A11.Zunqiu.Wang

Zunqiu Wang

11/18/2021

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

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

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

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v tibble  3.1.5     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.0.2     v forcats 0.5.1
## v purrr   0.3.4
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x ggplot2::%+%()   masks psych::%+%()
## x ggplot2::alpha() masks psych::alpha()
## x dplyr::filter()  masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library(dplyr)
library(reshape2)
```

```
##
## Attaching package: 'reshape2'
```

```
## The following object is masked from 'package:tidyr':
##
##     smiths
```

```
library(GPArotation)
```

```
data(bfi)
dim(bfi)
```

```
## [1] 2800   28
```

```
colSums(is.na(bfi))
```

```
##        A1        A2        A3        A4        A5        C1        C2        C3
##        16        27        26        19        16        21        24        20
##        C4        C5        E1        E2        E3        E4        E5        N1
##        26        16        23        16        25         9        21        22
```

```
##        N2        N3        N4        N5        O1        O2        O3        O4
##        21        11        36        29        22         0        28        14
##        O5    gender education       age
##        20         0       223         0
```

```
str(bfi)
```

```
## 'data.frame':    2800 obs. of  28 variables:
##  $ A1       : int  2 2 5 4 2 6 2 4 4 2 ...
##  $ A2       : int  4 4 4 4 3 6 5 3 3 5 ...
##  $ A3       : int  3 5 5 6 3 5 5 1 6 6 ...
##  $ A4       : int  4 2 4 5 4 6 3 5 3 6 ...
##  $ A5       : int  4 5 4 5 5 5 5 1 3 5 ...
##  $ C1       : int  2 5 4 4 4 6 5 3 6 6 ...
##  $ C2       : int  3 4 5 4 4 6 4 2 6 5 ...
##  $ C3       : int  3 4 4 3 5 6 4 4 3 6 ...
##  $ C4       : int  4 3 2 5 3 1 2 2 4 2 ...
##  $ C5       : int  4 4 5 5 2 3 3 4 5 1 ...
##  $ E1       : int  3 1 2 5 2 2 4 3 5 2 ...
##  $ E2       : int  3 1 4 3 2 1 3 6 3 2 ...
##  $ E3       : int  3 6 4 4 5 6 4 4 NA 4 ...
##  $ E4       : int  4 4 4 4 4 5 5 2 4 5 ...
##  $ E5       : int  4 3 5 4 5 6 5 1 3 5 ...
##  $ N1       : int  3 3 4 2 2 3 1 6 5 5 ...
##  $ N2       : int  4 3 5 5 3 5 2 3 5 5 ...
##  $ N3       : int  2 3 4 2 4 2 2 2 2 5 ...
##  $ N4       : int  2 5 2 4 4 2 1 6 3 2 ...
##  $ N5       : int  3 5 3 1 3 3 1 4 3 4 ...
##  $ O1       : int  3 4 4 3 3 4 5 3 6 5 ...
##  $ O2       : int  6 2 2 3 3 3 2 2 6 1 ...
##  $ O3       : int  3 4 5 4 4 5 5 4 6 5 ...
##  $ O4       : int  4 3 5 3 3 6 6 5 6 5 ...
##  $ O5       : int  3 3 2 5 3 1 1 3 1 2 ...
##  $ gender   : int  1 2 2 2 1 2 1 1 1 2 ...
##  $ education: int  NA NA NA NA NA 3 NA 2 1 NA ...
##  $ age      : int  16 18 17 17 17 21 18 19 19 17 ...
```

