

## Hw3-Part 1

Azat Dovgeldiyev

8/14/2020

### R Markdown

#### Dependency

```
library(car)
```

```
## Loading required package: carData
```

```
library(DAAG)
```

```
## Loading required package: lattice
```

```
##
```

```
## Attaching package: 'DAAG'
```

```
## The following object is masked from 'package:car':
```

```
##
```

```
##      vif
```

```
library(MASS)
```

```
##
```

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

```
## The following object is masked from 'package:DAAG':
```

```
##
```

```
##      hills
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.2
```

```
## -- Attaching packages -----  
----- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2      v purrr   0.3.4
```

```
## v tibble  3.0.1      v dplyr   0.8.5
```

```
## v tidyr   1.1.0      v stringr 1.4.0
```

```
## v readr   1.3.1      v forcats 0.5.0
```

```
## Warning: package 'ggplot2' was built under R version 4.0.2
```

```
## Warning: package 'tidyr' was built under R version 4.0.2
```

```
## Warning: package 'stringr' was built under R version 4.0.2
```

```

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()
## x dplyr::recode() masks car::recode()
## x dplyr::select() masks MASS::select()
## x purrr::some() masks car::some()

library(clusterGeneration)

## Warning: package 'clusterGeneration' was built under R version 4.0.2

library(corrplot)

## Warning: package 'corrplot' was built under R version 4.0.2

## corrplot 0.84 loaded

library(Hmisc) #Describe Function

## Warning: package 'Hmisc' was built under R version 4.0.2

## Loading required package: survival

##
## Attaching package: 'survival'

## The following object is masked from 'package:DAAG':
##
## lung

## Loading required package: Formula

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:dplyr':
##
## src, summarize

## The following objects are masked from 'package:base':
##
## format.pval, units

library(psych) #Multiple Functions for Statistics and Multivariate Analysis

## Warning: package 'psych' was built under R version 4.0.2

##
## Attaching package: 'psych'

## The following object is masked from 'package:Hmisc':
##
## describe

```

```
## The following objects are masked from 'package:ggplot2':
##
##      %+%, alpha

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

library(GGally) #ggpairs Function

## Warning: package 'GGally' was built under R version 4.0.2

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

##
## Attaching package: 'GGally'

## The following object is masked from 'package:dplyr':
##
##      nasa

library(ggplot2) #ggplot2 Functions
library(vioplot) #Violin Plot Function

## Warning: package 'vioplot' was built under R version 4.0.2

## Loading required package: sm

## Warning: package 'sm' was built under R version 4.0.2

## Package 'sm', version 2.2-5.6: type help(sm) for summary information

##
## Attaching package: 'sm'

## The following object is masked from 'package:MASS':
##
##      muscle

## The following object is masked from 'package:DAAG':
##
##      pause

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 4.0.2

##
## Attaching package: 'zoo'
```

```

## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric

library(corrplot) #Plot Correlations
library(REdaS) #Bartlett's Test of Sphericity

## Warning: package 'REdaS' was built under R version 4.0.2

## Loading required package: grid

library(psych) #PCA/FA functions
library(factoextra) #PCA Visualizations

## Warning: package 'factoextra' was built under R version 4.0.2

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

library("FactoMineR") #PCA functions

## Warning: package 'FactoMineR' was built under R version 4.0.2

library(ade4) #PCA Visualizations

## Warning: package 'ade4' was built under R version 4.0.2

##
## Attaching package: 'ade4'

## The following object is masked from 'package:FactoMineR':
##
##      reconst

coffee_data <- read.csv(file = "coffee_dataset.csv", header = TRUE, sep = ",")
head(coffee_data)

##      X Species      Owner Country.of.Origin
## 1 0 Arabica      metad plc      Ethiopia
## 2 1 Arabica      metad plc      Ethiopia
## 3 2 Arabica grounds for health admin      Guatemala
## 4 3 Arabica      yidnekachew dabessa      Ethiopia
## 5 4 Arabica      metad plc      Ethiopia
## 6 5 Arabica      ji-ae ahn      Brazil
##
##      Farm.Name Lot.Number      Mill ICO.Number
## 1      metad plc      metad plc 2014/2015
## 2      metad plc      metad plc 2014/2015
## 3 san marcos barrancas "san cristobal cuch
## 4      yidnekachew dabessa coffee plantation      wolensu
## 5      metad plc      metad plc 2014/2015
## 6
##
##      Company      Altitude      Region
## 1      metad agricultural developmet plc 1950-2200 guji-hambela

```

## 2	metad agricultural developmet plc	1950-2200	guji-hambela
## 3		1600 - 1800 m	
## 4	yidnekachew debessa coffee plantation	1800-2200	oromia
## 5	metad agricultural developmet plc	1950-2200	guji-hambela
## 6			
##	Producer	Number.of.Bags	Bag.Weight
## 1	METAD PLC	300	60 kg
## 2	METAD PLC	300	60 kg
## 3		5	1
## 4	Yidnekachew Dabessa Coffee Plantation	320	60 kg
## 5	METAD PLC	300	60 kg
## 6		100	30 kg
##	In.Country.Partner	Harvest.Year	Grading.Date
## 1	METAD Agricultural Development plc	2014	April 4th, 2015
## 2	METAD Agricultural Development plc	2014	April 4th, 2015
## 3	Specialty Coffee Association		May 31st, 2010
## 4	METAD Agricultural Development plc	2014	March 26th, 2015
## 5	METAD Agricultural Development plc	2014	April 4th, 2015
## 6	Specialty Coffee Institute of Asia	2013	September 3rd, 2013
##	Owner.1	Variety	Processing.Method Aroma Flavor Aftertaste
## 1	metad plc	Washed / Wet	8.67 8.83 8.67
## 2	metad plc Other	Washed / Wet	8.75 8.67 8.50
## 3	Grounds for Health Admin Bourbon		8.42 8.50 8.42
## 4	Yidnekachew Dabessa	Natural / Dry	8.17 8.58 8.42
## 5	metad plc Other	Washed / Wet	8.25 8.50 8.25
## 6	Ji-Ae Ahn	Natural / Dry	8.58 8.42 8.42
##	Acidity	Body	Balance Uniformity Clean.Cup Sweetness Cupper.Points
## 1	8.75	8.50	8.42 10 10 10 8.75
## 2	8.58	8.42	8.42 10 10 10 8.58
## 3	8.42	8.33	8.42 10 10 10 9.25
## 4	8.42	8.50	8.25 10 10 10 8.67
## 5	8.50	8.42	8.33 10 10 10 8.58
## 6	8.50	8.25	8.33 10 10 10 8.33
##	Total.Cup.Points	Moisture	Category.One.Defects Quakers Color
## 1	90.58	0.12	0 0 Green
## 2	89.92	0.12	0 0 Green
## 3	89.75	0.00	0 0
## 4	89.00	0.11	0 0 Green
## 5	88.83	0.12	0 0 Green
## 6	88.83	0.11	0 0 Bluish-Green
##	Category.Two.Defects	Expiration	Certification.Body
## 1	0	April 3rd, 2016	METAD Agricultural Development

```

plc
## 2          1      April 3rd, 2016 METAD Agricultural Development
plc
## 3          0      May 31st, 2011      Specialty Coffee Associat
ion
## 4          2      March 25th, 2016 METAD Agricultural Development
plc
## 5          2      April 3rd, 2016 METAD Agricultural Development
plc
## 6          1 September 3rd, 2014 Specialty Coffee Institute of A
sia
##          Certification.Address
## 1 309fcf77415a3661ae83e027f7e5f05dad786e44
## 2 309fcf77415a3661ae83e027f7e5f05dad786e44
## 3 36d0d00a3724338ba7937c52a378d085f2172daa
## 4 309fcf77415a3661ae83e027f7e5f05dad786e44
## 5 309fcf77415a3661ae83e027f7e5f05dad786e44
## 6 726e4891cf2c9a4848768bd34b668124d12c4224
##          Certification.Contact unit_of_measurement
## 1 19fef5a731de2db57d16da10287413f5f99bc2dd      m
## 2 19fef5a731de2db57d16da10287413f5f99bc2dd      m
## 3 0878a7d4b9d35ddb0fe2ce69a2062cceb45a660      m
## 4 19fef5a731de2db57d16da10287413f5f99bc2dd      m
## 5 19fef5a731de2db57d16da10287413f5f99bc2dd      m
## 6 b70da261fcc84831e3e9620c30a8701540abc200      m
## altitude_low_meters altitude_high_meters altitude_mean_meters
## 1          1950          2200          2075
## 2          1950          2200          2075
## 3          1600          1800          1700
## 4          1800          2200          2000
## 5          1950          2200          2075
## 6          NA          NA          NA

dim(coffee_data)

## [1] 1339  44

sum(is.na(coffee_data))

## [1] 691

coffee_clean<- na.omit(coffee_data)
dim(coffee_clean)

## [1] 1108  44

head(coffee_clean)

##   X Species          Owner Country.of.Origin
## 1 0 Arabica          metad plc      Ethiopia
## 2 1 Arabica          metad plc      Ethiopia
## 3 2 Arabica grounds for health admin  Guatemala

