Chapter 4 BADA

Barycentric discriminant analysis is a robust version of discriminant analysis that is used when multiple measurements describe a set of observations in which each observation belongs to one category (i.e.,group) from a set of a priori defined categories. BADA combines the original variables to create new variables that best separate the groups and that can also be used to optimally assign old or new observations to these categories. The quality of the performance is evaluated by cross-validation techniques that estimate the performance of the classification model for new observations. BADA is a very versatile technique that can be declined in several different varieties that can handle, for example, qualitative data and data structured in blocks.This versatility make BADA particularly suited for the analysis of multi-modal and Bigdata.

NOTE(In this project we are using bada for analysis not for prediction)

LETS START BY CLEARING THE ENVIRONMENT.

```
knitr::opts_chunk$set(echo = TRUE)
rm(list = ls())
graphics.off()
```

4.1 DATASET

The same dataset used for pca is used in BADA.

```
load("C:/Users/jeevan/Desktop/RM2/R-M/SmartphoneUsage (1).RData")
#install.packages('tidyverse')
library(tidyverse)
```

```
## -- Attaching packages ------------------------------------------------------------- tidy
## v tibble  2.1.3       v purrr   0.3.2
## v tidyr   0.8.3       v dplyr   0.8.0.1
## v readr   1.3.1       v stringr 1.4.0
## v tibble  2.1.3       v forcats 0.4.0
```

```
## -- Conflicts ---------------------------------------------------------------- tidyverse_
## x dplyr::combine() masks gridExtra::combine()
## x dplyr::filter()  masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library(ExPosition)
#install.packages('TExPosition') # if needed
library(TExPosition)
library(TInPosition)
library(PTCA4CATA)
# devtools::install_github('HerveAbdi/data4PCCAR')
library(data4PCCAR)
b=data.more
b["C"]=c("Am","E","Am","A","E","E","E","E","E","A","A","E","E","A","A","Am","E","E","A","A'

rawdata=data.pca[-c(22),]
Xmat <-  rawdata[, 2:4]
Ymat <-  rawdata[, 5:7]
XYmat <- rawdata[, 2:7]
```

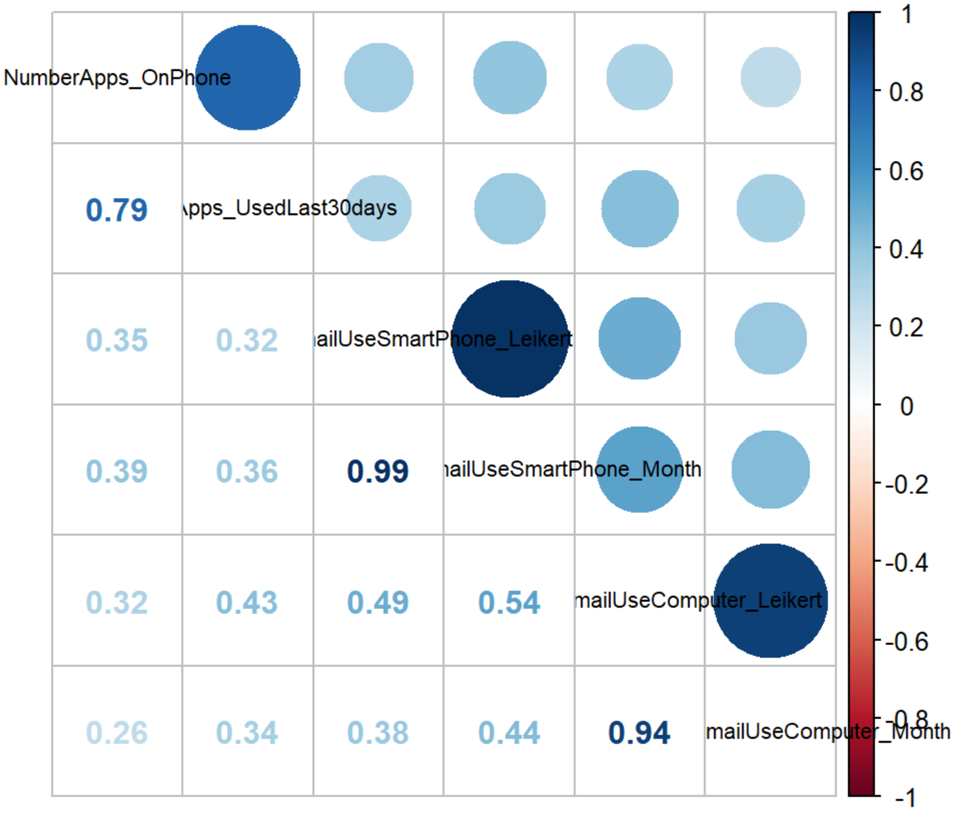
4.2 COR PLOT

In bada we will find the correlation between the variables just as pca but we won’t include the variable purchase decision.

```
library(corrplot)
```

```
cor.res <- cor(XYmat)
```

```
corrplot.mixed(cor.res, tl.cex = 0.7, tl.col = "black")
```