```
summary(bfi)
```

```
##        A1              A2              A3              A4              A5
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.0   Min.   :1.00
##  1st Qu.:1.000   1st Qu.:4.000   1st Qu.:4.000   1st Qu.:4.0   1st Qu.:4.00
##  Median :2.000   Median :5.000   Median :5.000   Median :5.0   Median :5.00
##  Mean   :2.413   Mean   :4.802   Mean   :4.604   Mean   :4.7   Mean   :4.56
##  3rd Qu.:3.000   3rd Qu.:6.000   3rd Qu.:6.000   3rd Qu.:6.0   3rd Qu.:5.00
##  Max.   :6.000   Max.   :6.000   Max.   :6.000   Max.   :6.0   Max.   :6.00
##  NA's   :16      NA's   :27      NA's   :26      NA's   :19    NA's   :16
##        C1              C2              C3              C4              C5
##  Min.   :1.000   Min.   :1.00   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:4.000   1st Qu.:4.00   1st Qu.:4.000   1st Qu.:1.000   1st Qu.:2.000
##  Median :5.000   Median :5.00   Median :5.000   Median :2.000   Median :3.000
##  Mean   :4.502   Mean   :4.37   Mean   :4.304   Mean   :2.553   Mean   :3.297
##  3rd Qu.:5.000   3rd Qu.:5.00   3rd Qu.:5.000   3rd Qu.:4.000   3rd Qu.:5.000
##  Max.   :6.000   Max.   :6.00   Max.   :6.000   Max.   :6.000   Max.   :6.000
##  NA's   :21      NA's   :24     NA's   :20      NA's   :26      NA's   :16
```

```
##        E1              E2              E3              E4
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:2.000   1st Qu.:2.000   1st Qu.:3.000   1st Qu.:4.000
##  Median :3.000   Median :3.000   Median :4.000   Median :5.000
##  Mean   :2.974   Mean   :3.142   Mean   :4.001   Mean   :4.422
##  3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:5.000   3rd Qu.:6.000
##  Max.   :6.000   Max.   :6.000   Max.   :6.000   Max.   :6.000
##  NA's   :23      NA's   :16      NA's   :25      NA's   :9
##        E5              N1              N2              N3
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:4.000   1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000
##  Median :5.000   Median :3.000   Median :4.000   Median :3.000
##  Mean   :4.416   Mean   :2.929   Mean   :3.508   Mean   :3.217
##  3rd Qu.:5.000   3rd Qu.:4.000   3rd Qu.:5.000   3rd Qu.:4.000
##  Max.   :6.000   Max.   :6.000   Max.   :6.000   Max.   :6.000
##  NA's   :21      NA's   :22      NA's   :21      NA's   :11
##        N4              N5             O1              O2             O3
##  Min.   :1.000   Min.   :1.00   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:2.000   1st Qu.:2.00   1st Qu.:4.000   1st Qu.:1.000   1st Qu.:4.000
##  Median :3.000   Median :3.00   Median :5.000   Median :2.000   Median :5.000
##  Mean   :3.186   Mean   :2.97   Mean   :4.816   Mean   :2.713   Mean   :4.438
##  3rd Qu.:4.000   3rd Qu.:4.00   3rd Qu.:6.000   3rd Qu.:4.000   3rd Qu.:5.000
##  Max.   :6.000   Max.   :6.00   Max.   :6.000   Max.   :6.000   Max.   :6.000
##  NA's   :36      NA's   :29     NA's   :22                      NA's   :28
##        O4              O5           gender        education         age
##  Min.   :1.000   Min.   :1.00   Min.   :1.000   Min.   :1.00   Min.   : 3.00
##  1st Qu.:4.000   1st Qu.:1.00   1st Qu.:1.000   1st Qu.:3.00   1st Qu.:20.00
##  Median :5.000   Median :2.00   Median :2.000   Median :3.00   Median :26.00
##  Mean   :4.892   Mean   :2.49   Mean   :1.672   Mean   :3.19   Mean   :28.78
##  3rd Qu.:6.000   3rd Qu.:3.00   3rd Qu.:2.000   3rd Qu.:4.00   3rd Qu.:35.00
##  Max.   :6.000   Max.   :6.00   Max.   :2.000   Max.   :5.00   Max.   :86.00
##  NA's   :14      NA's   :20                     NA's   :223
```

```
# impute NA with median of that column var values since all are numeric
bfi[,sapply(bfi, is.numeric)] <- lapply(bfi[,sapply(bfi, is.numeric)],
                                        function(x){
                                         x <- ifelse(is.na(x), median(x, na.rm  = TRUE), x)
                                          }
                                        )
summary(bfi)
```

