneSelection: iteration 231
## GeneSelection: iteration 232
## GeneSelection: iteration 233
## GeneSelection: iteration 234
## GeneSelection: iteration 235
## GeneSelection: iteration 236
## GeneSelection: iteration 237
## GeneSelection: iteration 238
## GeneSelection: iteration 239
## GeneSelection: iteration 240
## GeneSelection: iteration 241
## GeneSelection: iteration 242
## GeneSelection: iteration 243
## GeneSelection: iteration 244
## GeneSelection: iteration 245
## GeneSelection: iteration 246
## GeneSelection: iteration 247
```

```
## GeneSelection: iteration 248
## GeneSelection: iteration 249
## GeneSelection: iteration 250
```

```
#Classification is performed using LASSO for the first 5, 10, 15 and 20 most selected variables:

class_lasso5<-classification(X = dataX, y = Y, learningsets = CVdat, classifier = LassoCMA, genesel = va
```

```
## iteration 1
## iteration 2
## iteration 3
## iteration 4
## iteration 5
## iteration 6
## iteration 7
## iteration 8
## iteration 9
## iteration 10
## iteration 11
## iteration 12
## iteration 13
## iteration 14
## iteration 15
## iteration 16
## iteration 17
## iteration 18
## iteration 19
## iteration 20
## iteration 21
## iteration 22
## iteration 23
## iteration 24
## iteration 25
## iteration 26
## iteration 27
## iteration 28
## iteration 29
## iteration 30
## iteration 31
## iteration 32
## iteration 33
## iteration 34
## iteration 35
## iteration 36
## iteration 37
## iteration 38
## iteration 39
## iteration 40
## iteration 41
## iteration 42
## iteration 43
## iteration 44
## iteration 45
## iteration 46
```

```
## iteration 47
## iteration 48
## iteration 49
## iteration 50
## iteration 51
## iteration 52
## iteration 53
## iteration 54
## iteration 55
## iteration 56
## iteration 57
## iteration 58
## iteration 59
## iteration 60
## iteration 61
## iteration 62
## iteration 63
## iteration 64
## iteration 65
## iteration 66
## iteration 67
## iteration 68
## iteration 69
## iteration 70
## iteration 71
## iteration 72
## iteration 73
## iteration 74
## iteration 75
## iteration 76
## iteration 77
## iteration 78
## iteration 79
## iteration 80
## iteration 81
## iteration 82
## iteration 83
## iteration 84
## iteration 85
## iteration 86
## iteration 87
## iteration 88
## iteration 89
## iteration 90
## iteration 91
## iteration 92
## iteration 93
## iteration 94
## iteration 95
## iteration 96
## iteration 97
## iteration 98
## iteration 99
## iteration 100
```

```
## iteration 101
## iteration 102
## iteration 103
## iteration 104
## iteration 105
## iteration 106
## iteration 107
## iteration 108
## iteration 109
## iteration 110
## iteration 111
## iteration 112
## iteration 113
## iteration 114
## iteration 115
## iteration 116
## iteration 117
## iteration 118
## iteration 119
## iteration 120
## iteration 121
## iteration 122
## iteration 123
## iteration 124
## iteration 125
## iteration 126
## iteration 127
## iteration 128
## iteration 129
## iteration 130
## iteration 131
## iteration 132
## iteration 133
## iteration 134
## iteration 135
## iteration 136
## iteration 137
## iteration 138
## iteration 139
## iteration 140
## iteration 141
## iteration 142
## iteration 143
## iteration 144
## iteration 145
## iteration 146
## iteration 147
## iteration 148
## iteration 149
## iteration 150
## iteration 151
## iteration 152
## iteration 153
## iteration 154
```

```
## iteration 155
## iteration 156
## iteration 157
## iteration 158
## iteration 159
## iteration 160
## iteration 161
## iteration 162
## iteration 163
## iteration 164
## iteration 165
## iteration 166
## iteration 167
## iteration 168
## iteration 169
## iteration 170
## iteration 171
## iteration 172
## iteration 173
## iteration 174
## iteration 175
## iteration 176
## iteration 177
## iteration 178
## iteration 179
## iteration 180
## iteration 181
## iteration 182
## iteration 183
## iteration 184
## iteration 185
## iteration 186
## iteration 187
## iteration 188
## iteration 189
## iteration 190
## iteration 191
## iteration 192
## iteration 193
## iteration 194
## iteration 195
## iteration 196
## iteration 197
## iteration 198
## iteration 199
## iteration 200
## iteration 201
## iteration 202
## iteration 203
## iteration 204
## iteration 205
## iteration 206
## iteration 207
## iteration 208
```

