PCA-R

November 19, 2021

[ ]:
```
library("tidyverse")
library("FactoMineR")
library("factoextra")
library("FactoInvestigate")
library("corrplot")
```

[ ]:
```
df = read_csv('datasets/data.csv')
head(df)
```

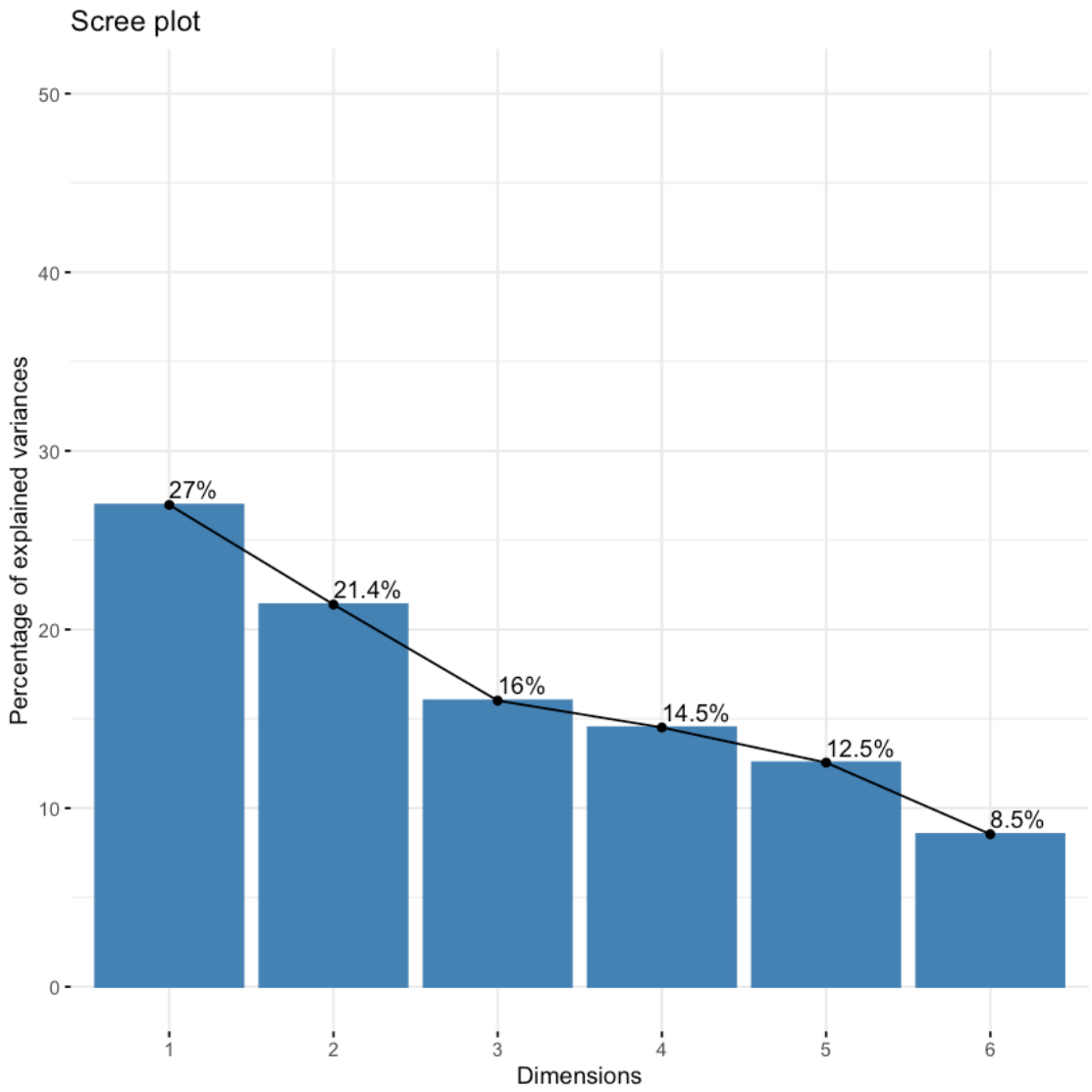
1 PCA

[3]:
```
df_pca = df %>% select('score', 'n_matches', 'n_updates_photo', 'n_photos',␣
↪'sent_ana', 'length_prof', 'account_age')

# Apply PCA
res.pca = PCA(df_pca, scale.unit=TRUE, quanti.sup=1)
```

