

# Hw3-Part 1

Azat Dovgeldiyev

8/14/2020

## R Markdown

### Dependency

### pca only with numbers that have direct impact on coffee

```
pca_numbers <- coffee_clean[,c(17,18,19,20,21,22,23,24,25,26,28,  
29,30,32)]
```

```
KMO(pca_numbers)
```

```
## Kaiser-Meyer-Olkin factor adequacy  
## Call: KMO(r = pca_numbers)  
## Overall MSA = 0.93  
## MSA for each item =  
##          Aroma        Flavor      Aftertaste  
##          0.97        0.92        0.95  
##          Acidity      Body       Balance  
##          0.96        0.96        0.96  
##          Uniformity   Clean.Cup  Sweetness  
##          0.91        0.88        0.88  
##          Cupper.Points Moisture Category.One.Defects  
##          0.97        0.79        0.57  
##          Quakers Category.Two.Defects  
##          0.52        0.69
```

```
bart_spher(pca_numbers) #samples are not equal to zero, we can run model
```

```
## Bartlett's Test of Sphericity  
##  
## Call: bart_spher(x = pca_numbers)  
##  
##      X2 = 10390.438  
##      df = 91  
## p-value < 2.22e-16
```

```
alpha(pca_numbers, check.keys = TRUE) # test for reliability analysis using  
Cronbach's alpha
```

```
## Warning in alpha(pca_numbers, check.keys = TRUE): Some items were  
negatively correlated with total scale and were automatically reversed.  
## This is indicated by a negative sign for the variable name.
```

```

##  

## Reliability analysis  

## Call: alpha(x = pca_numbers, check.keys = TRUE)  

##  

##   raw_alpha std.alpha G6(smc) average_r S/N    ase mean    sd median_r  

##      0.47      0.88      0.91      0.34 7.2 0.017    16 0.56      0.34  

##  

##   lower alpha upper      95% confidence boundaries  

## 0.44 0.47 0.5  

##  

## Reliability if an item is dropped:  

##  

##          raw_alpha std.alpha G6(smc) average_r S/N alpha se  

var.r  

## Aroma           0.45     0.86     0.90     0.32 6.1  0.017  

0.084  

## Flavor          0.44     0.85     0.89     0.31 5.8  0.017  

0.078  

## Aftertaste      0.44     0.85     0.89     0.31 5.8  0.017  

0.079  

## Acidity         0.45     0.86     0.89     0.32 6.0  0.017  

0.083  

## Body            0.45     0.86     0.89     0.32 6.0  0.017  

0.083  

## Balance         0.44     0.86     0.89     0.31 5.9  0.017  

0.082  

## Uniformity      0.45     0.87     0.90     0.34 6.7  0.017  

0.096  

## Clean.Cup        0.43     0.87     0.90     0.34 6.7  0.018  

0.097  

## Sweetness        0.45     0.87     0.90     0.34 6.8  0.017  

0.094  

## Cupper.Points   0.44     0.86     0.89     0.32 6.0  0.017  

0.083  

## Moisture-        0.47     0.89     0.92     0.38 7.9  0.017  

0.090  

## Category.One.Defects- 0.40     0.89     0.92     0.38 8.0  0.020  

0.088  

## Quakers          0.48     0.89     0.92     0.39 8.5  0.017  

0.080  

## Category.Two.Defects- 0.71     0.88     0.91     0.37 7.7  0.014  

0.093  

##  

##          med.r  

## Aroma            0.21  

## Flavor           0.21  

## Aftertaste       0.20  

## Acidity          0.21  

## Body             0.21  

## Balance          0.20  

## Uniformity       0.21  

## Clean.Cup         0.20