```
##        A1              A2              A3              A4
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:1.000   1st Qu.:4.000   1st Qu.:4.000   1st Qu.:4.000
##  Median :2.000   Median :5.000   Median :5.000   Median :5.000
##  Mean   :2.411   Mean   :4.804   Mean   :4.607   Mean   :4.702
##  3rd Qu.:3.000   3rd Qu.:6.000   3rd Qu.:6.000   3rd Qu.:6.000
##  Max.   :6.000   Max.   :6.000   Max.   :6.000   Max.   :6.000
##        A5              C1              C2              C3
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:4.000   1st Qu.:4.000   1st Qu.:4.000   1st Qu.:4.000
##  Median :5.000   Median :5.000   Median :5.000   Median :5.000
##  Mean   :4.563   Mean   :4.506   Mean   :4.375   Mean   :4.309
##  3rd Qu.:5.000   3rd Qu.:5.000   3rd Qu.:5.000   3rd Qu.:5.000
##  Max.   :6.000   Max.   :6.000   Max.   :6.000   Max.   :6.000
```

```
##        C4              C5              E1              E2
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:1.000   1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.000
##  Median :2.000   Median :3.000   Median :3.000   Median :3.000
##  Mean   :2.548   Mean   :3.295   Mean   :2.975   Mean   :3.141
##  3rd Qu.:4.000   3rd Qu.:5.000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :6.000   Max.   :6.000   Max.   :6.000   Max.   :6.000
##        E3              E4              E5              N1             N2
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.00   Min.   :1.000
##  1st Qu.:3.000   1st Qu.:4.000   1st Qu.:4.000   1st Qu.:2.00   1st Qu.:2.000
##  Median :4.000   Median :5.000   Median :5.000   Median :3.00   Median :4.000
##  Mean   :4.001   Mean   :4.424   Mean   :4.421   Mean   :2.93   Mean   :3.511
##  3rd Qu.:5.000   3rd Qu.:6.000   3rd Qu.:5.000   3rd Qu.:4.00   3rd Qu.:5.000
##  Max.   :6.000   Max.   :6.000   Max.   :6.000   Max.   :6.00   Max.   :6.000
##        N3              N4              N5             O1             O2
##  Min.   :1.000   Min.   :1.000   Min.   :1.00   Min.   :1.000   Min.   :1.000
##  1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.00   1st Qu.:4.000   1st Qu.:1.000
##  Median :3.000   Median :3.000   Median :3.00   Median :5.000   Median :2.000
##  Mean   :3.216   Mean   :3.183   Mean   :2.97   Mean   :4.817   Mean   :2.713
##  3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.00   3rd Qu.:6.000   3rd Qu.:4.000
##  Max.   :6.000   Max.   :6.000   Max.   :6.00   Max.   :6.000   Max.   :6.000
##        O3              O4              O5            gender
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:4.000   1st Qu.:4.000   1st Qu.:1.000   1st Qu.:1.000
##  Median :5.000   Median :5.000   Median :2.000   Median :2.000
##  Mean   :4.444   Mean   :4.893   Mean   :2.486   Mean   :1.672
##  3rd Qu.:5.000   3rd Qu.:6.000   3rd Qu.:3.000   3rd Qu.:2.000
##  Max.   :6.000   Max.   :6.000   Max.   :6.000   Max.   :2.000
##    education          age
##  Min.   :1.000   Min.   : 3.00
##  1st Qu.:3.000   1st Qu.:20.00
##  Median :3.000   Median :26.00
##  Mean   :3.175   Mean   :28.78
##  3rd Qu.:4.000   3rd Qu.:35.00
##  Max.   :5.000   Max.   :86.00
```

```
bfi.scale <- scale(bfi)
bfi.df <- data.frame(bfi.scale)
names(bfi.df)
```

```
##  [1] "A1"        "A2"        "A3"        "A4"        "A5"        "C1"
##  [7] "C2"        "C3"        "C4"        "C5"        "E1"        "E2"
## [13] "E3"        "E4"        "E5"        "N1"        "N2"        "N3"
## [19] "N4"        "N5"        "O1"        "O2"        "O3"        "O4"
## [25] "O5"        "gender"    "education" "age"
```

Q1

