Modeling Data - Breast Cancer Analysis

Howard Nguyen

2022-09-28

INSTALL AND LOAD PACKAGES

```
# Install pacman ("package manager") if needed
if (!require("pacman")) install.packages("pacman")
```

```
## Loading required package: pacman
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(dplyr)
library(MASS)
```

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

```
library(caTools)
library(caret)
```

```
## Loading required package: lattice
##
## Attaching package: 'caret'
##
## The following object is masked from 'package:purrr':
##
##     lift
```

```
library(readr, dplyr)
library(ggplot2)
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```
library(gridExtra)
```

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

```
library(pROC)
```

```
## Type 'citation("pROC")' for a citation.
##
## Attaching package: 'pROC'
##
## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```

```
library(psych) # for pc
```

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

```
library(GPArotation)
library(magrittr)
```

```
##
## Attaching package: 'magrittr'
##
## The following object is masked from 'package:purrr':
##
##     set_names
##
## The following object is masked from 'package:tidyr':
##
##     extract
```

```
library(rio)
```

LOAD AND PREPARE DATA

```
df <- read_csv("../Data/breast_cancer_data.csv")
```

```
## Rows: 1138 Columns: 32
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (1): diagnosis
## dbl (31): id, radius_mean, texture_mean, perimeter_mean, area_mean, smoothne...
##
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
```

```
head(df)
```

```
## # A tibble: 6 x 32
##        id diagn~1 radiu~2 textu~3 perim~4 area_~5 smoot~6 compa~7 conca~8 conca~9
##     <dbl> <chr>     <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1 8.42e5 M          18.0    10.4   123.     1001   0.118   0.278   0.300   0.147
## 2 8.43e5 M          20.6    17.8   133.     1326   0.0847  0.0786  0.0869  0.0702
## 3 8.43e7 M          19.7    21.2   130      1203   0.110   0.160   0.197   0.128
## 4 8.43e7 M          11.4    20.4    77.6     386.  0.142   0.284   0.241   0.105
## 5 8.44e7 M          20.3    14.3   135.     1297   0.100   0.133   0.198   0.104
## 6 8.44e5 M          12.4    15.7    82.6     477.  0.128   0.17    0.158   0.0809
## # ... with 22 more variables: symmetry_mean <dbl>,
## #   fractal_dimension_mean <dbl>, radius_se <dbl>, texture_se <dbl>,
## #   perimeter_se <dbl>, area_se <dbl>, smoothness_se <dbl>,
## #   compactness_se <dbl>, concavity_se <dbl>, ‘concave points_se‘ <dbl>,
## #   symmetry_se <dbl>, fractal_dimension_se <dbl>, radius_worst <dbl>,
## #   texture_worst <dbl>, perimeter_worst <dbl>, area_worst <dbl>,
## #   smoothness_worst <dbl>, compactness_worst <dbl>, concavity_worst <dbl>, ...
```

```
tail(df)
```

```
## # A tibble: 6 x 32
##        id diagn~1 radiu~2 textu~3 perim~4 area_~5 smoot~6 compa~7 conca~8 conca~9
##     <dbl> <chr>     <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1 926125 M         20.9     25.1    143     1347   0.110   0.224   0.317   0.147
## 2 926424 M         21.6     22.4    142     1479   0.111   0.116   0.244   0.139
## 3 926682 M         20.1     28.2    131.    1261   0.0978  0.103   0.144   0.0979
## 4 926954 M         16.6     28.1    108.     858.  0.0846  0.102   0.0925  0.0530
## 5 927241 M         20.6     29.3    140.    1265   0.118   0.277   0.351   0.152
## 6  92751 B          7.76    24.5     47.9    181   0.0526  0.0436  0       0
## # ... with 22 more variables: symmetry_mean <dbl>,
## #   fractal_dimension_mean <dbl>, radius_se <dbl>, texture_se <dbl>,
## #   perimeter_se <dbl>, area_se <dbl>, smoothness_se <dbl>,
## #   compactness_se <dbl>, concavity_se <dbl>, ‘concave points_se‘ <dbl>,
## #   symmetry_se <dbl>, fractal_dimension_se <dbl>, radius_worst <dbl>,
## #   texture_worst <dbl>, perimeter_worst <dbl>, area_worst <dbl>,
## #   smoothness_worst <dbl>, compactness_worst <dbl>, concavity_worst <dbl>, ...
```

```
# review the dataset
str(df)
```

```
## spec_tbl_df [1,138 x 32] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ id                     : num [1:1138] 842302 842517 84300903 84348301 84358402 ...
##  $ diagnosis              : chr [1:1138] "M" "M" "M" "M" ...
##  $ radius_mean            : num [1:1138] 18 20.6 19.7 11.4 20.3 ...
##  $ texture_mean           : num [1:1138] 10.4 17.8 21.2 20.4 14.3 ...
##  $ perimeter_mean         : num [1:1138] 122.8 132.9 130 77.6 135.1 ...
##  $ area_mean              : num [1:1138] 1001 1326 1203 386 1297 ...
##  $ smoothness_mean        : num [1:1138] 0.1184 0.0847 0.1096 0.1425 0.1003 ...
##  $ compactness_mean       : num [1:1138] 0.2776 0.0786 0.1599 0.2839 0.1328 ...
##  $ concavity_mean         : num [1:1138] 0.3001 0.0869 0.1974 0.2414 0.198 ...
##  $ concave points_mean    : num [1:1138] 0.1471 0.0702 0.1279 0.1052 0.1043 ...
##  $ symmetry_mean          : num [1:1138] 0.242 0.181 0.207 0.26 0.181 ...
##  $ fractal_dimension_mean : num [1:1138] 0.0787 0.0567 0.06 0.0974 0.0588 ...
##  $ radius_se              : num [1:1138] 1.095 0.543 0.746 0.496 0.757 ...
##  $ texture_se             : num [1:1138] 0.905 0.734 0.787 1.156 0.781 ...
##  $ perimeter_se           : num [1:1138] 8.59 3.4 4.58 3.44 5.44 ...
##  $ area_se                : num [1:1138] 153.4 74.1 94 27.2 94.4 ...
##  $ smoothness_se          : num [1:1138] 0.0064 0.00522 0.00615 0.00911 0.01149 ...
##  $ compactness_se         : num [1:1138] 0.049 0.0131 0.0401 0.0746 0.0246 ...
##  $ concavity_se           : num [1:1138] 0.0537 0.0186 0.0383 0.0566 0.0569 ...
##  $ concave points_se      : num [1:1138] 0.0159 0.0134 0.0206 0.0187 0.0188 ...
##  $ symmetry_se            : num [1:1138] 0.03 0.0139 0.0225 0.0596 0.0176 ...
##  $ fractal_dimension_se   : num [1:1138] 0.00619 0.00353 0.00457 0.00921 0.00511 ...
##  $ radius_worst           : num [1:1138] 25.4 25 23.6 14.9 22.5 ...
##  $ texture_worst          : num [1:1138] 17.3 23.4 25.5 26.5 16.7 ...
##  $ perimeter_worst        : num [1:1138] 184.6 158.8 152.5 98.9 152.2 ...
##  $ area_worst             : num [1:1138] 2019 1956 1709 568 1575 ...
##  $ smoothness_worst       : num [1:1138] 0.162 0.124 0.144 0.21 0.137 ...
##  $ compactness_worst      : num [1:1138] 0.666 0.187 0.424 0.866 0.205 ...
##  $ concavity_worst        : num [1:1138] 0.712 0.242 0.45 0.687 0.4 ...
##  $ concave points_worst   : num [1:1138] 0.265 0.186 0.243 0.258 0.163 ...
##  $ symmetry_worst         : num [1:1138] 0.46 0.275 0.361 0.664 0.236 ...
##  $ fractal_dimension_worst: num [1:1138] 0.1189 0.089 0.0876 0.173 0.0768 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   id = col_double(),
##   ..   diagnosis = col_character(),
##   ..   radius_mean = col_double(),
##   ..   texture_mean = col_double(),
##   ..   perimeter_mean = col_double(),
##   ..   area_mean = col_double(),
##   ..   smoothness_mean = col_double(),
##   ..   compactness_mean = col_double(),
##   ..   concavity_mean = col_double(),
##   ..   ‘concave points_mean‘ = col_double(),
##   ..   symmetry_mean = col_double(),
##   ..   fractal_dimension_mean = col_double(),
##   ..   radius_se = col_double(),
##   ..   texture_se = col_double(),
##   ..   perimeter_se = col_double(),
##   ..   area_se = col_double(),
##   ..   smoothness_se = col_double(),
##   ..   compactness_se = col_double(),
##   ..   concavity_se = col_double(),
##   ..   ‘concave points_se‘ = col_double(),
##   ..   symmetry_se = col_double(),
##   ..   fractal_dimension_se = col_double(),
##   ..   radius_worst = col_double(),
##   ..   texture_worst = col_double(),
##   ..   perimeter_worst = col_double(),
##   ..   area_worst = col_double(),
##   ..   smoothness_worst = col_double(),
##   ..   compactness_worst = col_double(),
##   ..   concavity_worst = col_double(),
##   ..   ‘concave points_worst‘ = col_double(),
##   ..   symmetry_worst = col_double(),
##   ..   fractal_dimension_worst = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