4.3 SCREE PLOT

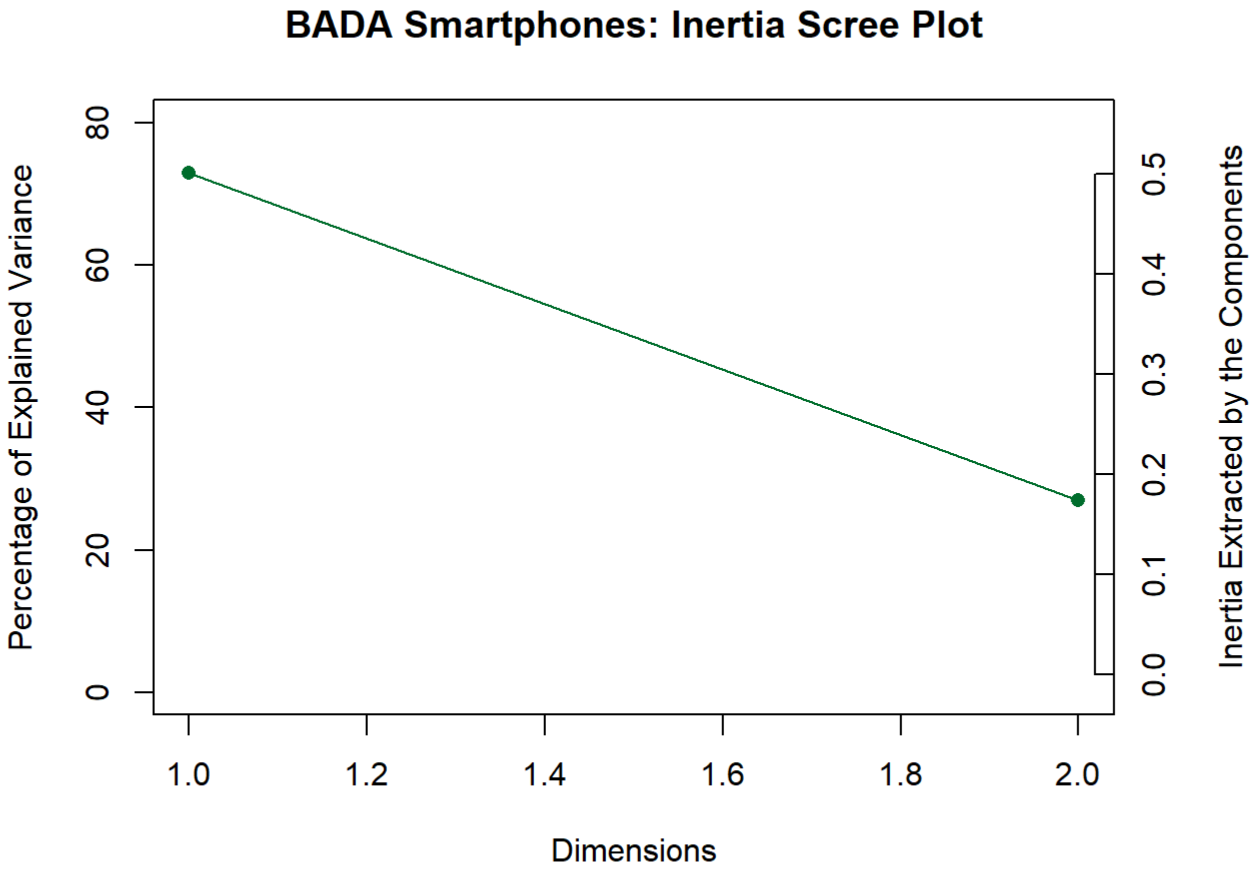
component 1 showed max variance.

component 2 does not explain much variance.

```r
resBADA <- tepBADA(XYmat, DESIGN =  b$C[which(b$C!="AF")], make_design_nominal = TRUE,
                   graphs = FALSE)
# Inferences ----
set.seed(70301) # we have a problem with the inference part
  # it will be addressed soon. In the meantime we fix the seed
  # for random
nIter = 50
resBADA.inf <- tepBADA.inference.battery(XYmat,
                  DESIGN =b$C[which(b$C!="AF")],
                  make_design_nominal = TRUE,
                  test.iters = nIter,
                  graphs = FALSE)
```

```
## [1] "It is estimated that your iterations will take 0.02 minutes."
## [1] "R is not in interactive() mode. Resample-based tests will be conducted. Please take
## ===========================================================================
```

```r
PlotScree(ev = resBADA$TExPosition.Data$eigs,
          title = 'BADA Smartphones: Inertia Scree Plot',
          plotKaiser = FALSE,
          color4Kaiser = ggplot2::alpha('darkorchid4', .5),
          lwd4Kaiser  = 2)
```



```
# Save the plot
```

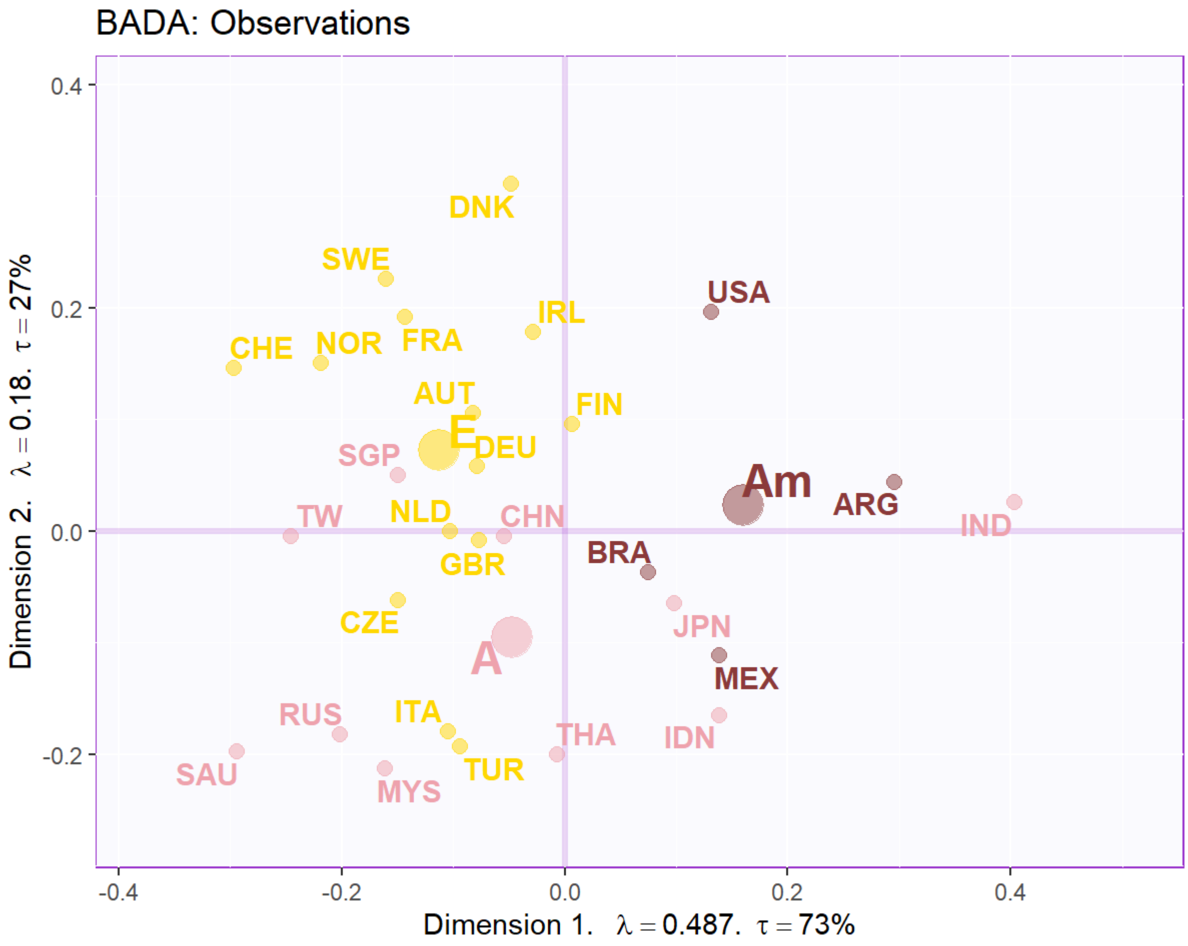
4.4 PROJECTIONS.