```
pcaA <- prcomp(bfi.df)
str(pcaA)
```

```
## List of 5
##  $ sdev    : num [1:28] 2.25 1.67 1.46 1.37 1.28 ...
##  $ rotation: num [1:28, 1:28] 0.117 -0.22 -0.246 -0.196 -0.27 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:28] "A1" "A2" "A3" "A4" ...
```

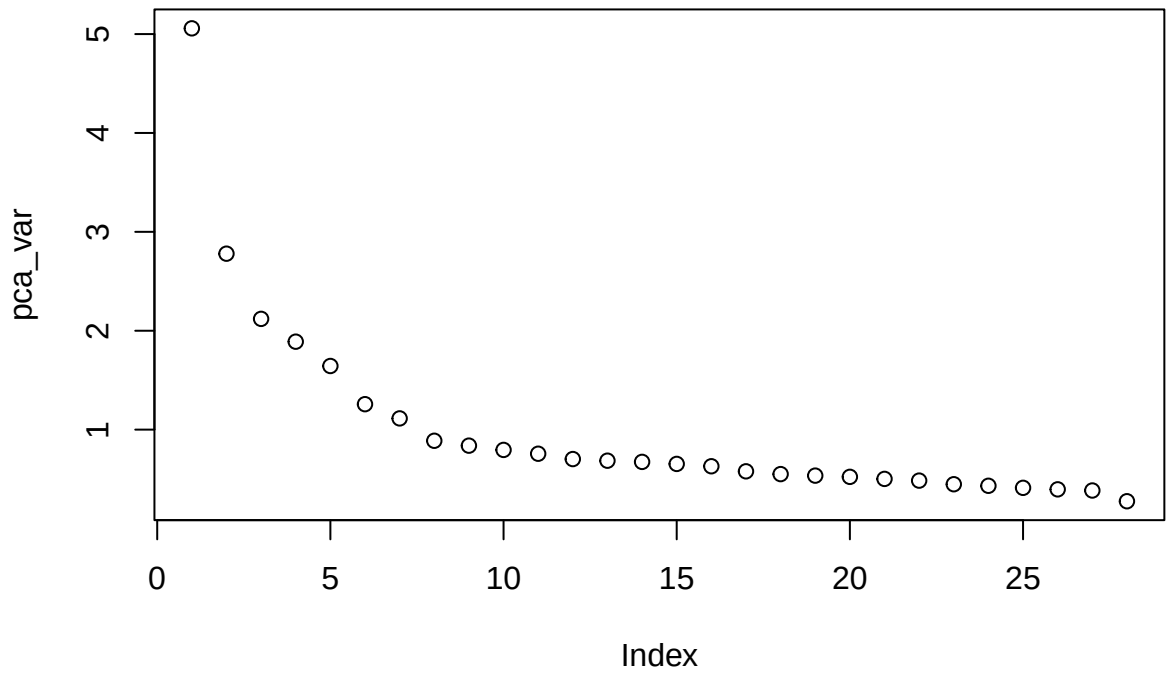
```
##   .. ..$ : chr [1:28] "PC1" "PC2" "PC3" "PC4" ...
##  $ center  : Named num [1:28] -1.03e-16 3.46e-17 4.39e-17 2.32e-17 -2.17e-16 ...
##   ..- attr(*, "names")= chr [1:28] "A1" "A2" "A3" "A4" ...
##  $ scale   : logi FALSE
##  $ x       : num [1:2800, 1:28] 2.566 0.466 0.942 1.591 0.552 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:2800] "61617" "61618" "61620" "61621" ...
##   .. ..$ : chr [1:28] "PC1" "PC2" "PC3" "PC4" ...
##  - attr(*, "class")= chr "prcomp"
```