SUMMARIZE DATAFRAME

```
# Convert diagnosis field to factor
df$diagnosis <- as.factor(df$diagnosis)
# Summary of all variables.
summary(df)
```

```
##        id            diagnosis  radius_mean      texture_mean
##  Min.   :     8670   B:714     Min.   : 6.981   Min.   : 9.71
##  1st Qu.:   869218   M:424     1st Qu.:11.700   1st Qu.:16.17
##  Median :   906024             Median :13.370   Median :18.84
##  Mean   : 30371831             Mean   :14.127   Mean   :19.29
##  3rd Qu.:  8813129             3rd Qu.:15.780   3rd Qu.:21.80
##  Max.   :911320502             Max.   :28.110   Max.   :39.28
##  perimeter_mean     area_mean      smoothness_mean   compactness_mean
##  Min.   : 43.79   Min.   : 143.5   Min.   :0.05263   Min.   :0.01938
##  1st Qu.: 75.17   1st Qu.: 420.3   1st Qu.:0.08637   1st Qu.:0.06492
##  Median : 86.24   Median : 551.1   Median :0.09587   Median :0.09263
##  Mean   : 91.97   Mean   : 654.9   Mean   :0.09636   Mean   :0.10434
##  3rd Qu.:104.10   3rd Qu.: 782.7   3rd Qu.:0.10530   3rd Qu.:0.13040
##  Max.   :188.50   Max.   :2501.0   Max.   :0.16340   Max.   :0.34540
##  concavity_mean    concave points_mean symmetry_mean    fractal_dimension_mean
##  Min.   :0.00000   Min.   :0.00000     Min.   :0.1060   Min.   :0.04996
##  1st Qu.:0.02956   1st Qu.:0.02031     1st Qu.:0.1619   1st Qu.:0.05770
##  Median :0.06154   Median :0.03350     Median :0.1792   Median :0.06154
##  Mean   :0.08880   Mean   :0.04892     Mean   :0.1812   Mean   :0.06280
##  3rd Qu.:0.13070   3rd Qu.:0.07400     3rd Qu.:0.1957   3rd Qu.:0.06612
##  Max.   :0.42680   Max.   :0.20120     Max.   :0.3040   Max.   :0.09744
##    radius_se        texture_se       perimeter_se       area_se
##  Min.   :0.1115   Min.   :0.3602   Min.   : 0.757   Min.   :  6.802
##  1st Qu.:0.2324   1st Qu.:0.8339   1st Qu.: 1.606   1st Qu.: 17.850
##  Median :0.3242   Median :1.1080   Median : 2.287   Median : 24.530
##  Mean   :0.4052   Mean   :1.2169   Mean   : 2.866   Mean   : 40.337
##  3rd Qu.:0.4789   3rd Qu.:1.4740   3rd Qu.: 3.357   3rd Qu.: 45.190
##  Max.   :2.8730   Max.   :4.8850   Max.   :21.980   Max.   :542.200
##  smoothness_se      compactness_se      concavity_se     concave points_se
##  Min.   :0.001713   Min.   :0.002252   Min.   :0.00000   Min.   :0.000000
##  1st Qu.:0.005169   1st Qu.:0.013080   1st Qu.:0.01509   1st Qu.:0.007638
##  Median :0.006380   Median :0.020450   Median :0.02589   Median :0.010930
##  Mean   :0.007041   Mean   :0.025478   Mean   :0.03189   Mean   :0.011796
##  3rd Qu.:0.008146   3rd Qu.:0.032450   3rd Qu.:0.04205   3rd Qu.:0.014710
##  Max.   :0.031130   Max.   :0.135400   Max.   :0.39600   Max.   :0.052790
##   symmetry_se       fractal_dimension_se  radius_worst   texture_worst
##  Min.   :0.007882   Min.   :0.0008948    Min.   : 7.93   Min.   :12.02
##  1st Qu.:0.015160   1st Qu.:0.0022480    1st Qu.:13.01   1st Qu.:21.08
##  Median :0.018730   Median :0.0031870    Median :14.97   Median :25.41
##  Mean   :0.020542   Mean   :0.0037949    Mean   :16.27   Mean   :25.68
##  3rd Qu.:0.023480   3rd Qu.:0.0045580    3rd Qu.:18.79   3rd Qu.:29.72
##  Max.   :0.078950   Max.   :0.0298400    Max.   :36.04   Max.   :49.54
##  perimeter_worst    area_worst     smoothness_worst  compactness_worst
##  Min.   : 50.41   Min.   : 185.2   Min.   :0.07117   Min.   :0.02729
##  1st Qu.: 84.11   1st Qu.: 515.3   1st Qu.:0.11660   1st Qu.:0.14720
##  Median : 97.66   Median : 686.5   Median :0.13130   Median :0.21190
##  Mean   :107.26   Mean   : 880.6   Mean   :0.13237   Mean   :0.25427
##  3rd Qu.:125.40   3rd Qu.:1084.0   3rd Qu.:0.14600   3rd Qu.:0.33910
##  Max.   :251.20   Max.   :4254.0   Max.   :0.22260   Max.   :1.05800
##  concavity_worst  concave points_worst symmetry_worst   fractal_dimension_worst
##  Min.   :0.0000   Min.   :0.00000      Min.   :0.1565   Min.   :0.05504
##  1st Qu.:0.1145   1st Qu.:0.06493      1st Qu.:0.2504   1st Qu.:0.07146
##  Median :0.2267   Median :0.09993      Median :0.2822   Median :0.08004
##  Mean   :0.2722   Mean   :0.11461      Mean   :0.2901   Mean   :0.08395
##  3rd Qu.:0.3829   3rd Qu.:0.16140      3rd Qu.:0.3179   3rd Qu.:0.09208
##  Max.   :1.2520   Max.   :0.29100      Max.   :0.6638   Max.   :0.20750
```

```
prop.table(table(df$diagnosis))
```

