Project

2023-12-26

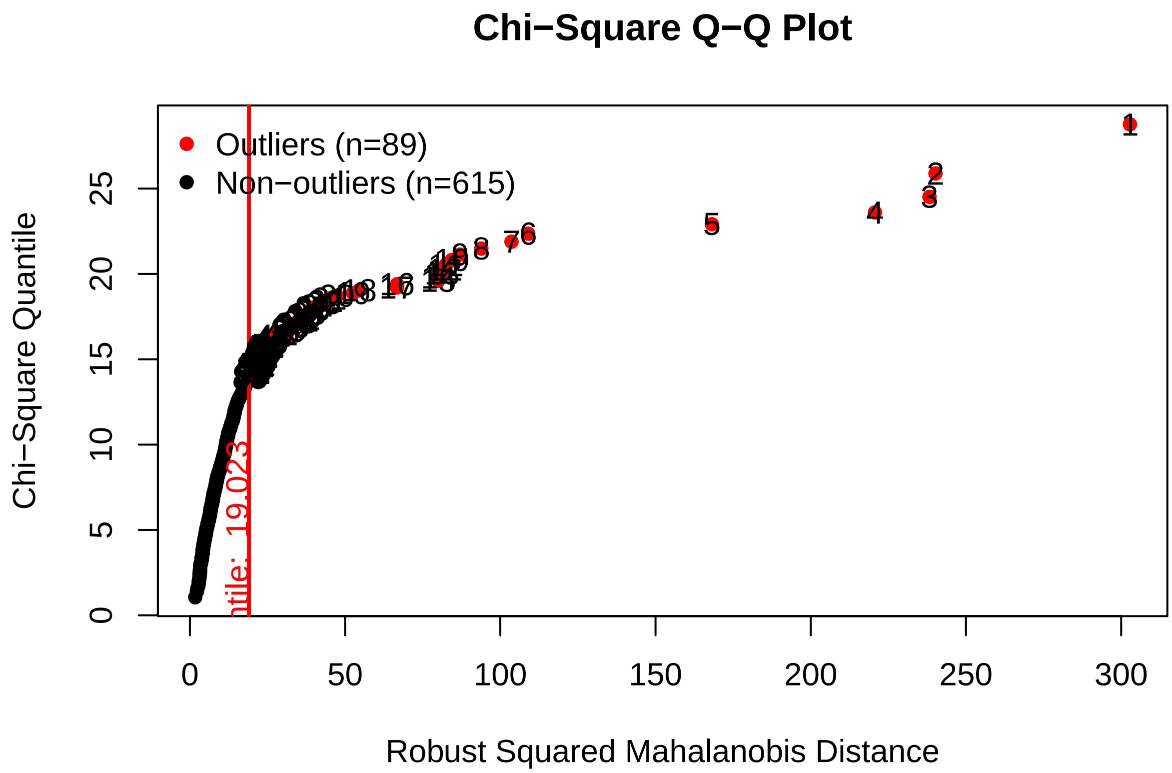
```
library(readxl)
xlfile <- "C:/Users/berka/OneDrive/Masaüstü/hintsağlıkyeni.xlsx"
data <- read_excel(xlfile)
```

```
data <- na.omit(data)


for (col in names(data)[-c(1, 2)]) {
  data[[col]] <- as.numeric(data[[col]])
}
```

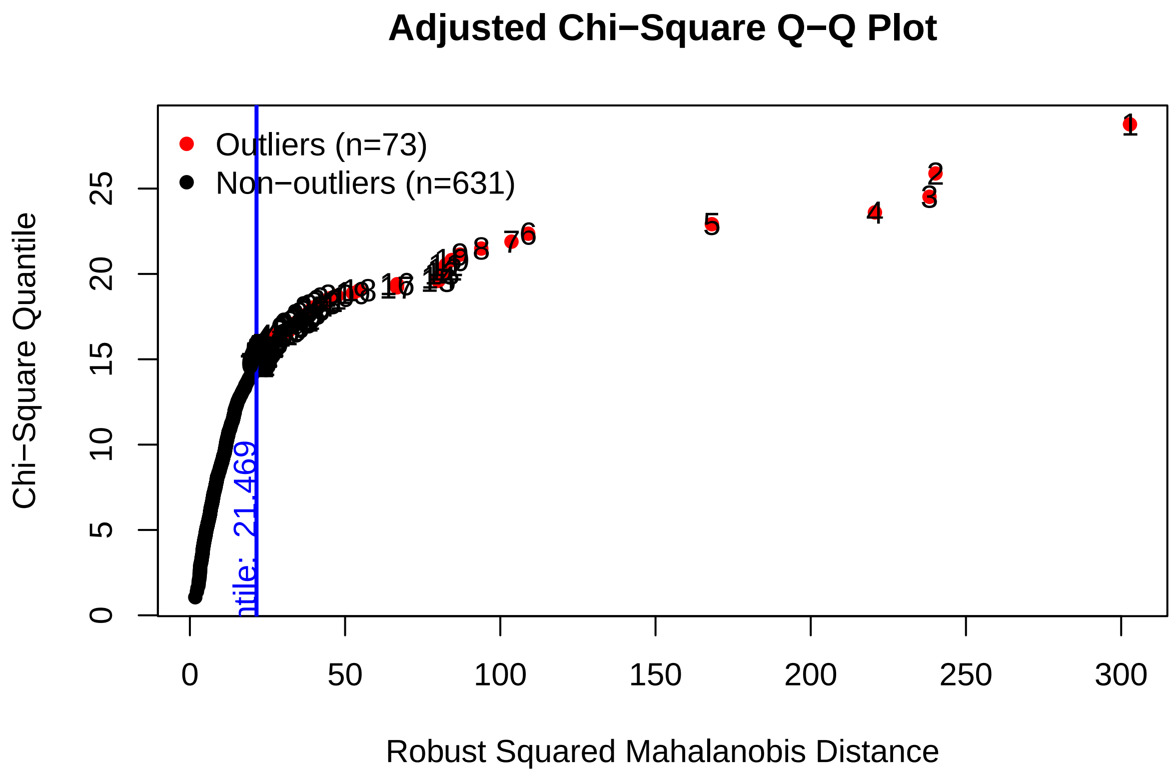
```
## Warning: NAs introduced by coercion

## Warning: NAs introduced by coercion

## Warning: NAs introduced by coercion
```

```
data_all <- data[, c("State/UT",'LiterateWomen (%)',
                              'Women schooling +10(%)',
                              'Current Use of Family Planning Methods (%)',
                              'Total Unmet need for Family Planning (%)',
                              'Unmet need for spacing (%)',
                              'Health worker t.a family planning w.n.u.(%)',
                              'Children <5 who are stunted (%)',
                              'Children <5 who are underweight  (%)',
                              'User mention s.e of family planning (%)' )]


data_numeric<- data[, c('Women schooling +10(%)','LiterateWomen (%)',
                              'Current Use of Family Planning Methods (%)',
                              'Total Unmet need for Family Planning (%)',
                              'Unmet need for spacing (%)',
                              'Health worker t.a family planning w.n.u.(%)',
                              'Children <5 who are stunted (%)',
                              'Children <5 who are underweight  (%)',
                              'User mention s.e of family planning (%)')]


data_all$`State/UT` <- as.factor(data_all$`State/UT`)
```

```
library(MVN)

#Mahalanobis distance
multi_result <- mvn(data = data_numeric, mvnTest = "royston", multivariateOutlierMethod = "quan")
```




```r
#As seen that, we have 189 outliers observations proved by Mahalanobis Distance in this dataset

#Adjusted Mahalanobis distance
multi_result <- mvn(data = data_numeric, mvnTest = "royston", multivariateOutlierMethod = "adj")
```

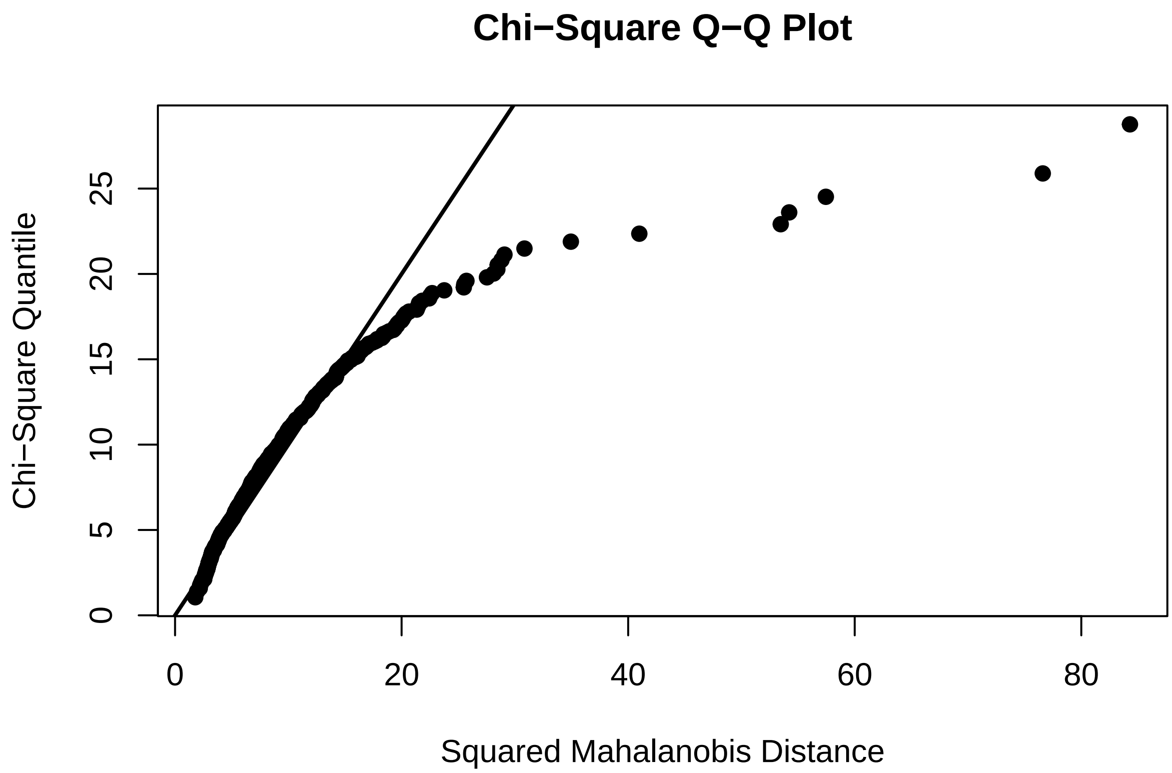



```r
#As seen that, we have 172 outliers observations proved by Adjusted Mahalanobis Distance in this datase
#Because our outliers make up more than 10% of the data we do not omit them.
```

```r
### EXPLORATORY DATA ANALYSIS ###

library(MVN)

cor_matrix <- cor(data_numeric)
high_corr_pairs <- which(upper.tri(cor_matrix, diag = TRUE) & cor_matrix > abs(0.7) & cor_matrix < abs(
high_corr_pairs
```

```
##                                          row col
## Women schooling +10(%)                     1   2
## Total Unmet need for Family Planning (%)   4   5
## Children <5 who are stunted (%)            7   8
```

```r
## corelation between (1 and 2), (4,5), (7,8) are high

## formal test

mvn(data_numeric, multivariatePlot = "qq")$plot
```