These plots show us which countries show similar usage of smartphones.

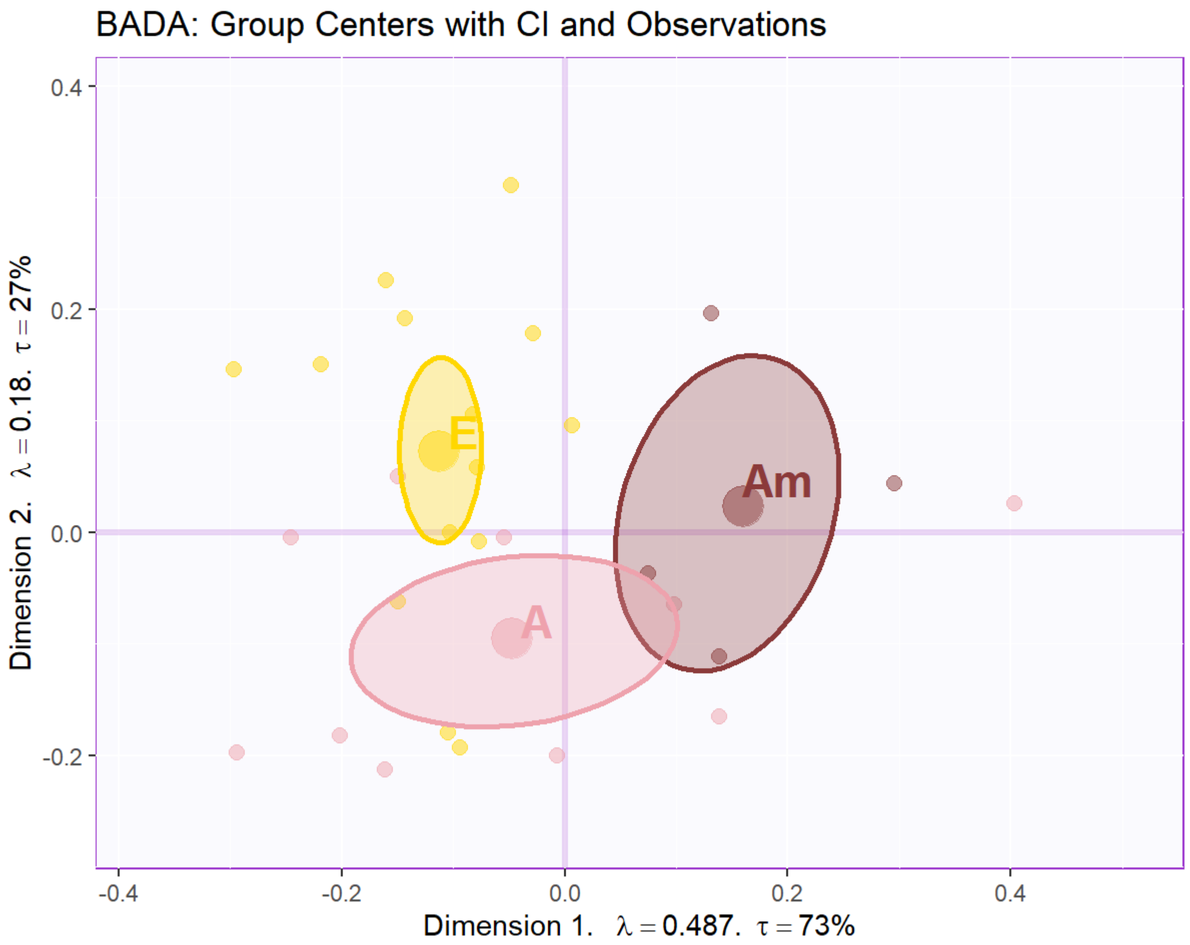
interpretation can be difficult so it is a good practice to group them.

```r
wineColors <-  b$C[which(b$C!="AF")]
wineColors <- recode(wineColors, Am = 'indianred4',
                        E = 'gold', A = 'lightpink2',)

Imap <- PTCA4CATA::createFactorMap(
  resBADA$TExPosition.Data$fii,
  col.points = wineColors,
  col.labels = wineColors,
  alpha.points = .5
)

label4Map <- createxyLabels.gen(1,2,
                                lambda = resBADA$TExPosition.Data$eigs,
                                tau = resBADA$TExPosition.Data$t)



wineMeans <- PTCA4CATA::getMeans(resBADA$TExPosition.Data$fii,
                                                  b$C[which(b$C!="AF")])

col4Means <- recode(rownames(wineMeans),Am = 'indianred4',
                        E = 'gold', A = 'lightpink2', )


names(col4Means) <- rownames(wineMeans)


MapGroup <- PTCA4CATA::createFactorMap(wineMeans,
                             # use the constraint from the main map
                             constraints = Imap$constraints,
                             col.points = col4Means,
                             cex = 7,  # size of the dot (bigger)
                             col.labels = col4Means,
                             text.cex = 6)
```