```

```

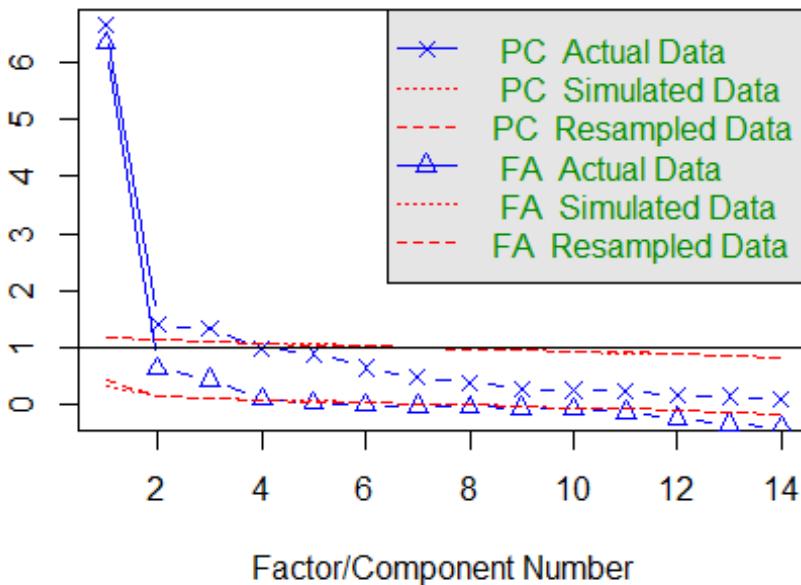
## Sweetness          0.21
## Cupper.Points    0.21
## Moisture-         0.41
## Category.One.Defects- 0.41
## Quakers           0.41
## Category.Two.Defects- 0.41
##
## Item statistics
##                               n raw.r std.r r.cor  r.drop  mean   sd
## Aroma                 1083 0.496  0.81  0.82  0.4579  7.57 0.386
## Flavor                1083 0.555  0.89  0.92  0.5179  7.52 0.403
## Aftertaste             1083 0.576  0.89  0.92  0.5400  7.39 0.407
## Acidity               1083 0.491  0.82  0.83  0.4521  7.52 0.388
## Body                  1083 0.457  0.82  0.83  0.4199  7.50 0.361
## Balance                1083 0.535  0.86  0.87  0.4959  7.50 0.417
## Uniformity             1083 0.400  0.62  0.59  0.3413  9.87 0.526
## Clean.Cup              1083 0.489  0.60  0.56  0.4063  9.85 0.792
## Sweetness              1083 0.334  0.59  0.55  0.2739  9.92 0.511
## Cupper.Points          1083 0.521  0.83  0.84  0.4751  7.48 0.471
## Moisture-              1083 0.186  0.26  0.17  0.1802  46.91 0.045
## Category.One.Defects- 1083 0.531  0.24  0.15  0.3263  46.62 1.870
## Quakers                1083 0.087  0.12  0.01 -0.0069  0.14 0.732
## Category.Two.Defects- 1083 0.852  0.34  0.26  0.3159  43.42 5.314

comp <- fa.parallel(pca_numbers)

```

eigenvalues of principal components and factor analysis

### Parallel Analysis Scree Plots



```

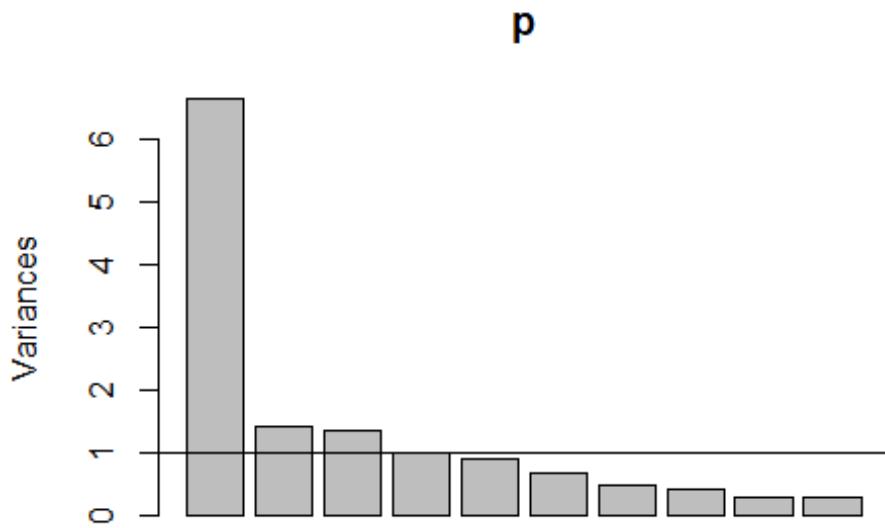
## Parallel analysis suggests that the number of factors = 3 and the number
## of components = 3

comp

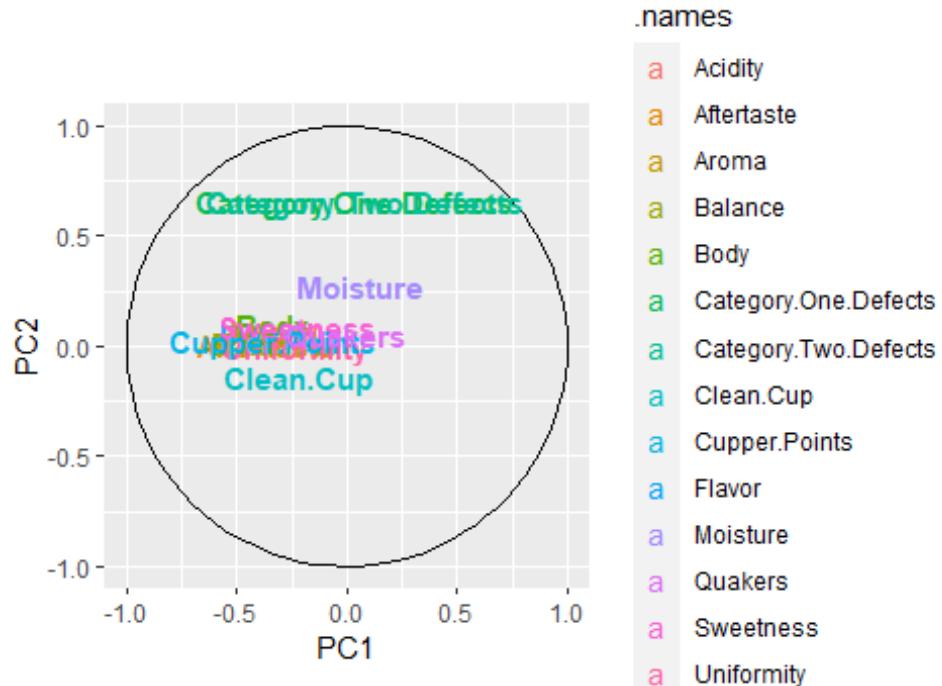
## Call: fa.parallel(x = pca_numbers)
## Parallel analysis suggests that the number of factors = 3 and the number
## of components = 3
##
## Eigen Values of
##   Original factors Resampled data Simulated data Original components
## 1      6.34          0.45       0.33        6.63
## 2      0.64          0.16       0.16        1.40
## 3      0.45          0.12       0.13        1.33
##   Resampled components Simulated components
## 1            1.19         1.19
## 2            1.14         1.15
## 3            1.11         1.11

p <- prcomp(pca_numbers, center = T, scale = T)
plot(p)
abline(1,0)

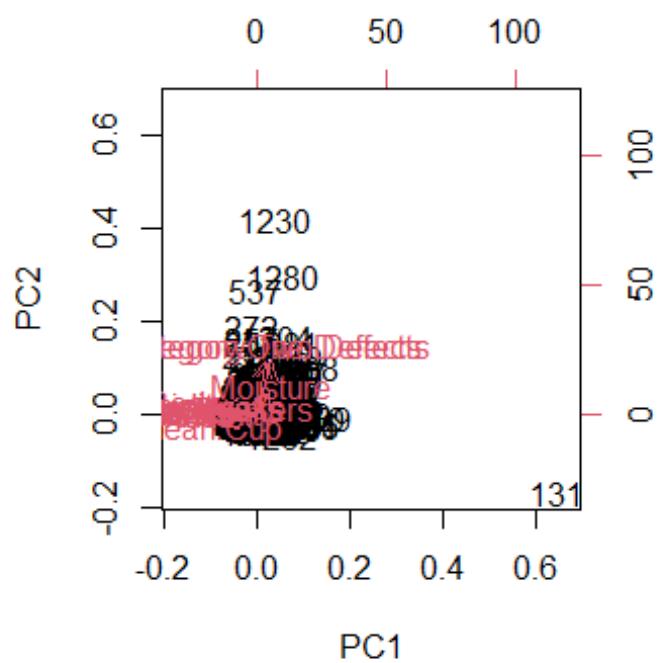
```



```
PCA_Plot(p)
```