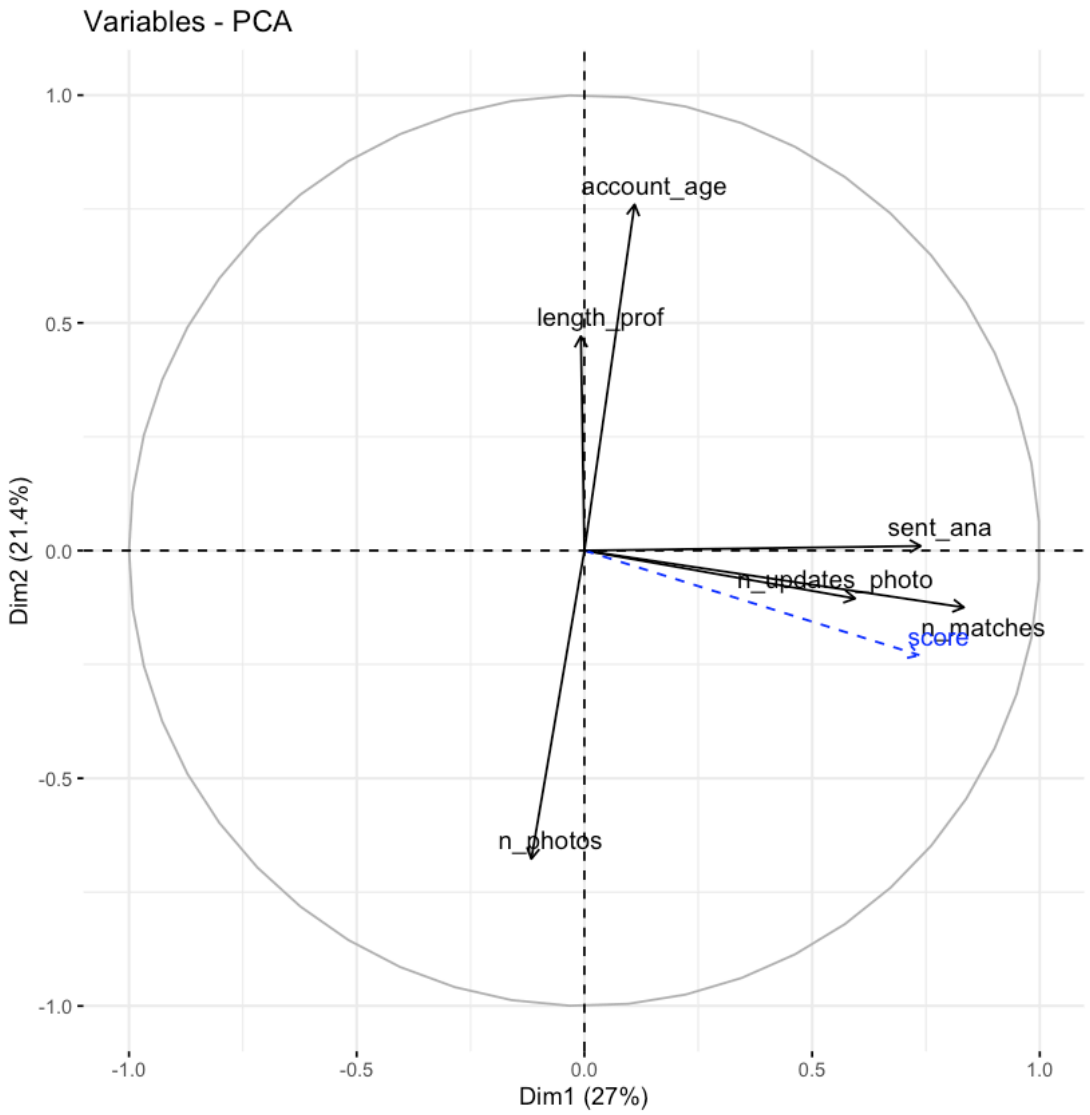
1.0.1 Scree plot

[4]:
```
fviz_eig(res.pca, addlabels = TRUE, ylim = c(0, 50), repel = TRUE)
```



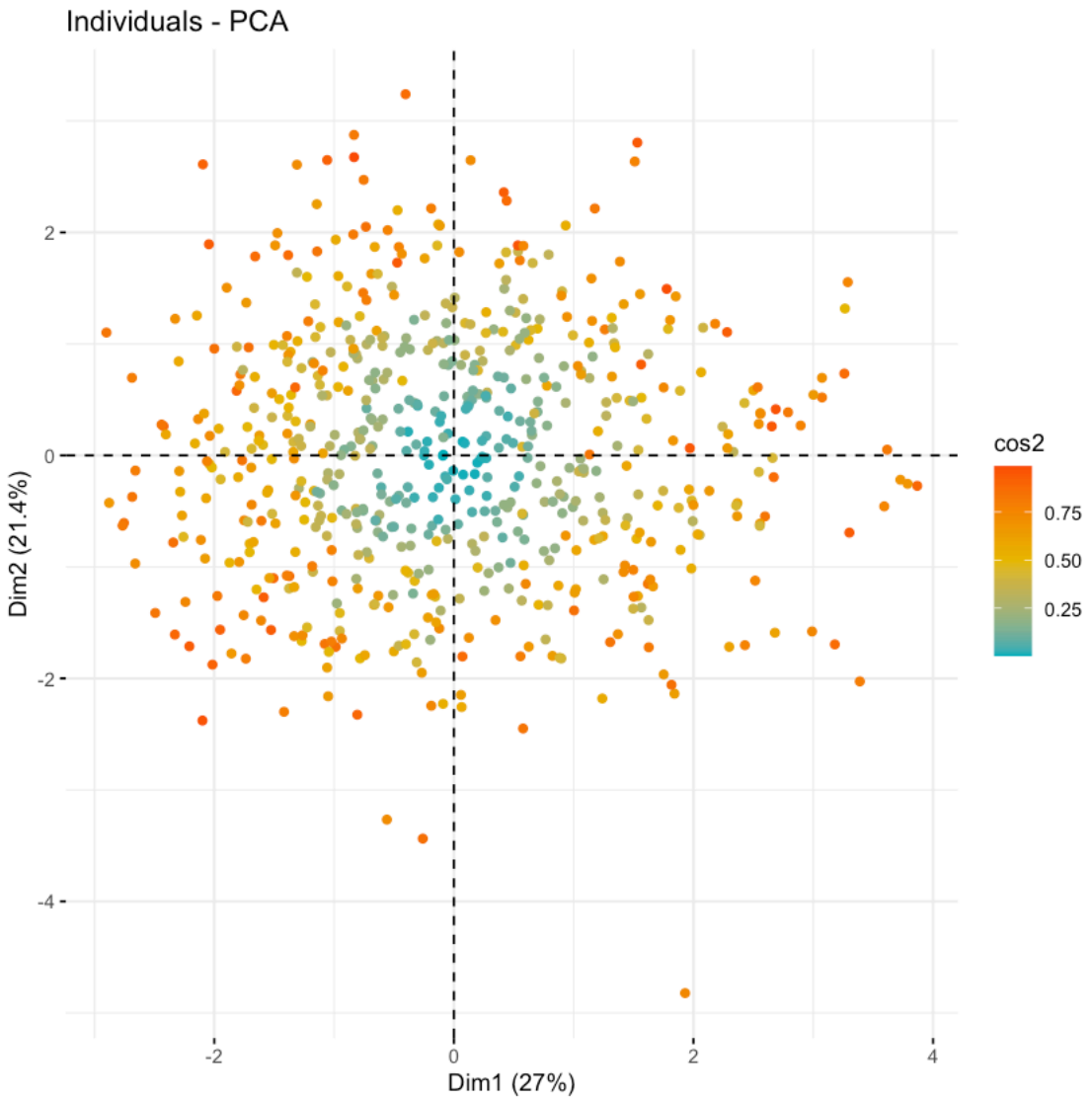
1.0.2 Correlation circle

[5]:
```
fviz_pca_var(res.pca, repel = TRUE)
```



1.0.3 Individual plot

[6]:
```
set.seed(123)
# Graph of individuals
fviz_pca_ind(res.pca, col.ind = "cos2",
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             label = "none",
             select.ind = list(name = sample(1:2931, 750, replace=TRUE)), # 25% sample
             )
```