```
## iteration 209
## iteration 210
## iteration 211
## iteration 212
## iteration 213
## iteration 214
## iteration 215
## iteration 216
## iteration 217
## iteration 218
## iteration 219
## iteration 220
## iteration 221
## iteration 222
## iteration 223
## iteration 224
## iteration 225
## iteration 226
## iteration 227
## iteration 228
## iteration 229
## iteration 230
## iteration 231
## iteration 232
## iteration 233
## iteration 234
## iteration 235
## iteration 236
## iteration 237
## iteration 238
## iteration 239
## iteration 240
## iteration 241
## iteration 242
## iteration 243
## iteration 244
## iteration 245
## iteration 246
## iteration 247
## iteration 248
## iteration 249
## iteration 250
```

```
class_lasso10<-classification(X = dataX, y = Y, learningsets = CVdat, classifier = LassoCMA, genesel = 
```

```
## iteration 1
## iteration 2
## iteration 3
## iteration 4
## iteration 5
## iteration 6
## iteration 7
## iteration 8
## iteration 9
```

```
## iteration 10
## iteration 11
## iteration 12
## iteration 13
## iteration 14
## iteration 15
## iteration 16
## iteration 17
## iteration 18
## iteration 19
## iteration 20
## iteration 21
## iteration 22
## iteration 23
## iteration 24
## iteration 25
## iteration 26
## iteration 27
## iteration 28
## iteration 29
## iteration 30
## iteration 31
## iteration 32
## iteration 33
## iteration 34
## iteration 35
## iteration 36
## iteration 37
## iteration 38
## iteration 39
## iteration 40
## iteration 41
## iteration 42
## iteration 43
## iteration 44
## iteration 45
## iteration 46
## iteration 47
## iteration 48
## iteration 49
## iteration 50
## iteration 51
## iteration 52
## iteration 53
## iteration 54
## iteration 55
## iteration 56
## iteration 57
## iteration 58
## iteration 59
## iteration 60
## iteration 61
## iteration 62
## iteration 63
```

```
## iteration 64
## iteration 65
## iteration 66
## iteration 67
## iteration 68
## iteration 69
## iteration 70
## iteration 71
## iteration 72
## iteration 73
## iteration 74
## iteration 75
## iteration 76
## iteration 77
## iteration 78
## iteration 79
## iteration 80
## iteration 81
## iteration 82
## iteration 83
## iteration 84
## iteration 85
## iteration 86
## iteration 87
## iteration 88
## iteration 89
## iteration 90
## iteration 91
## iteration 92
## iteration 93
## iteration 94
## iteration 95
## iteration 96
## iteration 97
## iteration 98
## iteration 99
## iteration 100
## iteration 101
## iteration 102
## iteration 103
## iteration 104
## iteration 105
## iteration 106
## iteration 107
## iteration 108
## iteration 109
## iteration 110
## iteration 111
## iteration 112
## iteration 113
## iteration 114
## iteration 115
## iteration 116
## iteration 117
```

```
## iteration 118
## iteration 119
## iteration 120
## iteration 121
## iteration 122
## iteration 123
## iteration 124
## iteration 125
## iteration 126
## iteration 127
## iteration 128
## iteration 129
## iteration 130
## iteration 131
## iteration 132
## iteration 133
## iteration 134
## iteration 135
## iteration 136
## iteration 137
## iteration 138
## iteration 139
## iteration 140
## iteration 141
## iteration 142
## iteration 143
## iteration 144
## iteration 145
## iteration 146
## iteration 147
## iteration 148
## iteration 149
## iteration 150
## iteration 151
## iteration 152
## iteration 153
## iteration 154
## iteration 155
## iteration 156
## iteration 157
## iteration 158
## iteration 159
## iteration 160
## iteration 161
## iteration 162
## iteration 163
## iteration 164
## iteration 165
## iteration 166
## iteration 167
## iteration 168
## iteration 169
## iteration 170
## iteration 171
```

```
## iteration 172
## iteration 173
## iteration 174
## iteration 175
## iteration 176
## iteration 177
## iteration 178
## iteration 179
## iteration 180
## iteration 181
## iteration 182
## iteration 183
## iteration 184
## iteration 185
## iteration 186
## iteration 187
## iteration 188
## iteration 189
## iteration 190
## iteration 191
## iteration 192
## iteration 193
## iteration 194
## iteration 195
## iteration 196
## iteration 197
## iteration 198
## iteration 199
## iteration 200
## iteration 201
## iteration 202
## iteration 203
## iteration 204
## iteration 205
## iteration 206
## iteration 207
## iteration 208
## iteration 209
## iteration 210
## iteration 211
## iteration 212
## iteration 213
## iteration 214
## iteration 215
## iteration 216
## iteration 217
## iteration 218
## iteration 219
## iteration 220
## iteration 221
## iteration 222
## iteration 223
## iteration 224
## iteration 225
```