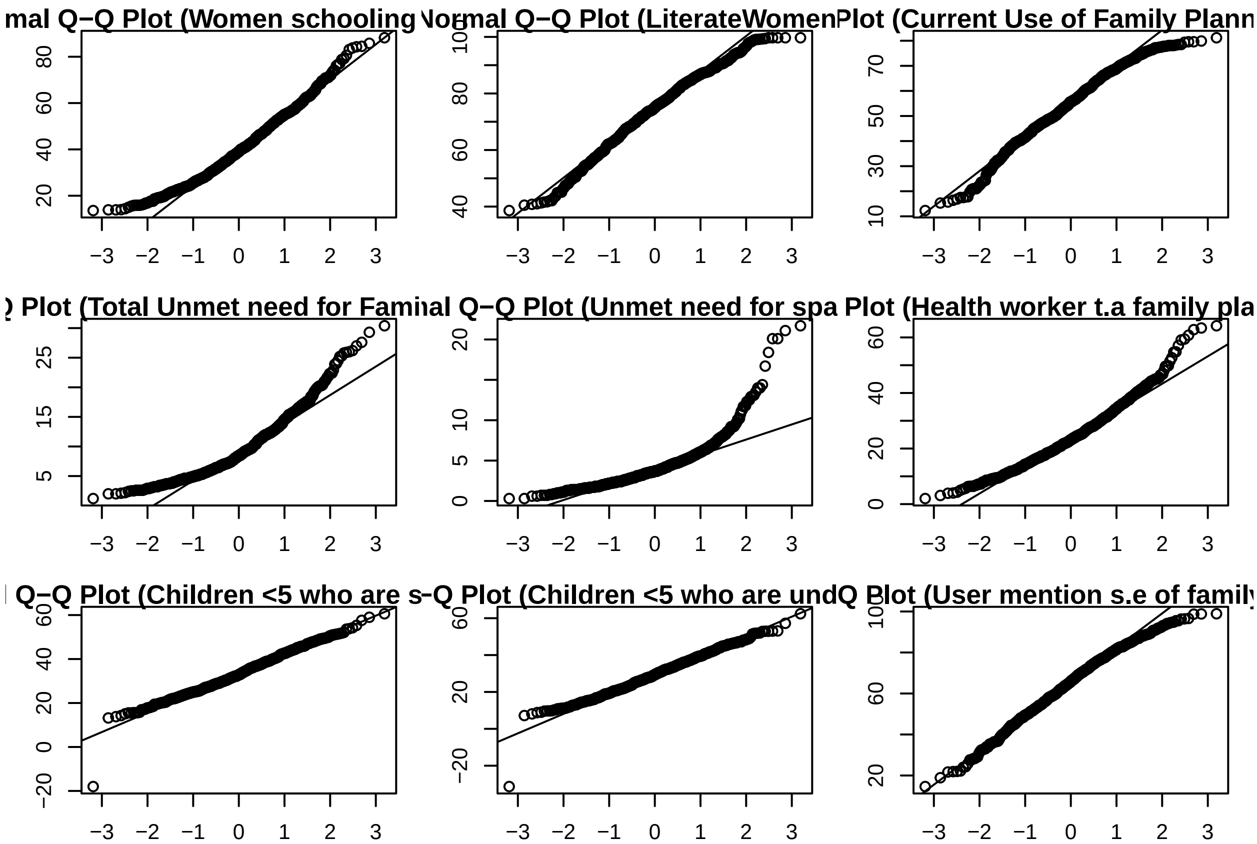



```
## NULL
```

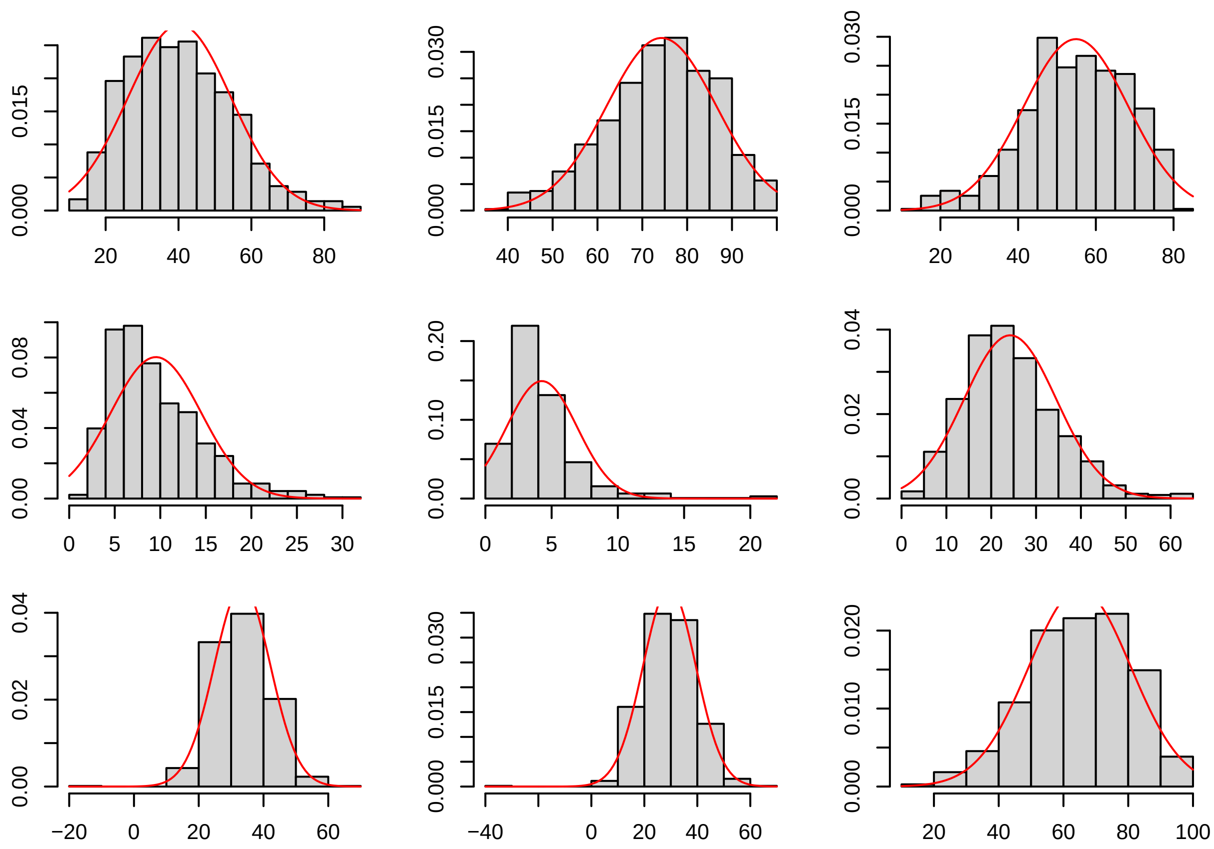
```
# left skewed

# our data observation is bigger than 20 so we used roytson

multi_result <- mvn(data = data_numeric, mvnTest = "royston")
multi_result$multivariateNormality
```

```
##      Test        H      p value MVN
## 1 Royston 316.0259 3.661439e-64  NO
```

```
# p value is less than alpha,
#we reject H0 and we can say that we don't have enough evidence to conclude that the data follow normal

# univariate Plots


par(mar = c(3, 3, 1, 1))
multi_result <- mvn(data = data_numeric, mvnTest = "royston", univariatePlot = "qqplot")
```




```
# univariate histograms



multi_result <- mvn(data = data_numeric, mvnTest = "royston", univariatePlot = "histogram", univariateT
```




```
multi_result$univariateNormality
```

```
##           Test                                      Variable Statistic   p value
## 1 Shapiro-Wilk           Women schooling +10(%)                0.9769  <0.001
## 2 Shapiro-Wilk              LiterateWomen (%)                  0.9864  <0.001
## 3 Shapiro-Wilk Current Use of Family Planning Methods (%)      0.9807  <0.001
## 4 Shapiro-Wilk  Total Unmet need for Family Planning (%)       0.9238  <0.001
## 5 Shapiro-Wilk            Unmet need for spacing (%)           0.8043  <0.001
## 6 Shapiro-Wilk Health worker t.a family planning w.n.u.(%)     0.9715  <0.001
## 7 Shapiro-Wilk         Children <5 who are stunted (%)         0.9853  <0.001
## 8 Shapiro-Wilk      Children <5 who are underweight  (%)       0.9850  <0.001
## 9 Shapiro-Wilk     User mention s.e of family planning (%)     0.9885  <0.001
##   Normality
## 1    NO
## 2    NO
## 3    NO
## 4    NO
## 5    NO
## 6    NO
## 7    NO
## 8    NO
## 9    NO
```

```
# As you see, all the variables which violates the
# multivariate normality
```

```r
#### INFERENCE ABOUT MEAN ####

#the responses we want to model are Total Unmet need for Family Planning (%) and Unmet need for spacing
library(magrittr)
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
```

```r
library(MVN)

y <- data_numeric %>% select(`Total Unmet need for Family Planning (%)`, `Unmet need for spacing (%)`)

#calculate the mean of the dependent variables
xbar = colMeans(y)
xbar
```

```
## Total Unmet need for Family Planning (%)
##                                 9.526420
##               Unmet need for spacing (%)
##                                 4.260653
```

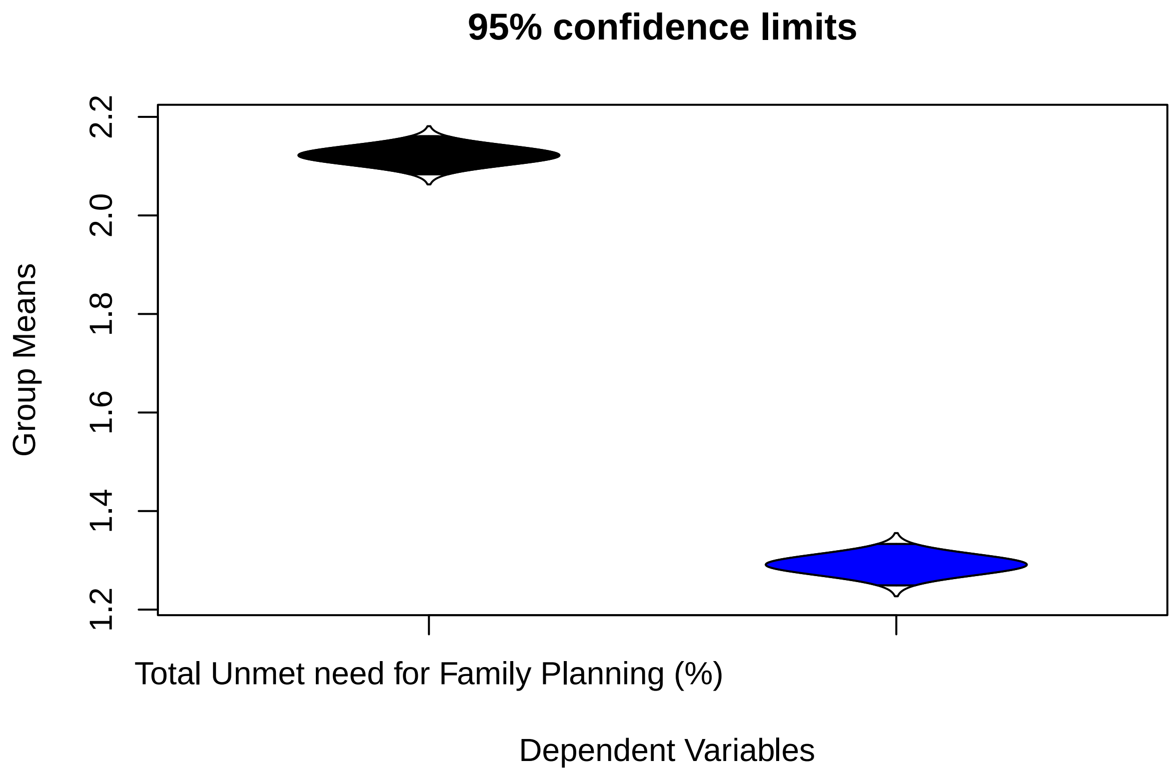
```r
#n>20 we use royston
test <- mvn(y, mvnTest = "royston")
test$multivariateNormality
```

```
##      Test        H      p value MVN
## 1 Royston 164.7273 7.047211e-37  NO
```

```r
#create MU0 vector
mu0 <- c(10, 5)
#The response matrix does not follow normal distribution so we can consider the log of the data
log_y <- log(y)
test2 <- mvn(log_y, mvnTest = "royston")
test2$univariateNormality
```

```
##               Test                                 Variable Statistic   p value
## 1 Anderson-Darling Total Unmet need for Family Planning (%)    0.8206    0.0339
## 2 Anderson-Darling        Unmet need for spacing (%)           1.5692    0.0005
##   Normality
## 1    NO
## 2    NO
```

```r
#we assumpiton normality is satisfied in this case, note that we will not test mu0 we will test log(mu0
library(psych)
error.bars(log_y, ylab="Group Means", xlab=" Dependent Variables")
```




```r
library(ICSNP)
```

```
## Loading required package: mvtnorm

## Loading required package: ICS
```

```r
library(DescTools)
```

```
##
## Attaching package: 'DescTools'

## The following objects are masked from 'package:psych':
##
##     AUC, ICC, SD
```

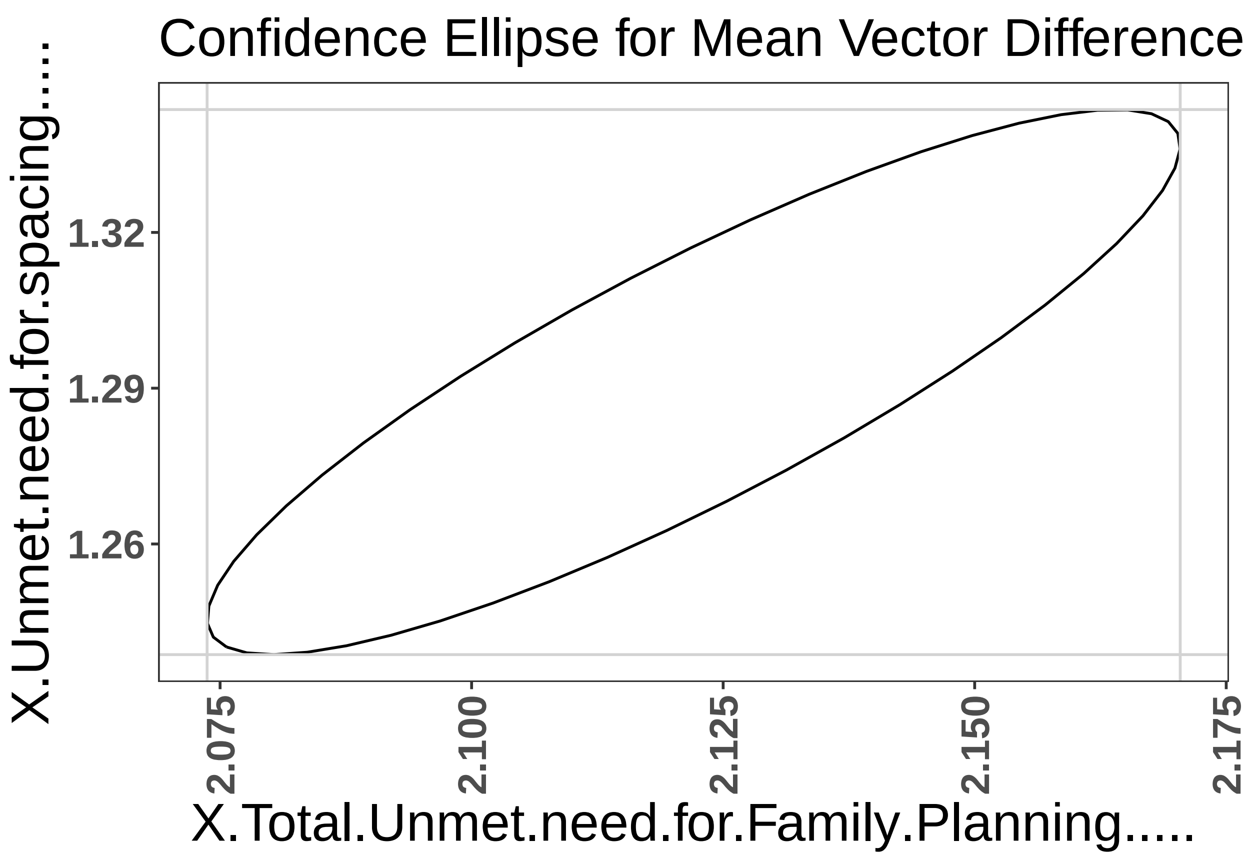
```r
HotellingsT2Test(log_y, mu = log(mu0))
```

```
##
##  Hotelling's one sample T2-test
##
## data:  log_y
## T.2 = 139.01, df1 = 2, df2 = 702, p-value < 2.2e-16
## alternative hypothesis: true location is not equal to c(2.30258509299405,1.6094379124341)
```

```
#since p < alpha we reject H0. Therefore, we do not have enough evidence to conclude that the log of th

library(mvdalab)
```

```
##
## Attaching package: 'mvdalab'

## The following object is masked from 'package:psych':
##
##     smc
```

```
MVcis(log_y)
```




```
##                                                [,1]     [,2]
## X.Total.Unmet.need.for.Family.Planning..... 2.073709 2.170430
## X.Unmet.need.for.spacing.....               1.238694 1.343663
```

```r
#while mu0 values are in the simultaneous confidence intervals for each variable since it is not in the
```

```r
###### COMPARISONS OF SEVERAL MULTIVARIATE MEANS ######

#here we will test whether the response variables (Total Unmet need for Family Planning (%),Unmet need

library(rstatix)
```

```
##
## Attaching package: 'rstatix'

## The following object is masked from 'package:stats':
##
##     filter
```

```r
library(dplyr)
library(ggplot2)
```

```
##
## Attaching package: 'ggplot2'

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

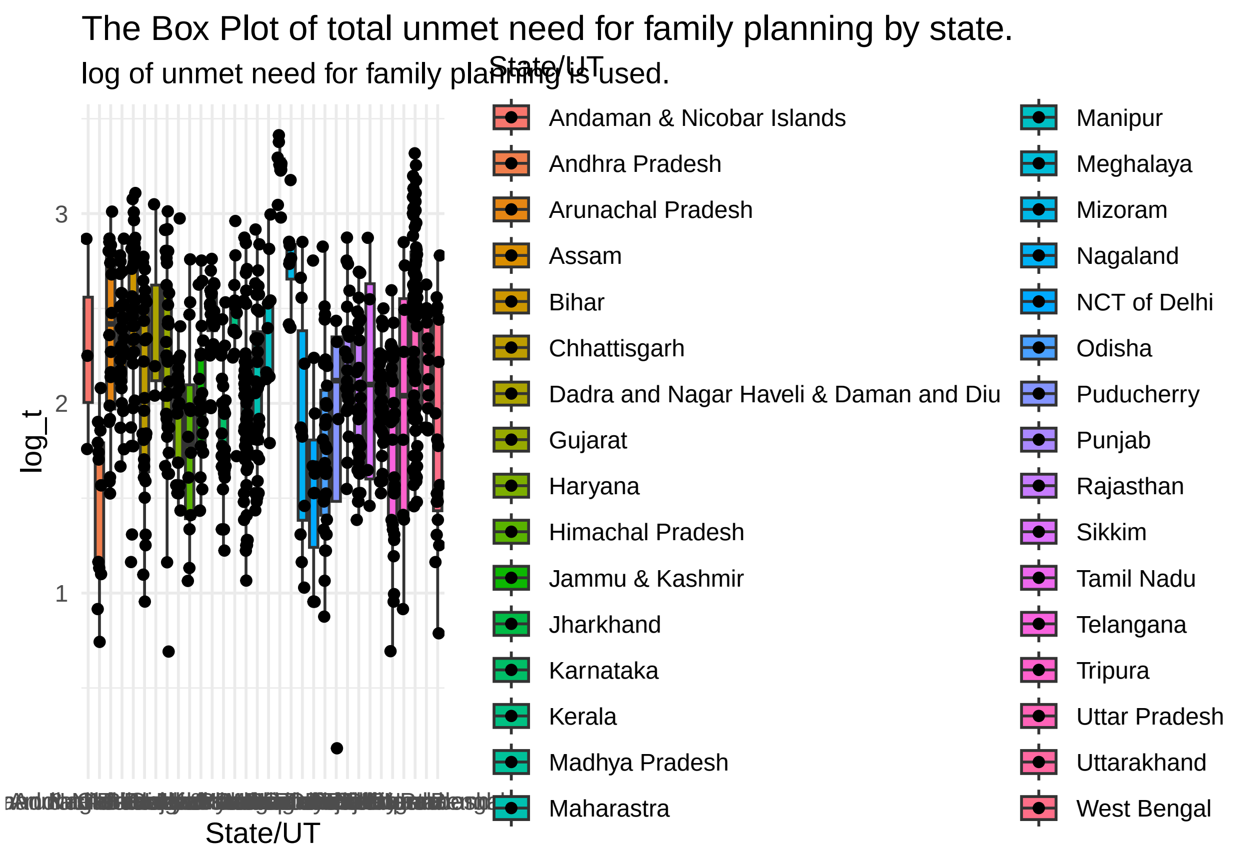
```r
mysubset <- data_all %>% select(`Total Unmet need for Family Planning (%)`, `Unmet need for spacing (%)