```
pca_var <- pcaA$sdev^2
# pcaA$x
pcaA$rotation == pca_var
```

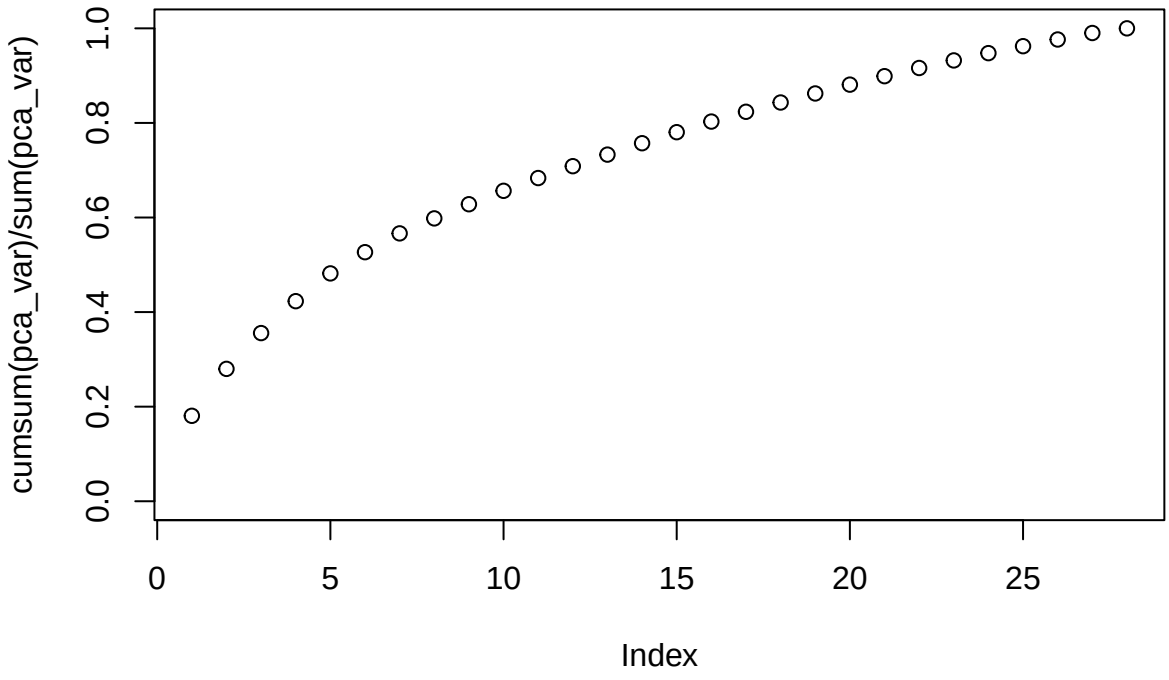
```
##             PC1   PC2   PC3   PC4   PC5   PC6   PC7   PC8   PC9  PC10  PC11
## A1        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## A2        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## A3        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## A4        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## A5        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## C1        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## C2        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## C3        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## C4        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## C5        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## E1        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## E2        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## E3        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## E4        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## E5        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## N1        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## N2        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## N3        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## N4        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## N5        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## O1        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## O2        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## O3        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## O4        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## O5        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## gender    FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## education FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## age       FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##             PC12  PC13  PC14  PC15  PC16  PC17  PC18  PC19  PC20  PC21  PC22
## A1        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## A2        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## A3        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## A4        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## A5        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## C1        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## C2        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## C3        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## C4        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## C5        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## E1        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
## E2        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## E3        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## E4        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## E5        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## N1        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## N2        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## N3        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## N4        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## N5        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## O1        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## O2        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## O3        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## O4        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## O5        FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## gender    FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## education FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## age       FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##             PC23  PC24  PC25  PC26  PC27  PC28
## A1        FALSE FALSE FALSE FALSE FALSE FALSE
## A2        FALSE FALSE FALSE FALSE FALSE FALSE
## A3        FALSE FALSE FALSE FALSE FALSE FALSE
## A4        FALSE FALSE FALSE FALSE FALSE FALSE
## A5        FALSE FALSE FALSE FALSE FALSE FALSE
## C1        FALSE FALSE FALSE FALSE FALSE FALSE
## C2        FALSE FALSE FALSE FALSE FALSE FALSE
## C3        FALSE FALSE FALSE FALSE FALSE FALSE
## C4        FALSE FALSE FALSE FALSE FALSE FALSE
## C5        FALSE FALSE FALSE FALSE FALSE FALSE
## E1        FALSE FALSE FALSE FALSE FALSE FALSE
## E2        FALSE FALSE FALSE FALSE FALSE FALSE
## E3        FALSE FALSE FALSE FALSE FALSE FALSE
## E4        FALSE FALSE FALSE FALSE FALSE FALSE
## E5        FALSE FALSE FALSE FALSE FALSE FALSE
## N1        FALSE FALSE FALSE FALSE FALSE FALSE
## N2        FALSE FALSE FALSE FALSE FALSE FALSE
## N3        FALSE FALSE FALSE FALSE FALSE FALSE
## N4        FALSE FALSE FALSE FALSE FALSE FALSE
## N5        FALSE FALSE FALSE FALSE FALSE FALSE
## O1        FALSE FALSE FALSE FALSE FALSE FALSE
## O2        FALSE FALSE FALSE FALSE FALSE FALSE
## O3        FALSE FALSE FALSE FALSE FALSE FALSE
## O4        FALSE FALSE FALSE FALSE FALSE FALSE
## O5        FALSE FALSE FALSE FALSE FALSE FALSE
## gender    FALSE FALSE FALSE FALSE FALSE FALSE
## education FALSE FALSE FALSE FALSE FALSE FALSE
## age       FALSE FALSE FALSE FALSE FALSE FALSE
```