```

## 4	3 Arabica	yidnekachew dabessa	Ethiopia
## 5	4 Arabica	metad plc	Ethiopia
## 8	7 Arabica	ethiopia commodity exchange	Ethiopia
##		Farm.Name	Lot.Number
## 1		metad plc	metad plc
## 2		metad plc	metad plc
## 3	san marcos barrancas "san cristobal cuch		
## 4	yidnekachew dabessa coffee plantation		wolensu
## 5		metad plc	metad plc
## 8		aolme	c.p.w.e
##		Company	Altitude
## 1		metad agricultural developmet plc	1950-2200
## 2		metad agricultural developmet plc	1950-2200
## 3			1600 - 1800 m
## 4		yidnekachew debessa coffee plantation	1800-2200
## 5		metad agricultural developmet plc	1950-2200
## 8			1570-1700
##		Producer	Number.of.Bags
## 1		METAD PLC	300
## 2		METAD PLC	300
## 3			5
## 4		Yidnekachew Dabessa Coffee Plantation	320
## 5		METAD PLC	300
## 8		Bazen Agricultural & Industrial Dev't Plc	300
##		In.Country.Partner	Harvest.Year
## 1		METAD Agricultural Development plc	2014
## 2		METAD Agricultural Development plc	2014
## 3		Specialty Coffee Association	May 31st, 2010
## 4		METAD Agricultural Development plc	2014
## 5		METAD Agricultural Development plc	2014
## 8		Ethiopia Commodity Exchange	March 2010
##		Owner.1 Variety	Processing.Method
## 1		metad plc	Washed / Wet
## 2		metad plc	Other
## 3		Grounds for Health Admin Bourbon	
## 4		Yidnekachew Dabessa	Natural / Dry
## 5		metad plc	Other
## 8		Ethiopia Commodity Exchange	
##		Acidity	Body
## 1		8.75	8.50
## 2		8.58	8.42
## 3		8.42	8.33
## 4		8.42	8.50
##		Balance	Uniformity
## 1		8.42	10
## 2		8.42	10
## 3		8.42	10
## 4		8.25	10
##		Clean.Cup	Sweetness
## 1		10	10.00
## 2		10	10.00
## 3		10	10.00
## 4		10	10.00
##		Cupper.Points	
## 1		8.75	
## 2		8.58	
## 3		9.25	
## 4		8.67	

```

## 5      8.50 8.42      8.33      10      10      10.00      8.58
## 8      8.42 8.33      8.50      10      10      9.33      9.00
##      Total.Cup.Points Moisture Category.One.Defects Quakers Color
## 1              90.58      0.12              0      0 Green
## 2              89.92      0.12              0      0 Green
## 3              89.75      0.00              0      0
## 4              89.00      0.11              0      0 Green
## 5              88.83      0.12              0      0 Green
## 8              88.67      0.03              0      0
##      Category.Two.Defects      Expiration      Certification.B
ody
## 1              0      April 3rd, 2016 METAD Agricultural Development
plc
## 2              1      April 3rd, 2016 METAD Agricultural Development
plc
## 3              0      May 31st, 2011      Specialty Coffee Associat
ion
## 4              2      March 25th, 2016 METAD Agricultural Development
plc
## 5              2      April 3rd, 2016 METAD Agricultural Development
plc
## 8              0 September 2nd, 2011      Ethiopia Commodity Excha
nge
##      Certification.Address
## 1 309fcf77415a3661ae83e027f7e5f05dad786e44
## 2 309fcf77415a3661ae83e027f7e5f05dad786e44
## 3 36d0d00a3724338ba7937c52a378d085f2172daa
## 4 309fcf77415a3661ae83e027f7e5f05dad786e44
## 5 309fcf77415a3661ae83e027f7e5f05dad786e44
## 8 a176532400aebdc345cf3d870f84ed3ecab6249e
##      Certification.Contact unit_of_measurement
## 1 19fef5a731de2db57d16da10287413f5f99bc2dd      m
## 2 19fef5a731de2db57d16da10287413f5f99bc2dd      m
## 3 0878a7d4b9d35ddb0fe2ce69a2062cceb45a660      m
## 4 19fef5a731de2db57d16da10287413f5f99bc2dd      m
## 5 19fef5a731de2db57d16da10287413f5f99bc2dd      m
## 8 61bbaf6a9f341e5782b8e7bd3ebf76aac89fe24b      m
##      altitude_low_meters altitude_high_meters altitude_mean_meters
## 1              1950              2200              2075
## 2              1950              2200              2075
## 3              1600              1800              1700
## 4              1800              2200              2000
## 5              1950              2200              2075
## 8              1570              1700              1635

```

`summary(coffee_clean)`

```

##      X      Species      Owner      Country.of.Origin
## Min. : 0.0 Length:1108 Length:1108 Length:1108
## 1st Qu.: 340.8 Class :character Class :character Class :character

```



```

## Median : 675.5   Mode :character   Mode :character   Mode :character
## Mean    : 670.8
## 3rd Qu.: 992.2
## Max.    :1336.0
## Farm.Name      Lot.Number      Mill      ICO.Number
## Length:1108    Length:1108    Length:1108    Length:1108
## Class :character    Class :character    Class :character    Class :character
## Mode  :character    Mode  :character    Mode  :character    Mode  :character
##
##
## Company        Altitude        Region        Producer
## Length:1108    Length:1108    Length:1108    Length:1108
## Class :character    Class :character    Class :character    Class :character
## Mode  :character    Mode  :character    Mode  :character    Mode  :character
##
##
## Number.of.Bags  Bag.Weight      In.Country.Partner  Harvest.Year
## Min.   : 0.0    Length:1108      Length:1108      Length:1108
## 1st Qu.: 20.0   Class :character    Class :character    Class :character
## Median : 200.0   Mode  :character    Mode  :character    Mode  :character
## Mean   : 158.8
## 3rd Qu.: 275.0
## Max.   :1062.0
## Grading.Date    Owner.1        Variety        Processing.Metho
d
## Length:1108    Length:1108    Length:1108    Length:1108
## Class :character    Class :character    Class :character    Class :character
## Mode  :character    Mode  :character    Mode  :character    Mode  :character
##
##
## Aroma          Flavor          Aftertaste        Acidity
## Min.   :0.000   Min.   :0.000   Min.   :0.000   Min.   :0.000
## 1st Qu.:7.420   1st Qu.:7.330   1st Qu.:7.250   1st Qu.:7.330
## Median :7.580   Median :7.580   Median :7.420   Median :7.500
## Mean   :7.571   Mean   :7.521   Mean   :7.394   Mean   :7.529
## 3rd Qu.:7.750   3rd Qu.:7.750   3rd Qu.:7.580   3rd Qu.:7.750
## Max.   :8.750   Max.   :8.830   Max.   :8.670   Max.   :8.750
## Body          Balance          Uniformity        Clean.Cup
## Min.   :0.000   Min.   :0.000   Min.   : 0.000   Min.   : 0.000
## 1st Qu.:7.330   1st Qu.:7.330   1st Qu.:10.000   1st Qu.:10.000
## Median :7.500   Median :7.500   Median :10.000   Median :10.000
## Mean   :7.507   Mean   :7.506   Mean   : 9.868   Mean   : 9.851
## 3rd Qu.:7.670   3rd Qu.:7.750   3rd Qu.:10.000   3rd Qu.:10.000
## Max.   :8.580   Max.   :8.750   Max.   :10.000   Max.   :10.000
## Sweetness      Cupper.Points    Total.Cup.Points    Moisture
## Min.   : 0.000   Min.   : 0.000   Min.   : 0.00   Min.   :0.00000
## 1st Qu.:10.000   1st Qu.: 7.250   1st Qu.:81.17   1st Qu.:0.10000