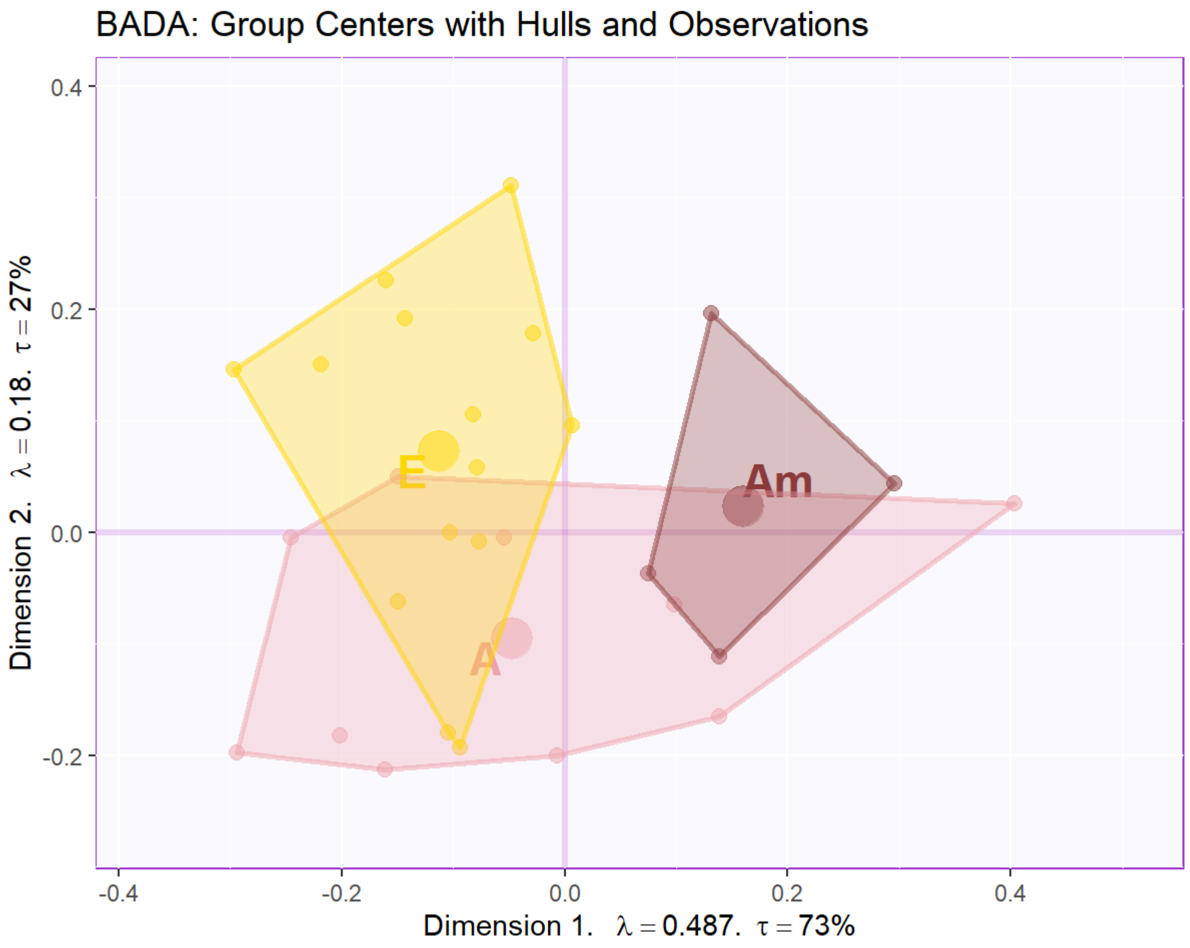
```
# The map with observations and group means
a003.bada <- Imap$zeMap + label4Map +
    MapGroup$zeMap_dots + MapGroup$zeMap_text+ggtitle('BADA: Observations')
print(a003.bada)
```



```r
#_______________________Bootstrapped CI ----
#______________________________________________________________________
# Create Confidence Interval Plots
fi.boot <- resBADA.inf$Inference.Data$boot.data$fi.boot.data$boots
# We want to use the rownames of fi.boot as reference to get the correct
# color. However, the original rownames include "." and don't match with
# the original row names. So, the `sub` function was used to get rid of
# the "." by replacing all "." in the rownames of fi.boot as an empty
# string.
rownames(fi.boot) <- sub("[[:punct:]]","",rownames(fi.boot))
# use function MakeCIEllipses from package PTCA4CATA
GraphElli <- PTCA4CATA::MakeCIEllipses(
            resBADA.inf$Inference.Data$boot.data$fi.boot.data$boots,
                                        col = col4Means[rownames(fi.boot)], # use rownames
                                        p.level = .95
)
#______________________________________________________________________
# create the I-map with Observations, means and confidence intervals
#
a004.bada.withCI <-  Imap$zeMap_background + Imap$zeMap_dots +
                     MapGroup$zeMap_dots + MapGroup$zeMap_text +
                     GraphElli + label4Map +
                     ggtitle('BADA: Group Centers with CI and Observations')
#______________________________________________________________________
                                                                                           #

a004.bada.withCI
```


BADA: Group Centers with CI and Observations
0.4
0.2
0.0
-0.2
Dimension 2. λ = 0.18. τ = 27%
E
Am
A
-0.4
-0.2
0.0
0.2
0.4
Dimension 1. λ = 0.487. τ = 73%