```
# scree plot
plot(pca_var)
```




```
# cumulative variance
plot(cumsum(pca_var)/sum(pca_var), ylim = c(0, 1))
```




```
# looking at elbow position is corresponding to index=7-8, and 7-8 factors explains about 60% of total
```

Q2

```
fact <- fa(bfi.df, nfactors=2)
# extracts first factor loadings from output
fact1 <- fact$loadings[,1]
fact1[order(fact1)]
```

```
##           E2           E1           C5           C4           A1           N4
## -0.517224255 -0.441872386 -0.311119456 -0.298320447 -0.200989671 -0.183066380
##           O5           O2           N5           N1           N2    education
## -0.174274537 -0.093692246 -0.006842212  0.018666342  0.029530227  0.037377595
##           N3           O4          age       gender           C3           C1
##  0.052786771  0.092569056  0.123212616  0.197555676  0.284519815  0.313021522
##           C2           O1           A4           O3           A2           E4
##  0.357575931  0.359511114  0.406787100  0.459376129  0.549639145  0.583150005
##           E5           A5           A3           E3
##  0.588820280  0.591662669  0.610269549  0.624450875
```

```
# extracts second factor loadings from output
fact2 <- fact$loadings[,2]
fact2[order(fact2)]
```

```
##         age          C3          A5          E4   education          A4
## -0.11400561 -0.06464183 -0.05534240 -0.05377030 -0.05183733 -0.04312097
##          C1          E1          O1          C2          E5          O3
## -0.03612536 -0.01143654  0.02136659  0.03106549  0.04489666  0.05697494
```

```
##          A3          O5          A2          E3          A1      gender
##  0.06123484  0.06977837  0.07627858  0.07786475  0.07908426  0.15112230
##          O2          E2          O4          C4          C5          N5
##  0.17100877  0.19999005  0.23486142  0.27785843  0.31674075  0.55060360
##          N4          N2          N1          N3
##  0.57832710  0.74126736  0.74157263  0.76609169
```

For factor 1, it has distinct/separation of two meanings of the variables. The denoted negative loadings represent a person with unhappy, introvert person whereas the positive loadings represent happy, extrovert, confident person.
For factor 2, it has distinct/separation of two meanings of the variables. The negative loadings represents calm, peaceful, patient person while positive loadings represents a hostile and impatient person.

Q3

```
kout <- kmeans(bfi.df,centers=2,nstart=25)
kout$tot.withinss
```

```
## [1] 69224.89
```

```
centroids <- kout$centers
topvars_centroid1 <- centroids[1,order(centroids[1,])]
topvars_centroid2 <- centroids[2,order(centroids[2,])]
tail(topvars_centroid1)
```