**biplot(p)**



```

p2 = psych::principal(pca_numbers, rotate="varimax", nfactors=3, scores=TRUE)
p2

## Principal Components Analysis
## Call: psych::principal(r = pca_numbers, nfactors = 3, rotate = "varimax",
##   scores = TRUE)
## Standardized loadings (pattern matrix) based upon correlation matrix
##          RC1    RC3    RC2     h2    u2 com
## Aroma      0.86  0.14 -0.06  0.755  0.25 1.1
## Flavor     0.93  0.15 -0.08  0.894  0.11 1.1
## Aftertaste 0.92  0.15 -0.13  0.877  0.12 1.1
## Acidity    0.88  0.13 -0.05  0.784  0.22 1.0
## Body       0.89  0.10 -0.01  0.795  0.20 1.0
## Balance    0.89  0.12 -0.11  0.822  0.18 1.1
## Uniformity 0.41  0.69 -0.12  0.652  0.35 1.7
## Clean.Cup   0.33  0.68 -0.28  0.652  0.35 1.8
## Sweetness   0.38  0.75 -0.02  0.706  0.29 1.5
## Cupper.Points 0.88  0.09 -0.10  0.787  0.21 1.0
## Moisture    -0.25  0.37  0.33  0.307  0.69 2.7
## Category.One.Defects 0.04 -0.07  0.79  0.628  0.37 1.0
## Quakers    -0.03  0.17  0.06  0.033  0.97 1.3
## Category.Two.Defects -0.12  0.05  0.81  0.668  0.33 1.1
##
##          RC1    RC3    RC2
## SS loadings 6.05 1.78 1.53
## Proportion Var 0.43 0.13 0.11
## Cumulative Var 0.43 0.56 0.67
## Proportion Explained 0.65 0.19 0.16
## Cumulative Proportion 0.65 0.84 1.00
##
## Mean item complexity = 1.3
## Test of the hypothesis that 3 components are sufficient.
##
## The root mean square of the residuals (RMSR) is 0.06
## with the empirical chi square 690.58 with prob < 2.2e-112
##
## Fit based upon off diagonal values = 0.98

print(p2$loadings, cutoff=.4, sort=T)

##
## Loadings:
##          RC1    RC3    RC2
## Aroma      0.855
## Flavor     0.929
## Aftertaste 0.916
## Acidity    0.875
## Body       0.887
## Balance    0.892
## Cupper.Points 0.877

```

```

## Uniformity          0.409  0.686
## Clean.Cup           0.680
## Sweetness            0.752
## Category.One.Defects      0.788
## Category.Two.Defects      0.807
## Moisture
## Quakers
##
##                  RC1    RC3    RC2
## SS loadings     6.049  1.783  1.528
## Proportion Var  0.432  0.127  0.109
## Cumulative Var 0.432  0.559  0.669

p2$values

##  [1] 6.63455442 1.39818012 1.32752646 0.99268189 0.89937953 0.65368845
##  [7] 0.47761388 0.39917806 0.28474718 0.26009117 0.24347682 0.18154655
## [13] 0.15465657 0.09267889