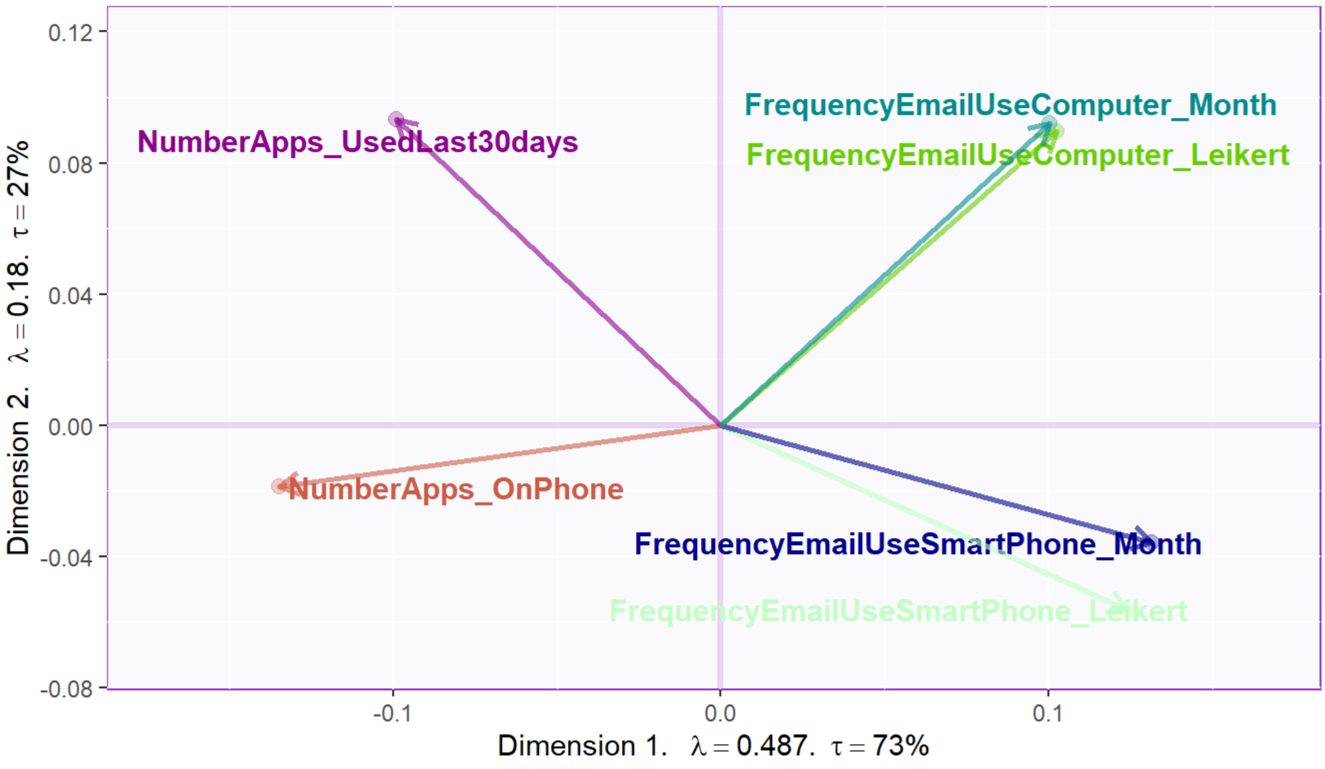
```
Fii <- resBADA$TExPosition.Data$fii
colnames(Fii) <- paste0('Dimension ', 1:ncol(Fii))
# getting the color correct: an ugly trick
col4Hull <- col4Means[match(names(col4Means),
                              levels( b$C[which(b$C!="AF")]) )]
GraphHull <- PTCA4CATA::MakeToleranceIntervals(Fii,
                                 design =  b$C[which(b$C!="AF")],
                                 col = col4Means,
                                 # the next line is required
                                 # for some strange unknown reasons
                                 names.of.factors =  c("Dim1","Dim2"),
                                 p.level = 1.00)
#
a006.bada.withHull <-  Imap$zeMap_background + Imap$zeMap_dots +
            MapGroup$zeMap_dots + MapGroup$zeMap_text +
            GraphHull + label4Map +
            ggtitle('BADA: Group Centers with Hulls and Observations')



col4X <- prettyGraphsColorSelection(n.colors = ncol(Xmat),
                                    starting.color = 42)

col4Y <- prettyGraphsColorSelection(n.colors = ncol(Ymat),
                                    starting.color = 13)
col4Var = c(col4X,col4Y)
#_____________________________________________________________________
a006.bada.withHull
```



4.5 LOADINGS

Use cosine interpretations between the variables

```
Fj <- resBADA$TExPosition.Data$fj
baseMap.j <- PTCA4CATA::createFactorMap(Fj,
                                        col.points   = col4Var,
                                        alpha.points =  .3,
                                        col.labels   = col4Var)

zeArrows <- addArrows(Fj, color = col4Var)
# A graph for the J-set
# A graph for the J-set
b001.aggMap.j <- baseMap.j$zeMap_background + # background layer
  baseMap.j$zeMap_dots + baseMap.j$zeMap_text +  # dots & labels
  label4Map
b002.aggMap.j <- b001.aggMap.j + zeArrows
b002.aggMap.j
```



4.6 ANALYSIS/HYPOTHESIS

BADA: Group Centers with CI and Observations

This plot shows that asian and american countries slightly overlap but europian countries are away from them. (NOTE we have grouped north american and south american countries together because of bootstrap code error)

BADA: Group Centers with Hulls and Observations

Few european and american countries overlap with asian countries.