mysubset <- mysubset %>% mutate(log_t = log(`Total Unmet need for Family Planning (%)`), log_s = log(`Un

#sample size for at least one group is outside the acceptable range for the Shapiro-Wilk test so we fil

valid_groups <- mysubset %>%
  group_by(`State/UT`) %>%
  summarise(sample_size = n()) %>%
  filter(sample_size >= 3 & sample_size <= 5000) %>%
  pull(`State/UT`) %>%
  as.character()

filtered_mysubset <- mysubset %>%
  filter(`State/UT` %in% valid_groups)

filtered_mysubset %>%
  group_by(`State/UT`) %>%
  summarise(n = n(),
    mean_log_total = mean(log_t),
    mean_log_spacing = mean(log_s),
    sd_log_total = sd(log_t),
    sd_log_spacing = sd(log_s)
  )
```

```
## # A tibble: 32 x 6
##    `State/UT`      n mean_log_total mean_log_spacing sd_log_total sd_log_spacing
##    <fct>       <int>          <dbl>            <dbl>        <dbl>          <dbl>
##  1 Andaman & ~     3           2.29            1.20         0.556          0.945
##  2 Andhra Pra~    13           1.48            0.895        0.422          0.507
##  3 Arunachal ~    20           2.36            1.81         0.481          0.520
##  4 Assam          33           2.30            1.30         0.296          0.409
##  5 Bihar          38           2.46            1.68         0.452          0.437
##  6 Chhattisga~    27           2.01            1.12         0.539          0.516
##  7 Dadra and ~     3           2.43            1.62         0.543          0.654
##  8 Gujarat        33           2.21            1.38         0.536          0.573
##  9 Haryana        22           1.99            1.13         0.360          0.469
## 10 Himachal P~    12           1.82            0.829        0.552          0.469
## # i 22 more rows
```

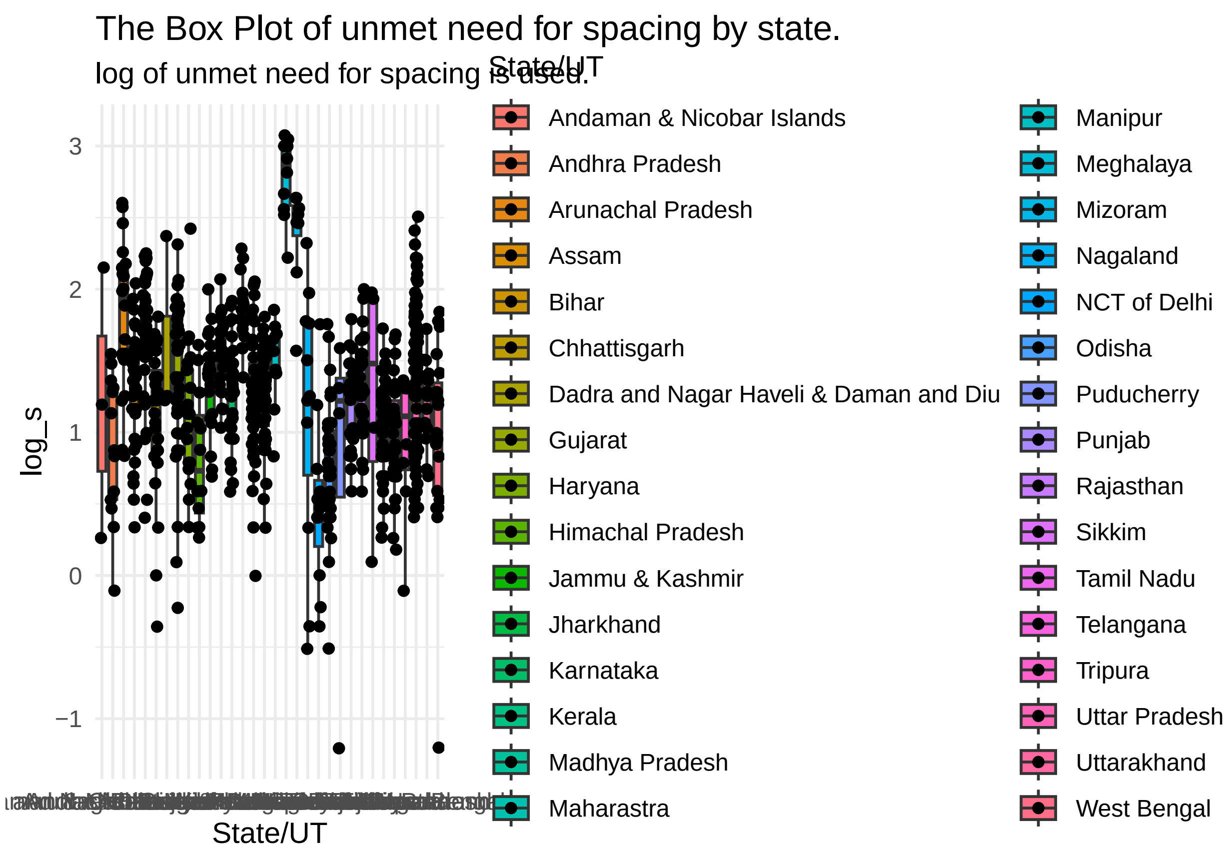
```
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'

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

```
p1 <- ggplot(filtered_mysubset, aes(x = `State/UT`, y = log_t, fill = `State/UT`)) + geom_boxplot(outli
  labs(title = "The Box Plot of total unmet need for family planning by state." ,subtitle = "log of unme
p2 <- ggplot(filtered_mysubset, aes(x = `State/UT`, y = log_s, fill = `State/UT`)) + geom_boxplot(outli
  labs(title = "The Box Plot of unmet need for spacing by state." ,subtitle = "log of unmet need for spa

grid.arrange(p1)
```




```
grid.arrange(p2)
```




```r
result <- filtered_mysubset %>%
  group_by(`State/UT`) %>%
  shapiro_test(log_t, log_s)

#we assume the normality is satisfied to find homogeneity of the variance we us Box's M-test

library(broom)
library(heplots)

boxM(Y = cbind(filtered_mysubset$log_t,filtered_mysubset$log_s), group = factor(filtered_mysubset$`Stat
```

```
##
##  Box's M-test for Homogeneity of Covariance Matrices
##
## data:  cbind(filtered_mysubset$log_t, filtered_mysubset$log_s)
## Chi-Sq (approx.) = 247.46, df = 93, p-value = 5.756e-16
```

```r
#we assume that we fail to reject the null hypothesis and conclude that variance-covariance  matrices a

#conduct manova

m1 <- manova(cbind(log_t,log_s)~`State/UT`,data =filtered_mysubset)
summary(m1)
```

```
##               Df  Pillai approx F num Df den Df    Pr(>F)
```

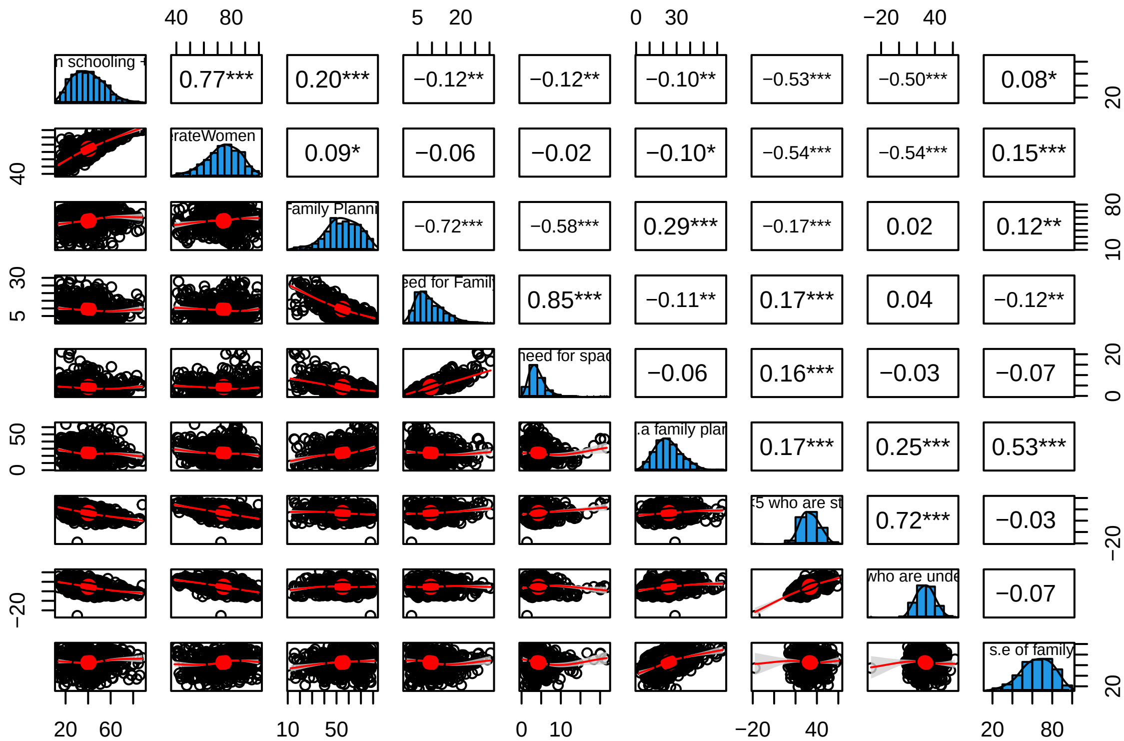
```
## `State/UT`  31 0.66126   10.612     62   1332 < 2.2e-16 ***
## Residuals  666
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# we are 95% confident that at least one state is
# significantly different than others since p < alpha
```

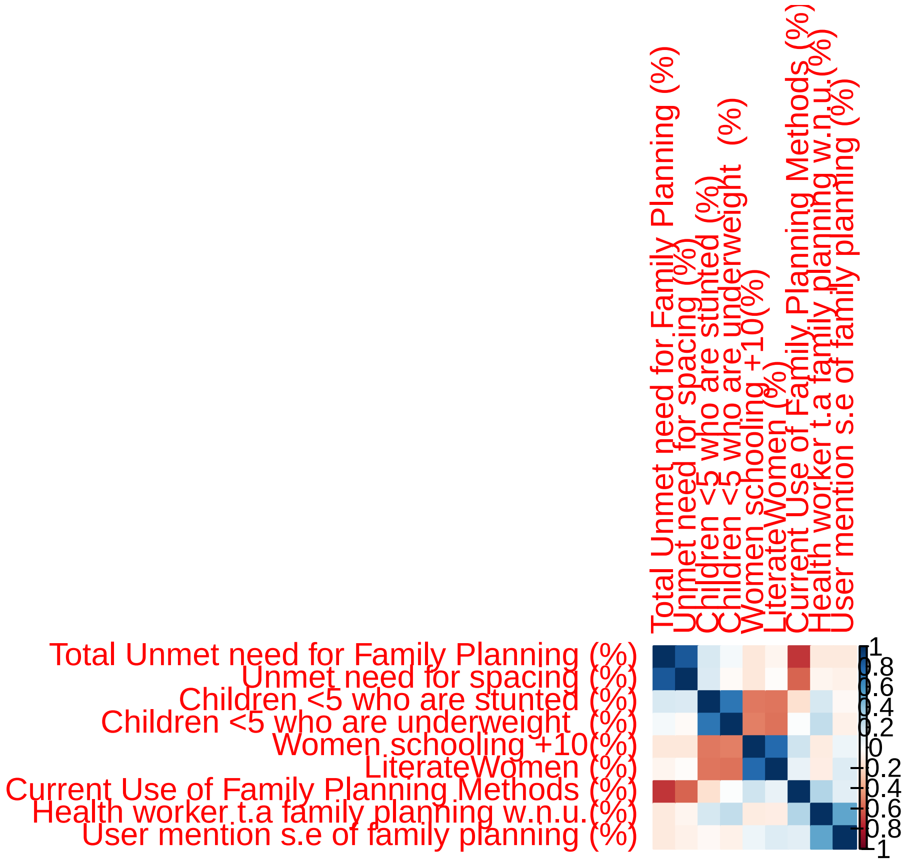
```
### Principal Component Analysis

library(psych)
library(ggplot2)
library(car)
```

```
## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:DescTools':
##
##     Recode

## The following object is masked from 'package:psych':
##
##     logit

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