```
## iteration 226
## iteration 227
## iteration 228
## iteration 229
## iteration 230
## iteration 231
## iteration 232
## iteration 233
## iteration 234
## iteration 235
## iteration 236
## iteration 237
## iteration 238
## iteration 239
## iteration 240
## iteration 241
## iteration 242
## iteration 243
## iteration 244
## iteration 245
## iteration 246
## iteration 247
## iteration 248
## iteration 249
## iteration 250
```

```
class_lasso15<-classification(X = dataX, y = Y, learningsets = CVdat, classifier = LassoCMA, genesel = v
```

```
## iteration 1
## iteration 2
## iteration 3
## iteration 4
## iteration 5
## iteration 6
## iteration 7
## iteration 8
## iteration 9
## iteration 10
## iteration 11
## iteration 12
## iteration 13
## iteration 14
## iteration 15
## iteration 16
## iteration 17
## iteration 18
## iteration 19
## iteration 20
## iteration 21
## iteration 22
## iteration 23
## iteration 24
## iteration 25
## iteration 26
```

```
## iteration 27
## iteration 28
## iteration 29
## iteration 30
## iteration 31
## iteration 32
## iteration 33
## iteration 34
## iteration 35
## iteration 36
## iteration 37
## iteration 38
## iteration 39
## iteration 40
## iteration 41
## iteration 42
## iteration 43
## iteration 44
## iteration 45
## iteration 46
## iteration 47
## iteration 48
## iteration 49
## iteration 50
## iteration 51
## iteration 52
## iteration 53
## iteration 54
## iteration 55
## iteration 56
## iteration 57
## iteration 58
## iteration 59
## iteration 60
## iteration 61
## iteration 62
## iteration 63
## iteration 64
## iteration 65
## iteration 66
## iteration 67
## iteration 68
## iteration 69
## iteration 70
## iteration 71
## iteration 72
## iteration 73
## iteration 74
## iteration 75
## iteration 76
## iteration 77
## iteration 78
## iteration 79
## iteration 80
```

```
## iteration 81
## iteration 82
## iteration 83
## iteration 84
## iteration 85
## iteration 86
## iteration 87
## iteration 88
## iteration 89
## iteration 90
## iteration 91
## iteration 92
## iteration 93
## iteration 94
## iteration 95
## iteration 96
## iteration 97
## iteration 98
## iteration 99
## iteration 100
## iteration 101
## iteration 102
## iteration 103
## iteration 104
## iteration 105
## iteration 106
## iteration 107
## iteration 108
## iteration 109
## iteration 110
## iteration 111
## iteration 112
## iteration 113
## iteration 114
## iteration 115
## iteration 116
## iteration 117
## iteration 118
## iteration 119
## iteration 120
## iteration 121
## iteration 122
## iteration 123
## iteration 124
## iteration 125
## iteration 126
## iteration 127
## iteration 128
## iteration 129
## iteration 130
## iteration 131
## iteration 132
## iteration 133
## iteration 134
```

```
## iteration 135
## iteration 136
## iteration 137
## iteration 138
## iteration 139
## iteration 140
## iteration 141
## iteration 142
## iteration 143
## iteration 144
## iteration 145
## iteration 146
## iteration 147
## iteration 148
## iteration 149
## iteration 150
## iteration 151
## iteration 152
## iteration 153
## iteration 154
## iteration 155
## iteration 156
## iteration 157
## iteration 158
## iteration 159
## iteration 160
## iteration 161
## iteration 162
## iteration 163
## iteration 164
## iteration 165
## iteration 166
## iteration 167
## iteration 168
## iteration 169
## iteration 170
## iteration 171
## iteration 172
## iteration 173
## iteration 174
## iteration 175
## iteration 176
## iteration 177
## iteration 178
## iteration 179
## iteration 180
## iteration 181
## iteration 182
## iteration 183
## iteration 184
## iteration 185
## iteration 186
## iteration 187
## iteration 188
```