BADA:LOADINGS

If apps are used more on the phone email app will be used less and vice versa

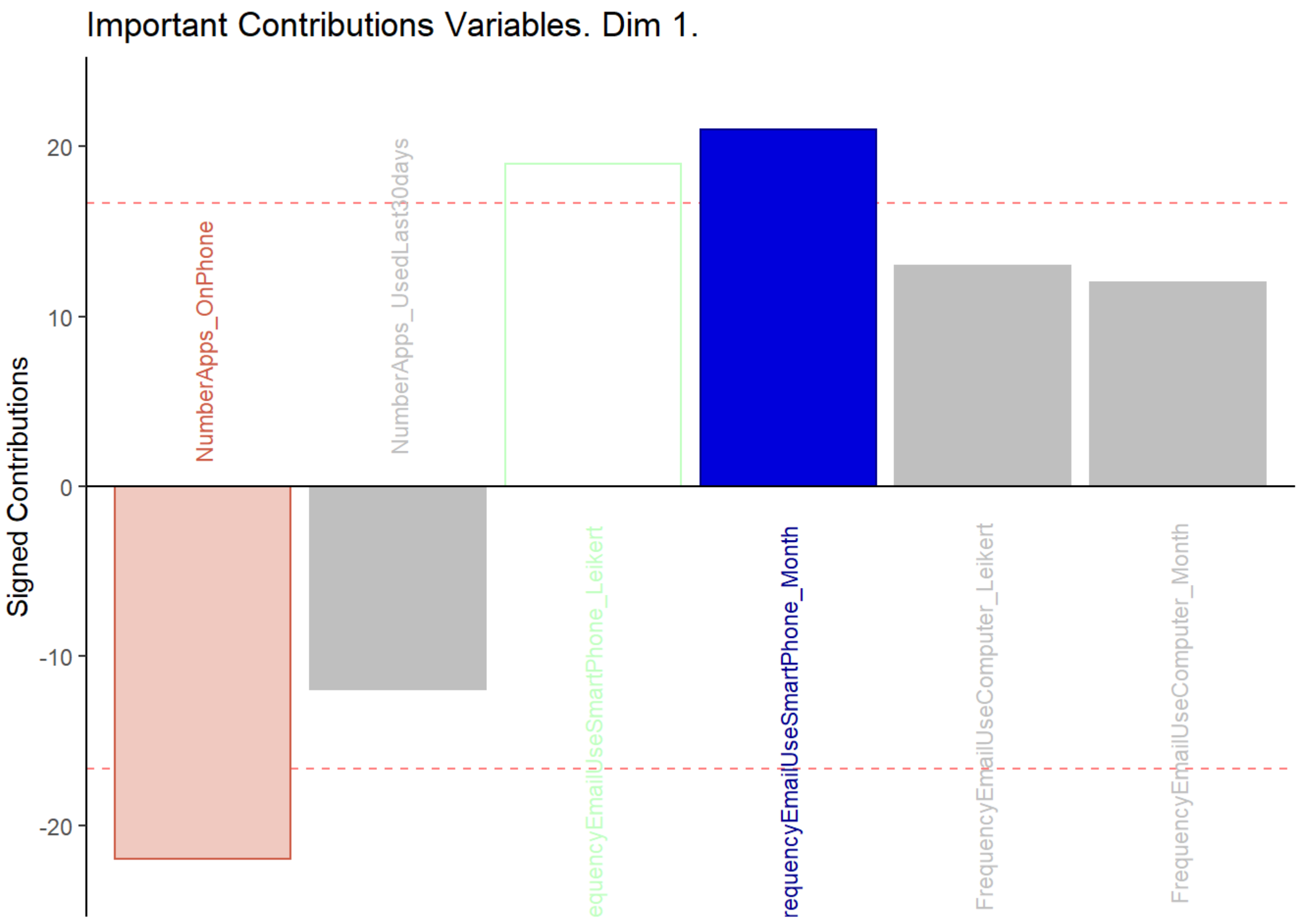
If more apps are there on the phone computer email usage would be less and vice versa.

4.7 CONTRIBUTIONS

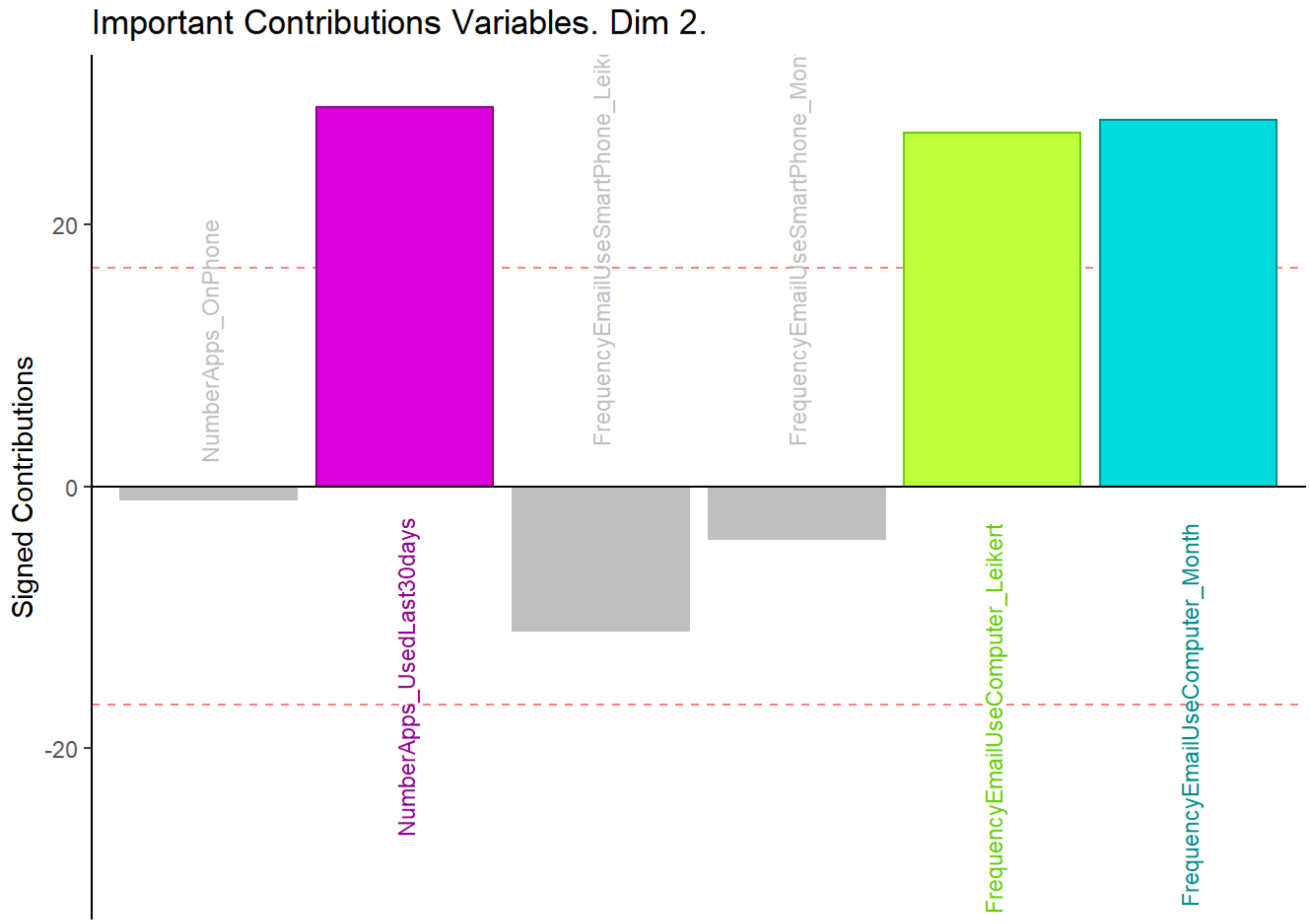
For component 1, 3 variables show significant contributions.