```
pairs.panels(data_numeric,
             smooth = TRUE,
             scale = FALSE,
             density = TRUE,
             ellipses = TRUE,
             method = "pearson",
             pch = 21,
             lm = FALSE,
             cor = TRUE,
             jiggle = FALSE,
             factor = 2,
             hist.col = 4,
             stars = TRUE,
             ci = TRUE)
```



```r
res <- cor(data_numeric, method="pearson")
corrplot::corrplot(res, method= "color", order = "hclust")
```



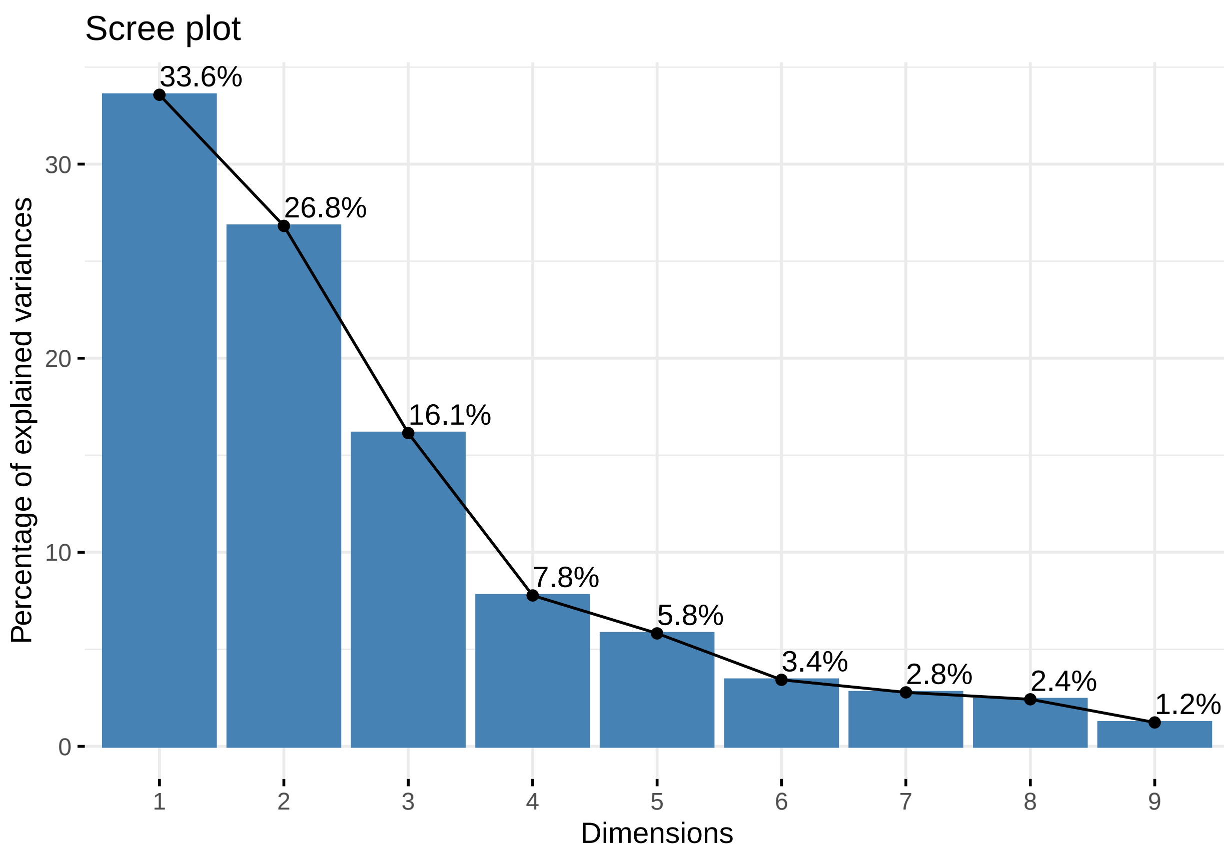

```r
scaled_data_numeric <- scale(data_numeric)

pcares <- prcomp(scaled_data_numeric)

summary(pcares)
```

```
## Importance of components:
##                           PC1    PC2    PC3     PC4     PC5     PC6     PC7
## Standard deviation     1.7383 1.5536 1.2053 0.83656 0.72383 0.55555 0.50035
## Proportion of Variance 0.3357 0.2682 0.1614 0.07776 0.05821 0.03429 0.02782
## Cumulative Proportion  0.3357 0.6039 0.7653 0.84310 0.90132 0.93561 0.96343
##                            PC8     PC9
## Standard deviation     0.46721 0.33299
## Proportion of Variance 0.02425 0.01232
## Cumulative Proportion  0.98768 1.00000
```

```r
library(factoextra)
```

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

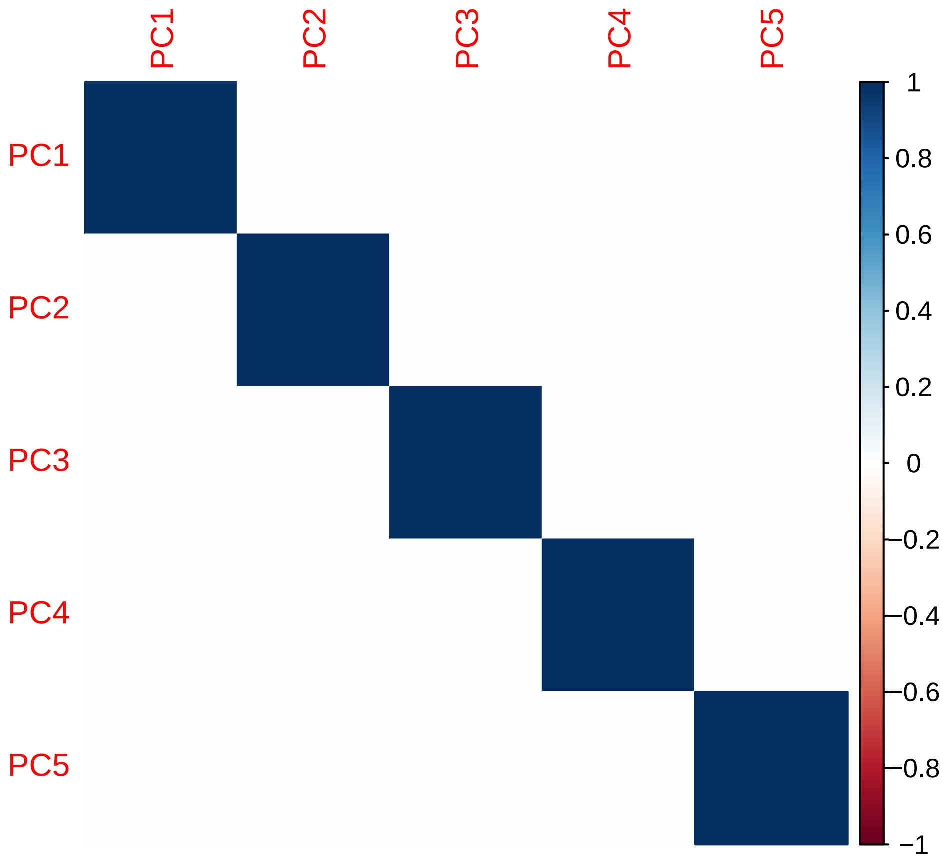
```r
fviz_eig(pcares,addlabels=TRUE) #represent the proportion of the values
```




```
#We extracted the first 5 from the rest of it because the first 5 explains 90% of variability in the da

pcares2 <- pcares$x[,1:5]
head(pcares2)
```

```
##             PC1        PC2        PC3        PC4        PC5
## [1,] -1.6725430 -0.2912808 -0.4148674  0.5195832 -1.3933076
## [2,] -1.2107778  1.6854198 -0.2808335  0.4366350  0.5148882
## [3,] -1.1962695 -2.3576824 -2.4621662 -0.2929252 -0.5881610
## [4,] -1.3167773  0.2067751  1.7941007 -0.8543695 -1.4066189
## [5,]  0.6116768  0.8587644  1.7878326  0.3399913 -1.4205090
## [6,] -0.5634287  0.7032105  2.3021959  0.6643093 -0.6837106
```

```
#to check orthogonality we construct a correlation plot

res1 <- cor(pcares2, method="pearson")
corrplot::corrplot(res1, method= "color", order = "hclust")
```




```
#it can be seen that components are linearly independent so there is no collinearity
```

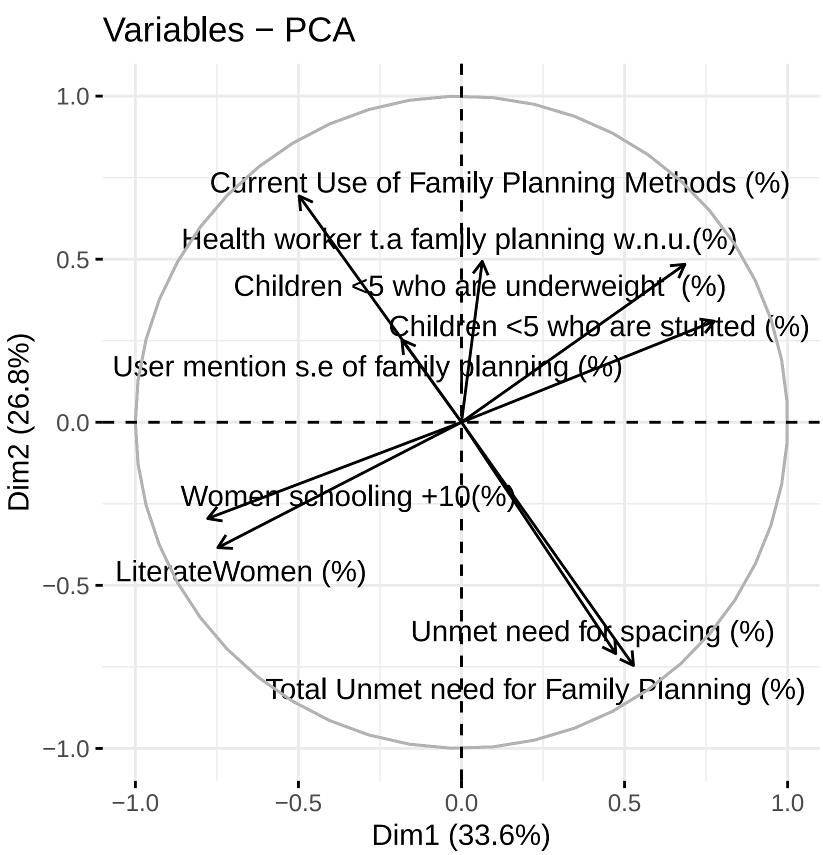
```
#correlation between PC's and our numerical variables

cor(scaled_data_numeric, pcares2)
```

```
##                                               PC1        PC2         PC3
## Women schooling +10(%)                -0.77833110 -0.2952574 -0.11391267
## LiterateWomen (%)                     -0.74715796 -0.3845246 -0.20908377
## Current Use of Family Planning Methods (%)  -0.49788112  0.6939213  0.05946638
## Total Unmet need for Family Planning (%)     0.52728978 -0.7457821 -0.26023221
## Unmet need for spacing (%)             0.47293006 -0.7094759 -0.32393377
## Health worker t.a family planning w.n.u.(%)  0.06362427  0.4933147 -0.74340325
## Children <5 who are stunted (%)        0.77536040  0.3098296 -0.05109300
## Children <5 who are underweight  (%)   0.68508196  0.4841905  0.01914394
## User mention s.e of family planning (%) -0.18274750  0.2526387 -0.81501206
##                                               PC4         PC5
## Women schooling +10(%)                 0.43536608  0.05852881
## LiterateWomen (%)                      0.34800212  0.13383142
## Current Use of Family Planning Methods (%)   0.15369829 -0.34977286
## Total Unmet need for Family Planning (%)     0.08181528 -0.12005311
## Unmet need for spacing (%)             0.07138377 -0.25509609
## Health worker t.a family planning w.n.u.(%)  0.04964010 -0.32798254
## Children <5 who are stunted (%)        0.34507233  0.24447087
## Children <5 who are underweight  (%)   0.41159375  0.07584513
```

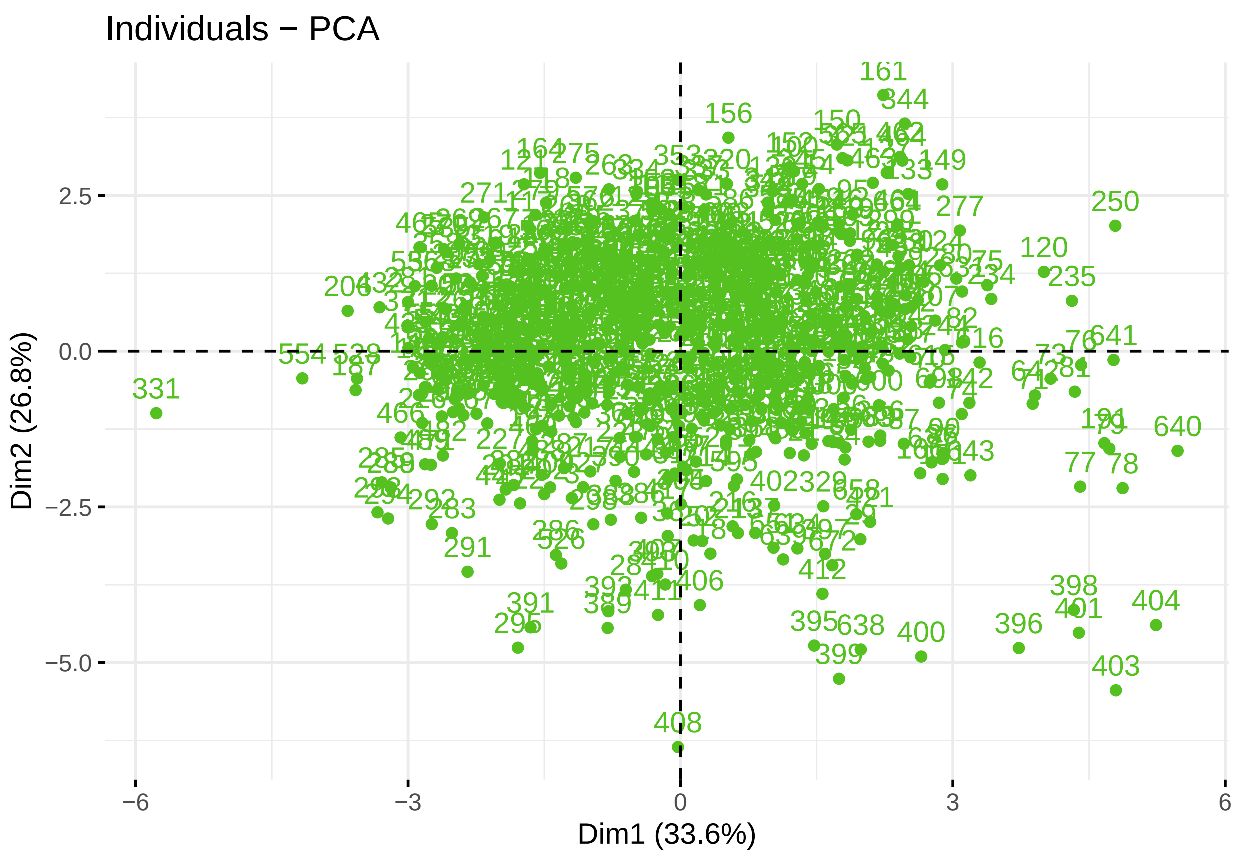
```
## User mention s.e of family planning (%)     -0.25065136  0.35730951
```