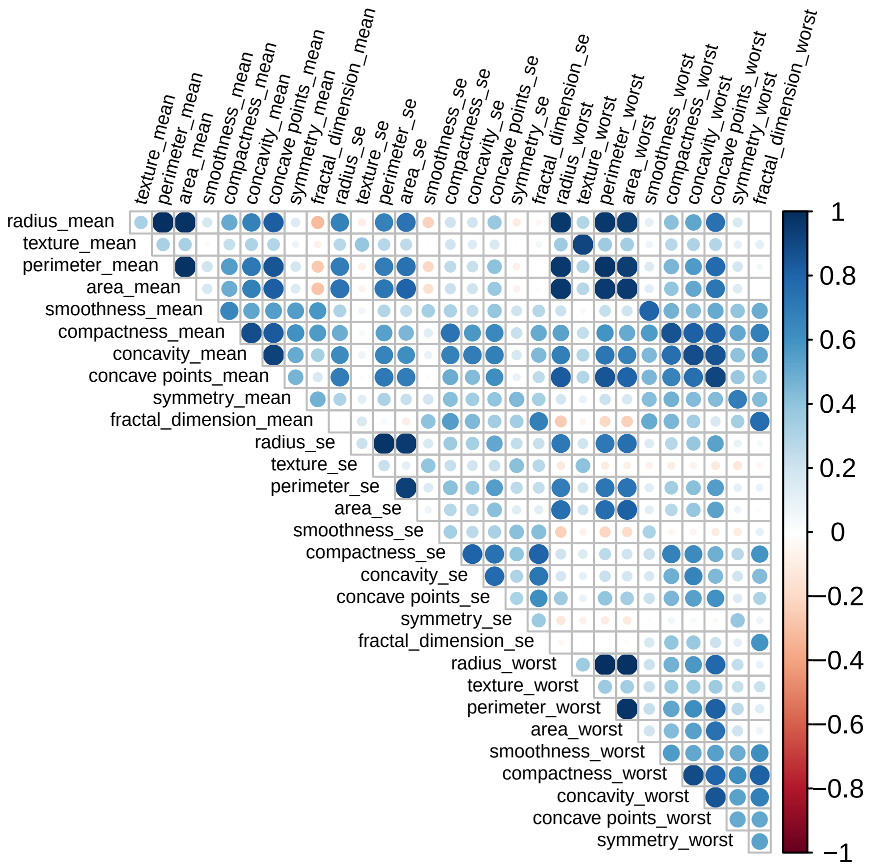
```
## 
##         B         M 
## 0.6274165 0.3725835
```

```
dfii <- df[,3:ncol(df)]
```

```
# evaluate correlation all attributes in the dataset
dfii%>%
  cor(dfii) %>%
  round(2) %>%
  print()
```

```
##                          radius_mean texture_mean perimeter_mean area_mean
## radius_mean                     1.00         0.32           1.00      0.99
## texture_mean                    0.32         1.00           0.33      0.32
## perimeter_mean                  1.00         0.33           1.00      0.99
## area_mean                       0.99         0.32           0.99      1.00
## smoothness_mean                 0.17        -0.02           0.21      0.18
## compactness_mean                0.51         0.24           0.56      0.50
## concavity_mean                  0.68         0.30           0.72      0.69
## concave points_mean             0.82         0.29           0.85      0.82
## symmetry_mean                   0.15         0.07           0.18      0.15
## fractal_dimension_mean         -0.31        -0.08          -0.26     -0.28
## radius_se                       0.68         0.28           0.69      0.73
## texture_se                     -0.10         0.39          -0.09     -0.07
## perimeter_se                    0.67         0.28           0.69      0.73
## area_se                         0.74         0.26           0.74      0.80
## smoothness_se                  -0.22         0.01          -0.20     -0.17
## compactness_se                  0.21         0.19           0.25      0.21
## concavity_se                    0.19         0.14           0.23      0.21
## concave points_se               0.38         0.16           0.41      0.37
## symmetry_se                    -0.10         0.01          -0.08     -0.07
## fractal_dimension_se           -0.04         0.05          -0.01     -0.02
## radius_worst                    0.97         0.35           0.97      0.96
## texture_worst                   0.30         0.91           0.30      0.29
## perimeter_worst                 0.97         0.36           0.97      0.96
## area_worst                      0.94         0.34           0.94      0.96
## smoothness_worst                0.12         0.08           0.15      0.12
## compactness_worst               0.41         0.28           0.46      0.39
## concavity_worst                 0.53         0.30           0.56      0.51
## concave points_worst            0.74         0.30           0.77      0.72
## symmetry_worst                  0.16         0.11           0.19      0.14
## fractal_dimension_worst         0.01         0.12           0.05      0.00
##                         smoothness_mean compactness_mean concavity_mean
## radius_mean                         0.17             0.51           0.68
## texture_mean                       -0.02             0.24           0.30
## perimeter_mean                      0.21             0.56           0.72
## area_mean                           0.18             0.50           0.69
## smoothness_mean                     1.00             0.66           0.52
## compactness_mean                    0.66             1.00           0.88
## concavity_mean                      0.52             0.88           1.00
## concave points_mean                 0.55             0.83           0.92
## symmetry_mean                       0.56             0.60           0.50
## fractal_dimension_mean              0.58             0.57           0.34
## radius_se                           0.30             0.50           0.63
## texture_se                          0.07             0.05           0.08
## perimeter_se                        0.30             0.55           0.66
## area_se                             0.25             0.46           0.62
## smoothness_se                       0.33             0.14           0.10
## compactness_se                      0.32             0.74           0.67
## concavity_se                        0.25             0.57           0.69
## concave points_se                   0.38             0.64           0.68
## symmetry_se                         0.20             0.23           0.18
## fractal_dimension_se                0.28             0.51           0.45
## radius_worst                        0.21             0.54           0.69
## texture_worst                       0.04             0.25           0.30
## perimeter_worst                     0.24             0.59           0.73
## area_worst                          0.21             0.51           0.68
## smoothness_worst                    0.81             0.57           0.45
## compactness_worst                  0.47             0.87           0.75
## concavity_worst                    0.43             0.82           0.88
## concave points_worst               0.50             0.82           0.86
## symmetry_worst                     0.39             0.51           0.41
## fractal_dimension_worst            0.50             0.69           0.51
##                         concave points_mean symmetry_mean
## radius_mean                            0.82          0.15
## texture_mean                           0.29          0.07
## perimeter_mean                         0.85          0.18
## area_mean                              0.82          0.15
## smoothness_mean                        0.55          0.56
## compactness_mean                       0.83          0.60
## concavity_mean                         0.92          0.50
## concave points_mean                    1.00          0.46
## symmetry_mean                          0.46          1.00
## fractal_dimension_mean                 0.17          0.48
## radius_se                              0.70          0.30
## texture_se                             0.02          0.13
## perimeter_se                           0.71          0.31
## area_se                                0.69          0.22
## smoothness_se                          0.03          0.19
## compactness_se                         0.49          0.42
## concavity_se                           0.44          0.34
## concave points_se                      0.62          0.39
## symmetry_se                            0.10          0.45
## fractal_dimension_se                   0.26          0.33
## radius_worst                           0.83          0.19
## texture_worst                          0.29          0.09
## perimeter_worst                        0.86          0.22
## area_worst                             0.81          0.18
## smoothness_worst                       0.45          0.43
## compactness_worst                      0.67          0.47
## concavity_worst                        0.75          0.43
## concave points_worst                   0.91          0.43
## symmetry_worst                         0.38          0.70
## fractal_dimension_worst                0.37          0.44
##                         fractal_dimension_mean radius_se texture_se
## radius_mean                              -0.31      0.68      -0.10
## texture_mean                             -0.08      0.28       0.39
## perimeter_mean                           -0.26      0.69      -0.09
## area_mean                                -0.28      0.73      -0.07
## smoothness_mean                           0.58      0.30       0.07
## compactness_mean                          0.57      0.50       0.05
## concavity_mean                            0.34      0.63       0.08
## concave points_mean                       0.17      0.70       0.02
## symmetry_mean                             0.48      0.30       0.13
## fractal_dimension_mean                    1.00      0.00       0.16
## radius_se                                 0.00      1.00       0.21
## texture_se                                0.16      0.21       1.00
## perimeter_se                              0.04      0.97       0.22
## area_se                                  -0.09      0.95       0.11
## smoothness_se                             0.40      0.16       0.40
## compactness_se                            0.56      0.36       0.23
## concavity_se                              0.45      0.33       0.19
## concave points_se                         0.34      0.51       0.23
## symmetry_se                               0.35      0.24       0.41
## fractal_dimension_se                      0.69      0.23       0.28
## radius_worst                             -0.25      0.72      -0.11
## texture_worst                            -0.05      0.19       0.41
## perimeter_worst                          -0.21      0.72      -0.10
## area_worst                               -0.23      0.75      -0.08
## smoothness_worst                          0.50      0.14      -0.07
## compactness_worst                         0.46      0.29      -0.09
## concavity_worst                           0.35      0.38      -0.07
## concave points_worst                      0.18      0.53      -0.12
## symmetry_worst                            0.33      0.09      -0.13
## fractal_dimension_worst                   0.77      0.05      -0.05
##                         perimeter_se area_se smoothness_se compactness_se
## radius_mean                     0.67    0.74         -0.22           0.21
## texture_mean                    0.28    0.26          0.01           0.19
## perimeter_mean                  0.69    0.74         -0.20           0.25
## area_mean                       0.73    0.80         -0.17           0.21
## smoothness_mean                 0.30    0.25          0.33           0.32
## compactness_mean                0.55    0.46          0.14           0.74
## concavity_mean                  0.66    0.62          0.10           0.67
## concave points_mean             0.71    0.69          0.03           0.49
## symmetry_mean                   0.31    0.22          0.19           0.42
## fractal_dimension_mean          0.04   -0.09          0.40           0.56
## radius_se                       0.97    0.95          0.16           0.36
## texture_se                      0.22    0.11          0.40           0.23
## perimeter_se                    1.00    0.94          0.15           0.42
## area_se                         0.94    1.00          0.08           0.28
## smoothness_se                   0.15    0.08          1.00           0.34
## compactness_se                  0.42    0.28          0.34           1.00
## concavity_se                    0.36    0.27          0.27           0.80
## concave points_se               0.56    0.42          0.33           0.74
## symmetry_se                     0.27    0.13          0.41           0.39
## fractal_dimension_se            0.24    0.13          0.43           0.80
## radius_worst                    0.70    0.76         -0.23           0.20
## texture_worst                   0.20    0.20         -0.07           0.14
## perimeter_worst                 0.72    0.76         -0.22           0.26
## area_worst                      0.73    0.81         -0.18           0.20
## smoothness_worst                0.13    0.13          0.31           0.23
## compactness_worst               0.34    0.28         -0.06           0.68
## concavity_worst                 0.42    0.39         -0.06           0.64
## concave points_worst            0.55    0.54         -0.10           0.48
## symmetry_worst                  0.11    0.07         -0.11           0.28
## fractal_dimension_worst         0.09    0.02          0.10           0.59
##                         concavity_se concave points_se symmetry_se
## radius_mean                     0.19              0.38       -0.10
## texture_mean                    0.14              0.16        0.01
## perimeter_mean                  0.23              0.41       -0.08
## area_mean                       0.21              0.37       -0.07
## smoothness_mean                 