```
## iteration 189
## iteration 190
## iteration 191
## iteration 192
## iteration 193
## iteration 194
## iteration 195
## iteration 196
## iteration 197
## iteration 198
## iteration 199
## iteration 200
## iteration 201
## iteration 202
## iteration 203
## iteration 204
## iteration 205
## iteration 206
## iteration 207
## iteration 208
## iteration 209
## iteration 210
## iteration 211
## iteration 212
## iteration 213
## iteration 214
## iteration 215
## iteration 216
## iteration 217
## iteration 218
## iteration 219
## iteration 220
## iteration 221
## iteration 222
## iteration 223
## iteration 224
## iteration 225
## iteration 226
## iteration 227
## iteration 228
## iteration 229
## iteration 230
## iteration 231
## iteration 232
## iteration 233
## iteration 234
## iteration 235
## iteration 236
## iteration 237
## iteration 238
## iteration 239
## iteration 240
## iteration 241
## iteration 242
```

```
## iteration 243
## iteration 244
## iteration 245
## iteration 246
## iteration 247
## iteration 248
## iteration 249
## iteration 250
```

```
class_lasso20<-classification(X = dataX, y = Y, learningsets = CVdat, classifier = LassoCMA, genesel = 
```

```
## iteration 1
## iteration 2
## iteration 3
## iteration 4
## iteration 5
## iteration 6
## iteration 7
## iteration 8
## iteration 9
## iteration 10
## iteration 11
## iteration 12
## iteration 13
## iteration 14
## iteration 15
## iteration 16
## iteration 17
## iteration 18
## iteration 19
## iteration 20
## iteration 21
## iteration 22
## iteration 23
## iteration 24
## iteration 25
## iteration 26
## iteration 27
## iteration 28
## iteration 29
## iteration 30
## iteration 31
## iteration 32
## iteration 33
## iteration 34
## iteration 35
## iteration 36
## iteration 37
## iteration 38
## iteration 39
## iteration 40
## iteration 41
## iteration 42
## iteration 43
```

```
## iteration 44
## iteration 45
## iteration 46
## iteration 47
## iteration 48
## iteration 49
## iteration 50
## iteration 51
## iteration 52
## iteration 53
## iteration 54
## iteration 55
## iteration 56
## iteration 57
## iteration 58
## iteration 59
## iteration 60
## iteration 61
## iteration 62
## iteration 63
## iteration 64
## iteration 65
## iteration 66
## iteration 67
## iteration 68
## iteration 69
## iteration 70
## iteration 71
## iteration 72
## iteration 73
## iteration 74
## iteration 75
## iteration 76
## iteration 77
## iteration 78
## iteration 79
## iteration 80
## iteration 81
## iteration 82
## iteration 83
## iteration 84
## iteration 85
## iteration 86
## iteration 87
## iteration 88
## iteration 89
## iteration 90
## iteration 91
## iteration 92
## iteration 93
## iteration 94
## iteration 95
## iteration 96
## iteration 97
```

```
## iteration 98
## iteration 99
## iteration 100
## iteration 101
## iteration 102
## iteration 103
## iteration 104
## iteration 105
## iteration 106
## iteration 107
## iteration 108
## iteration 109
## iteration 110
## iteration 111
## iteration 112
## iteration 113
## iteration 114
## iteration 115
## iteration 116
## iteration 117
## iteration 118
## iteration 119
## iteration 120
## iteration 121
## iteration 122
## iteration 123
## iteration 124
## iteration 125
## iteration 126
## iteration 127
## iteration 128
## iteration 129
## iteration 130
## iteration 131
## iteration 132
## iteration 133
## iteration 134
## iteration 135
## iteration 136
## iteration 137
## iteration 138
## iteration 139
## iteration 140
## iteration 141
## iteration 142
## iteration 143
## iteration 144
## iteration 145
## iteration 146
## iteration 147
## iteration 148
## iteration 149
## iteration 150
## iteration 151
```

```
## iteration 152
## iteration 153
## iteration 154
## iteration 155
## iteration 156
## iteration 157
## iteration 158
## iteration 159
## iteration 160
## iteration 161
## iteration 162
## iteration 163
## iteration 164
## iteration 165
## iteration 166
## iteration 167
## iteration 168
## iteration 169
## iteration 170
## iteration 171
## iteration 172
## iteration 173
## iteration 174
## iteration 175
## iteration 176
## iteration 177
## iteration 178
## iteration 179
## iteration 180
## iteration 181
## iteration 182
## iteration 183
## iteration 184
## iteration 185
## iteration 186
## iteration 187
## iteration 188
## iteration 189
## iteration 190
## iteration 191
## iteration 192
## iteration 193
## iteration 194
## iteration 195
## iteration 196
## iteration 197
## iteration 198
## iteration 199
## iteration 200
## iteration 201
## iteration 202
## iteration 203
## iteration 204
## iteration 205
```