```

```

## Aftertaste          0.897
## Acidity            0.844
## Body               0.837
## Balance            0.844
## Copper.Points      0.840
## Uniformity         0.623
## Clean.Cup          0.604
## Sweetness          0.760
## Category.Two.Defects 0.766
## Moisture
## Category.One.Defects 0.466
## Quakers
##
##             Factor1 Factor2 Factor3
## SS loadings     5.497   1.654   0.979
## Proportion Var  0.393   0.118   0.070
## Cumulative Var 0.393   0.511   0.581

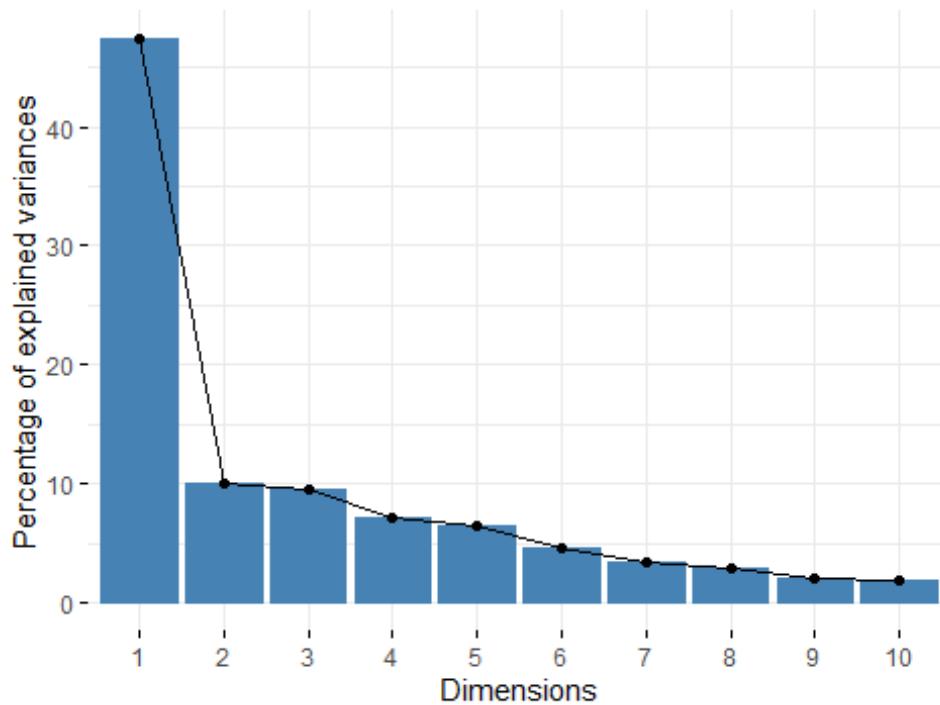
summary(fit)

##           Length Class    Mode
## converged     1   -none- logical
## loadings      42   loadings numeric
## uniquenesses  14   -none- numeric
## correlation   196  -none- numeric
## criteria      3   -none- numeric
## factors       1   -none- numeric
## dof            1   -none- numeric
## method         1   -none- character
## rotmat        9   -none- numeric
## STATISTIC     1   -none- numeric
## PVAL           1   -none- numeric
## n.obs          1   -none- numeric
## call           3   -none- call

p3 <- prcomp(pca_numbers, scale = TRUE)
fviz_eig(p3)

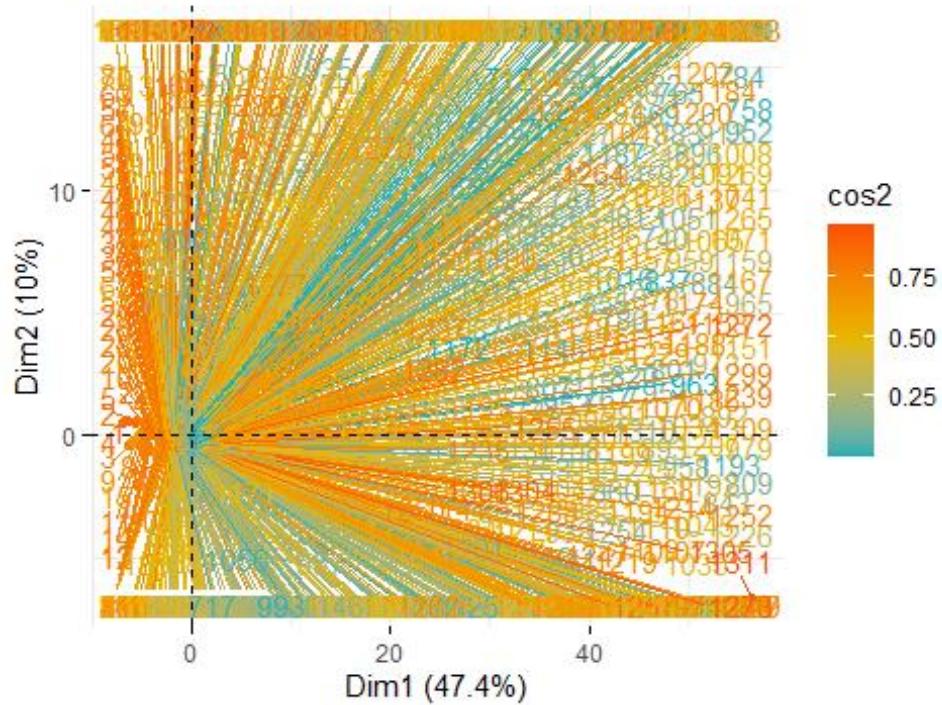
```