0.25              0.38        0.20
## compactness_mean                0.57              0.64        0.23
## concavity_mean                  0.69              0.68        0.18
## concave points_mean             0.44              0.62        0.10
## symmetry_mean                   0.34              0.39        0.45
## fractal_dimension_mean          0.45              0.34        0.35
## radius_se                       0.33              0.51        0.24
## texture_se                      0.19              0.23        0.41
## perimeter_se                    0.36              0.56        0.27
## area_se                         0.27              0.42        0.13
## smoothness_se                   0.27              0.33        0.41
## compactness_se                  0.80              0.74        0.39
## concavity_se                    1.00              0.77        0.31
## concave points_se               0.77              1.00        0.31
## symmetry_se                     0.31              0.31        1.00
## fractal_dimension_se            0.73              0.61        0.37
## radius_worst                    0.19              0.36       -0.13
## texture_worst                   0.10              0.09       -0.08
## perimeter_worst                 0.23              0.39       -0.10
## area_worst                      0.19              0.34       -0.11
## smoothness_worst                0.17              0.22       -0.01
## compactness_worst               0.48              0.45        0.06
## concavity_worst                 0.66              0.55        0.04
## concave points_worst            0.44              0.60       -0.03
## symmetry_worst                  0.20              0.14        0.39
## fractal_dimension_worst         0.44              0.31        0.08
##                         fractal_dimension_se radius_worst texture_worst
## radius_mean                            -0.04         0.97          0.30
## texture_mean                            0.05         0.35          0.91
## perimeter_mean                         -0.01         0.97          0.30
## area_mean                              -0.02         0.96          0.29
## smoothness_mean                         0.28         0.21          0.04
## compactness_mean                        0.51         0.54          0.25
## concavity_mean                          0.45         0.69          0.30
## concave points_mean                     0.26         0.83          0.29
## symmetry_mean                           0.33         0.19          0.09
## fractal_dimension_mean                  0.69        -0.25         -0.05
## radius_se                               0.23         0.72          0.19
## texture_se                              0.28        -0.11          0.41
## perimeter_se                            0.24         0.70          0.20
## area_se                                 0.13         0.76          0.20
## smoothness_se                           0.43        -0.23         -0.07
## compactness_se                          0.80         0.20          0.14
## concavity_se                            0.73         0.19          0.10
## concave points_se                       0.61         0.36          0.09
## symmetry_se                             0.37        -0.13         -0.08
## fractal_dimension_se                    1.00        -0.04          0.00
## radius_worst                           -0.04         1.00          0.36
## texture_worst                           0.00         0.36          1.00
## perimeter_worst                         0.00         0.99          0.37
## area_worst                             -0.02         0.98          0.35
## smoothness_worst                        0.17         0.22          0.23
## compactness_worst                       0.39         0.48          0.36
## concavity_worst                         0.38         0.57          0.37
## concave points_worst                    0.22         0.79          0.36
## symmetry_worst                          0.11         0.24          0.23
## fractal_dimension_worst                 0.59         0.09          0.22
##                         perimeter_worst area_worst smoothness_worst
## radius_mean                        0.97       0.94             0.12
## texture_mean                       0.36       0.34             0.08
## perimeter_mean                     0.97       0.94             0.15
## area_mean                          0.96       0.96             0.12
## smoothness_mean                    0.24       0.21             0.81
## compactness_mean                   0.59       0.51             0.57
## concavity_mean                     0.73       0.68             0.45
## concave points_mean                0.86       0.81             0.45
## symmetry_mean                      0.22       0.18             0.43
## fractal_dimension_mean            -0.21      -0.23             0.50
## radius_se                          0.72       0.75             0.14
## texture_se                        -0.10      -0.08            -0.07
## perimeter_se                       0.72       0.73             0.13
## area_se                            0.76       0.81             0.13
## smoothness_se                     -0.22      -0.18             0.31
## compactness_se                     0.26       0.20             0.23
## concavity_se                       0.23       0.19             0.17
## concave points_se                  0.39       0.34             0.22
## symmetry_se                       -0.10      -0.11            -0.01
## fractal_dimension_se               0.00      -0.02             0.17
## radius_worst                       0.99       0.98             0.22
## texture_worst                      0.37       0.35             0.23
## perimeter_worst                    1.00       0.98             0.24
## area_worst                         0.98       1.00             0.21
## smoothness_worst                   0.24       0.21             1.00
## compactness_worst                  0.53       0.44             0.57
## concavity_worst                    0.62       0.54             0.52
## concave points_worst               0.82       0.75             0.55
## symmetry_worst                     0.27       0.21             0.49
## fractal_dimension_worst            0.14       0.08             0.62
##                         compactness_worst concavity_worst concave points_worst
## radius_mean                          0.41            0.53                 0.74
## texture_mean                         0.28            0.30                 0.30
## perimeter_mean                       0.46            0.56                 0.77
## area_mean                            0.39            0.51                 0.72
## smoothness_mean                      0.47            0.43                 0.50
## compactness_mean                     0.87            0.82                 0.82
## concavity_mean                       0.75            0.88                 0.86
## concave points_mean                  0.67            0.75                 0.91
## symmetry_mean                        0.47            0.43                 0.43
## fractal_dimension_mean               0.46            0.35                 0.18
## radius_se                            0.29            0.38                 0.53
## texture_se                          -0.09           -0.07                -0.12
## perimeter_se                         0.34            0.42                 0.55
## area_se                              0.28            0.39                 0.54
## smoothness_se                       -0.06           -0.06                -0.10
## compactness_se                       0.68            0.64                 0.48
## concavity_se                         0.48            0.66                 0.44
## concave points_se                    0.45            0.55                 0.60
## symmetry_se                          0.06            0.04                -0.03
## fractal_dimension_se                 0.39            0.38                 0.22
## radius_worst                         0.48            0.57                 0.79
## texture_worst                        0.36            0.37                 0.36
## perimeter_worst                      0.53            0.62                 0.82
## area_worst                           0.44            0.54                 0.75
## smoothness_worst                     0.57            0.52                 0.55
## compactness_worst                    1.00            0.89                 0.80
## concavity_worst                      0.89            1.00                 0.86
## concave points_worst                 0.80            0.86                 1.00
## symmetry_worst                       0.61            0.53                 0.50
## fractal_dimension_worst              0.81            0.69                 0.51
##                         symmetry_worst fractal_dimension_worst
## radius_mean                       0.16                    0.01
## texture_mean                      0.11                    0.12
## perimeter_mean                    0.19                    0.05
## area_mean                         0.14                    0.00
## smoothness_mean                   0.39                    0.50
## compactness_mean                  0.51                    0.69
## concavity_mean                    0.41                    0.51
## concave points_mean               0.38                    0.37
## symmetry_mean                     0.70                    0.44
## fractal_dimension_mean            0.33                    0.77
## radius_se                         0.09                    0.05
## texture_se                       -0.13                   -0.05
## perimeter_se                      0.11                    0.09
## area_se                           0.07                    0.02
## smoothness_se                    -0.11                    0.10
## compactness_se                    0.28                    0.59
## concavity_se                      0.20                    0.44
## concave points_se                 0.14                    0.31
## symmetry_se                       0.39                    0.08
## fractal_dimension_se              0.11                    0.59
## radius_worst                      0.24                    0.09
## texture_worst                     0.23                    0.22
## perimeter_worst                   0.27                    0.14
## area_worst                        0.21                    0.08
## smoothness_worst                  0.49                    0.62
## compactness_worst                 0.61                    0.81
## concavity_worst                   0.53                    0.69
## concave points_worst              0.50                    0.51
## symmetry_worst                    1.00                    0.54
## fractal_dimension_worst           0.54                    1.00
```

```
# view correlation visually
dfii%>%
  cor(dfii) %>%
  corrplot(
    type = "upper",      # matrix: full, upper, or lower
    diag = F,            # remove diagonal
    order = "original",  # order for labels
    tl.col = "black",    # font color
    tl.srt = 75,         # label angel
    tl.cex = 0.6         # font size
  )
```



```
dfii %$% cor.test(radius_mean, perimeter_mean)
```

```
##
##  Pearson's product-moment correlation
##
## data:  radius_mean and perimeter_mean
## t = 513.79, df = 1136, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.9975910 0.9980906
## sample estimates:
##       cor
## 0.9978553
```

```
dfii %$% cor.test(radius_mean, fractal_dimension_worst)
```

```
##
##  Pearson's product-moment correlation
##
## data:  radius_mean and fractal_dimension_worst
## t = 0.23816, df = 1136, p-value = 0.8118
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.05106638  0.06515043
## sample estimates:
```