```r
#by the table it can be said that Component 1 is strongly negatively related with Women schooling +10 a
# Component 2 is strongly negatively correlated with Total unmet need for family planning and unmet nee
#Component 3 is strongly negatively correlated with Health worker t.a family planning w.n.u and User me
```

```r
#Plots of PC's
fviz_pca_var(pcares,axes = c(1, 2), repel = TRUE)
```




```r
fviz_pca_ind(pcares, col.ind = "#54C121")
```




```
### Principal Component Regression ###

ols.data <- data.frame(scaled_data_numeric[, 5], pcares2)

xlmodel <- lm(scaled_data_numeric ~ ., data = ols.data)
summary(xlmodel)
```

```
## Response Women schooling +10(%) :
##
## Call:
## lm(formula = `Women schooling +10(%)` ~ scaled_data_numeric...5. +
##     PC1 + PC2 + PC3 + PC4 + PC5, data = ols.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.01824 -0.21753  0.03157  0.23222  0.73590
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)              4.643e-17  1.190e-02   0.000    1.000
## scaled_data_numeric...5. -1.494e-01  3.807e-02  -3.925 9.53e-05 ***
## PC1                      -4.071e-01  1.242e-02 -32.778  < 2e-16 ***
## PC2                      -2.583e-01  1.900e-02 -13.593  < 2e-16 ***
## PC3                      -1.347e-01  1.423e-02  -9.467  < 2e-16 ***
## PC4                       5.332e-01  1.460e-02  36.507  < 2e-16 ***
## PC5                       2.819e-02  2.123e-02   1.328    0.185
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3158 on 697 degrees of freedom
## Multiple R-squared:  0.9011, Adjusted R-squared:  0.9003
## F-statistic:  1059 on 6 and 697 DF,  p-value: < 2.2e-16
##
##
## Response LiterateWomen (%) :
##
## Call:
## lm(formula = `LiterateWomen (%)` ~ scaled_data_numeric...5. +
##     PC1 + PC2 + PC3 + PC4 + PC5, data = ols.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.77190 -0.24581 -0.03068  0.23831  1.15011
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -1.281e-16  1.261e-02   0.000    1.000
## scaled_data_numeric...5.  3.301e-02  4.035e-02   0.818    0.414
## PC1                      -4.388e-01  1.316e-02 -33.340   <2e-16 ***
## PC2                      -2.324e-01  2.014e-02 -11.542   <2e-16 ***
## PC3                      -1.646e-01  1.508e-02 -10.918   <2e-16 ***
## PC4                       4.132e-01  1.548e-02  26.697   <2e-16 ***
## PC5                       1.965e-01  2.250e-02   8.734   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3347 on 697 degrees of freedom
## Multiple R-squared:  0.8889, Adjusted R-squared:  0.888
## F-statistic: 929.8 on 6 and 697 DF,  p-value: < 2.2e-16
##
##
## Response Current Use of Family Planning Methods (%) :
##
## Call:
## lm(formula = `Current Use of Family Planning Methods (%)` ~ scaled_data_numeric...5. +
##     PC1 + PC2 + PC3 + PC4 + PC5, data = ols.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.04880 -0.20039  0.01415  0.19793  0.82783
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -2.839e-16  1.064e-02   0.000        1
## scaled_data_numeric...5.  6.554e-01  3.404e-02  19.253   <2e-16 ***
## PC1                      -4.647e-01  1.110e-02 -41.852   <2e-16 ***
## PC2                       7.459e-01  1.699e-02  43.907   <2e-16 ***
## PC3                       2.255e-01  1.272e-02  17.728   <2e-16 ***
## PC4                       1.278e-01  1.306e-02   9.788   <2e-16 ***
## PC5                      -2.523e-01  1.898e-02 -13.288   <2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2824 on 697 degrees of freedom
## Multiple R-squared:  0.921,  Adjusted R-squared:  0.9203
## F-statistic:  1353 on 6 and 697 DF,  p-value: < 2.2e-16
##
##
## Response Total Unmet need for Family Planning (%) :
##
## Call:
## lm(formula = `Total Unmet need for Family Planning (%)` ~ scaled_data_numeric...5. +
##     PC1 + PC2 + PC3 + PC4 + PC5, data = ols.data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8598 -0.1373  0.0116  0.1535  0.6256
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -3.014e-16  8.337e-03    0.00        1
## scaled_data_numeric...5. -5.389e-01  2.667e-02  -20.21   <2e-16 ***
## PC1                       4.500e-01  8.700e-03   51.72   <2e-16 ***
## PC2                      -7.261e-01  1.331e-02  -54.56   <2e-16 ***
## PC3                      -3.607e-01  9.964e-03  -36.20   <2e-16 ***
## PC4                       1.438e-01  1.023e-02   14.06   <2e-16 ***
## PC5                      -3.558e-01  1.487e-02  -23.92   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2212 on 697 degrees of freedom
## Multiple R-squared:  0.9515, Adjusted R-squared:  0.9511
## F-statistic:  2278 on 6 and 697 DF,  p-value: < 2.2e-16
##
##
## Response Unmet need for spacing (%) :
##
## Call:
## lm(formula = `Unmet need for spacing (%)` ~ scaled_data_numeric...5. +
##     PC1 + PC2 + PC3 + PC4 + PC5, data = ols.data)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -8.865e-16 -8.550e-17 -4.330e-17  1.100e-18  2.752e-14
##
## Coefficients:
##                           Estimate Std. Error    t value Pr(>|t|)
## (Intercept)              -1.041e-31  3.947e-17  0.000e+00    1.000
## scaled_data_numeric...5.  1.000e+00  1.262e-16  7.922e+15   <2e-16 ***
## PC1                      -2.544e-17  4.118e-17 -6.180e-01    0.537
## PC2                       4.395e-17  6.300e-17  6.980e-01    0.486
## PC3                       3.688e-18  4.717e-17  7.800e-02    0.938
## PC4                       2.155e-17  4.842e-17  4.450e-01    0.656
## PC5                       5.515e-17  7.040e-17  7.830e-01    0.434
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.047e-15 on 697 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:      1
## F-statistic: 1.069e+32 on 6 and 697 DF,  p-value: < 2.2e-16
##
##
## Response Health worker t.a family planning w.n.u.(%) :
##
## Call:
## lm(formula = `Health worker t.a family planning w.n.u.(%)` ~
##     scaled_data_numeric...5. + PC1 + PC2 + PC3 + PC4 + PC5, data = ols.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.63896 -0.15145 -0.00936  0.13162  0.89700
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -4.709e-16  8.366e-03   0.000    1.000
## scaled_data_numeric...5. -6.476e-01  2.676e-02 -24.201   <2e-16 ***
## PC1                       2.128e-01  8.730e-03  24.377   <2e-16 ***
## PC2                       2.178e-02  1.336e-02   1.631    0.103
## PC3                      -7.908e-01  9.999e-03 -79.093   <2e-16 ***
## PC4                       1.146e-01  1.027e-02  11.164   <2e-16 ***
## PC5                      -6.814e-01  1.492e-02 -45.655   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.222 on 697 degrees of freedom
## Multiple R-squared:  0.9511, Adjusted R-squared:  0.9507
## F-statistic:  2262 on 6 and 697 DF,  p-value: < 2.2e-16
##
##
## Response Children <5 who are stunted (%) :
##
## Call:
## lm(formula = `Children <5 who are stunted (%)` ~ scaled_data_numeric...5. +
##     PC1 + PC2 + PC3 + PC4 + PC5, data = ols.data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.4430 -0.2462  0.0028  0.2274  0.9706
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)              2.104e-16  1.265e-02   0.000  1.00000
## scaled_data_numeric...5. 3.134e-01  4.047e-02   7.742 3.45e-14 ***
## PC1                      3.608e-01  1.320e-02  27.326  < 2e-16 ***
## PC2                      3.425e-01  2.020e-02  16.956  < 2e-16 ***
## PC3                      4.183e-02  1.512e-02   2.766  0.00583 **
## PC4                      3.858e-01  1.553e-02  24.846  < 2e-16 ***
## PC5                      4.482e-01  2.257e-02  19.855  < 2e-16 ***
```

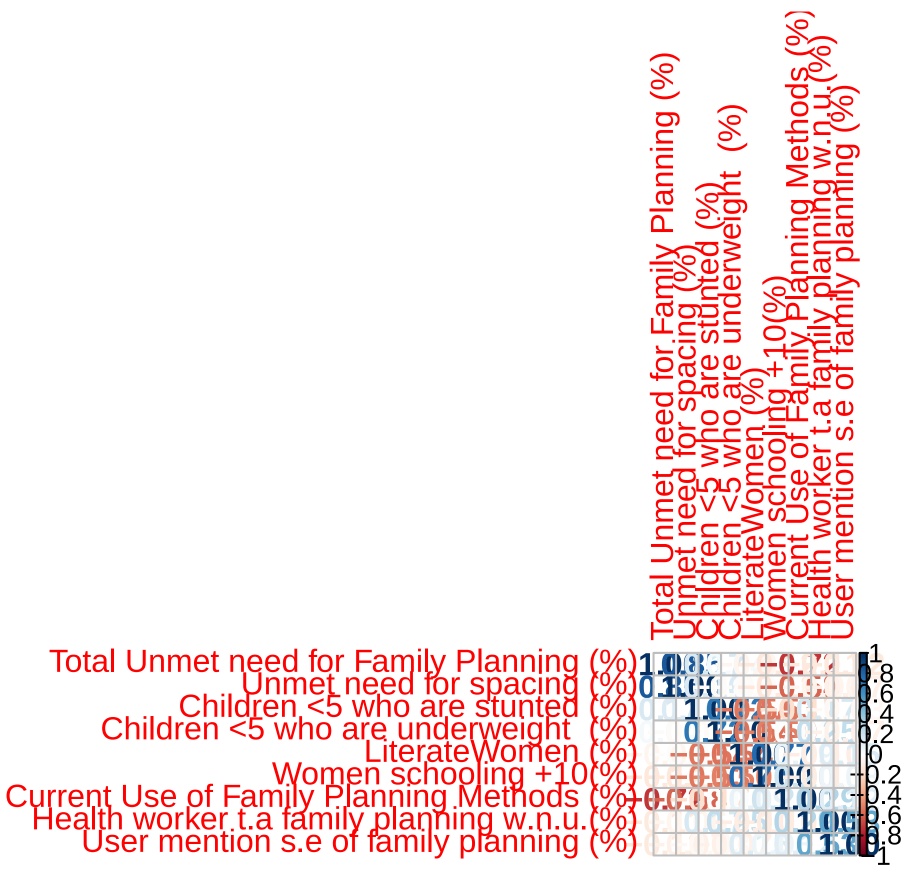
```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3357 on 697 degrees of freedom
## Multiple R-squared:  0.8882, Adjusted R-squared:  0.8873
## F-statistic: 923.3 on 6 and 697 DF,  p-value: < 2.2e-16
##
##
## Response Children <5 who are underweight  (%) :
##
## Call:
## lm(formula = `Children <5 who are underweight  (%)` ~ scaled_data_numeric...5. +
##     PC1 + PC2 + PC3 + PC4 + PC5, data = ols.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.97260 -0.22973  0.00004  0.24407  1.53509
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -2.178e-17  1.307e-02   0.000  1.00000
## scaled_data_numeric...5. -1.200e-01  4.182e-02  -2.869  0.00424 **
## PC1                       4.268e-01  1.364e-02  31.286  < 2e-16 ***
## PC2                       2.569e-01  2.087e-02  12.308  < 2e-16 ***
## PC3                      -1.636e-02  1.562e-02  -1.047  0.29530
## PC4                       5.022e-01  1.604e-02  31.312  < 2e-16 ***
## PC5                       6.250e-02  2.332e-02   2.680  0.00754 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3469 on 697 degrees of freedom
## Multiple R-squared:  0.8807, Adjusted R-squared:  0.8797
## F-statistic: 857.7 on 6 and 697 DF,  p-value: < 2.2e-16
##
##
## Response User mention s.e of family planning (%) :
##
## Call:
## lm(formula = `User mention s.e of family planning (%)` ~ scaled_data_numeric...5. +
##     PC1 + PC2 + PC3 + PC4 + PC5, data = ols.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.83330 -0.10527  0.00774  0.11579  0.54462
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)               1.744e-17  6.785e-03    0.00        1
## scaled_data_numeric...5.  4.031e-01  2.170e-02   18.57   <2e-16 ***
## PC1                      -2.148e-01  7.079e-03  -30.34   <2e-16 ***
## PC2                       3.467e-01  1.083e-02   32.01   <2e-16 ***
## PC3                      -5.678e-01  8.109e-03  -70.03   <2e-16 ***
## PC4                      -3.340e-01  8.325e-03  -40.12   <2e-16 ***
## PC5                       6.357e-01  1.210e-02   52.52   <2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.18 on 697 degrees of freedom
## Multiple R-squared:  0.9679, Adjusted R-squared:  0.9676
## F-statistic:  3499 on 6 and 697 DF,  p-value: < 2.2e-16
```