```

```

## Median :10.000 Median : 7.500 Median :82.50 Median :0.11000
## Mean : 9.874 Mean : 7.489 Mean :82.11 Mean :0.09203
## 3rd Qu.:10.000 3rd Qu.: 7.750 3rd Qu.:83.58 3rd Qu.:0.12000
## Max. :10.000 Max. :10.000 Max. :90.58 Max. :0.20000
## Category.One.Defects Quakers Color Category.Two.Defects
## Min. : 0.0000 Min. : 0.000 Length:1108 Min. : 0.000
## 1st Qu.: 0.0000 1st Qu.: 0.000 Class :character 1st Qu.: 0.000
## Median : 0.0000 Median : 0.000 Mode :character Median : 2.000
## Mean : 0.3673 Mean : 0.139 Mean : 3.537
## 3rd Qu.: 0.0000 3rd Qu.: 0.000 3rd Qu.: 4.000
## Max. :31.0000 Max. :11.000 Max. :47.000
## Expiration Certification.Body Certification.Address
## Length:1108 Length:1108 Length:1108
## Class :character Class :character Class :character
## Mode :character Mode :character Mode :character
##
##
## Certification.Contact unit_of_measurement altitude_low_meters
## Length:1108 Length:1108 Min. : 1
## Class :character Class :character 1st Qu.: 1100
## Mode :character Mode :character Median : 1311
## Mean : 1751
## 3rd Qu.: 1599
## Max. :190164
## altitude_high_meters altitude_mean_meters
## Min. : 1 Min. : 1
## 1st Qu.: 1100 1st Qu.: 1100
## Median : 1350 Median : 1311
## Mean : 1799 Mean : 1775
## 3rd Qu.: 1650 3rd Qu.: 1600
## Max. :190164 Max. :190164

coffee_clean <- coffee_clean[!(coffee_clean$Species=="Robusta"),]
table(coffee_clean$Species)

##
## Arabica
## 1083

coffee_clean %>%
  group_by(Country.of.Origin) %>%
  tally() %>%
  top_n(., 10) %>%
  arrange(desc(n))

## Selecting by n

## # A tibble: 10 x 2
## Country.of.Origin n

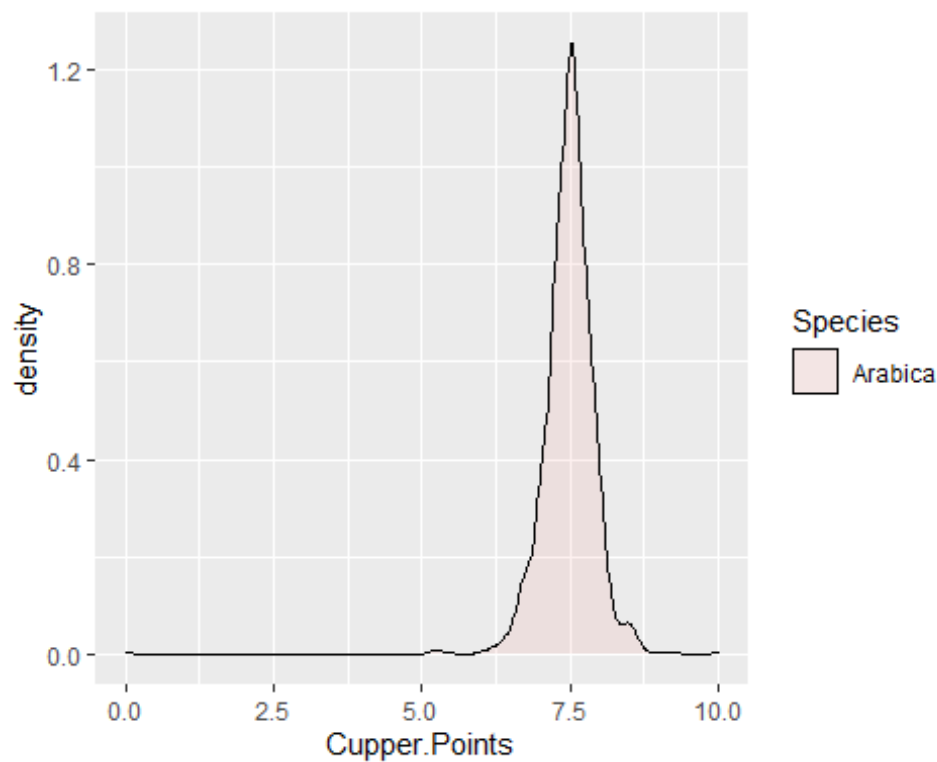
```

```
##      <chr>                <int>
##  1 Mexico                232
##  2 Guatemala             156
##  3 Colombia              149
##  4 Brazil                105
##  5 Taiwan                70
##  6 Honduras              51
##  7 Costa Rica            46
##  8 Tanzania, United Republic Of 37
##  9 Ethiopia              30
## 10 Uganda                24
```

```
table(coffee_clean$Color)
```

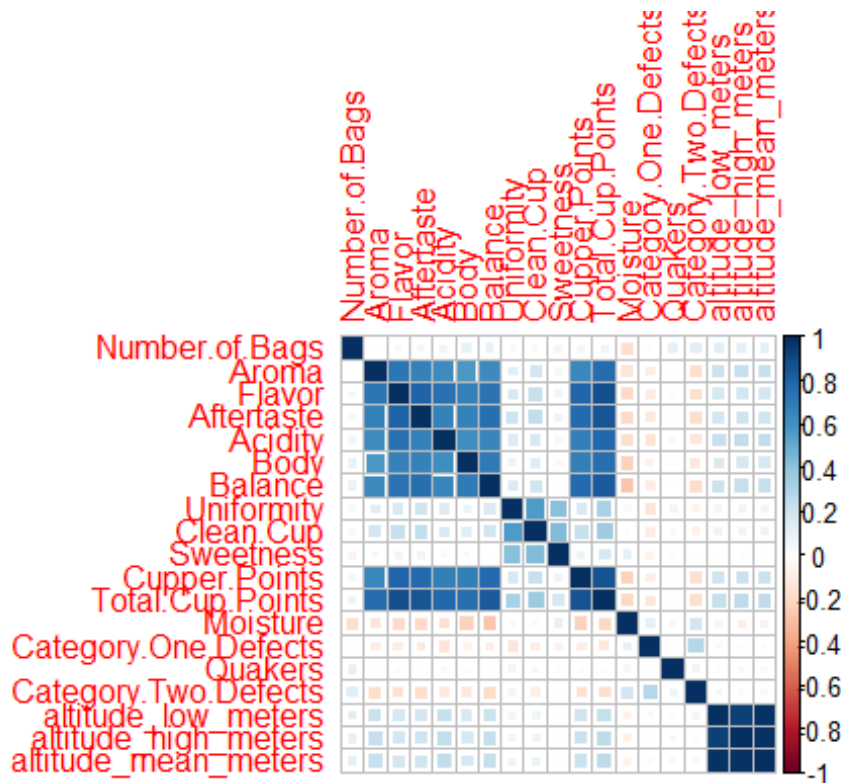
```
##
##           Blue-Green Bluish-Green           Green           None
##           139           69           81           750           44
```

```
library(dplyr)
coffee_clean %>%
  mutate(id = row_number()) %>%
  select(X, Species, Cupper.Points) %>%
  ggplot(aes(Cupper.Points, fill = Species)) +
  geom_density(alpha = 0.1)
```



```
# geom_histogram(position = "identity", alpha = 0.7, show.legend = FALSE)
```

```
library(corrplot)
numbers <- coffee_clean[,c(13,21,22,23,24,25,26,27,28,29,30,31,32,33,34,36,42,43,44)]
corrplot(cor(numbers, method = "spearman"), method = "square")
```



```
coffee_clean %>%
  group_by(Certification.Body) %>%
  tally() %>%
  top_n(., 10) %>%
  arrange(desc(n))
```

```
## Selecting by n
```

```
## # A tibble: 10 x 2
```

Certification.Body	n
<chr>	<int>
1 AMECAFE	204
2 Specialty Coffee Association	177
3 AlmacafÃ©	148
4 Asociacion Nacional Del CafÃ©	135
5 Instituto HondureÃ±o del CafÃ©	57
6 Blossom Valley International	53
7 Brazil Specialty Coffee Association	48
8 Africa Fine Coffee Association	47
9 Specialty Coffee Association of Costa Rica	42
10 NUCOFFEE	31

```

coffee_clean$Species <- factor(coffee_clean$Species,exclude = NULL)
coffee_clean$Owner <- factor(coffee_clean$Owner,exclude = NULL)
coffee_clean$Lot.Number <- factor(coffee_clean$Lot.Number,exclude = NULL)
coffee_clean$Country.of.Origin <- factor(coffee_clean$Country.of.Origin,excl
ude = NULL)
coffee_clean$Color <- factor(coffee_clean$Color,exclude = NULL)
coffee_clean$Farm.Name <- factor(coffee_clean$Farm.Name,exclude = NULL)
coffee_clean$Mill <- factor(coffee_clean$Mill,exclude = NULL)
coffee_clean$ICO.Number <- factor(coffee_clean$ICO.Number,exclude = NULL)
coffee_clean$Company <- factor(coffee_clean$Company,exclude = NULL)
coffee_clean$Altitude <- factor(coffee_clean$Altitude,exclude = NULL)
coffee_clean$Region <- factor(coffee_clean$Region,exclude = NULL)
coffee_clean$Producer <- factor(coffee_clean$Producer,exclude = NULL)
coffee_clean$Bag.Weight <- factor(coffee_clean$Bag.Weight,exclude = NULL)
coffee_clean$In.Country.Partner <- factor(coffee_clean$In.Country.Partner,excl
ude = NULL)
coffee_clean$Harvest.Year <- factor(coffee_clean$Harvest.Year,exclude = NULL)
coffee_clean$Grading.Date <- factor(coffee_clean$Grading.Date,exclude = NULL)
coffee_clean$Owner.1 <- factor(coffee_clean$Owner.1,exclude = NULL)
coffee_clean$Variety <- factor(coffee_clean$Variety,exclude = NULL)
coffee_clean$Processing.Method <- factor(coffee_clean$Processing.Method,exclu
de = NULL)
coffee_clean$Expiration <- factor(coffee_clean$Expiration,exclude = NULL)
coffee_clean$Certification.Body <- factor(coffee_clean$Certification.Body,excl
ude = NULL)
coffee_clean$Certification.Address <- factor(coffee_clean$Certification.Addre
ss,exclude = NULL)
coffee_clean$Certification.Contact <- factor(coffee_clean$Certification.Conta
ct,exclude = NULL)
coffee_clean$unit_of_measurement <- factor(coffee_clean$unit_of_measurement,e
xclude = NULL)

coffee_clean <- subset(coffee_clean, select = -c(Species))
coffee_clean <- subset(coffee_clean, select = -c(Certification.Address,Certif
ication.Contact))
coffee_clean <- subset(coffee_clean, select = -c(X))
coffee_clean <- subset(coffee_clean, select = -c(Owner, Owner.1))