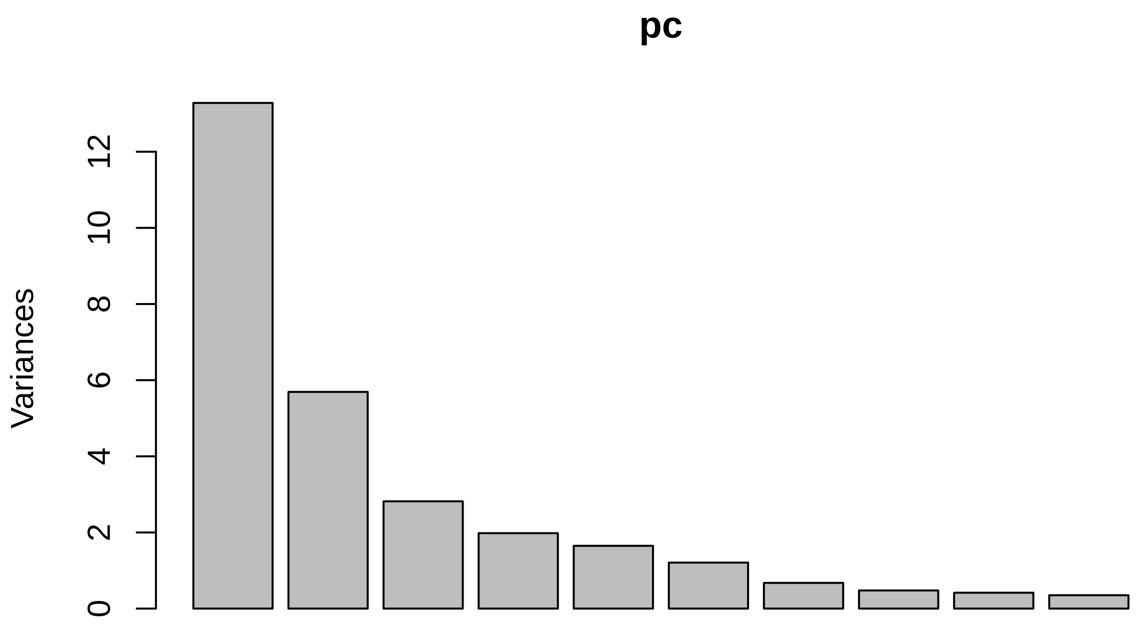
```
##         cor
## 0.007065886
```

PRINCIPAL COMPONENT MODEL

```
pc <- dfii %>%
  prcomp(
    center = TRUE,
    scale = TRUE
  )
summary(pc)
```

```
## Importance of components:
##                           PC1    PC2     PC3     PC4     PC5     PC6     PC7
## Standard deviation     3.6444 2.3857 1.67867 1.40735 1.28403 1.09880 0.82172
## Proportion of Variance 0.4427 0.1897 0.09393 0.06602 0.05496 0.04025 0.02251
## Cumulative Proportion  0.4427 0.6324 0.72636 0.79239 0.84734 0.88759 0.91010
##                            PC8    PC9    PC10   PC11    PC12    PC13    PC14
## Standard deviation     0.69037 0.6457 0.59219 0.5421 0.51104 0.49128 0.39624
## Proportion of Variance 0.01589 0.0139 0.01169 0.0098 0.00871 0.00805 0.00523
## Cumulative Proportion  0.92598 0.9399 0.95157 0.9614 0.97007 0.97812 0.98335
##                           PC15    PC16    PC17    PC18    PC19    PC20   PC21
## Standard deviation     0.30681 0.28260 0.24372 0.22939 0.22244 0.17652 0.1731
## Proportion of Variance 0.00314 0.00266 0.00198 0.00175 0.00165 0.00104 0.0010
## Cumulative Proportion  0.98649 0.98915 0.99113 0.99288 0.99453 0.99557 0.9966
##                           PC22    PC23   PC24    PC25    PC26    PC27    PC28
## Standard deviation     0.16565 0.15602 0.1344 0.12442 0.09043 0.08307 0.03987
## Proportion of Variance 0.00091 0.00081 0.0006 0.00052 0.00027 0.00023 0.00005
## Cumulative Proportion  0.99749 0.99830 0.9989 0.99942 0.99969 0.99992 0.99997
##                           PC29    PC30
## Standard deviation     0.02736 0.01153
## Proportion of Variance 0.00002 0.00000
## Cumulative Proportion  1.00000 1.00000
```

```
# plot the principal component
plot(pc)
```

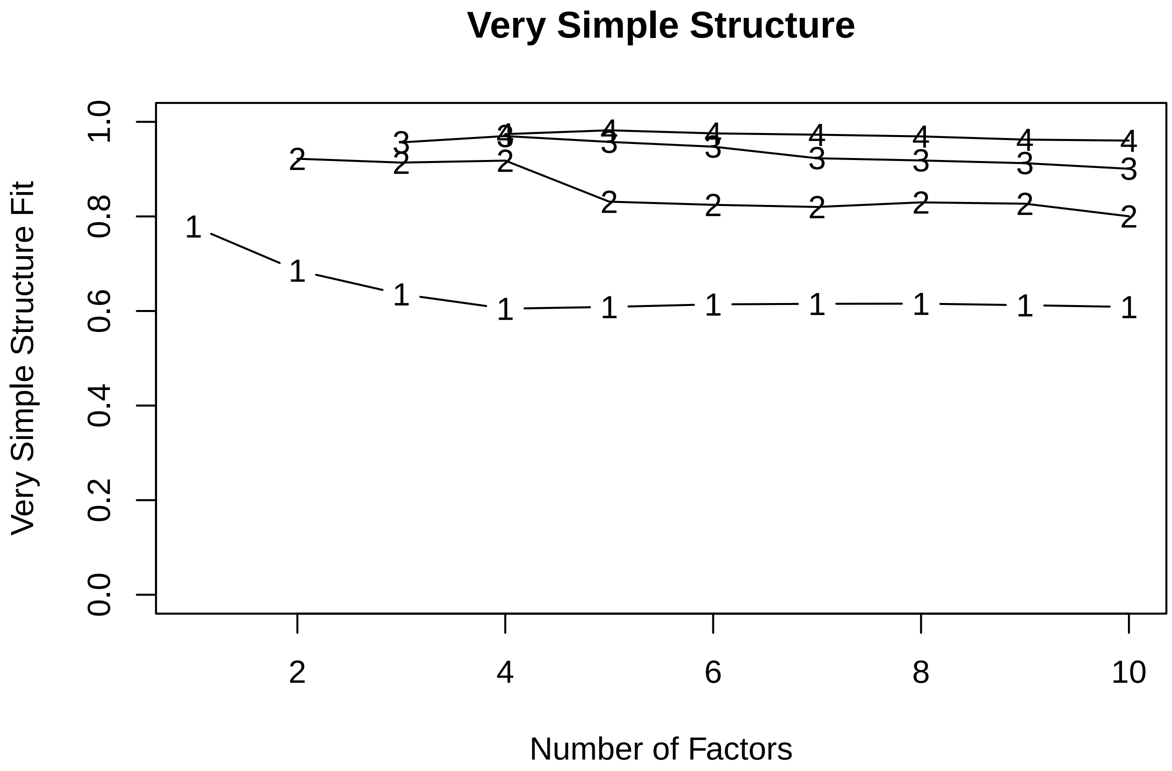


VERY SIMPLE STRUCTURE

Use "Very Simple Structure" to suggest number of factors Note: MAP = Minimum Absolute Partial correlation; n is the proposed maximum number of factors

```
dfii %>%
  dplyr::select(1:30) %>%
  vss(n = 10)
```

```
## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.

## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.

## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.

## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.
## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate, : An
## ultra-Heywood case was detected. Examine the results carefully

## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.

## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate, : An
## ultra-Heywood case was detected. Examine the results carefully

## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.

## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate, : An
## ultra-Heywood case was detected. Examine the results carefully

## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.

## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate, : An
## ultra-Heywood case was detected. Examine the results carefully

## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.

## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate, : An
## ultra-Heywood case was detected. Examine the results carefully

## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.

## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate, : An
## ultra-Heywood case was detected. Examine the results carefully

## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.

## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate, : An
## ultra-Heywood case was detected. Examine the results carefully
```



```
##
## Very Simple Structure
## Call: vss(x = ., n = 10)
## VSS complexity 1 achieves a maximimum of 0.78  with  1  factors
## VSS complexity 2 achieves a maximimum of 0.92  with  2  factors
##
## The Velicer MAP achieves a minimum of 0.07  with  9  factors
## BIC achieves a minimum of  13154.76  with  10  factors
## Sample Size adjusted BIC achieves a minimum of  13726.5  with  10  factors
##
## Statistics by number of factors
##    vss1 vss2   map dof chisq prob sqresid  fit RMSEA   BIC SABIC complex eChisq
## 1  0.78 0.00 0.171 405 58779    0   49.94 0.78  0.36 55929 57216     1.0  42948
## 2  0.69 0.92 0.087 376 39285    0   17.64 0.92  0.30 36640 37834     1.3  12052
## 3  0.64 0.91 0.073 348 33490    0    9.82 0.96  0.29 31041 32146     1.5   5812
## 4  0.60 0.92 0.079 321 30570    0    5.86 0.97  0.29 28311 29330     1.5   3345
## 5  0.61 0.83 0.071 295 26904    0    3.13 0.99  0.28 24828 25765     1.7   1439
## 6  0.61 0.82 0.072 270 25412    0    1.58 0.99  0.29 23512 24370     1.9    597
## 7  0.62 0.82 0.075 246 21868    0    1.11 1.00  0.28 20137 20918     2.0    290
## 8  0.62 0.83 0.080 223 19217    0    0.93 1.00  0.27 17648 18356     2.1    184
## 9  0.61 0.83 0.068 201 16330    0    0.74 1.00  0.27 14915 15554     2.1    109
## 10 0.61 0.80 0.077 180 14421    0    0.63 1.00  0.26 13155 13726     2.2     64
##      SRMR eCRMS  eBIC
## 1  0.2083 0.216 40098
## 2  0.1103 0.119  9406
## 3  0.0766 0.086  3363
## 4  0.0581 0.068  1086
## 5  0.0381 0.046  -637
## 6  0.0246 0.031 -1303
## 7  0.0171 0.023 -1441
## 8  0.0136 0.019 -1385
## 9  0.0105 0.015 -1305
## 10 0.0081 0.013 -1202
```

```
# Or use "nfactors" to do the same; includes VSS
dfii %>%
  dplyr::select(1:30) %>%
  nfactors(n = 10)
```

```
## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.

## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.

## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.

## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.

## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate, : An
## ultra-Heywood case was detected. Examine the results carefully

## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.

## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate, : An
## ultra-Heywood case was detected. Examine the results carefully

## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.

## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate, : An
## ultra-Heywood case was detected. Examine the results carefully

## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.

## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate, : An
## ultra-Heywood case was detected. Examine the results carefully
## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.
```