```
## iteration 206
## iteration 207
## iteration 208
## iteration 209
## iteration 210
## iteration 211
## iteration 212
## iteration 213
## iteration 214
## iteration 215
## iteration 216
## iteration 217
## iteration 218
## iteration 219
## iteration 220
## iteration 221
## iteration 222
## iteration 223
## iteration 224
## iteration 225
## iteration 226
## iteration 227
## iteration 228
## iteration 229
## iteration 230
## iteration 231
## iteration 232
## iteration 233
## iteration 234
## iteration 235
## iteration 236
## iteration 237
## iteration 238
## iteration 239
## iteration 240
## iteration 241
## iteration 242
## iteration 243
## iteration 244
## iteration 245
## iteration 246
## iteration 247
## iteration 248
## iteration 249
## iteration 250
```

```
result_list <- list(class_lasso5, class_lasso10, class_lasso15, class_lasso20)

#Classification accuracy is compared:

comparison_lasso<- compare(result_list,plot = F, measure = c("misclassification","auc"))
print(comparison_lasso)
```

```
##        misclassification    auc
```

```
## Lasso             0.3013 0.7448
## Lasso2            0.2894 0.7681
## Lasso3            0.3149 0.7452
## Lasso4            0.3299 0.7196
```

```
#          misclassification   auc
# Lasso          0.3011429 0.7456723
# Lasso2         0.2908571 0.7672543
# Lasso3         0.3134286 0.7465282
# Lasso4         0.3298571 0.7196486

#Based on results, the method Lasso2 with the best classification accuracy (maximum AUC=0.7672543) is t
#Thus, we print the 10 most selected variables in the iterative process, and this is the model proposed

ntop<-10

seliter <- numeric()
for (i in 1:iterations) seliter <- c(seliter, toplist(varsel_lasso, iter = i, top = ntop, show = FALSE)$

selected_lasso<-sort(table(seliter), dec = TRUE)

index_lasso<-as.numeric(names(selected_lasso[1:ntop]))

topselection_lasso<-data.frame(colnames(dataX)[index_lasso], selected_lasso[1:ntop], 100*selected_lasso
colnames(topselection_lasso)<-c("variable", "frequency of selection", "percentage of selection")
topselection_lasso
```

```
##            variable frequency of selection percentage of selection NA NA
## 1        DIAPH3.2_s                     35                      49 35 98
## 2           MMP9_s                      3                      42  3 84
## 3         ALDH4A1_s                     6                      33  6 66
## 4           GNAZ_s                     49                      33 49 66
## 5          AYTL2_s                      9                      32  9 64
## 6        CDC42BPA_s                     17                      27 17 54
## 7  Contig63649_RC_s                     42                      22 42 44
## 8            DTL_s                     23                      21 23 42
## 9          TSPYL5_s                     10                      18 10 36
## 10           PECI_s                     34                      18 34 36
```

```
#The following is the list of selected variables for my fitted logistical model using LASSO:

#           variable frequency of selection percentage of selection NA NA
# 1        DIAPH3.2_s                     5                      49  5 98
# 2           MMP9_s                     24                      42 24 84
# 3         ALDH4A1_s                     25                      33 25 66
# 4           GNAZ_s                     21                      32 21 64
# 5          AYTL2_s                     48                      32 48 64
# 6        CDC42BPA_s                     22                      28 22 56
# 7  Contig63649_RC_s                     27                      22 27 44
# 8            DTL_s                     40                      21 40 42
# 9          TSPYL5_s                      3                      18  3 36
# 10           PECI_s                     41                      18 41 36
```

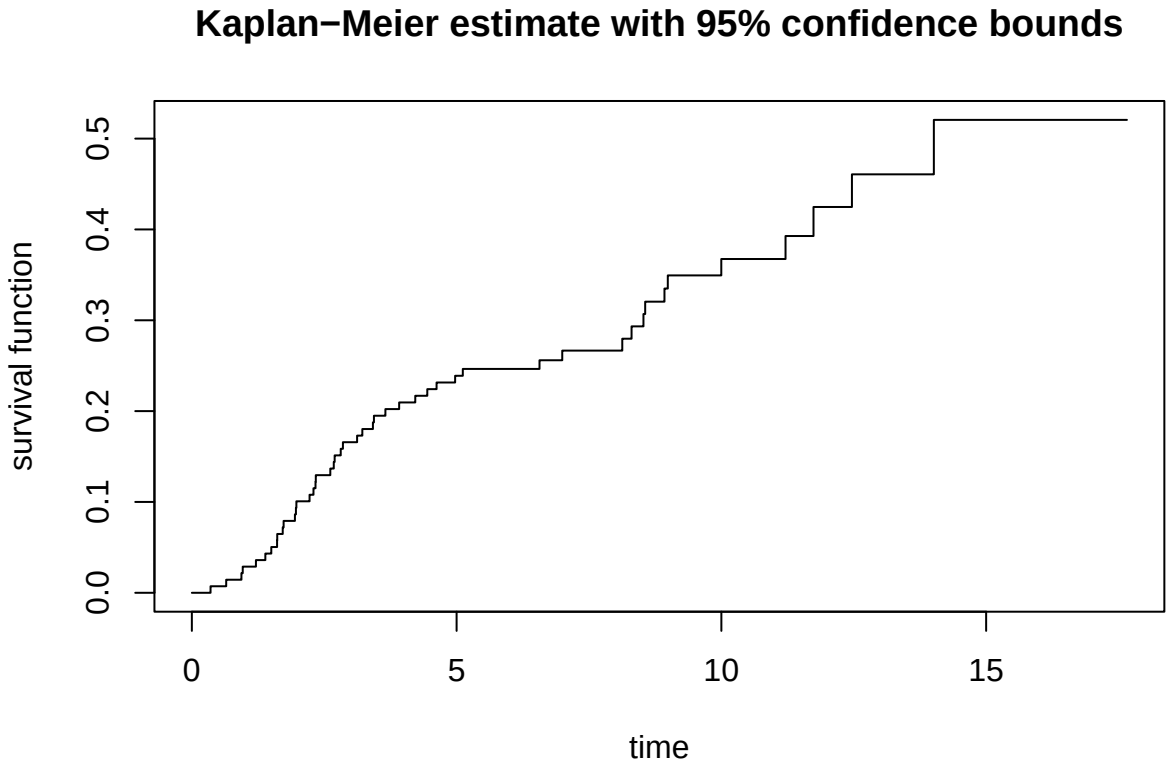
QUESTION 16 Obtain Kaplan-Meier survival curves for the time of symptoms as a function of the kind of infection and test for the significance of the difference in duration of symptoms. Discuss the results.

```
#BACKGROUND: In survival analysis the outcome of interest requires information on two variables, a time
#The indicator variables is 1 when the event of interest has occurred or 0 otherwise. This two variable
#and this object is used as the outcome in the analysis.

#stime: Time with symptoms (days).
#sind: Indicator of symptoms: (1 = symptoms finished; 0 = symptoms remain)
#hosp: Indicator of hospitalization risk event (1= hospitalization, 0 = no hospitalization) : THIS WILL



#Function survfit() applied to a survival object Surv(,)provides tables and plots of Kaplan-Meier survi

#Kaplan-Meier curves for the time of symptoms.
kmcurve1<-survfit(Surv(viral34_ca$stime,viral34_ca$sind)~ 1)
summary(kmcurve1)
```

```
## Call: survfit(formula = Surv(viral34_ca$stime, viral34_ca$sind) ~ 1)
##
##    time n.risk n.event Pr((s0)) Pr(symptoms_finished)
##   0.353    139       1    0.993                0.00719
##   0.649    138       1    0.986                0.01439
##   0.936    137       1    0.978                0.02158
##   0.961    136       1    0.971                0.02878
##   1.210    135       1    0.964                0.03597
##   1.388    134       1    0.957                0.04317
##   1.500    133       1    0.950                0.05036
##   1.610    132       1    0.942                0.05755
##   1.613    131       1    0.935                0.06475
##   1.717    130       1    0.928                0.07194
##   1.733    129       1    0.921                0.07914
##   1.947    128       1    0.914                0.08633
##   1.966    127       1    0.906                0.09353
##   1.974    126       1    0.899                0.10072
##   2.223    125       1    0.892                0.10791
##   2.297    124       1    0.885                0.11511
##   2.335    123       1    0.878                0.12230
##   2.341    122       1    0.871                0.12950
##   2.615    120       1    0.863                0.13675
##   2.680    119       1    0.856                0.14400
##   2.697    118       1    0.849                0.15126
##   2.812    117       1    0.841                0.15851
##   2.853    116       1    0.834                0.16577
##   3.121    115       1    0.827                0.17302
##   3.220    114       1    0.820                0.18028
##   3.420    112       1    0.812                0.18759
##   3.439    111       1    0.805                0.19491
##   3.655    110       1    0.798                0.20223
##   3.915    109       1    0.790                0.20955
##   4.219    108       1    0.783                0.21687
##   4.446    107       1    0.776                0.22419
##   4.621    106       1    0.768                0.23151
```