```r
#you can see the MSE value is 1.67

mean((ols.data$scaled_data_numeric - predict(xlmodel))^2)
```

```
## [1] 1.674971
```

```r
### Factor Analysis and Factor Rotation ###

# we will consider numeric variables of our data set

myfactordata <- data_all[,2:10]
cm <- cor(myfactordata, method = "pearson")
corrplot::corrplot(cm, method= "number", order = "hclust")
```




```r
# we can observe that there are some correlated variables

KMO(r=cm)
```

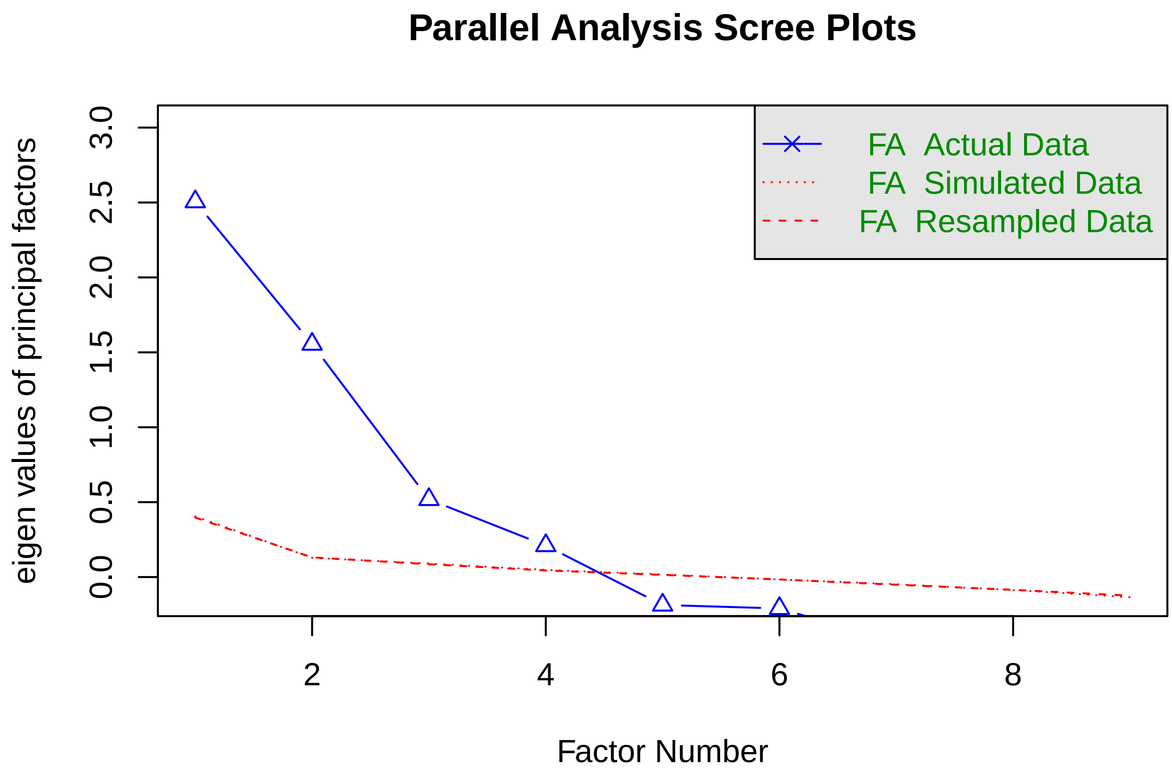
```
## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = cm)
## Overall MSA =  0.64
## MSA for each item =
##                          LiterateWomen (%)
##                                       0.70
##                     Women schooling +10(%)
##                                       0.69
##  Current Use of Family Planning Methods (%)
##                                       0.67
##    Total Unmet need for Family Planning (%)
##                                       0.58
##                 Unmet need for spacing (%)
##                                       0.61
## Health worker t.a family planning w.n.u.(%)
##                                       0.51
##             Children <5 who are stunted (%)
##                                       0.72
##          Children <5 who are underweight  (%)
##                                       0.67
##     User mention s.e of family planning (%)
##                                       0.47
```

```
#Since MSA > 0.5, we can run Factor Analysis on this data.
#Besides, Bartletts test of sphericity should be significant.

cortest.bartlett(cm,nrow(myfactordata))
```

```
## $chisq
## [1] 3443.101
##
## $p.value
## [1] 0
##
## $df
## [1] 36
```

```
#The approximate of Chi-square is 3443.101 with 36 degrees of freedom,
#which is significant at 0.05 Level of significance.
#The KMO statistic of 0.64 is also large (greater than 0.50).
#Hence Factor Analysis is considered as an appropriate technique for further analysis of the data.
```

```
parallel <- fa.parallel(myfactordata, fm = "minres", fa = "fa")
```




```
## Parallel analysis suggests that the number of factors =  4  and the number of components =  NA
```

```
#We can see from the graph that after
#factor 4 there is a sharp change in the curvature
#of the scree plot. This shows that after factor 7
#the total variance accounts for smaller amounts

# let see whether 4 factor is enough to group variables

factanal(myfactordata, factors = 4)$PVAL
```

```
##    objective
## 1.168781e-16
```

```
# 4 factor solution is not  adequate
# lets check whether 5 factor is is enough to group variables

factanal(myfactordata, factors = 5)$PVAL
```

```
##    objective
## 1.890441e-06
```

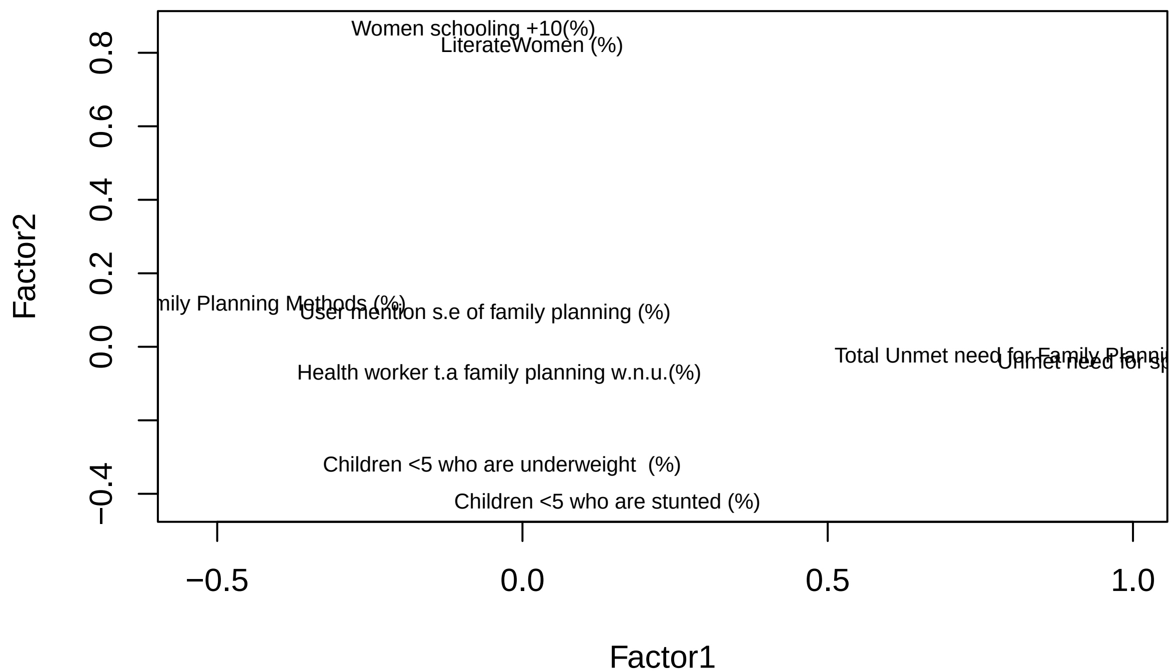
```r
#It is not enough when the number of factors is 5,
#but it has the largest p value among all possibilities,
#so we choose the number of factors as 5.

f <- factanal(myfactordata, factors = 5)

f
```

```
##
## Call:
## factanal(x = myfactordata, factors = 5)
##
## Uniquenesses:
##                           LiterateWomen (%)
##                                       0.241
##                      Women schooling +10(%)
##                                       0.185
##  Current Use of Family Planning Methods (%)
##                                       0.005
##    Total Unmet need for Family Planning (%)
##                                       0.196
##                    Unmet need for spacing (%)
##                                       0.005
## Health worker t.a family planning w.n.u.(%)
##                                       0.565
##               Children <5 who are stunted (%)
##                                       0.400
##          Children <5 who are underweight  (%)
##                                       0.005
##     User mention s.e of family planning (%)
##                                       0.005
##
## Loadings:
##                                             Factor1 Factor2 Factor3 Factor4
## LiterateWomen (%)                                    0.818  -0.295
## Women schooling +10(%)                               0.862  -0.243
## Current Use of Family Planning Methods (%)  -0.536   0.114           0.149
## Total Unmet need for Family Planning (%)     0.833
## Unmet need for spacing (%)                   0.995
## Health worker t.a family planning w.n.u.(%)                  0.194   0.578
## Children <5 who are stunted (%)              0.139  -0.425   0.626
## Children <5 who are underweight  (%)                -0.325   0.940
## User mention s.e of family planning (%)                              0.983
##                                             Factor5
## LiterateWomen (%)
## Women schooling +10(%)
## Current Use of Family Planning Methods (%)   0.820
## Total Unmet need for Family Planning (%)    -0.309
## Unmet need for spacing (%)
## Health worker t.a family planning w.n.u.(%)  0.238
## Children <5 who are stunted (%)
## Children <5 who are underweight  (%)
## User mention s.e of family planning (%)
##
```

```
##                  Factor1 Factor2 Factor3 Factor4 Factor5
## SS loadings        2.003   1.727   1.476   1.341   0.846
## Proportion Var     0.223   0.192   0.164   0.149   0.094
## Cumulative Var     0.223   0.414   0.578   0.727   0.821
##
## Test of the hypothesis that 5 factors are sufficient.
## The chi square statistic is 22.7 on 1 degree of freedom.
## The p-value is 1.89e-06
```

```
# you can see that factor1 is dominated by the
#Total Unmet need for Family Planning (%) and
#Unmet need for spacing (%) also factor2 is dominated by
# the Women schooling +10(%) and LiterateWomen (%)
# Moreover factors explain 82% of the variance
```

```
load <- f$loadings[,1:2]
plot(load,type="n")
text(load,labels=names(myfactordata),cex=.7)
```



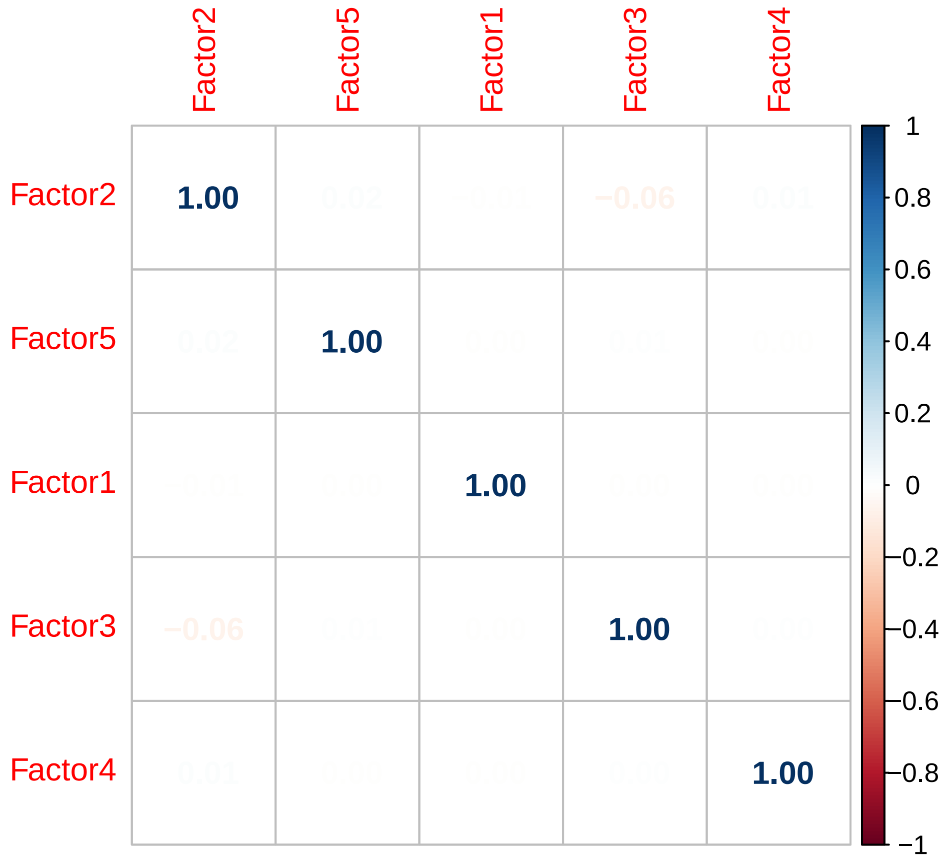

```
# according to the plot Total Unmet need for Family Planning (%) and
# Unmet need for spacing (%) dominate factor 1,while
# Women schooling +10(%) and LiterateWomen (%) dominate factor2

#Let's check the consistency of the first factor.
names(f$loadings[,1])[abs(f$loadings[,1])>0.4]
```

```
## [1] "Current Use of Family Planning Methods (%)"
## [2] "Total Unmet need for Family Planning (%)"
## [3] "Unmet need for spacing (%)"
```

```
f1 <- myfactordata[,names(f$loadings[,1])[abs(f$loadings[,1])>0.4]]

#Hocam aşağıda öncesinde # bulunan 3 satır knitlerken problem yaratıyor ama outputları raporda var umar

#install.packages("psych")
#library(psych)
#summary(alpha(f1, check.keys = TRUE))

# the alpha is 0.64 so we can repeat this process for the
# rest of the factors.

scores <- factanal(myfactordata, factors = 5, scores = "regression")$scores
head(scores)
```