Scree plot

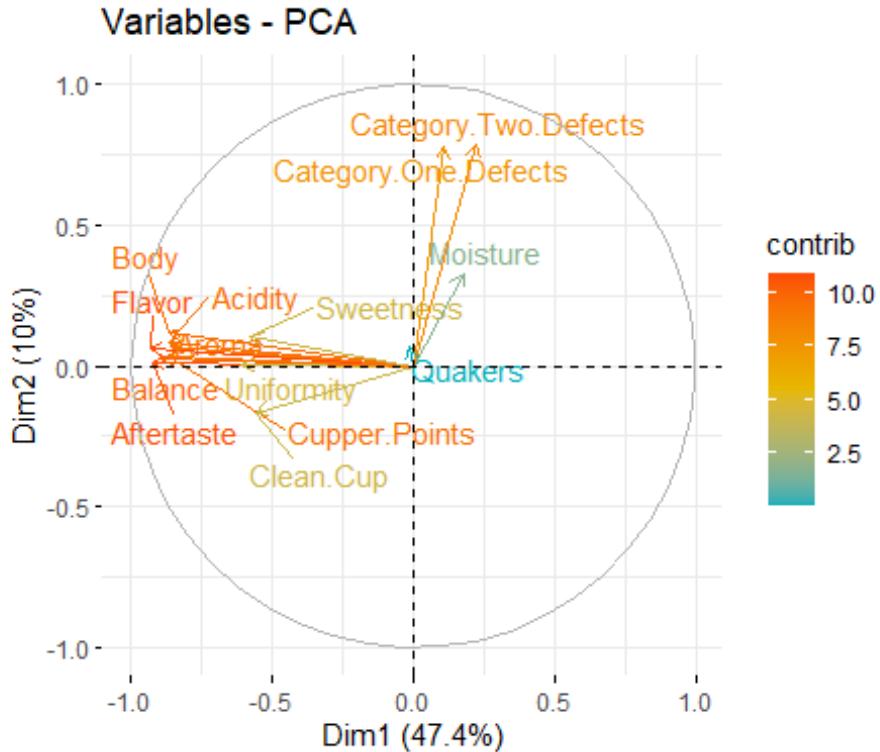


```
#PCA Individuals
pI<-fviz_pca_ind(p3,
  col.ind = "cos2", # Color by the quality of representation
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
  repel = TRUE      # Avoid text overlapping
)
pI
```

## Individuals - PCA



```
pca_var<-fviz_pca_var(p3,
                        col.var = "contrib", # Color by contributions to the PC
                        gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
                        repel = TRUE      # Avoid text overlapping
)
pca_var
```