For component 2, 3 variables show significant contributions.

```
ctrj <- resBADA$TExPosition.Data$cj
signed.ctrj <- ctrj * sign(Fj)
# BR1
c001.plotCtrj.1 <- PrettyBarPlot2(
                       bootratio = round(100*signed.ctrj[,1]),
                       threshold = 100 / nrow(signed.ctrj),
                       ylim = NULL,
                       color4bar = gplots::col2hex(col4Var),
                       color4ns = "gray75",
                       plotnames = TRUE,
                       main = 'Important Contributions Variables. Dim 1.',
                       ylab = "Signed Contributions")

c001.plotCtrj.1
```



```
c002.plotCtrj.2 <- PrettyBarPlot2(
  bootratio = round(100*signed.ctrj[,2]),
  threshold = 100 / nrow(signed.ctrj),
  ylim = NULL,
  color4bar = gplots::col2hex(col4Var),
  color4ns = "gray75",
  plotnames = TRUE,
  main = 'Important Contributions Variables. Dim 2.',
  ylab = "Signed Contributions")

c002.plotCtrj.2
```



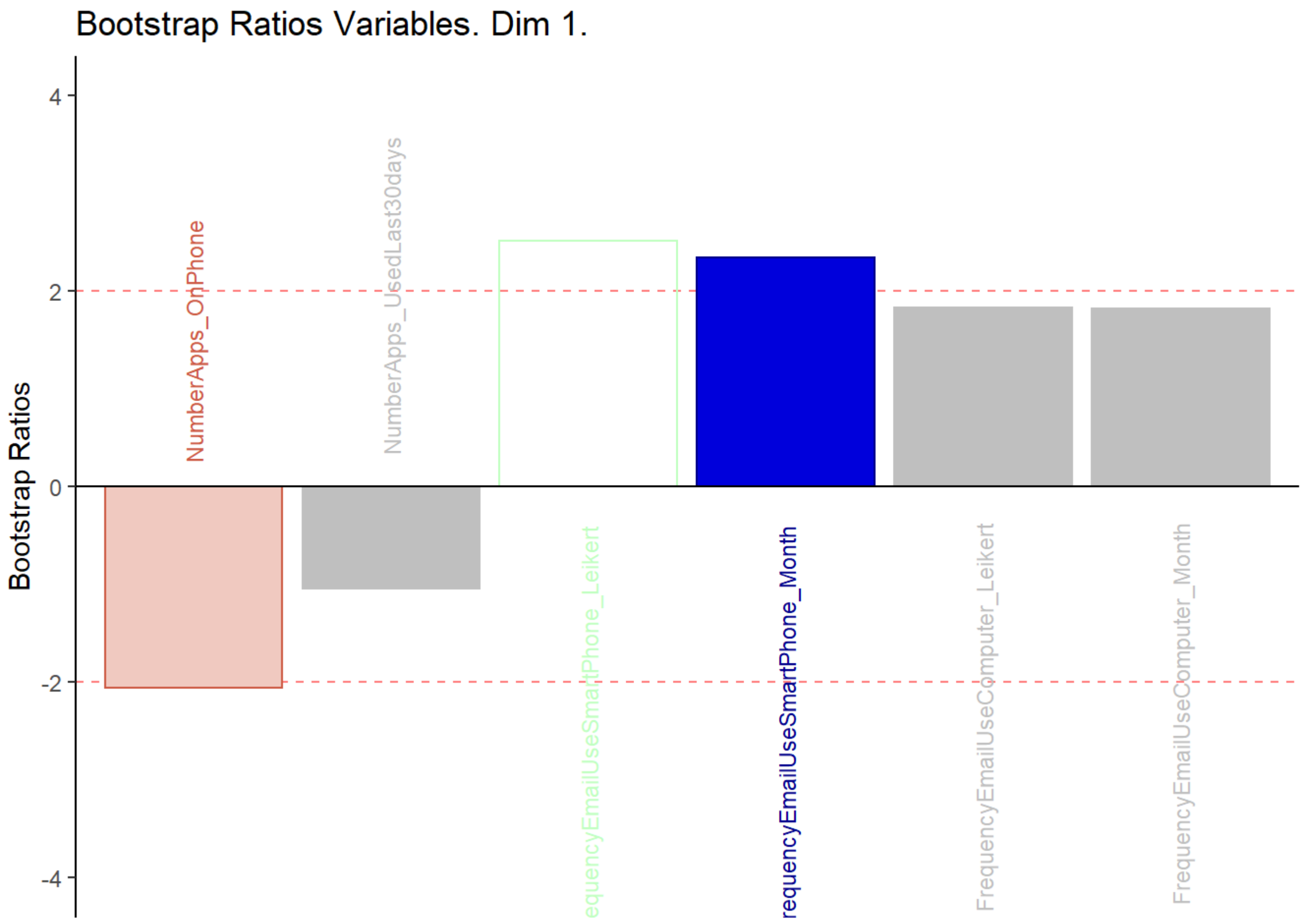