```
##         Factor1    Factor2    Factor3   Factor4    Factor5
## [1,] -0.3189249  0.9909258 -0.1358053 -1.110191 0.06901788
## [2,] -1.0169715  0.7712556  1.5313542  1.231457 0.66000926
## [3,]  1.7019653  1.0696794 -0.9450179  1.426938 0.28400801
## [4,] -0.2091638 -0.4765037 -1.0580617 -1.161133 1.68017045
## [5,]  0.2062547 -0.7439483  0.0253829 -1.579691 1.99066195
## [6,] -0.6624517  0.1303761  0.4867826 -1.823779 1.04271411
```

```
# it can be see that factors are uncorrelated

cm1 <- cor(scores, method="pearson")
corrplot::corrplot(cm1, method= "number", order = "hclust")
```




```
# As you see, they are almost uncorrelated which guarantees
#that no multicollinearity problem in linear regression
```

```
### Discrimination and Classification ###

library(MASS)
```

```
##
## Attaching package: 'MASS'

## The following object is masked from 'package:rstatix':
##
##     select

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

```
library(klaR)
library(ggplot2)
library(GGally)
```

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

```r
library(mlbench)


# Enable the r-universe repo
options(repos = c(
  fawda123 = 'https://fawda123.r-universe.dev',
  CRAN = 'https://cloud.r-project.org'))

#Install ggord
#install.packages('ggord')
library(ggord)

#It is seen that we do not have any class problem,
#so we can proceed
summary(data_all)
```

```
##            State/UT   LiterateWomen (%) Women schooling +10(%)
##  Uttar Pradesh : 75   Min.   :38.60     Min.   :13.60
##  Madhya Pradesh: 51   1st Qu.:66.80     1st Qu.:29.23
##  Bihar         : 38   Median :75.10     Median :39.15
##  Maharastra    : 36   Mean   :74.28     Mean   :40.28
##  Assam         : 33   3rd Qu.:83.70     3rd Qu.:49.83
##  Gujarat       : 33   Max.   :99.70     Max.   :88.20
##  (Other)       :438
##  Current Use of Family Planning Methods (%)
##  Min.   :12.30
##  1st Qu.:46.67
##  Median :55.60
##  Mean   :54.90
##  3rd Qu.:65.62
##  Max.   :81.20
##
##  Total Unmet need for Family Planning (%) Unmet need for spacing (%)
##  Min.   : 1.200                           Min.   : 0.300
##  1st Qu.: 5.800                           1st Qu.: 2.600
##  Median : 8.450                           Median : 3.700
##  Mean   : 9.526                           Mean   : 4.261
##  3rd Qu.:12.300                           3rd Qu.: 5.125
##  Max.   :30.400                           Max.   :21.700
##
##  Health worker t.a family planning w.n.u.(%) Children <5 who are stunted (%)
##  Min.   : 2.00                               Min.   :-18.00
##  1st Qu.:16.90                               1st Qu.: 27.30
##  Median :23.15                               Median : 32.85
##  Mean   :24.23                               Mean   : 33.46
##  3rd Qu.:30.23                               3rd Qu.: 39.20
##  Max.   :64.20                               Max.   : 60.60
##
##  Children <5 who are underweight  (%) User mention s.e of family planning (%)
##  Min.   :-31.30                       Min.   :14.60
##  1st Qu.: 22.07                       1st Qu.:54.38
##  Median : 29.35                       Median :66.05
##  Mean   : 29.47                       Mean   :65.04
```

```
##  3rd Qu.: 36.30                    3rd Qu.:76.92
##  Max.   : 62.40                    Max.   :98.90
##
```

```
# 75 people live in uttar pradesh, 51 people live in madhya pradesh
# 38 people live in bihar, 36 people live maharastra, 33 people live in
# assam, 33 people live in gujarat and 438 people live in other state
```

```
library(MASS)
set.seed(467)

#Use 80% of dataset as training set and remaining 20% as testing set
sample <- sample(c(TRUE, FALSE), nrow(data_all), replace=TRUE, prob=c(0.8,0.2))
train <- data_all[sample, ]
test <- data_all[!sample, ]

mymodel <- lda(`State/UT`~.,data = train)
```

```
## Warning in lda.default(x, grouping, ...): groups Chandigarh Lakshadweep are
## empty
```

```
mymodel
```

```
## Call:
## lda(‘State/UT‘ ~ ., data = train)
##
## Prior probabilities of groups:
##              Andaman & Nicobar Islands                         Andhra Pradesh
##                            0.005272408                            0.017574692
##                      Arunachal Pradesh                                  Assam
##                            0.024604569                            0.040421793
##                                  Bihar                           Chhattisgarh
##                            0.052724077                            0.036906854
## Dadra and Nagar Haveli & Daman and Diu                                    Goa
##                            0.005272408                            0.003514938
##                                Gujarat                                Haryana
##                            0.040421793                            0.031634446
##                       Himachal Pradesh                        Jammu & Kashmir
##                            0.021089631                            0.029876977
##                              Jharkhand                              Karnataka
##                            0.035149385                            0.036906854
##                                 Kerala                                 Ladakh
##                            0.019332162                            0.001757469
##                         Madhya Pradesh                            Maharastra
##                            0.066783831                            0.056239016
##                                Manipur                              Meghalaya
##                            0.015817223                            0.014059754
##                                Mizoram                               Nagaland
##                            0.010544815                            0.019332162
##                           NCT of Delhi                                 Odisha
##                            0.014059754                            0.047451670
##                             Puducherry                                 Punjab
```

```
##                              0.007029877                             0.033391916
##                                Rajasthan                                  Sikkim
##                              0.047451670                             0.005272408
##                               Tamil Nadu                               Telangana
##                              0.045694200                             0.043936731
##                                  Tripura                           Uttar Pradesh
##                              0.010544815                             0.112478032
##                              Uttarakhand                             West Bengal
##                              0.017574692                             0.029876977
##
## Group means:
##                                         `LiterateWomen (%)`
## Andaman & Nicobar Islands                          86.06667
## Andhra Pradesh                                     66.84000
## Arunachal Pradesh                                  71.95714
## Assam                                              77.32174
## Bihar                                              56.74667
## Chhattisgarh                                       68.87619
## Dadra and Nagar Haveli & Daman and Diu             83.16667
## Goa                                                93.10000
## Gujarat                                            75.52174
## Haryana                                            80.86111
## Himachal Pradesh                                   90.25000
## Jammu & Kashmir                                    75.88235
## Jharkhand                                          62.83500
## Karnataka                                          77.04286
## Kerala                                             97.52727
## Ladakh                                             77.20000
## Madhya Pradesh                                     67.48947
## Maharastra                                         82.45625
## Manipur                                            85.14444
## Meghalaya                                          85.40000
## Mizoram                                            91.91667
## Nagaland                                           84.50909
## NCT of Delhi                                       85.01250
## Odisha                                             69.58519
## Puducherry                                         91.07500
## Punjab                                             80.43684
## Rajasthan                                          64.39630
## Sikkim                                             87.33333
## Tamil Nadu                                         85.82308
## Telangana                                          62.98400
## Tripura                                            79.16667
## Uttar Pradesh                                      68.14219
## Uttarakhand                                        83.78000
## West Bengal                                        74.05294
##                                         `Women schooling +10(%)`
## Andaman & Nicobar Islands                               50.66667
## Andhra Pradesh                                          39.23000
## Arunachal Pradesh                                       38.55000
## Assam                                                   29.21739
## Bihar                                                   27.62000
## Chhattisgarh                                            33.91429
## Dadra and Nagar Haveli & Daman and Diu                  42.36667
```

```
## Goa                                                    71.45000
## Gujarat                                                32.24783
## Haryana                                                48.78889
## Himachal Pradesh                                       62.80833
## Jammu & Kashmir                                        49.31176
## Jharkhand                                              30.50500
## Karnataka                                              48.44286
## Kerala                                                 74.88182
## Ladakh                                                 50.30000
## Madhya Pradesh                                         27.96579
## Maharastra                                             46.49688
## Manipur                                                44.83333
## Meghalaya                                              30.36250
## Mizoram                                                41.85000
## Nagaland                                               39.70000
## NCT of Delhi                                           60.07500
## Odisha                                                 31.55926
## Puducherry                                             66.50000
## Punjab                                                 54.55263
## Rajasthan                                              30.76667
## Sikkim                                                 45.53333
## Tamil Nadu                                             56.09615
## Telangana                                              42.11600
## Tripura                                                21.48333
## Uttar Pradesh                                          37.97344
## Uttarakhand                                            52.51000
## West Bengal                                            33.26471
##                                         `Current Use of Family Planning Methods (%)`
## Andaman & Nicobar Islands                                                   60.26667
## Andhra Pradesh                                                              71.96000
## Arunachal Pradesh                                                           47.35000
## Assam                                                                       45.86522
## Bihar                                                                       43.96000
## Chhattisgarh                                                                57.85238
## Dadra and Nagar Haveli & Daman and Diu                                      55.13333
## Goa                                                                         61.30000
## Gujarat                                                                     54.12609
## Haryana                                                                     60.31111
## Himachal Pradesh                                                            65.21667
## Jammu & Kashmir                                                             52.32941
## Jharkhand                                                                   47.49500
## Karnataka                                                                   68.30000
## Kerala                                                                      56.66364
## Ladakh                                                                      41.60000
## Madhya Pradesh                                                              65.15263
## Maharastra                                                                  65.28750
## Manipur                                                                     18.98889
## Meghalaya                                                                   25.05000
## Mizoram                                                                     38.38333
## Nagaland                                                                    46.84545
## NCT of Delhi                                                                58.73750
## Odisha                                                                      49.34074
## Puducherry                                                                  66.10000
## Punjab                                                                      51.11579
```

```
## Rajasthan                                                               60.69630
## Sikkim                                                                  57.13333
## Tamil Nadu                                                              65.61538
## Telangana                                                               65.35600
## Tripura                                                                 50.06667
## Uttar Pradesh                                                           44.49688
## Uttarakhand                                                             60.21000
## West Bengal                                                             59.59412
##                                        `Total Unmet need for Family Planning (%)`
## Andaman & Nicobar Islands                                             10.966667
## Andhra Pradesh                                                         4.260000
## Arunachal Pradesh                                                     12.471429
## Assam                                                                 10.578261
## Bihar                                                                 13.026667
## Chhattisgarh                                                           9.109524
## Dadra and Nagar Haveli & Daman and Diu                                12.600000
## Goa                                                                    8.200000
## Gujarat                                                               10.630435
## Haryana                                                                7.822222
## Himachal Pradesh                                                       7.108333
## Jammu & Kashmir                                                        7.747059
## Jharkhand                                                             12.095000
## Karnataka                                                              6.404762
## Kerala                                                                11.072727
## Ladakh                                                                 8.800000
## Madhya Pradesh                                                         7.713158
## Maharastra                                                             9.043750
## Manipur                                                               11.611111
## Meghalaya                                                             25.162500
## Mizoram                                                               15.400000
## Nagaland                                                               7.881818
## NCT of Delhi                                                           5.687500
## Odisha                                                                 6.348148
## Puducherry                                                             7.400000
## Punjab                                                                10.036842
## Rajasthan                                                              8.137037
## Sikkim                                                                11.900000
## Tamil Nadu                                                             7.607692
## Telangana                                                              6.652000
## Tripura                                                                6.416667
## Uttar Pradesh                                                         12.660937
## Uttarakhand                                                            9.490000
## West Bengal                                                            7.735294
##                                        `Unmet need for spacing (%)`
## Andaman & Nicobar Islands                                 4.400000
## Andhra Pradesh                                            2.500000
## Arunachal Pradesh                                         7.164286
## Assam                                                     4.047826
## Bihar                                                     5.826667
## Chhattisgarh                                              3.552381
## Dadra and Nagar Haveli & Daman and Diu                    5.866667
## Goa                                                       3.900000
## Gujarat                                                   4.660870
## Haryana                                                   3.616667
```

```
## Himachal Pradesh                                       2.541667
## Jammu & Kashmir                                        3.911765
## Jharkhand                                              4.925000
## Karnataka                                              3.690476
## Kerala                                                 6.272727
## Ladakh                                                 4.900000
## Madhya Pradesh                                         3.873684
## Maharastra                                             3.753125
## Manipur                                                4.588889
## Meghalaya                                             16.362500
## Mizoram                                               11.883333
## Nagaland                                               4.190909
## NCT of Delhi                                           1.850000
## Odisha                                                 2.288889
## Puducherry                                             3.000000
## Punjab                                                 3.610526
## Rajasthan                                              3.951852
## Sikkim                                                 5.633333
## Tamil Nadu                                             2.984615
## Telangana                                              2.860000
## Tripura                                                2.466667
## Uttar Pradesh                                          4.737500
## Uttarakhand                                            3.500000
## West Bengal                                            3.194118
##                                        `Health worker t.a family planning w.n.u.(%)`
## Andaman & Nicobar Islands                                                   31.60000
## Andhra Pradesh                                                              19.89000
## Arunachal Pradesh                                                           18.97857
## Assam                                                                       22.35217
## Bihar                                                                       22.46667
## Chhattisgarh                                                                32.99048
## Dadra and Nagar Haveli & Daman and Diu                                      27.50000
## Goa                                                                         26.45000
## Gujarat                                                                     35.69565
## Haryana                                                                     25.07222
## Himachal Pradesh                                                            20.60000
## Jammu & Kashmir                                                             10.31765
## Jharkhand                                                                   29.43500
## Karnataka                                                                   37.91429
## Kerala                                                                      15.19091
## Ladakh                                                                       9.10000
## Madhya Pradesh                                                              29.05000
## Maharastra                                                                  22.54062
## Manipur                                                                      6.30000
## Meghalaya                                                                   31.35000
## Mizoram                                                                     16.83333
## Nagaland                                                                    10.32727
## NCT of Delhi                                                                17.83750
## Odisha                                                                      25.87037
## Puducherry                                                                  25.17500
## Punjab                                                                      20.92105
## Rajasthan                                                                   25.08148
## Sikkim                                                                      23.30000
## Tamil Nadu                                                                  28.51154
```

```
## Telangana                                                                    17.28800
## Tripura                                                                       9.00000
## Uttar Pradesh                                                                26.37344
## Uttarakhand                                                                  19.70000
## West Bengal                                                                  16.82353
##                                         `Children <5 who are stunted (%)`
## Andaman & Nicobar Islands                                        23.23333
## Andhra Pradesh                                                   30.49000
## Arunachal Pradesh                                                28.15000
## Assam                                                            36.08261
## Bihar                                                            42.55333
## Chhattisgarh                                                     35.89048
## Dadra and Nagar Haveli & Daman and Diu                           35.43333
## Goa                                                              26.25000
## Gujarat                                                          37.76522
## Haryana                                                          26.73889
## Himachal Pradesh                                                 31.55000
## Jammu & Kashmir                                                  26.59412
## Jharkhand                                                        40.23500
## Karnataka                                                        34.03810
## Kerala                                                           23.72727
## Ladakh                                                           36.50000
## Madhya Pradesh                                                   34.97368
## Maharastra                                                       35.04375
## Manipur                                                          24.62222
## Meghalaya                                                        43.71250
## Mizoram                                                          29.16667
## Nagaland                                                         33.16364
## NCT of Delhi                                                     30.57500
## Odisha                                                           31.70000
## Puducherry                                                       28.92500
## Punjab                                                           24.95789
## Rajasthan                                                        31.98519
## Sikkim                                                           24.80000
## Tamil Nadu                                                       25.50000
## Telangana                                                        32.57600
## Tripura                                                          33.95000
## Uttar Pradesh                                                    39.65781
## Uttarakhand                                                      27.22000
## West Bengal                                                      33.26471
##                                         `Children <5 who are underweight  (%)`
## Andaman & Nicobar Islands                                             28.26667
## Andhra Pradesh                                                        27.75000
## Arunachal Pradesh                                                     14.51429
## Assam                                                                 33.03913
## Bihar                                                                 42.84667
## Chhattisgarh                                                          34.34286
## Dadra and Nagar Haveli & Daman and Diu                                33.30000
## Goa                                                                   23.40000
## Gujarat                                                               39.58261
## Haryana                                                               21.42222
## Himachal Pradesh                                                      24.68333
## Jammu & Kashmir                                                       20.54706
## Jharkhand                                                             40.13000
```

```
## Karnataka                                                         31.30476
## Kerala                                                            19.93636
## Ladakh                                                            21.80000
## Madhya Pradesh                                                    33.28684
## Maharastra                                                        36.65313
## Manipur                                                           13.76667
## Meghalaya                                                         25.86250
## Mizoram                                                           13.66667
## Nagaland                                                          26.52727
## NCT of Delhi                                                      20.61250
## Odisha                                                            30.27037
## Puducherry                                                        21.02500
## Punjab                                                            17.22105
## Rajasthan                                                         28.62222
## Sikkim                                                            11.83333
## Tamil Nadu                                                        22.85769
## Telangana                                                         33.13600
## Tripura                                                           27.78333
## Uttar Pradesh                                                     32.71563
## Uttarakhand                                                       19.84000
## West Bengal                                                       33.39412
##                                        `User mention s.e of family planning (%)`
## Andaman & Nicobar Islands                                           73.60000
## Andhra Pradesh                                                      27.34000
## Arunachal Pradesh                                                   72.35000
## Assam                                                               70.28696
## Bihar                                                               51.07000
## Chhattisgarh                                                        84.33333
## Dadra and Nagar Haveli & Daman and Diu                              71.73333
## Goa                                                                 85.75000
## Gujarat                                                             73.96957
## Haryana                                                             69.04444
## Himachal Pradesh                                                    60.95000
## Jammu & Kashmir                                                     61.12941
## Jharkhand                                                           53.80000
## Karnataka                                                           72.33810
## Kerala                                                              63.09091
## Ladakh                                                              55.40000
## Madhya Pradesh                                                      68.81842
## Maharastra                                                          53.81562
## Manipur                                                             45.92222
## Meghalaya                                                           74.73750
## Mizoram                                                             64.70000
## Nagaland                                                            61.69091
## NCT of Delhi                                                        75.02500
## Odisha                                                              74.50370
## Puducherry                                                          66.90000
## Punjab                                                              76.83684
## Rajasthan                                                           60.87778
## Sikkim                                                              56.83333
## Tamil Nadu                                                          81.40769
## Telangana                                                           45.86400
## Tripura                                                             41.98333
## Uttar Pradesh                                                       71.40469
```

```
## Uttarakhand                                                          61.27000
## West Bengal                                                          52.08824
##
## Coefficients of linear discriminants:
##                                                      LD1         LD2
## `LiterateWomen (%)`                          0.158507270 -0.06530120
## `Women schooling +10(%)`                    -0.136482299  0.09831490
## `Current Use of Family Planning Methods (%)` -0.064417570 -0.03302861
## `Total Unmet need for Family Planning (%)`  -0.138080871 -0.08069874
## `Unmet need for spacing (%)`                 0.387474734  0.15342061
## `Health worker t.a family planning w.n.u.(%)` 0.009057177 -0.04182728
## `Children <5 who are stunted (%)`            0.031956453  0.02016710
## `Children <5 who are underweight  (%)`      -0.023265500 -0.10583735
## `User mention s.e of family planning (%)`    0.026547175  0.04689051
##                                                      LD3         LD4
## `LiterateWomen (%)`                          0.067280085  0.01895937
## `Women schooling +10(%)`                    -0.008348795  0.02164088
## `Current Use of Family Planning Methods (%)`  0.108868988  0.04273890
## `Total Unmet need for Family Planning (%)`  -0.063072784  0.25024341
## `Unmet need for spacing (%)`                 0.580302422 -0.27478400
## `Health worker t.a family planning w.n.u.(%)` -0.004249674  0.03371890
## `Children <5 who are stunted (%)`            0.009890523  0.02732148
## `Children <5 who are underweight  (%)`       0.010765855  0.04017371
## `User mention s.e of family planning (%)`   -0.029448603  0.05906472
##                                                      LD5          LD6
## `LiterateWomen (%)`                          0.115681701  0.052649283
## `Women schooling +10(%)`                    -0.084230252  0.060943995
## `Current Use of Family Planning Methods (%)`  0.021203943 -0.056422470
## `Total Unmet need for Family Planning (%)`   0.097868586  0.014835983
## `Unmet need for spacing (%)`                -0.476613447 -0.129686845
## `Health worker t.a family planning w.n.u.(%)` -0.039148458 -0.009626099
## `Children <5 who are stunted (%)`           -0.019405656  0.050259873
## `Children <5 who are underweight  (%)`      -0.001550101  0.088639156
## `User mention s.e of family planning (%)`    0.009659017 -0.026664356
##                                                      LD7          LD8
## `LiterateWomen (%)`                         -0.020137939  0.009920193
## `Women schooling +10(%)`                     0.006000685 -0.029498566
## `Current Use of Family Planning Methods (%)`  0.028360943  0.070863291
## `Total Unmet need for Family Planning (%)`  -0.049987928  0.372027158
## `Unmet need for spacing (%)`                 0.231951831 -0.412878564
## `Health worker t.a family planning w.n.u.(%)` -0.122118921 -0.039746514
## `Children <5 who are stunted (%)`           -0.032919928  0.111091596
## `Children <5 who are underweight  (%)`       0.089304085 -0.099042647
## `User mention s.e of family planning (%)`    0.056552012 -0.001398665
##                                                      LD9
## `LiterateWomen (%)`                         -0.002633246
## `Women schooling +10(%)`                    -0.012260914
## `Current Use of Family Planning Methods (%)`  0.008858125
## `Total Unmet need for Family Planning (%)`   0.209568968
## `Unmet need for spacing (%)`                -0.127790655
## `Health worker t.a family planning w.n.u.(%)`  0.001853185
## `Children <5 who are stunted (%)`           -0.135265196
## `Children <5 who are underweight  (%)`       0.051824658
## `User mention s.e of family planning (%)`   -0.004403395
```

```
##
## Proportion of trace:
##    LD1    LD2    LD3    LD4    LD5    LD6    LD7    LD8    LD9
## 0.3058 0.2105 0.1544 0.1174 0.0913 0.0746 0.0196 0.0183 0.0080
```