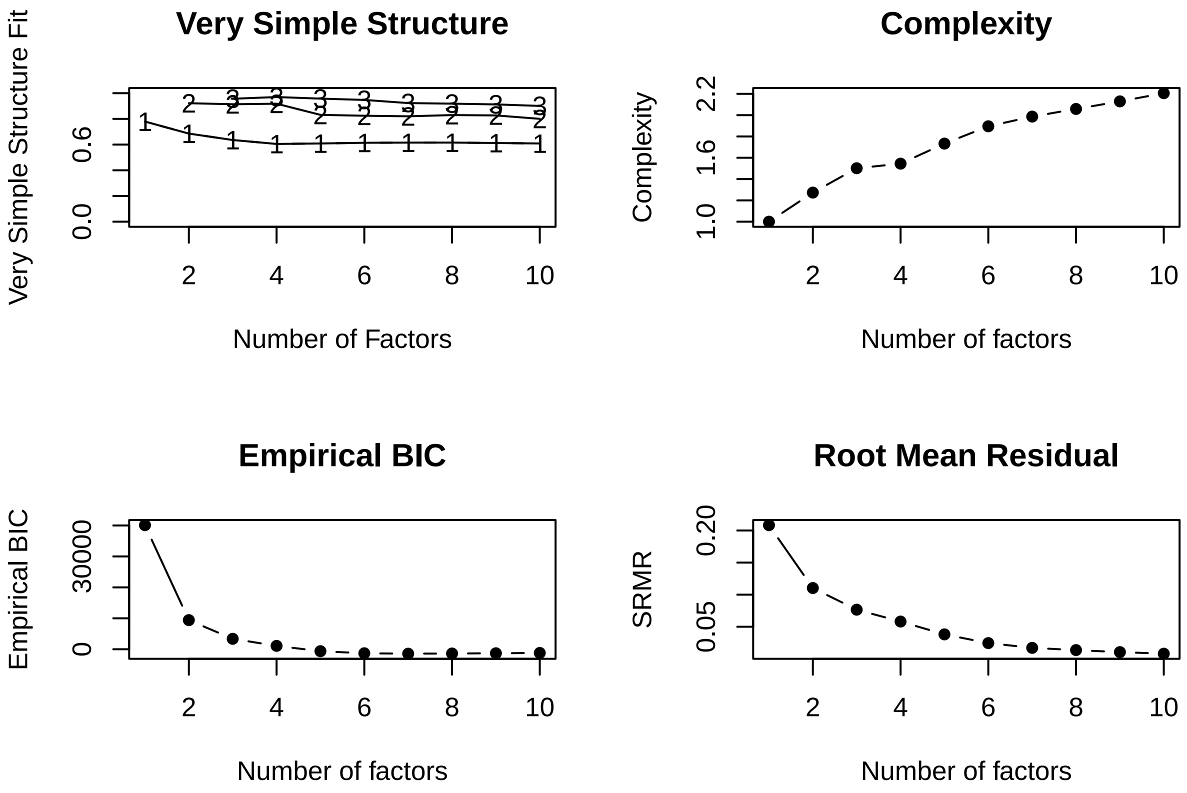
```
## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate, : An
## ultra-Heywood case was detected. Examine the results carefully
```

```
## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.
```

```
## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate, : An
## ultra-Heywood case was detected. Examine the results carefully
```

```
## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.
```

```
## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate, : An
## ultra-Heywood case was detected. Examine the results carefully
```



```
##
## Number of factors
## Call: vss(x = x, n = n, rotate = rotate, diagonal = diagonal, fm = fm,
##     n.obs = n.obs, plot = FALSE, title = title, use = use, cor = cor)
## VSS complexity 1 achieves a maximimum of 0.78  with  1  factors
## VSS complexity 2 achieves a maximimum of 0.92  with  2  factors
## The Velicer MAP achieves a minimum of 0.07  with  9  factors
## Empirical BIC achieves a minimum of  -1440.79  with  7  factors
## Sample Size adjusted BIC achieves a minimum of  13726.5  with  10  factors
##
## Statistics by number of factors
##    vss1 vss2   map dof chisq prob sqresid  fit RMSEA   BIC SABIC complex eChisq
## 1  0.78 0.00 0.171 405 58779    0   49.94 0.78  0.36 55929 57216     1.0  42948
## 2  0.69 0.92 0.087 376 39285    0   17.64 0.92  0.30 36640 37834     1.3  12052
## 3  0.64 0.91 0.073 348 33490    0    9.82 0.96  0.29 31041 32146     1.5   5812
## 4  0.60 0.92 0.079 321 30570    0    5.86 0.97  0.29 28311 29330     1.5   3345
## 5  0.61 0.83 0.071 295 26904    0    3.13 0.99  0.28 24828 25765     1.7   1439
## 6  0.61 0.82 0.072 270 25412    0    1.58 0.99  0.29 23512 24370     1.9    597
## 7  0.62 0.82 0.075 246 21868    0    1.11 1.00  0.28 20137 20918     2.0    290
## 8  0.62 0.83 0.080 223 19217    0    0.93 1.00  0.27 17648 18356     2.1    184
## 9  0.61 0.83 0.068 201 16330    0    0.74 1.00  0.27 14915 15554     2.1    109
## 10 0.61 0.80 0.077 180 14421    0    0.63 1.00  0.26 13155 13726     2.2     64
##      SRMR eCRMS  eBIC
## 1  0.2083 0.216 40098
## 2  0.1103 0.119  9406
## 3  0.0766 0.086  3363
## 4  0.0581 0.068  1086
## 5  0.0381 0.046  -637
## 6  0.0246 0.031 -1303
## 7  0.0171 0.023 -1441
## 8  0.0136 0.019 -1385
## 9  0.0105 0.015 -1305
## 10 0.0081 0.013 -1202
```

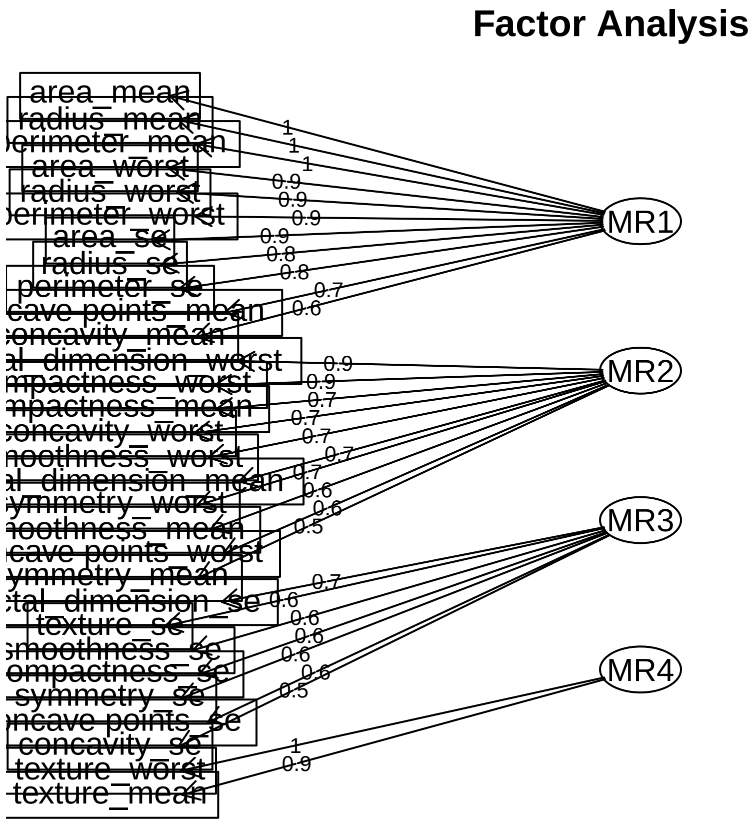
FACTOR ANALYSIS

Factor analysis using minimum residual (minres) method and oblimin rotation, which is useful for simple structure. Need to enter desired number of factors (from VSS or nfactors above).

```
# Calculate and plot factors with fa()
dfii %>%
  dplyr::select(1:30) %>%      # Select variables
  fa(                         # Use fa() function
    nfactors = 4,             # Use four factors
    rotate = "oblimin"        # Oblimin oblique rotation
  ) %T>%                      # T-pipe
  fa.diagram() %>%            # Diagram of factors and variables
  print()                     # Print results
```

```
## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.
```