4.8 BOOTSTRAP

The information conveyed by the scree plot is confirmed by the boostrap contributions

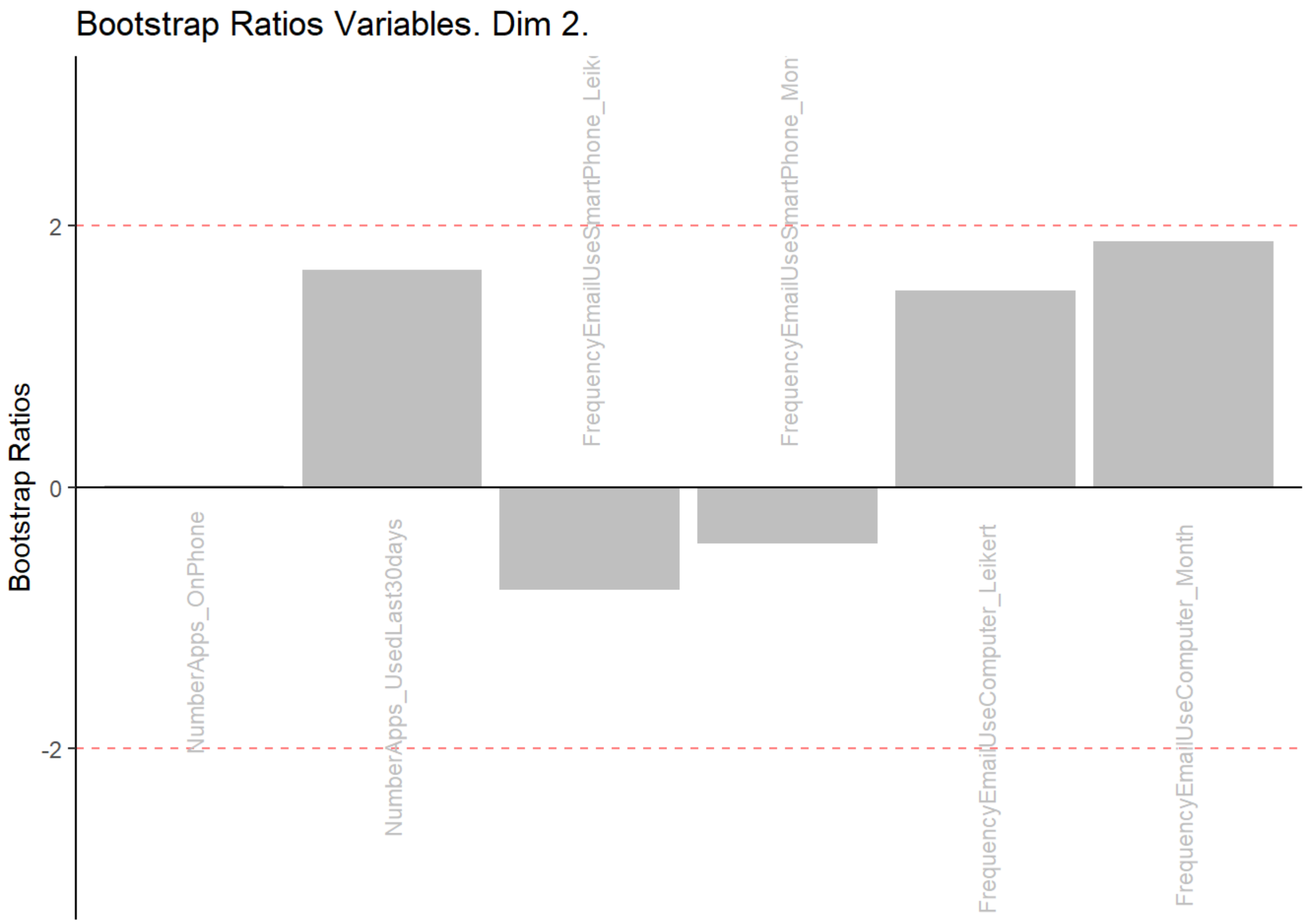
Variables have shown major contributions for component 1 than component 2

We can thereby reject the null hypothesis

```r
#
BRj <- resBADA.inf$Inference.Data$boot.data$fj.boot.data$tests$boot.ratios
# BR1
d001.plotBRj.1 <- PrettyBarPlot2(
  bootratio = BRj[,1],
  threshold = 2,
  ylim = NULL,
  color4bar = gplots::col2hex(col4Var),
  color4ns = "gray75",
  plotnames = TRUE,
  main = 'Bootstrap Ratios Variables. Dim 1.',
  ylab = "Bootstrap Ratios")

d001.plotBRj.1
```



```
#_____________________________________________________________________
###### 2 ====
#
d003.plotBRj.2 <- PrettyBarPlot2(
  bootratio = BRj[,2],
  threshold = 2,
  ylim = NULL,
  color4bar = gplots::col2hex(col4Var),
  color4ns = "gray75",
  plotnames = TRUE,
  main = 'Bootstrap Ratios Variables. Dim 2.',
  ylab = "Bootstrap Ratios")

d003.plotBRj.2
```



Bootstrap Ratios Variables. Dim 2.