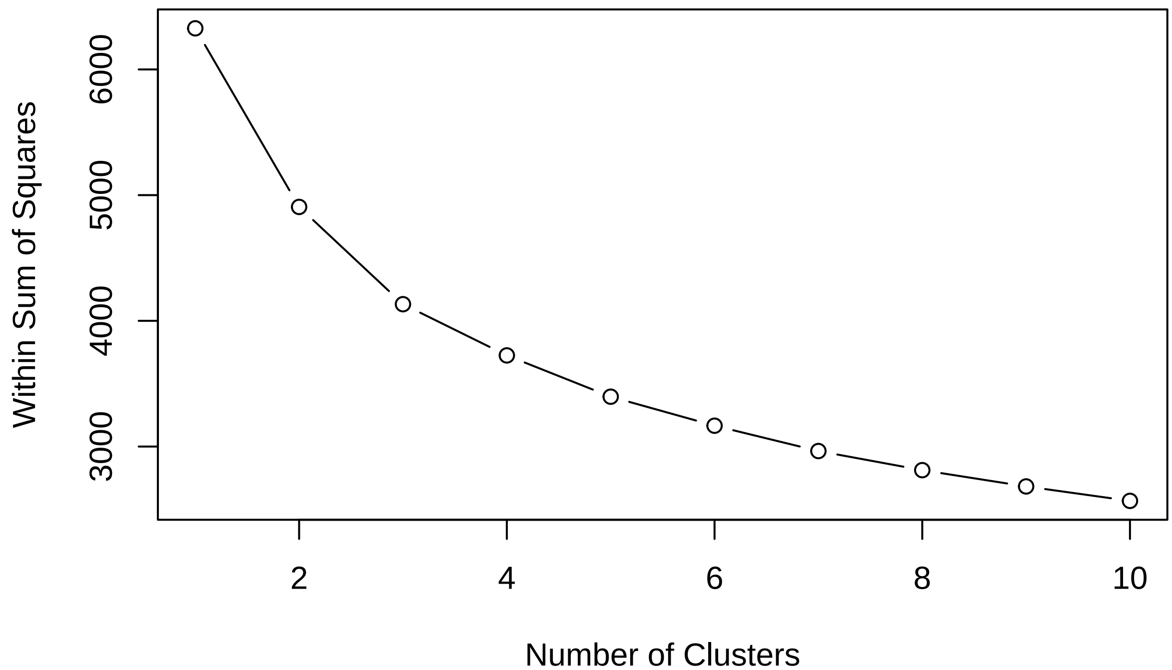
```r
# there are 36 levels in our data. The LDA output indicates that
#for example Uttar Prades = 0.10 that means 11% of the training observation
# corresponds to the patients that live in uttar Prades, 90% people live in other states
```

```r
model.values <- predict(mymodel)

train_predict <- predict(mymodel,train)$class
table_train <- table(Predicted = train_predict, Actual = train$`State/UT`)

sum(diag(table_train))/sum(table_train)
```

```
## [1] 0.6994728
```

```r
# the model correctly classifies the states where people live
#0.69 probability for the training data
# the classification error rate(misclassification) for training
# data is 1-0.69=0.31
```

```r
### Clustering ###

# Extracting the relevant numeric columns for clustering
data_for_clustering <- data_all[, c(2:10)]


standardized_data <- scale(data_for_clustering)

# Determine the optimal number of clusters using the elbow method
wss <- numeric(10)
for (i in 1:10) {
  kmeans_model <- kmeans(standardized_data, centers = i, nstart = 25)
  wss[i] <- sum(kmeans_model$withinss)
}

# Plot the elbow plot
plot(1:10, wss, type = "b", xlab = "Number of Clusters", ylab = "Within Sum of Squares")
```




```r
# Based on the plot, visually identify the "elbow" where the rate of decrease slows down
# The only "elbow" in the plot occurs for three groups, and so we will now look at the three-group solu

# Running k-means clustering with the chosen number of clusters
k_chosen <- 3  # Replace with the number identified from the elbow plot
kmeans_result <- kmeans(standardized_data, centers = k_chosen, nstart = 25)

data_all$Cluster <- as.factor(kmeans_result$cluster)

data_all$Cluster
```

```
##   [1] 2 2 1 2 3 2 2 2 2 2 2 2 3 3 3 2 1 1 3 1 1 1 2 2 2 2 2 1 1 1 1 2 1 2 1 1 3
##  [38] 3 1 3 3 2 1 3 1 2 1 3 3 3 3 1 2 1 3 1 1 3 3 3 1 1 3 1 3 2 3 3 3 3 1 3 3 1
##  [75] 3 3 1 1 1 3 3 3 3 1 1 1 1 3 1 1 3 3 3 3 3 3 1 3 3 3 3 3 3 3 3 3 3 2 3 3 3
## [112] 1 2 3 2 3 2 2 3 3 2 3 3 3 3 2 3 2 2 3 3 2 3 3 1 2 1 3 2 2 3 1 3 1 1 2 3 3
## [149] 3 3 1 3 3 3 2 3 2 3 1 1 3 1 3 2 1 1 3 3 3 3 3 3 1 2 2 2 2 2 2 2 2 2 2 2 2
## [186] 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 1 1 2 2 2 2 2
## [223] 1 2 2 2 2 3 3 3 3 3 3 3 3 1 3 3 3 3 3 1 3 3 3 2 3 3 3 3 3 2 3 3 3 3 3 3 3
## [260] 2 3 3 2 1 2 2 2 2 2 2 2 2 2 2 2 3 3 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 1 2
## [297] 1 1 3 1 1 1 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 3 3 3 3 2 1 2 2 3 1 3 2 2 3
## [334] 3 2 2 3 3 3 3 2 3 3 3 3 3 3 3 3 3 3 1 3 2 3 2 2 2 2 2 2 2 3 3 3 1 3 1 3 2
## [371] 2 2 2 2 3 2 2 3 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [408] 1 2 1 1 1 2 2 3 2 1 2 1 2 1 2 2 2 2 2 1 2 2 2 2 2 2 3 2 3 3 3 3 3 2 1 1
## [445] 2 2 2 3 2 2 2 2 1 3 3 3 2 3 3 3 3 3 3 3 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2
## [482] 2 2 2 2 2 2 2 2 2 2 2 2 2 1 3 3 3 1 3 2 2 2 2 3 3 3 3 3 2 3 3 3 1 2 3 3 2
## [519] 3 3 3 3 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
```

```
## [556] 2 2 2 3 2 2 1 2 3 3 3 2 2 3 3 3 2 3 2 3 2 3 3 2 2 2 3 2 2 3 3 3 2 2 3 1 3
## [593] 2 1 1 1 2 2 2 1 3 2 2 2 3 3 1 1 3 1 3 3 3 3 3 3 3 1 1 3 3 2 1 1 2 3 3 3 3
## [630] 3 3 3 3 1 3 1 1 1 1 1 3 1 1 3 3 3 3 3 1 1 1 1 1 1 3 1 2 1 1 3 3 3 1 3 2 2
## [667] 2 2 3 3 3 1 2 2 2 1 2 2 2 2 2 2 2 2 1 2 2 2 3 3 3 3 3 1 2 2 3 1 2 2 2 3 1
## [704] 3
## Levels: 1 2 3
```