```

```
summary(coffee_clean)
```

```

## Country.of.Origin      Farm.Name      Lot.Number
## Mexico :232           :202           :862
## Guatemala:156        various         : 46    1      : 11
## Colombia :149        rio verde         : 23    020/17 : 6
## Brazil :105          several         : 20    019/17 : 5
## Taiwan : 70          finca medina      : 15    102      : 3
## Honduras : 51        fazenda capoeirinha: 13    103      : 3
## (Other) :320          (Other)           :764    (Other):193
##                               Mill          ICO.Number
##                               :164          :113

```

```

## beneficio ixchel : 76 0 : 65
## trilladora boanza: 36 Taiwan : 31
## dry mill : 35 2222 : 11
## ipanema coffees : 16 - : 7
## neiva : 14 002/1660/0105: 7
## (Other) :742 (Other) :849
## Company Altitude Region
## :151 1100 : 43 huila : 93
## unex guatemala, s.a. : 73 1200 : 41 oriente : 66
## ipanema coffees : 50 1300 : 32 south of minas: 66
## exportadora de cafe condor s.a: 38 1400 : 32 veracruz : 31
## racafe & cia s.c.a : 26 4300 : 31 tarrazu : 18
## blossom valley® : 25 1250 : 30 comayagua : 17
## (Other) :720 (Other):874 (Other) :792
## Producer Number.of.Bags Bag.Weight
## : 96 Min. : 0.0 1 kg :307
## Ipanema Agr cola SA: 22 1st Qu.: 20.0 60 kg :212
## La Plata : 20 Median : 200.0 69 kg :187
## Ipanema Agr cola : 12 Mean : 158.1 70 kg :131
## VARIOS : 12 3rd Qu.: 275.0 2 kg : 97
## Ipanema Agr cola S.A: 11 Max. :1062.0 30 kg : 28
## (Other) :910 (Other):121
## In.Country.Partner Harvest.Year
## AMECAFE :204 2012 :293
## Specialty Coffee Association :177 2014 :199
## Almacaf  :148 2013 :142
## Asociacion Nacional Del Caf  :135 2015 :116
## Instituto Hondure o del Caf : 57 2016 : 91
## Blossom Valley International : 53 2017 : 62
## (Other) :309 (Other):180
## Grading.Date Variety Processing.Method
## July 11th, 2012 : 25 Caturra:237 : 79
## December 26th, 2013: 24 Typica :208 Natural / Dry :179
## June 6th, 2012 : 19 Bourbon:207 Other : 25
## August 30th, 2012 : 18 Other :100 Pulped natural / honey : 10
## July 26th, 2012 : 15 : 90 Semi-washed / Semi-pulped: 53
## March 29th, 2013 : 13 Catuai : 69 Washed / Wet :737
## (Other) :969 (Other):172
## Aroma Flavor Aftertaste Acidity
## Min. :0.000 Min. :0.000 Min. :0.000 Min. :0.000
## 1st Qu.:7.420 1st Qu.:7.330 1st Qu.:7.170 1st Qu.:7.330
## Median :7.580 Median :7.500 Median :7.420 Median :7.500
## Mean :7.566 Mean :7.516 Mean :7.389 Mean :7.525
## 3rd Qu.:7.750 3rd Qu.:7.750 3rd Qu.:7.580 3rd Qu.:7.710
## Max. :8.750 Max. :8.830 Max. :8.670 Max. :8.750
##
## Body Balance Uniformity Clean.Cup
## Min. :0.000 Min. :0.000 Min. : 0.000 Min. : 0.000
## 1st Qu.:7.330 1st Qu.:7.330 1st Qu.:10.000 1st Qu.:10.000
## Median :7.500 Median :7.500 Median :10.000 Median :10.000

```

```

## Mean :7.504 Mean :7.504 Mean : 9.866 Mean : 9.848
## 3rd Qu.:7.670 3rd Qu.:7.750 3rd Qu.:10.000 3rd Qu.:10.000
## Max. :8.580 Max. :8.750 Max. :10.000 Max. :10.000
##
## Sweetness Cupper.Points Total.Cup.Points Moisture
## Min. : 0.000 Min. : 0.000 Min. : 0.00 Min. :0.00000
## 1st Qu.:10.000 1st Qu.: 7.250 1st Qu.:81.21 1st Qu.:0.10000
## Median :10.000 Median : 7.500 Median :82.50 Median :0.11000
## Mean : 9.923 Mean : 7.482 Mean :82.12 Mean :0.09266
## 3rd Qu.:10.000 3rd Qu.: 7.750 3rd Qu.:83.67 3rd Qu.:0.12000
## Max. :10.000 Max. :10.000 Max. :90.58 Max. :0.20000
##
## Category.One.Defects Quakers Color Category.Two.De
fects
## Min. : 0.0000 Min. : 0.0000 :139 Min. : 0.00
## 1st Qu.: 0.0000 1st Qu.: 0.0000 Blue-Green : 69 1st Qu.: 0.00
## Median : 0.0000 Median : 0.0000 Bluish-Green: 81 Median : 2.00
## Mean : 0.3758 Mean : 0.1422 Green :750 Mean : 3.58
## 3rd Qu.: 0.0000 3rd Qu.: 0.0000 None : 44 3rd Qu.: 4.00
## Max. :31.0000 Max. :11.0000 Max. :47.00
##
## Expiration Certification.Body
## December 26th, 2014: 25 AMECAFE :204
## July 11th, 2013 : 25 Specialty Coffee Association :177
## June 6th, 2013 : 19 AlmacafÃ© :148
## August 30th, 2013 : 18 Asociacion Nacional Del CafÃ© :135
## July 26th, 2013 : 15 Instituto HondureÃ±o del CafÃ©: 57
## March 29th, 2014 : 13 Blossom Valley International : 53
## (Other) :968 (Other) :309
## unit_of_measurement altitude_low_meters altitude_high_meters
## ft:111 Min. : 1 Min. : 1
## m :972 1st Qu.: 1100 1st Qu.: 1100
## Median : 1311 Median : 1350
## Mean : 1760 Mean : 1809
## 3rd Qu.: 1600 3rd Qu.: 1650
## Max. :190164 Max. :190164
##
## altitude_mean_meters
## Min. : 1
## 1st Qu.: 1100
## Median : 1311
## Mean : 1784
## 3rd Qu.: 1600
## Max. :190164
##

```

```

coffee_clean <- transform(coffee_clean, Category.One.Defects = as.numeric(Category.One.Defects))
coffee_clean <- transform(coffee_clean, Category.Two.Defects = as.numeric(Category.Two.Defects))