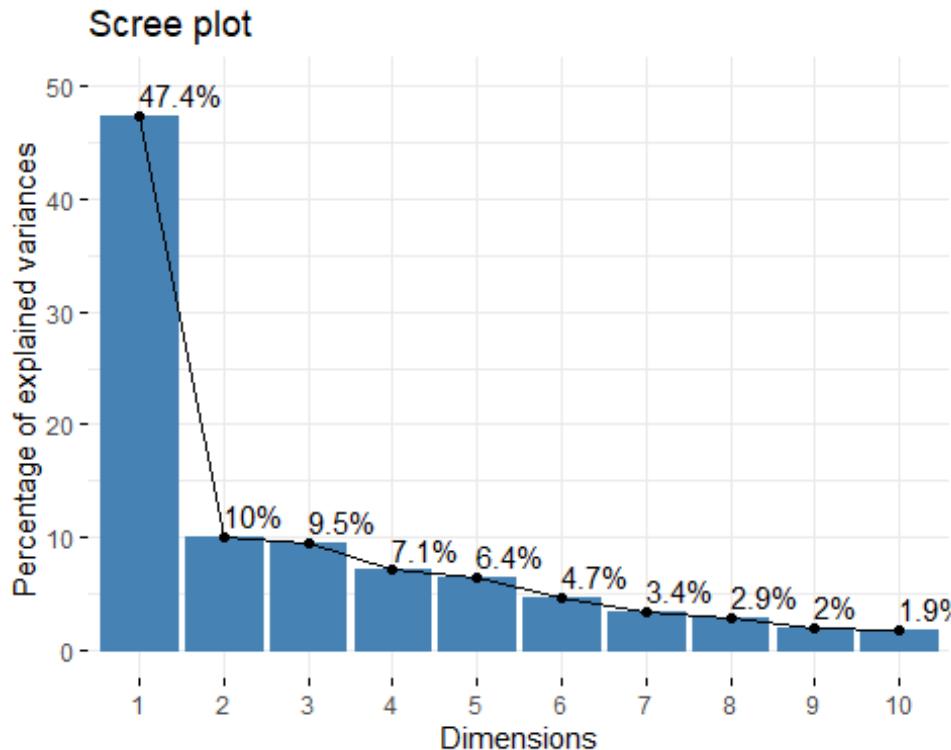


```
p4 <- PCA(pca_numbers, graph = FALSE)
#IF graph is set to true, it will provide the individual and variable maps

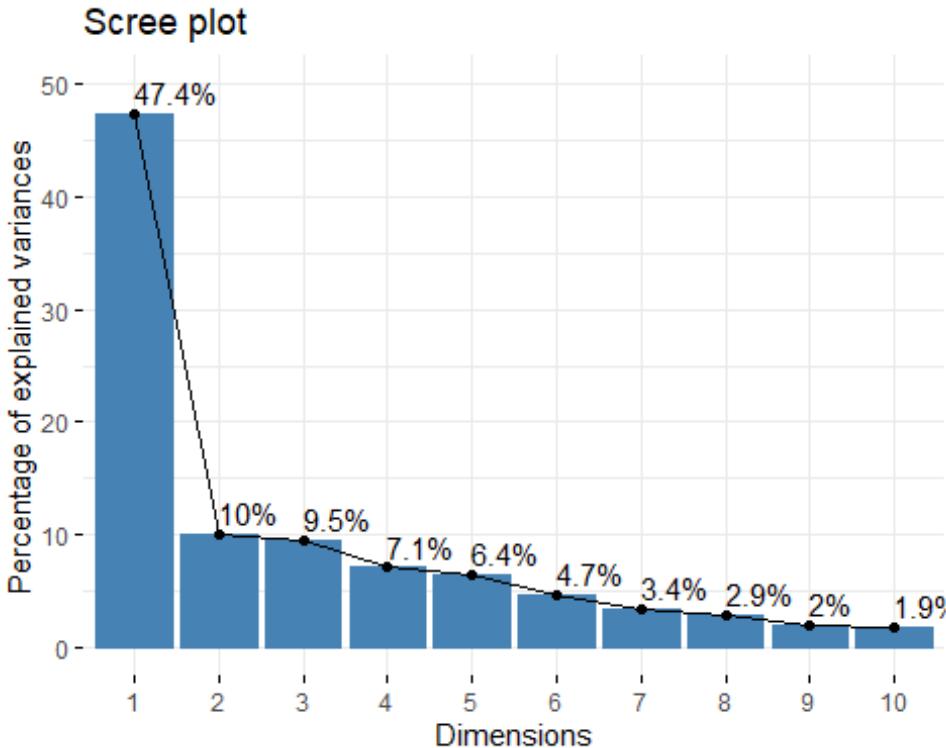
#Shows all the objects or functions available in PCA
print(p4)

## **Results for the Principal Component Analysis (PCA)**
## The analysis was performed on 1083 individuals, described by 14 variables
## *The results are available in the following objects:
## 
##      name           description
## 1  "$eig"         "eigenvalues"
## 2  "$var"          "results for the variables"
## 3  "$var$coord"   "coord. for the variables"
## 4  "$var$cor"      "correlations variables - dimensions"
## 5  "$var$cos2"     "cos2 for the variables"
## 6  "$var$contrib"  "contributions of the variables"
## 7  "$ind"          "results for the individuals"
## 8  "$ind$coord"   "coord. for the individuals"
## 9  "$ind$cos2"     "cos2 for the individuals"
## 10 "$ind$contrib"  "contributions of the individuals"
## 11 "$call"          "summary statistics"
## 12 "$call$centre"   "mean of the variables"
## 13 "$call$ecart.type" "standard error of the variables"
## 14 "$call$row.w"    "weights for the individuals"
## 15 "$call$col.w"    "weights for the variables"
```