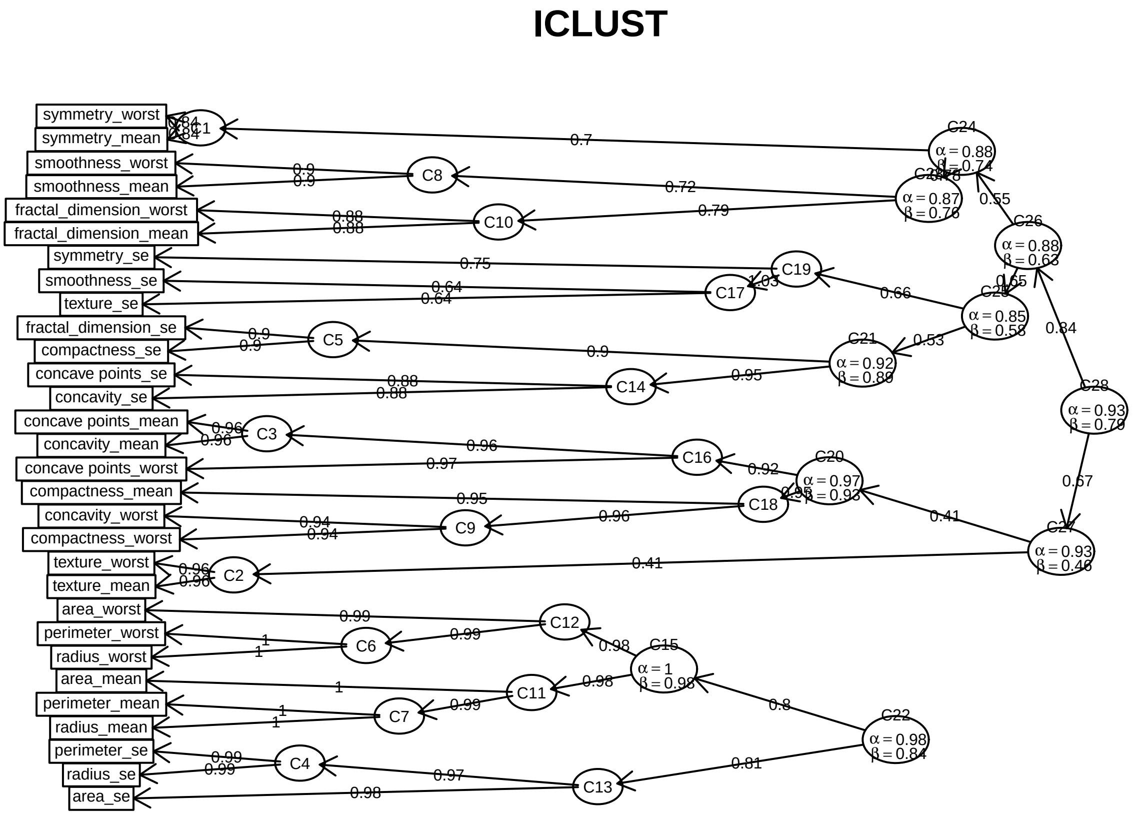
```
## Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate, : An
## ultra-Heywood case was detected. Examine the results carefully
```



```
## Factor Analysis using method =  minres
## Call: fa(r = ., nfactors = 4, rotate = "oblimin")
## Standardized loadings (pattern matrix) based upon correlation matrix
##                           MR1   MR2   MR3   MR4   h2     u2 com
## radius_mean              0.97  0.00 -0.13  0.04 0.95  0.047 1.0
## texture_mean             0.07 -0.05  0.10  0.89 0.83  0.171 1.0
## perimeter_mean           0.96  0.05 -0.10  0.04 0.96  0.043 1.0
## area_mean                0.98 -0.04 -0.06  0.03 0.96  0.036 1.0
## smoothness_mean          0.09  0.62  0.14 -0.15 0.47  0.531 1.3
## compactness_mean         0.33  0.74  0.21  0.01 0.92  0.082 1.6
## concavity_mean           0.55  0.53  0.24  0.05 0.91  0.090 2.4
## concave points_mean      0.74  0.41  0.09  0.00 0.92  0.082 1.6
## symmetry_mean            0.06  0.54  0.23 -0.05 0.42  0.585 1.4
## fractal_dimension_mean  -0.43  0.72  0.40 -0.10 0.81  0.191 2.3
## radius_se                0.81 -0.14  0.45 -0.01 0.84  0.160 1.6
## texture_se              -0.16 -0.25  0.62  0.46 0.55  0.452 2.4
## perimeter_se             0.80 -0.10  0.46 -0.01 0.85  0.152 1.6
## area_se                  0.87 -0.15  0.31 -0.03 0.81  0.189 1.3
## smoothness_se           -0.19  0.02  0.62 -0.03 0.40  0.597 1.2
## compactness_se           0.07  0.48  0.58  0.06 0.74  0.257 2.0
## concavity_se             0.11  0.37  0.54  0.02 0.58  0.423 1.9
## concave points_se        0.34  0.27  0.57 -0.04 0.66  0.343 2.1
## symmetry_se             -0.09  0.06  0.58 -0.03 0.35  0.650 1.1
## fractal_dimension_se    -0.13  0.38  0.66 -0.02 0.67  0.326 1.7
## radius_worst             0.94  0.08 -0.17  0.08 0.98  0.020 1.1
## texture_worst           -0.02  0.09 -0.05  1.00 1.02 -0.017 1.0
## perimeter_worst          0.93  0.13 -0.14  0.08 0.99  0.014 1.1
## area_worst               0.94  0.03 -0.11  0.07 0.95  0.053 1.0
## smoothness_worst        -0.02  0.72 -0.09  0.00 0.49  0.510 1.0
## compactness_worst        0.16  0.85 -0.08  0.15 0.88  0.120 1.2
## concavity_worst          0.32  0.74 -0.02  0.14 0.84  0.163 1.4
## concave points_worst     0.59  0.61 -0.12  0.06 0.92  0.076 2.1
## symmetry_worst          -0.01  0.67 -0.14  0.05 0.43  0.565 1.1
## fractal_dimension_worst -0.24  0.92  0.02  0.08 0.81  0.192 1.2
##
##                        MR1  MR2  MR3  MR4
## SS loadings           9.87 6.99 3.77 2.26
## Proportion Var        0.33 0.23 0.13 0.08
## Cumulative Var        0.33 0.56 0.69 0.76
## Proportion Explained  0.43 0.31 0.16 0.10
## Cumulative Proportion 0.43 0.74 0.90 1.00
##
##  With factor correlations of
##      MR1  MR2  MR3  MR4
## MR1 1.00 0.28 0.08 0.27
## MR2 0.28 1.00 0.24 0.15
## MR3 0.08 0.24 1.00 0.02
## MR4 0.27 0.15 0.02 1.00
##
## Mean item complexity =  1.5
## Test of the hypothesis that 4 factors are sufficient.
##
## The degrees of freedom for the null model are  435  and the objective function was  70.65 with Chi S
## The degrees of freedom for the model are 321  and the objective function was  27.21
##
## The root mean square of the residuals (RMSR) is  0.06
## The df corrected root mean square of the residuals is  0.07
##
## The harmonic number of observations is  1138 with the empirical chi square  3344.58  with prob <  0
## The total number of observations was  1138  with Likelihood Chi Square =  30569.59  with prob <  0
##
## Tucker Lewis Index of factoring reliability =  0.481
## RMSEA index =  0.288  and the 90 % confidence intervals are  0.285 0.291
## BIC =  28310.7
## Fit based upon off diagonal values = 0.99
```

HIERARCHICAL CLUSTERING

```
# Hierarchical clustering of items with iclust()
dfii %>%
  dplyr::select(1:30) %>%
  iclust()
```