```
##    4.972    104       1    0.761              0.23890
##    5.117    101       1    0.754              0.24643
##    6.565     79       1    0.744              0.25597
##    6.995     70       1    0.733              0.26660
##    8.129     56       1    0.720              0.27970
##    8.304     53       1    0.707              0.29329
##    8.528     52       1    0.693              0.30688
##    8.561     51       1    0.680              0.32047
##    8.925     47       1    0.665              0.33493
##    8.988     46       1    0.651              0.34939
##    9.999     36       1    0.633              0.36746
##   11.211     25       1    0.607              0.39276
##   11.740     19       1    0.575              0.42472
##   12.465     16       1    0.539              0.46068
##   14.012      9       1    0.479              0.52060
```

```
plot(kmcurve1, main="Kaplan-Meier estimate with 95% confidence bounds", xlab="time", ylab="survival fun
```



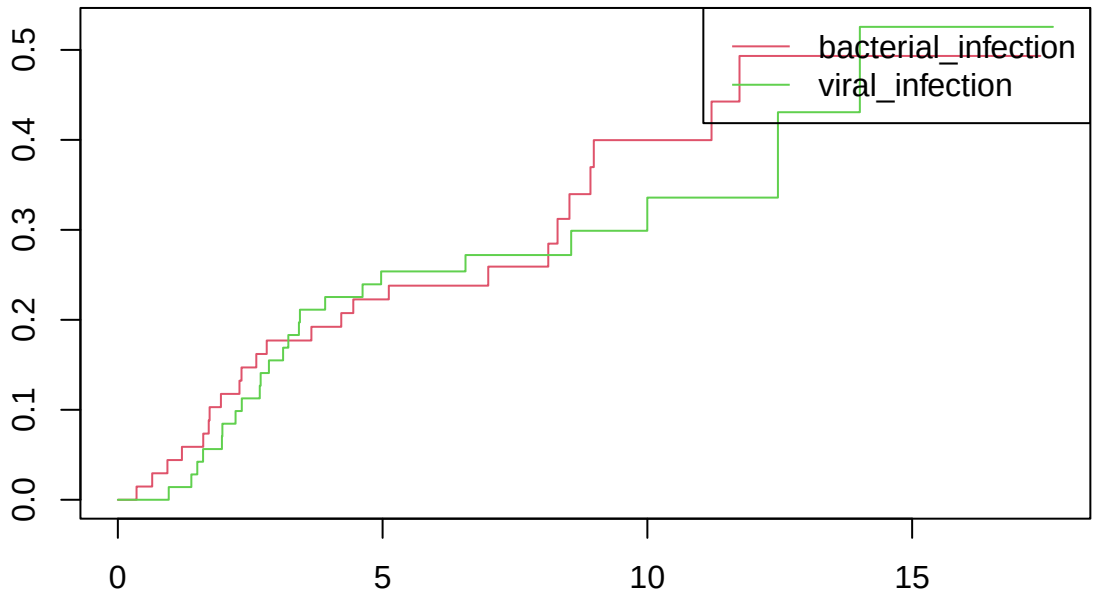

```
# #Kaplan-Meier curves for the time of symptoms.
# kmcurve2<-survfit(Surv(viral34_ca$stime,viral34_ca$hosp)~ 1)
# summary(kmcurve2)
# plot(kmcurve2, main="Kaplan-Meier estimate with 95% confidence bounds", xlab="time", ylab="survival f
# #"KM curve for hospitalization or no hospitalization

#Kaplan-Meier curves for the time of symptoms for the two levels of infection
```

```
kmcurve3<-survfit(Surv(viral34_ca$stime,viral34_ca$sind)~viral34_ca$infection)
summary(kmcurve3)
```