```
#Options for providing screeplot  
fviz_eig(p4, addlabels = TRUE, ylim = c(0, 50))
```



```
fviz_screeplot(p4, addlabels = TRUE, ylim = c(0, 50))
```



```

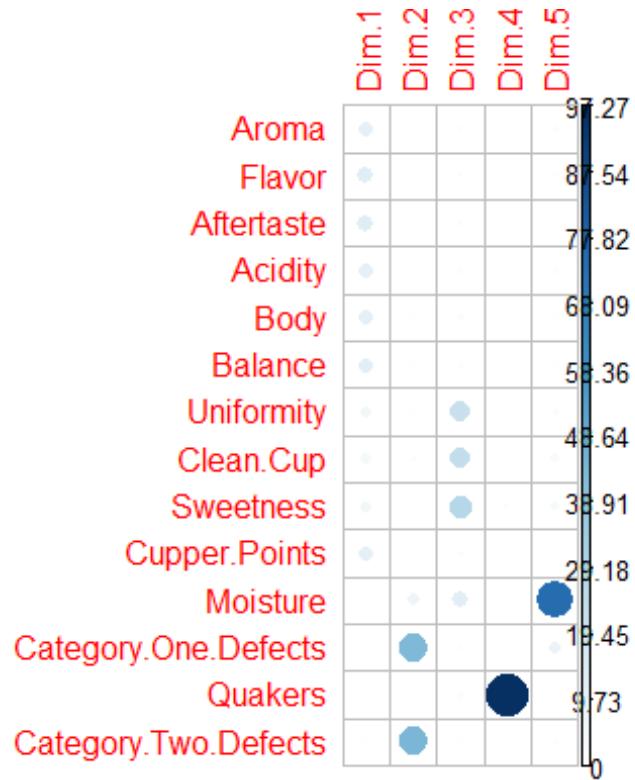
variables <- get_pca_var(p4)

head(variables$contrib,11)

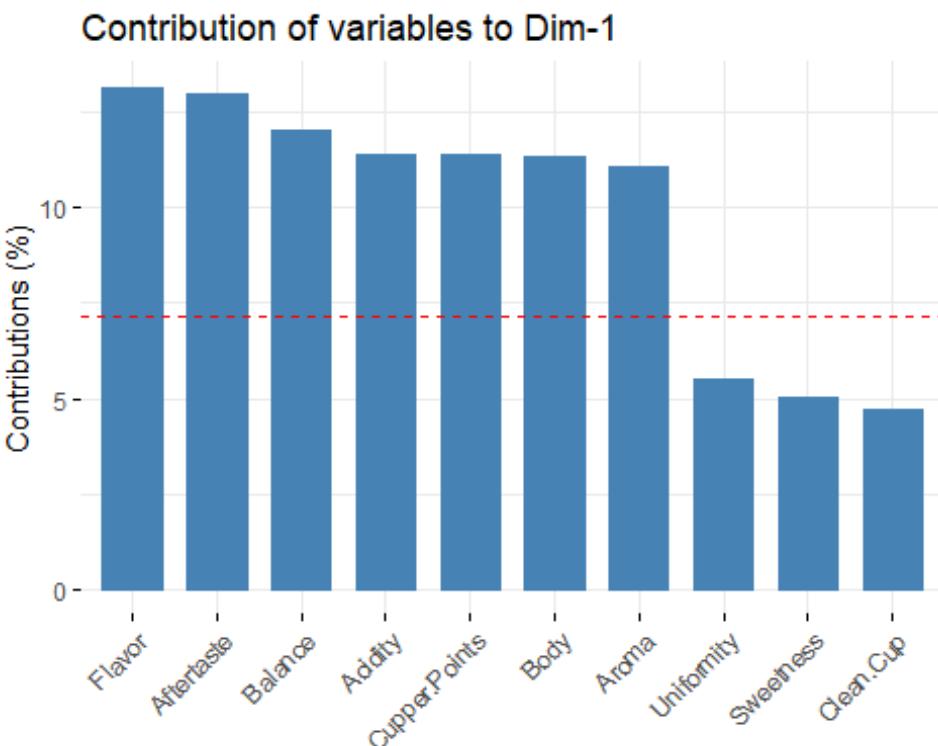
##           Dim.1      Dim.2      Dim.3      Dim.4      Dim.5
## Aroma     11.0662962 0.3191173091 1.199326 0.0503602102 2.230690044
## Flavor    13.1602789 0.2674716786 1.287451 0.0398045786 1.038522110
## Aftertaste 12.9606661 0.0143219958 1.279602 0.0487233735 0.411873539
## Acidity   11.4021321 0.5049620630 1.559561 0.0004357203 1.697706201
## Body      11.3167561 0.9506637739 2.358742 0.0072862443 0.002656628
## Balance   12.0321165 0.0410523890 1.735613 0.0232108950 0.031456445
## Uniformity 5.5273752 0.0003939713 21.485364 0.3177232130 2.359408193
## Clean.Cup  4.7025595 1.9740235130 23.554992 0.4020725830 3.587533322
## Sweetness  5.0184019 0.7591716628 27.325755 1.3109402052 4.274469584
## Cupper.Points 11.3923442 0.0777588252 2.247673 0.0833870105 0.098647001
## Moisture   0.4768461 7.7288403178 12.598053 0.1725573247 74.544677061

corrplot(variables$contrib, is.corr = FALSE)

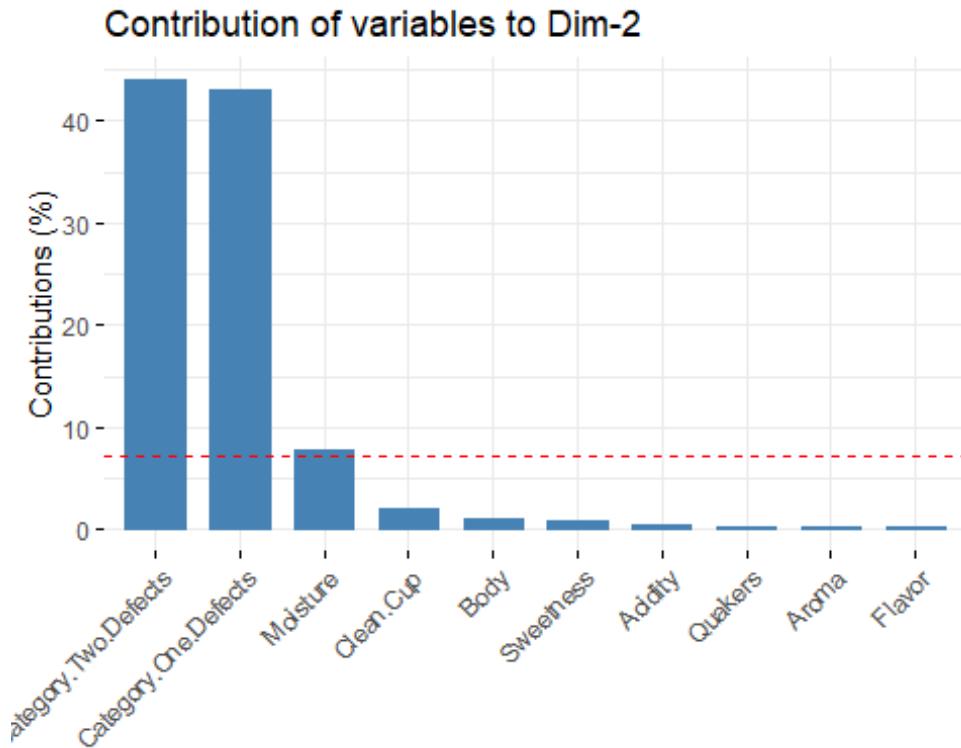
```