```

```

coffee_clean <- transform(coffee_clean, Number.of.Bags = as.numeric(Number.of
.Bags))
str(coffee_clean)

## 'data.frame':    1083 obs. of  38 variables:
## $ Country.of.Origin   : Factor w/ 35 levels "Brazil","Burundi",...: 9 9 10
9 9 9 9 9 9 31 ...
## $ Farm.Name           : Factor w/ 530 levels "", "-", "1", "200 farms",...: 3
76 376 443 521 376 31 31 507 170 141 ...
## $ Lot.Number          : Factor w/ 185 levels "", "0063/17", "007/16A",...: 1
1 1 1 1 1 1 1 1 1 1 ...
## $ Mill                : Factor w/ 426 levels "", "1", "17/18",...: 291 291 1
410 291 87 87 398 1 1 ...
## $ ICO.Number          : Factor w/ 680 levels "", "-", "??", "0",...: 392 392
1 1 392 77 77 391 1 677 ...
## $ Company             : Factor w/ 248 levels "", "â®, â¶$âæ<éš>",...: 159 15
9 1 244 159 1 1 88 111 71 ...
## $ Altitude            : Factor w/ 379 levels "-1", "~1000 meters",...: 224
224 166 203 224 158 158 199 214 374 ...
## $ Region              : Factor w/ 327 levels "", "52 narino (exact locatio
n: mattituy; municipal region: florida code 381",...: 117 117 1 220 117 220 22
1 267 220 22 ...
## $ Producer            : Factor w/ 626 levels "", "-", "1", "â¼µæ-†é\200²",..
: 409 409 1 615 409 64 64 142 181 27 ...
## $ Number.of.Bags      : num  300 300 5 320 300 300 300 50 300 10 ...
## $ Bag.Weight          : Factor w/ 46 levels "0 kg", "1", "1 kg",...: 39 39 2
39 39 39 39 39 39 3 ...
## $ In.Country.Partner  : Factor w/ 26 levels "Africa Fine Coffee Associati
on",...: 15 15 19 15 15 12 12 15 15 2 ...
## $ Harvest.Year        : Factor w/ 43 levels "", "08/09 crop",...: 15 15 1 1
5 15 39 39 15 15 15 ...
## $ Grading.Date        : Factor w/ 466 levels "April 10th, 2014",...: 35 35
357 300 35 457 457 307 302 282 ...
## $ Variety             : Factor w/ 29 levels "", "Arusha", "Blue Mountain",.
.: 1 16 4 1 16 1 1 16 1 16 ...
## $ Processing.Method   : Factor w/ 6 levels "", "Natural / Dry",...: 6 6 1 2
6 1 1 2 2 6 ...
## $ Aroma               : num  8.67 8.75 8.42 8.17 8.25 8.25 8.67 8.08 8.17
8.25 ...
## $ Flavor              : num  8.83 8.67 8.5 8.58 8.5 8.33 8.67 8.58 8.67 8
.42 ...
## $ Aftertaste          : num  8.67 8.5 8.42 8.42 8.25 8.5 8.58 8.5 8.25 8.
17 ...
## $ Acidity             : num  8.75 8.58 8.42 8.42 8.5 8.42 8.42 8.5 8.5 8.
33 ...
## $ Body                : num  8.5 8.42 8.33 8.5 8.42 8.33 8.33 7.67 7.75 8
.08 ...
## $ Balance             : num  8.42 8.42 8.42 8.25 8.33 8.5 8.42 8.42 8.17
8.17 ...
## $ Uniformity          : num  10 10 10 10 10 10 9.33 10 10 10 ...

```



```
## $ Clean.Cup          : num  10 10 10 10 10 10 10 10 10 10 ...
## $ Sweetness          : num  10 10 10 10 10 9.33 9.33 10 10 10 ...
## $ Cupper.Points      : num   8.75 8.58 9.25 8.67 8.58 9 8.67 8.5 8.58 8.5
...
## $ Total.Cup.Points   : num  90.6 89.9 89.8 89 88.8 ...
## $ Moisture           : num   0.12 0.12 0 0.11 0.12 0.03 0.03 0.1 0.1 0 ..
.
## $ Category.One.Defects: num   0 0 0 0 0 0 0 0 0 0 ...
## $ Quakers            : num   0 0 0 0 0 0 0 0 0 0 ...
## $ Color               : Factor w/ 5 levels "", "Blue-Green",...: 4 4 1 4 4
1 1 4 1 1 ...
## $ Category.Two.Defects: num   0 1 0 2 2 0 0 4 1 0 ...
## $ Expiration          : Factor w/ 465 levels "April 10th, 2015",...: 34 34
356 297 34 456 456 303 299 279 ...
## $ Certification.Body  : Factor w/ 25 levels "Africa Fine Coffee Associati
on",...: 15 15 18 15 15 12 12 15 15 2 ...
## $ unit_of_measurement : Factor w/ 2 levels "ft","m": 2 2 2 2 2 2 2 2 2 2
...
## $ altitude_low_meters : num  1950 1950 1600 1800 1950 ...
## $ altitude_high_meters: num  2200 2200 1800 2200 2200 ...
## $ altitude_mean_meters: num  2075 2075 1700 2000 2075 ...
```

## #Getting started with PCA

```
PCA_Plot = function(pcaData)
{
  library(ggplot2)

  theta = seq(0,2*pi,length.out = 100)
  circle = data.frame(x = cos(theta), y = sin(theta))
  p = ggplot(circle,aes(x,y)) + geom_path()

  loadings = data.frame(pcaData$rotation, .names = row.names(pcaData$rotation
))
  p + geom_text(data=loadings, mapping=aes(x = PC1, y = PC2, label = .names,
colour = .names, fontface="bold")) +
  coord_fixed(ratio=1) + labs(x = "PC1", y = "PC2")
}

PCA_Plot_Secondary = function(pcaData)
{
  library(ggplot2)

  theta = seq(0,2*pi,length.out = 100)
  circle = data.frame(x = cos(theta), y = sin(theta))
  p = ggplot(circle,aes(x,y)) + geom_path()

  loadings = data.frame(pcaData$rotation, .names = row.names(pcaData$rotation
))
  p + geom_text(data=loadings, mapping=aes(x = PC3, y = PC4, label = .names,
```

```

colour = .names, fontface="bold")) +
  coord_fixed(ratio=1) + labs(x = "PC3", y = "PC4")
}

PCA_Plot_Psyc = function(pcaData)
{
  library(ggplot2)

  theta = seq(0,2*pi,length.out = 100)
  circle = data.frame(x = cos(theta), y = sin(theta))
  p = ggplot(circle,aes(x,y)) + geom_path()

  loadings = as.data.frame(unclass(pcaData$loadings))
  s = rep(0, ncol(loadings))
  for (i in 1:ncol(loadings))
  {
    s[i] = 0
    for (j in 1:nrow(loadings))
      s[i] = s[i] + loadings[j, i]^2
    s[i] = sqrt(s[i])
  }

  for (i in 1:ncol(loadings))
    loadings[, i] = loadings[, i] / s[i]

  loadings$.names = row.names(loadings)

  p + geom_text(data=loadings, mapping=aes(x = PC1, y = PC2, label = .names,
colour = .names, fontface="bold")) +
  coord_fixed(ratio=1) + labs(x = "PC1", y = "PC2")
}

PCA_Plot_Psyc_Secondary = function(pcaData)
{
  library(ggplot2)

  theta = seq(0,2*pi,length.out = 100)
  circle = data.frame(x = cos(theta), y = sin(theta))
  p = ggplot(circle,aes(x,y)) + geom_path()

  loadings = as.data.frame(unclass(pcaData$loadings))
  s = rep(0, ncol(loadings))
  for (i in 1:ncol(loadings))
  {
    s[i] = 0
    for (j in 1:nrow(loadings))
      s[i] = s[i] + loadings[j, i]^2
    s[i] = sqrt(s[i])
  }
}

```

```

for (i in 1:ncol(loadings))
  loadings[, i] = loadings[, i] / s[i]

loadings$.names = row.names(loadings)

print(loadings)
p + geom_text(data=loadings, mapping=aes(x = PC3, y = PC4, label = .names,
colour = .names, fontface="bold")) +
  coord_fixed(ratio=1) + labs(x = "PC3", y = "PC4")
}

```

## 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.

```

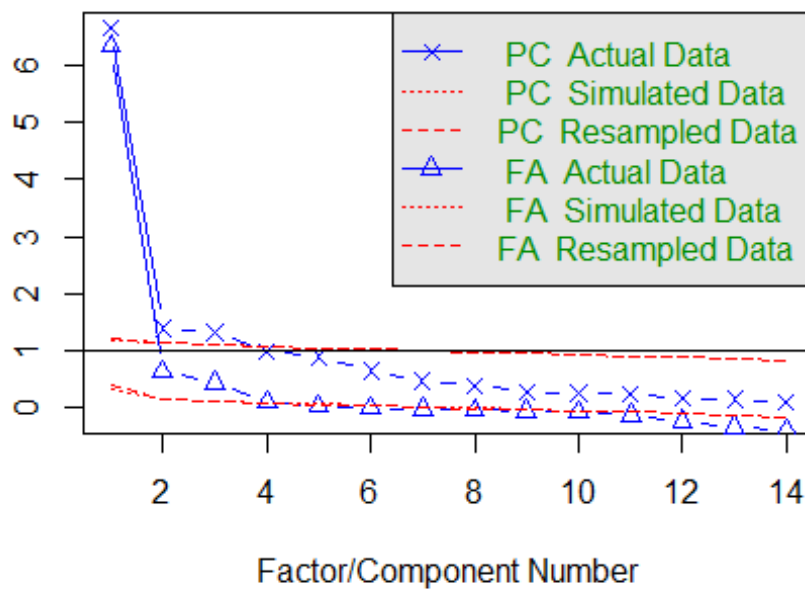
```
##
## Reliability analysis
## Call: alpha(x = pca_numbers, check.keys = TRUE)
##
##   raw_alpha std.alpha G6(smc) average_r S/N   ase mean   sd median_r
##       0.47      0.88    0.91      0.34 7.2 0.017   16 0.56      0.34
##
##   lower alpha upper      95% confidence boundaries
## 0.44 0.47 0.5
##
## Reliability if an item is dropped:
##               raw_alpha std.alpha G6(smc) average_r S/N alpha se v
ar.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