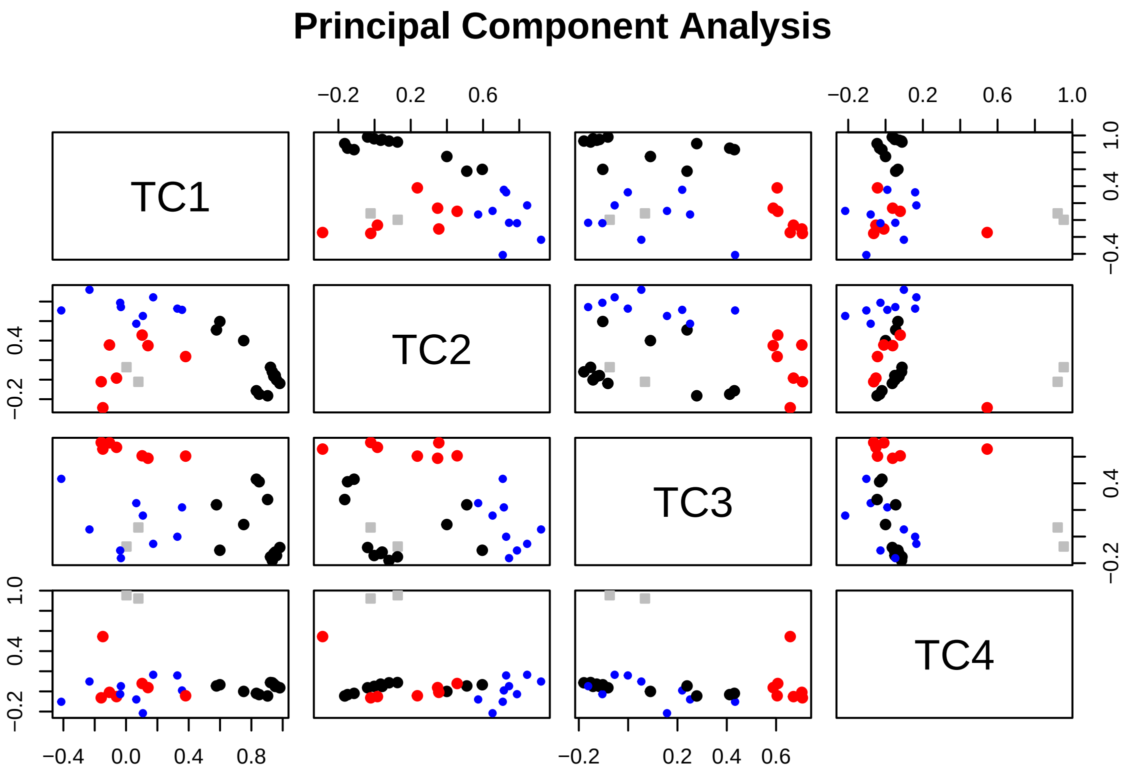
```
## ICLUST (Item Cluster Analysis)
## Call: iclust(r.mat = .)
##
## Purified Alpha:
##  C22  C28
## 0.98 0.92
##
## G6* reliability:
## C22 C28
##   1   1
##
## Original Beta:
##  C22  C28
## 0.84 0.79
##
## Cluster size:
## C22 C28
##  11  19
##
## Item by Cluster Structure matrix:
##                           O   P   C22   C28
## radius_mean             C22 C22  0.95  0.33
## texture_mean            C28 C28  0.34  0.36
## perimeter_mean          C22 C22  0.96  0.38
## area_mean               C22 C22  0.97  0.33
## smoothness_mean         C28 C28  0.28  0.65
## compactness_mean        C28 C28  0.60  0.90
## concavity_mean          C28 C28  0.76  0.84
## concave points_mean     C28 C22  0.87  0.72
## symmetry_mean           C28 C28  0.26  0.65
## fractal_dimension_mean  C28 C28 -0.15  0.66
## radius_se               C22 C22  0.86  0.46
## texture_se              C28 C28  0.00  0.26
## perimeter_se            C22 C22  0.85  0.49
## area_se                 C22 C22  0.88  0.38
## smoothness_se           C28 C28 -0.09  0.33
## compactness_se          C28 C28  0.31  0.81
## concavity_se            C28 C28  0.29  0.71
## concave points_se       C28 C28  0.47  0.71
## symmetry_se             C28 C28  0.01  0.39
## fractal_dimension_se    C28 C28  0.08  0.66
## radius_worst            C22 C22  0.96  0.37
## texture_worst           C28 C28  0.31  0.38
## perimeter_worst         C22 C22  0.97  0.43
## area_worst              C22 C22  0.96  0.36
## smoothness_worst        C28 C28  0.21  0.61
## compactness_worst       C28 C28  0.46  0.81
## concavity_worst         C28 C28  0.57  0.81
## concave points_worst    C28 C22  0.77  0.73
## symmetry_worst          C28 C28  0.20  0.57
## fractal_dimension_worst C28 C28  0.10  0.75
##
## With eigenvalues of:
## C22 C28
## 9.9 8.6
##
## Purified scale intercorrelations
##  reliabilities on diagonal
##  correlations corrected for attenuation above diagonal:
##      C22  C28
## C22 0.98 0.48
## C28 0.45 0.92
##
## Cluster fit =  0.81   Pattern fit =  0.95  RMSR =  0.12
```

PC WITH K FACTORS

```
# First PCA with no rotation, specify 5 factors
dfii %>% principal(nfactors = 4)
```

```
## Principal Components Analysis
## Call: principal(r = ., nfactors = 4)
## Standardized loadings (pattern matrix) based upon correlation matrix
##                           RC1   RC2   RC3   RC4   h2    u2 com
## radius_mean              0.95  0.12 -0.13  0.10 0.95 0.049 1.1
## texture_mean             0.25  0.06  0.05  0.91 0.90 0.105 1.2
## perimeter_mean           0.95  0.17 -0.10  0.10 0.95 0.045 1.1
## area_mean                0.97  0.09 -0.08  0.09 0.96 0.039 1.0
## smoothness_mean          0.16  0.65  0.26 -0.19 0.55 0.452 1.6
## compactness_mean         0.46  0.76  0.33  0.05 0.91 0.091 2.1
## concavity_mean           0.66  0.60  0.32  0.11 0.91 0.092 2.5
## concave points_mean      0.80  0.50  0.16  0.05 0.92 0.079 1.8
## symmetry_mean            0.14  0.58  0.34 -0.05 0.47 0.527 1.8
## fractal_dimension_mean  -0.31  0.66  0.54 -0.09 0.83 0.174 2.4
## radius_se                0.83 -0.01  0.40  0.03 0.85 0.151 1.4
## texture_se              -0.06 -0.22  0.60  0.54 0.70 0.298 2.3
## perimeter_se             0.82  0.03  0.42  0.04 0.85 0.145 1.5
## area_se                  0.87 -0.02  0.26  0.01 0.83 0.167 1.2
## smoothness_se           -0.14  0.00  0.70 -0.05 0.51 0.487 1.1
## compactness_se           0.21  0.51  0.67  0.11 0.76 0.237 2.1
## concavity_se             0.22  0.40  0.64  0.07 0.62 0.378 2.0
## concave points_se        0.43  0.32  0.64  0.00 0.70 0.303 2.3
## symmetry_se             -0.04  0.05  0.67 -0.03 0.45 0.545 1.0
## fractal_dimension_se    -0.03  0.38  0.75  0.02 0.71 0.287 1.5
## radius_worst             0.94  0.20 -0.16  0.13 0.97 0.028 1.2
## texture_worst            0.20  0.19 -0.07  0.94 0.95 0.045 1.2
## perimeter_worst          0.94  0.25 -0.13  0.14 0.98 0.022 1.2
## area_worst               0.95  0.16 -0.11  0.12 0.95 0.053 1.1
## smoothness_worst         0.06  0.76  0.01 -0.01 0.58 0.423 1.0
## compactness_worst        0.32  0.86  0.07  0.19 0.88 0.123 1.4
## concavity_worst          0.46  0.77  0.11  0.19 0.84 0.155 1.8
## concave points_worst     0.68  0.66 -0.01  0.11 0.92 0.082 2.1
## symmetry_worst           0.07  0.72 -0.05  0.06 0.53 0.475 1.0
## fractal_dimension_worst -0.08  0.88  0.19  0.11 0.83 0.173 1.1
##
##                          RC1  RC2  RC3  RC4
## SS loadings           10.19 6.93 4.40 2.26
## Proportion Var         0.34 0.23 0.15 0.08
## Cumulative Var         0.34 0.57 0.72 0.79
## Proportion Explained   0.43 0.29 0.19 0.09
## Cumulative Proportion  0.43 0.72 0.91 1.00
##
## Mean item complexity =  1.5
## Test of the hypothesis that 4 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.06
##  with the empirical chi square  3773.14  with prob <  0
##
## Fit based upon off diagonal values = 0.98
```

```
# Second PCA with oblimin (oblique) rotation
dfii %>%
  principal(
    nfactors = 4,
    rotate = "oblimin"
  ) %>%
  plot()  # Plot position of variables on components
```


Principal Component Analysis
TC1
TC2
TC3
TC4
−0.2
0.2
0.6
−0.2
0.2
0.6
1.0
1.0
0.4
−0.4
0.4
−0.2
0.4
−0.2
1.0
0.4
−0.2
−0.4
0.0
0.4
0.8
−0.2
0.2
0.4
0.6