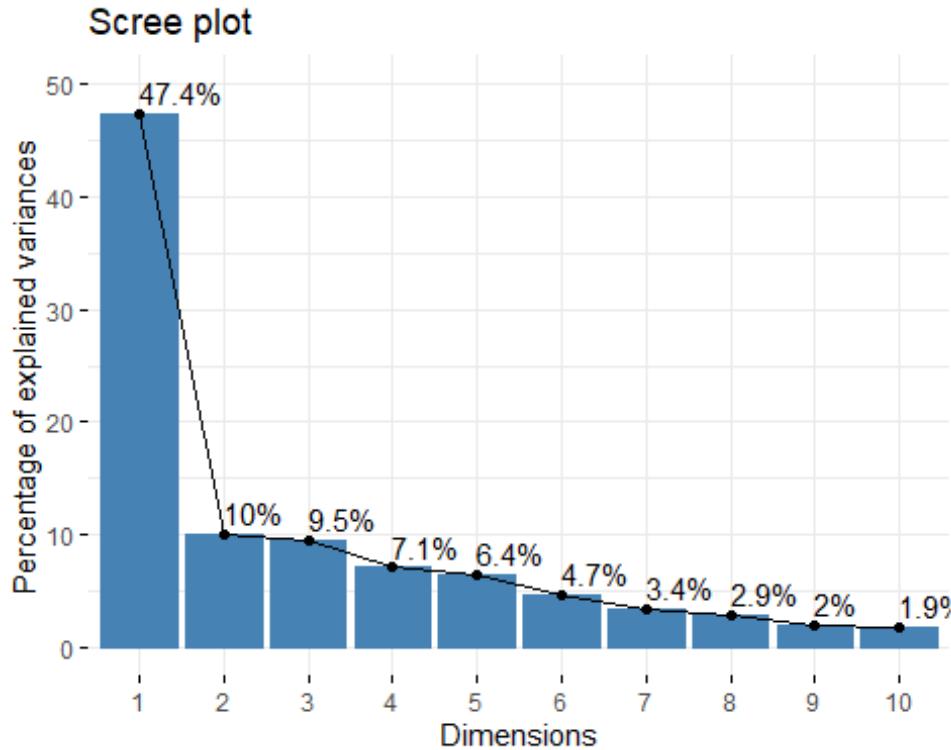
```
# Contributions of variables to PC1
fviz_contrib(p4, choice = "var", axes = 1, top = 10)
```



```
# Contributions of variables to PC2  
fviz_contrib(p4, choice = "var", axes = 2, top = 10)
```



```
p5 <- dudi.pca(pca_numbers,  
                  scannf = FALSE,    # Hide scree plot  
                  nf = 3            # Number of components kept in the results  
)  
fviz_screeplot(p5, addlabels = TRUE, ylim = c(0, 50))
```



```

variables2 <- get_pca_var(p5)
#Which variables contribute the most to the PCs?
#there are 11 variables
head(variables2$contrib, 11)

##           Dim.1      Dim.2      Dim.3
## Aroma     11.0662962 0.3191173091 1.199326
## Flavor    13.1602789 0.2674716786 1.287451
## Aftertaste 12.9606661 0.0143219958 1.279602
## Acidity   11.4021321 0.5049620630 1.559561
## Body      11.3167561 0.9506637739 2.358742
## Balance   12.0321165 0.0410523890 1.735613
## Uniformity 5.5273752 0.0003939713 21.485364
## Clean.Cup  4.7025595 1.9740235130 23.554992
## Sweetness  5.0184019 0.7591716628 27.325755
## Copper.Points 11.3923442 0.0777588252 2.247673
## Moisture   0.4768461 7.7288403178 12.598053

corrplot(variables2$contrib, is.corr = FALSE)

```