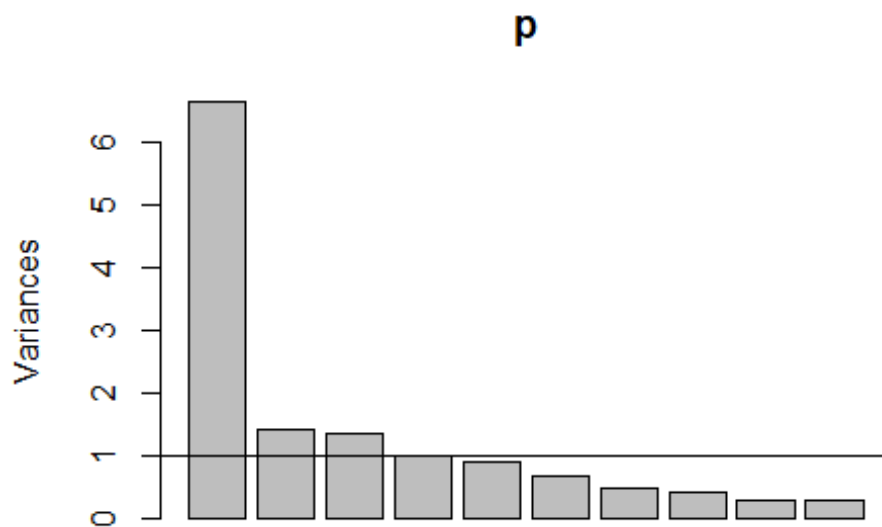


```
## Parallel analysis suggests that the number of factors = 3 and the number
of components = 3

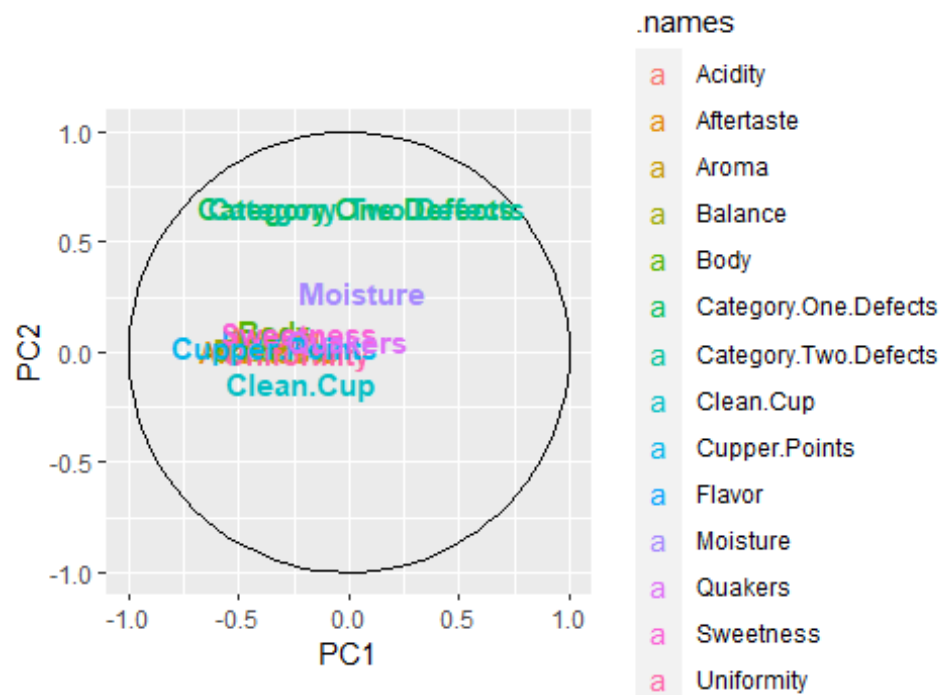
comp

## 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 0.39 0.33 6.63
## 2 0.64 0.16 0.16 1.40
## 3 0.45 0.12 0.12 1.33
## Resampled components Simulated components
## 1 1.21 1.20
## 2 1.15 1.15
## 3 1.11 1.11

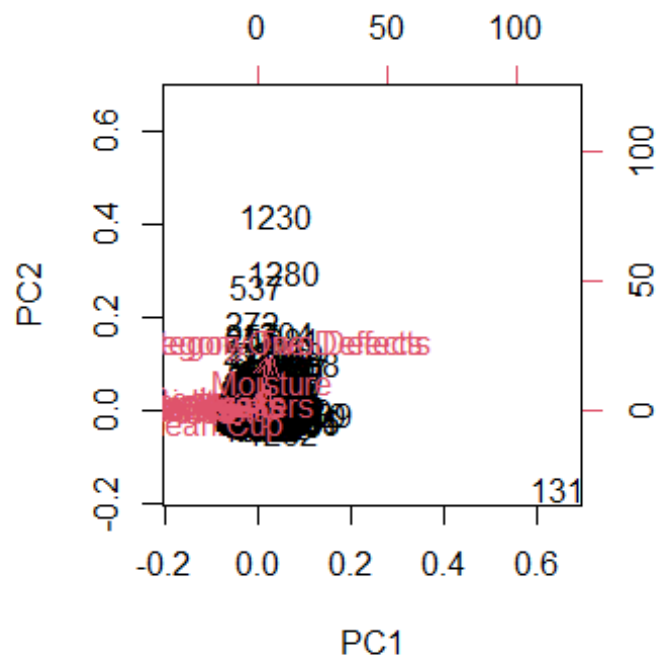
p <- prcomp(pca_numbers, center = T, scale = T)
plot(p)
abline(1,0)
```



```
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 = Taste
      RC2 = Defect classification
      RC3 = Cup-to-cup difference
##           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

table(p2$values>1)

##
## FALSE  TRUE
##     11     3

scores <- p2$scores
scores_1 <- scores[,1]

min_score <- min(scores_1)
min_score

## [1] -19.30803

max_score <- max(scores_1)
max_score

## [1] 3.20632

```

```
summary(scores_1)
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
## -19.30803 -0.49118   0.03328   0.00000   0.50845   3.20632
```

```
scores_2 <- scores[,2]
```

```
fit = factanal(pca_numbers, 3)
```

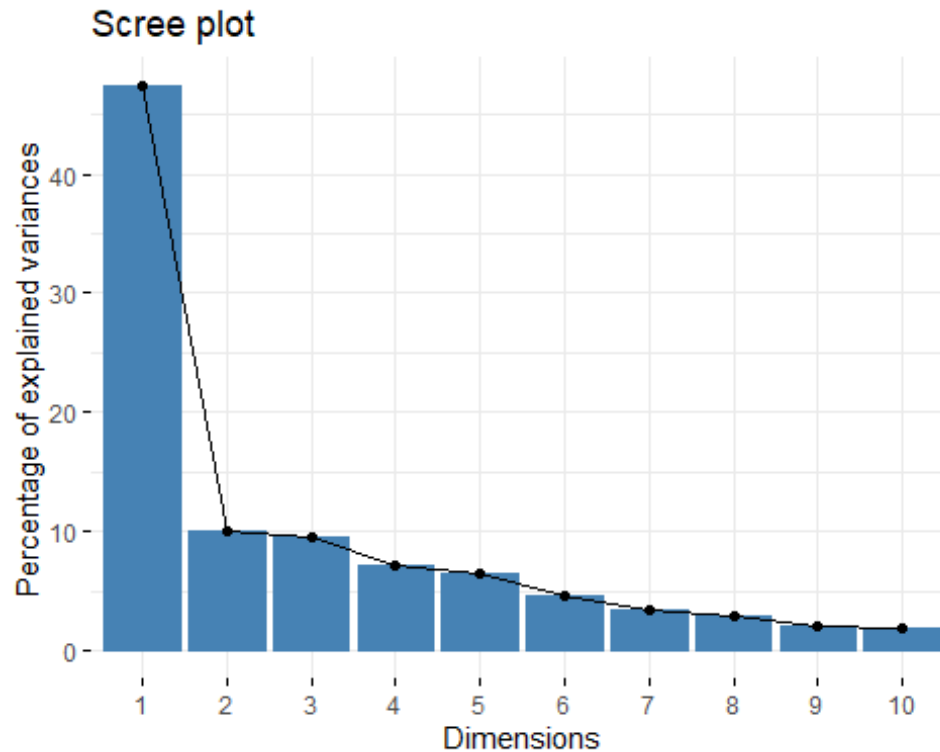
```
print(fit$loadings, cutoff=.4, sort=T)
```

```
##
## Loadings:
##               Factor1 Factor2 Factor3
## Aroma          0.824
## Flavor          0.918
## Aftertaste      0.897
## Acidity         0.844
## Body            0.837
## Balance         0.844
## Cupper.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
```