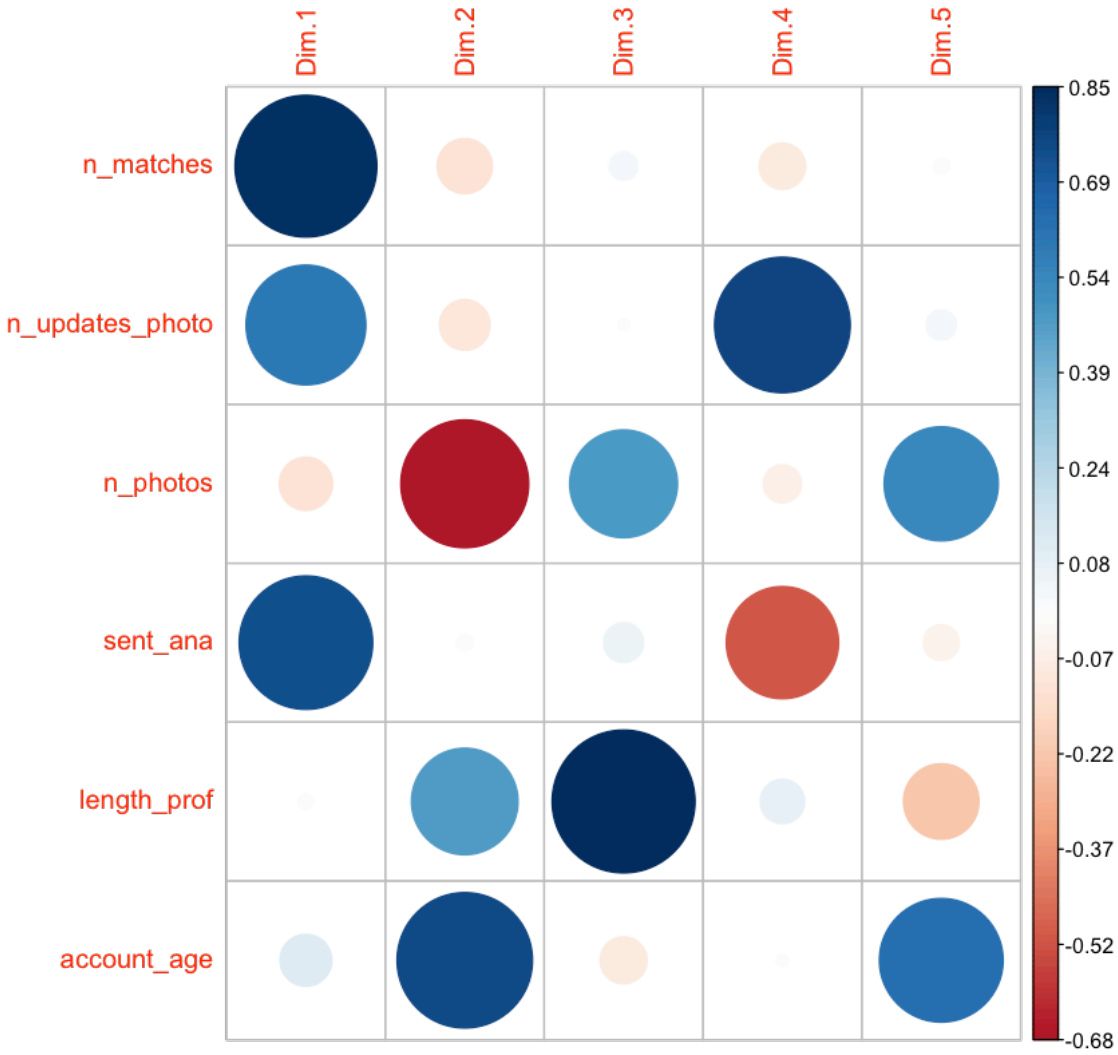


1.0.4 Biplot

[7]:
```
set.seed(123)
# Biplot of individuals and variables
fviz_pca_biplot(res.pca, label = "var",
                select.ind = list(name = sample(1:2931, 300, replace=TRUE)), #␣
 ↪10% sample
                repel = TRUE)
```

1.0.5 Loadings plot of each Principal Component

[8]:
```
corrplot(res.pca$var$coord, is.corr=FALSE)
```


Dim.1
Dim.2
Dim.3
Dim.4
Dim.5
n_matches
n_updates_photo
n_photos
sent_ana
length_prof
account_age
0.85
0.69
0.54
0.39
0.24
0.08
-0.07
-0.22
-0.37
-0.52
-0.68