```
##        A2        E5        A3        E3        E4        A5
## 0.3619903 0.3921397 0.4174190 0.4212761 0.4468109 0.4576639
```

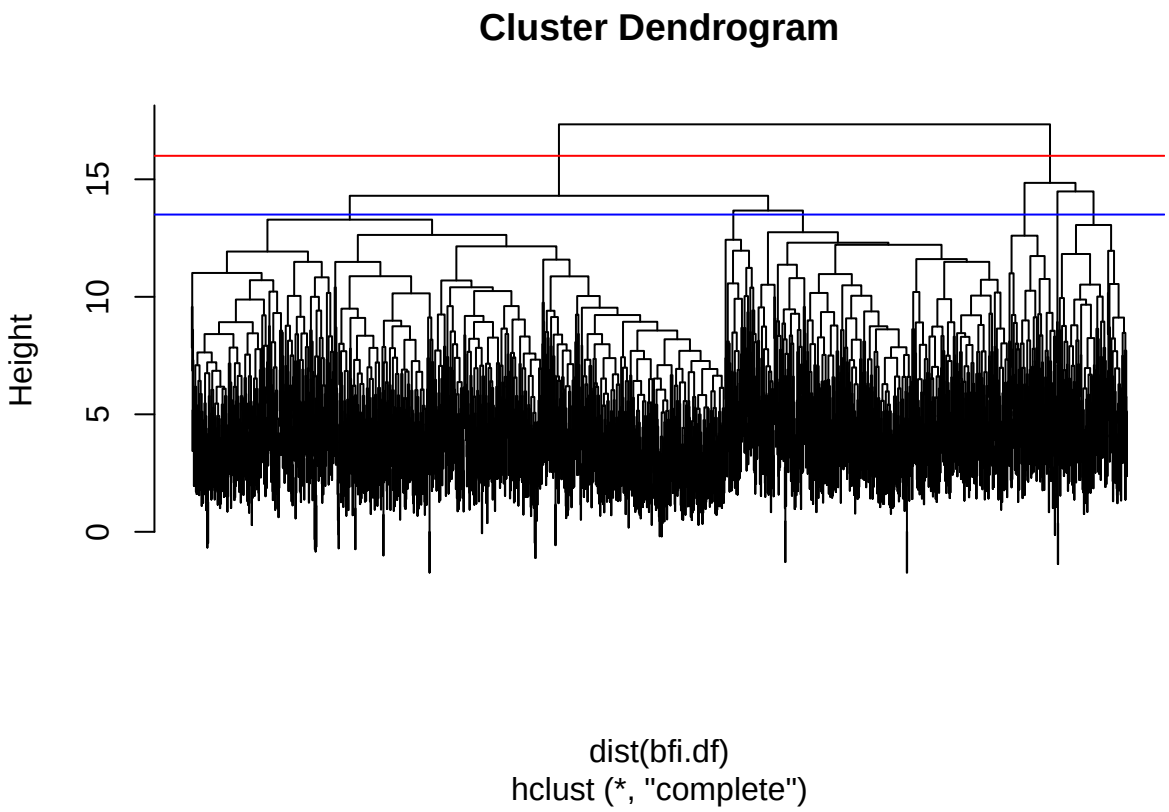
```
tail(topvars_centroid2)
```

```
##        E1        N1        C4        C5        N4        E2
## 0.3879479 0.3953162 0.4183931 0.4585234 0.4852927 0.5643027
```

The first cluster represents a positive, confident attitude; the second cluster represents angry, sad, negative feelings. There is an overlap between clusters and factors, but a obvious difference is that factors are dimensional and oppositional: there are two directions for every factor, and we often see clear oppositions at either end, such as unhappy-vs-happy (factor 1) and patient-vs-impatient (factor 2). Clusters are less oppositional: we care about variables that score highly that are near the cluster centroid.

Q4

```
hout <- hclust(dist(bfi.df), method="complete")
plot(hout, labels = F)
abline(a=16, b=0, col="red") # 2 clusters
abline(a=13.5, b=0, col="blue") # 6 clusters
```




```
cut <- as.vector(cutree(hout,k=2))
clust_means <- aggregate(bfi.df, by=list(cut), mean)
clust1 <- tail(unlist(sort(clust_means[1, -1])))
```

```
## Warning in xtfrm.data.frame(x): cannot xtfrm data frames
```

```
tail(topvars_centroid2)
```

```
##        E1        N1        C4        C5        N4        E2
## 0.3879479 0.3953162 0.4183931 0.4585234 0.4852927 0.5643027
```

```
clust2 <- tail(unlist(sort(clust_means[2, -1])))
```

```
## Warning in xtfrm.data.frame(x): cannot xtfrm data frames
```

```
tail(topvars_centroid1)
```

```
##        A2        E5        A3        E3        E4        A5
## 0.3619903 0.3921397 0.4174190 0.4212761 0.4468109 0.4576639
```

Although the high scoring of exact means and centers not same, but similar high scoring variables are observed using kmeans and hierachical clustering.

Q5
Conclusion:
- Factors have dimensional and oppositional in genral sense that each has 2 opposite directions
- In contrast, clusters exibit one direction and we emphasize variables that score highly. The 1st cluster represents angry, negative personality; the 2nd cluster represents happy, positive personality.