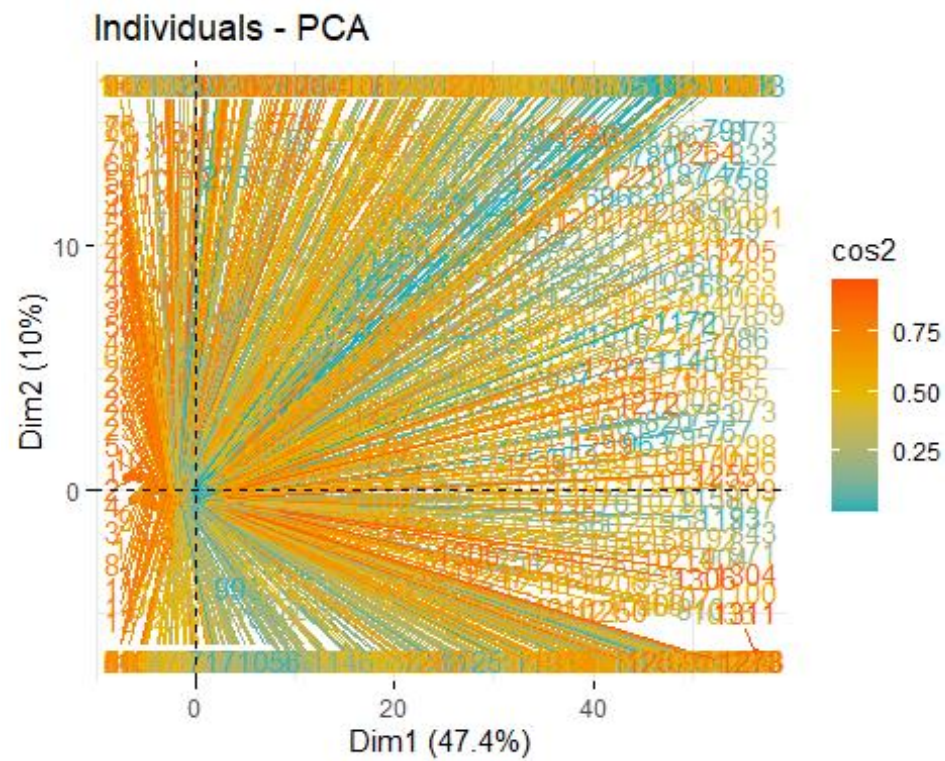
```
summary(fit)
```

```
##      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
```

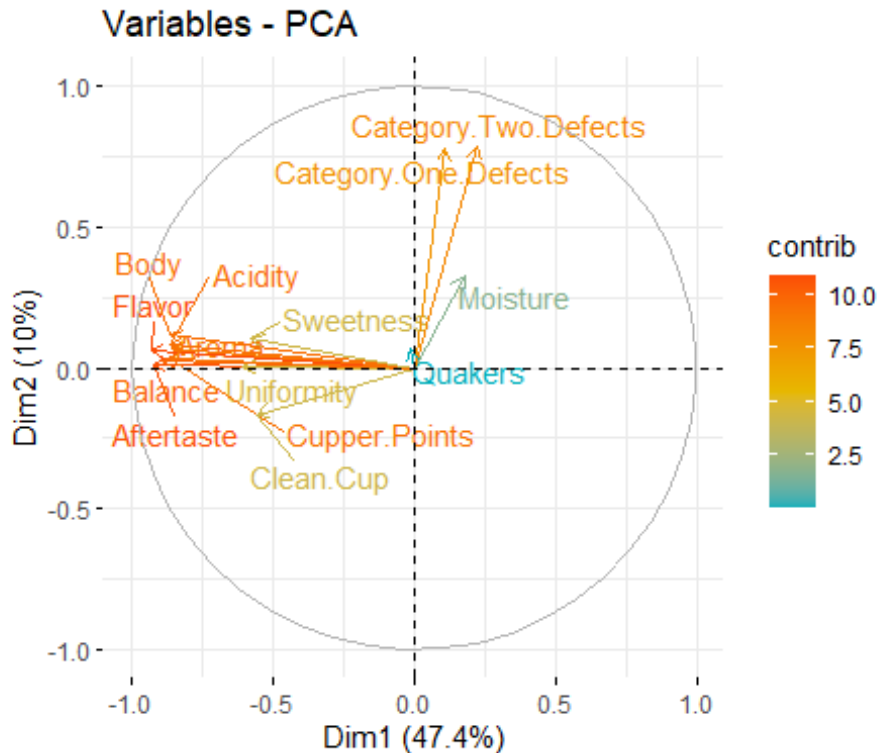
```
p3 <- prcomp(pca_numbers, scale = TRUE)
fviz_eig(p3)
```



```
#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
```



```
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
```

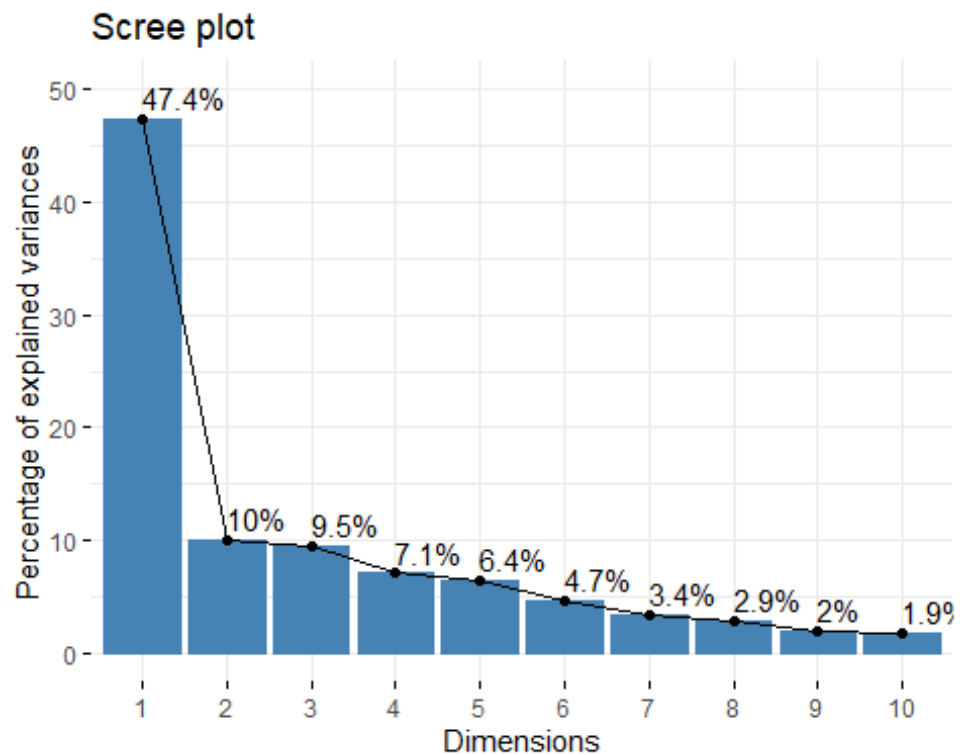


```
p4 <- PCA(pca_numbers, graph = FALSE)
#IF graph is set to true, it will provide the individual and variable maps

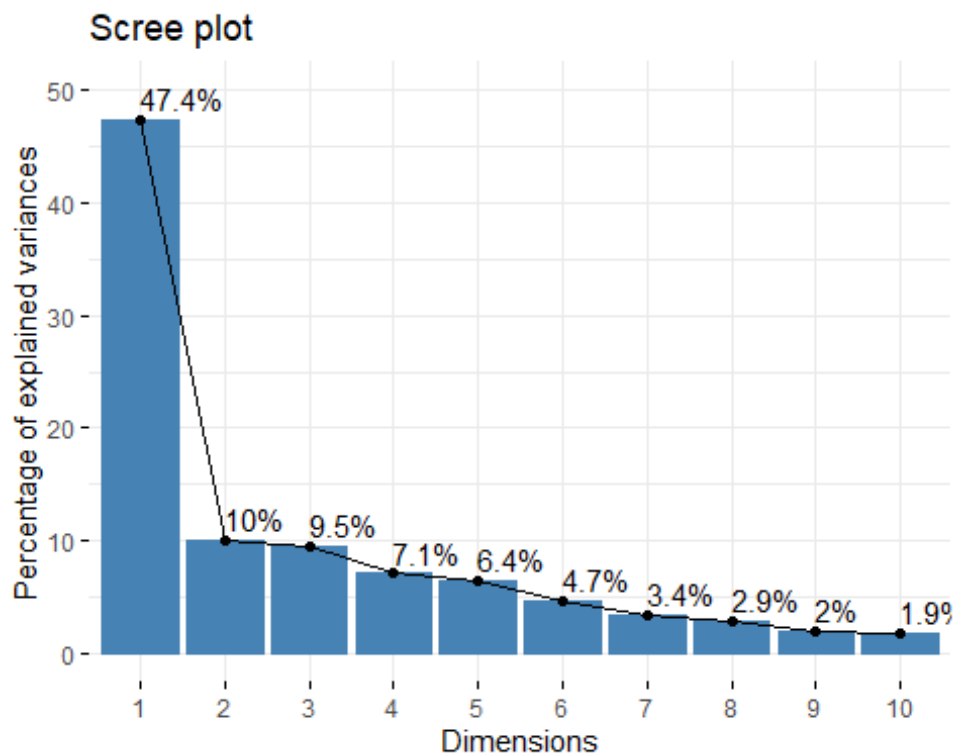
#Shows all the objects or functions available in PCA
print(p4)

## **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 "$call$secart.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_screepplot(p4, addlabels = TRUE, ylim = c(0, 50))
```