```
## Call: survfit(formula = Surv(viral34_ca$stime, viral34_ca$sind) ~ viral34_ca$infection)
##
##                 viral34_ca$infection=bacterial_infection
##    time n.risk n.event Pr((s0)) Pr(symptoms_finished)
##   0.353     68       1    0.985                0.0147
##   0.649     67       1    0.971                0.0294
##   0.936     66       1    0.956                0.0441
##   1.210     65       1    0.941                0.0588
##   1.613     64       1    0.926                0.0735
##   1.717     63       1    0.912                0.0882
##   1.733     62       1    0.897                0.1029
##   1.947     61       1    0.882                0.1176
##   2.297     60       1    0.868                0.1324
##   2.335     59       1    0.853                0.1471
##   2.615     57       1    0.838                0.1620
##   2.812     56       1    0.823                0.1770
##   3.655     54       1    0.808                0.1922
##   4.219     53       1    0.793                0.2075
##   4.446     52       1    0.777                0.2227
##   5.117     51       1    0.762                0.2380
##   6.995     36       1    0.741                0.2591
##   8.129     29       1    0.715                0.2847
##   8.304     26       1    0.688                0.3122
##   8.528     25       1    0.660                0.3397
##   8.925     22       1    0.630                0.3697
##   8.988     21       1    0.600                0.3997
##  11.211     14       1    0.557                0.4426
##  11.740     11       1    0.507                0.4933
##
##                 viral34_ca$infection=viral_infection
##    time n.risk n.event Pr((s0)) Pr(symptoms_finished)
##   0.961     71       1    0.986                0.0141
##   1.388     70       1    0.972                0.0282
##   1.500     69       1    0.958                0.0423
##   1.610     68       1    0.944                0.0563
##   1.966     67       1    0.930                0.0704
##   1.974     66       1    0.915                0.0845
##   2.223     65       1    0.901                0.0986
##   2.341     64       1    0.887                0.1127
##   2.680     63       1    0.873                0.1268
##   2.697     62       1    0.859                0.1408
##   2.853     61       1    0.845                0.1549
##   3.121     60       1    0.831                0.1690
##   3.220     59       1    0.817                0.1831
##   3.420     58       1    0.803                0.1972
##   3.439     57       1    0.789                0.2113
##   3.915     56       1    0.775                0.2254
##   4.621     55       1    0.761                0.2394
##   4.972     53       1    0.746                0.2538
##   6.565     41       1    0.728                0.2720
```

```
##    8.561     27       1    0.701                0.2990
##    9.999     19       1    0.664                0.3358
##   12.465      7       1    0.569                0.4307
##   14.012      6       1    0.474                0.5256
```

```
plot(kmcurve3, col=2:3)
legend("topright",col=2:3, legend=c("bacterial_infection","viral_infection"), lty=1)
```




```
# #Kaplan-Meier curves for the time of symptoms for the two levels of infection (1:2)
# kmcurve4<-survfit(Surv(viral34_ca$stime,viral34_ca$hosp)~ viral34_ca$infection)
# summary(kmcurve4)
# plot(kmcurve4, main="KM curve for for hospitalization or no hospitalization", col=(2:3))
# legend("topright",col=2:3, legend=c("bacterial_Infection","viral_Infection"), lty=1)


#Kaplan-Meier curves describe and summarize the survival times: estimation and interpretation of survivu
#estimator). The Kaplan and Meier (K-M) estimator of the survivor function is a step function with
#jumps at the observed event times. Based on the curves, it is apparent that the mean and median survivu
#for viral and bacterial infection "look" similar.

#The log-rank test is used to confirm if two survival curves are statistically different by testing fol
#The null hypothesis is H0: S1(t) =S2(t), for all t > 0
#The alternative hypothesis is: H1: S1(t) S2(t), for some t > 0

#Performing the log-rank test for equality of two survival functions according to type of infection
```

```
#survdiff(Surv(viral34_ca$stime,viral34_ca$sind)~viral34_ca$infection)
#Error in survdiff(Surv(viral34_ca$stime, viral34_ca$sind) ~ viral34_ca$infection):Right censored data

# #Performing the log-rank test for equality of two survival functions according to type of infection
# survdiff(Surv(viral34_ca$stime,viral34_ca$hosp)~viral34_ca$infection)
# #Error in survdiff(Surv(viral34_c$stime, viral34_c$hosp) ~ viral34_c$infection) : Right censored data
```

QUESTION 17 Perform a Cox regression model for duration symptoms as a function of the covariates (ignore gene expression levels). Discuss the results

```
#BACKGROUND: The Cox PH model is the most commonly used regression model for a survival time.
#The Cox model specifies the hazard at time t for an individual with covariates (e.g. infection type) x
#coxmodel1<-coxph(Surv(viral34_ca$stime,viral34_ca$sind)~ viral34_ca$infection+viral34_ca$sind+viral34_
#coxmodel1

#Running with hospitalization categorical variable:
#coxmodel2<-coxph(Surv(viral34_ca$stime,viral34_ca$hosp)~ viral34_ca$infection+viral34_ca$sind+viral34_
#coxmodel2
#Error in coxph(Surv(viral34_c$stime, viral34_c$hosp) ~ viral34_c$infection + : an id statement is requ

#Error in coxph(Surv(viral34_ca$stime, viral34_ca$hosp) ~ viral34_ca$infection +  :
#                 an id statement is required for multi-state models


#cox<-survfit(coxmodel1)
#cox
#plot(cox)

#Cox diagnostics
#1. Non overlaping survival curves
#2. log(-log(Surv) aproximately parallel lines

#plot(log(-log(kmcurve3$surv)))

#No strata(covariate) + logwbc was added to the original list of covariates
# No stratified Cox model was plotted
```