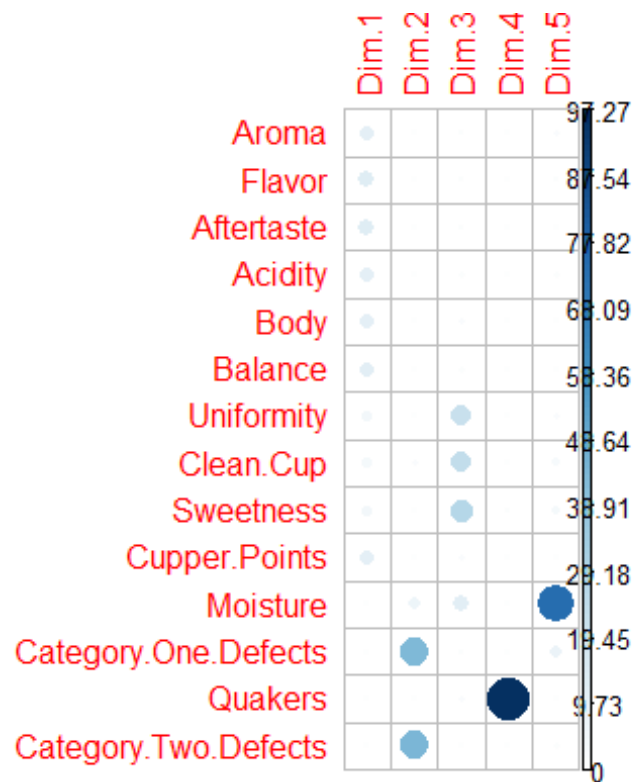


```
variables <- get_pca_var(p4)
```

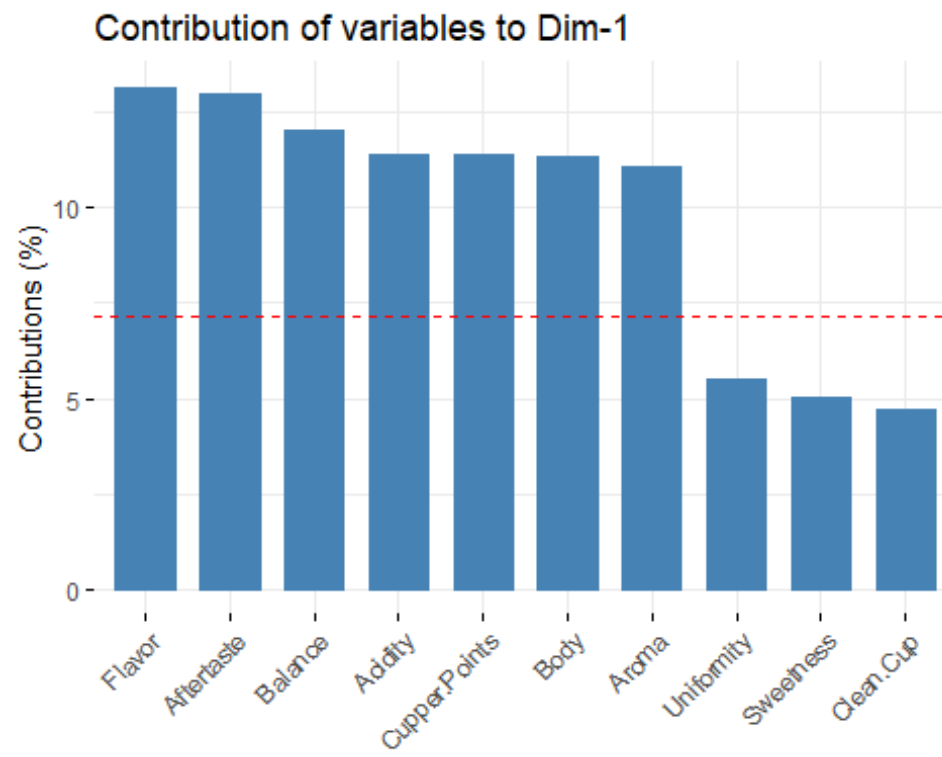
```
head(variables$contrib,11)
```

##	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)
```

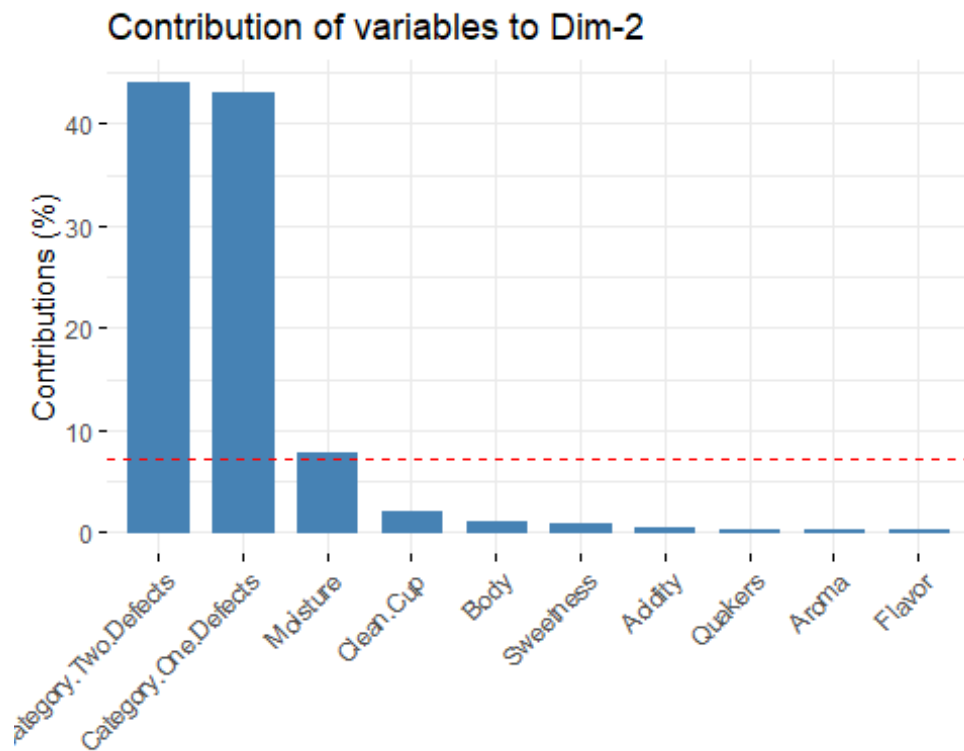


```
# Contributions of variables to PC1
fviz_contrib(p4, choice = "var", axes = 1, top = 10)
```

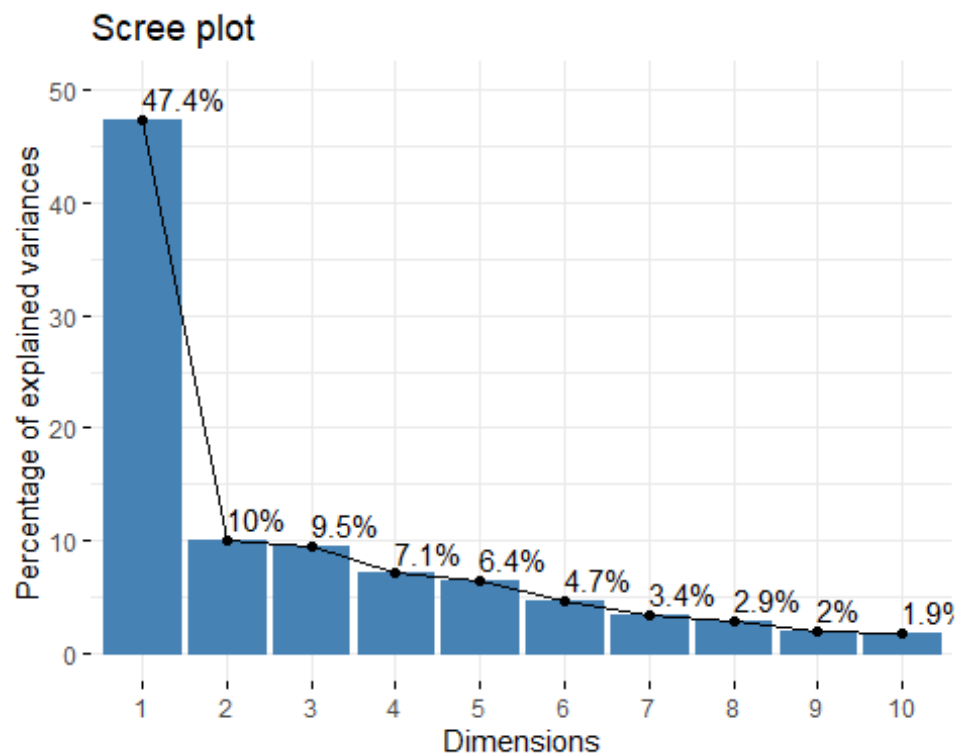




```
# Contributions of variables to PC2  
fviz_contrib(p4, choice = "var", axes = 2, top = 10)
```



```
p5 <- dudi.pca(pca_numbers,  
               scannf = FALSE, # Hide scree plot  
               nf = 3          # Number of components kept in the results  
)  
fviz_screplot(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)
```

