```
library(factoextra)

df<-read.csv("DataCountries.txt", sep="\t")

head(df)
```

| | Agriculture | Livestock | Mining | Industry | Country |
|---|---|---|---|---|---|
| 1 | 7.0 | 3.2 | 4.7 | 1.4 | Albania |
| 2 | 6.4 | 3.2 | 4.5 | 1.5 | Andorra |
| 3 | 6.9 | 3.1 | 4.9 | 1.5 | Austria |
| 4 | 5.5 | 2.3 | 4.0 | 1.3 | Belarus |
| 5 | 6.5 | 2.8 | 4.6 | 1.5 | Belgium |
| 6 | 5.7 | 2.8 | 4.5 | 1.3 | Bosnia |

```
> summary(df)
```

| Agriculture | Livestock | Mining | Industry | Country |
|---|---|---|---|---|
| Min. :4.900 | Min. :2.000 | Min. :3.300 | Min. :1.000 | Length:46 |
| 1st Qu.:5.600 | 1st Qu.:2.525 | 1st Qu.:4.000 | 1st Qu.:1.200 | Class :character |
| Median :5.950 | Median :2.800 | Median :4.400 | Median :1.300 | Mode :character |
| Mean :5.959 | Mean :2.770 | Mean :4.291 | Mean :1.333 | |
| 3rd Qu.:6.300 | 3rd Qu.:3.000 | 3rd Qu.:4.600 | 3rd Qu.:1.500 | |
| Max. :7.000 | Max. :3.400 | Max. :5.100 | Max. :1.800 | |

## PCA Analysis

Now we will run a PCA analysis on our dataset. Note that we need to include only the numeric variables. We will also set as row names the column `Country`.

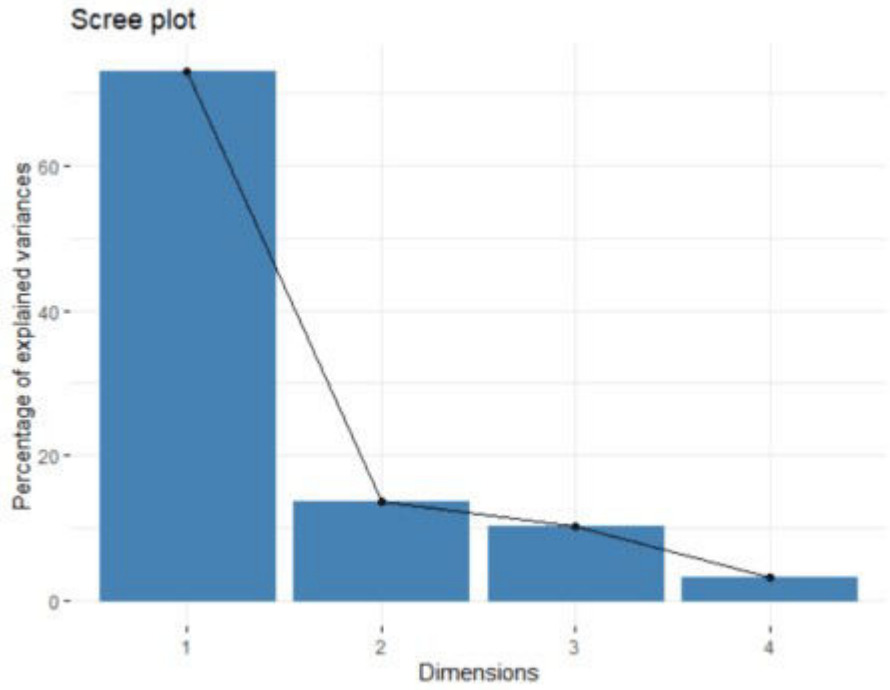
```
# set as rownames the column Country
rownames(df)<-df$Country

# remove the Countrly columns
df$Country<-NULL

# run a PCA Analysis
dfPCA <- prcomp(df, center = TRUE, scale. = TRUE)
```

Let's get Scree plot which shows the percentage of explained variance by Principal Component.

```
fviz_eig (dfPCA)
```

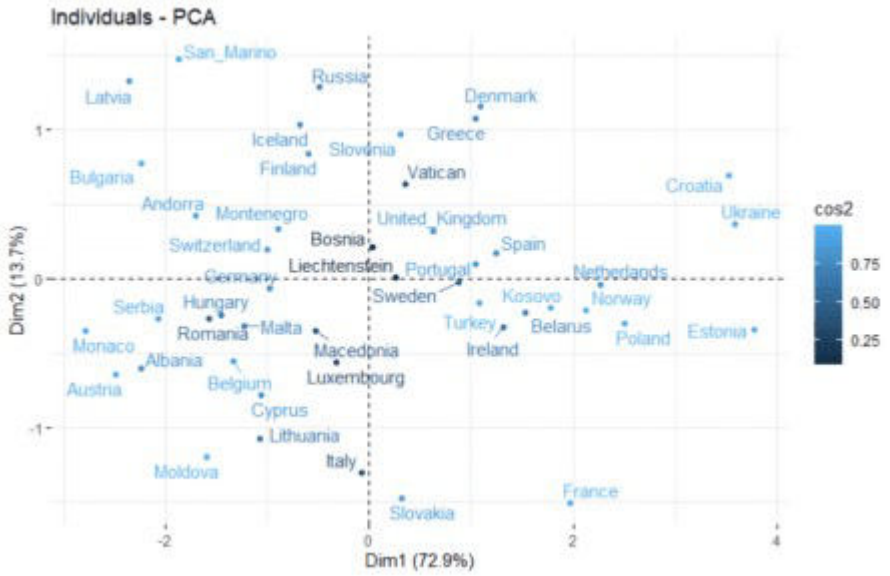



## Graph of Individual

Let's plot all the countries into two dimensions by taking into consideration the quality of the individuals on the factor map.

```
# cos2 = the quality of the individuals on the factor map
# Select and visualize some individuals (ind) with select.ind argument.
 # - ind with cos2 >= 0.96: select.ind = list(cos2 = 0.96)
 # - Top 20 ind according to the cos2: select.ind = list(cos2 = 20)
 # - Top 20 contributing individuals: select.ind = list(contrib = 20)
 # - Select ind by names: select.ind = list(name = c("23", "42", "119")
)

fviz_pca_ind(dfPCA, col.ind = "cos2" , repel = TRUE)
```



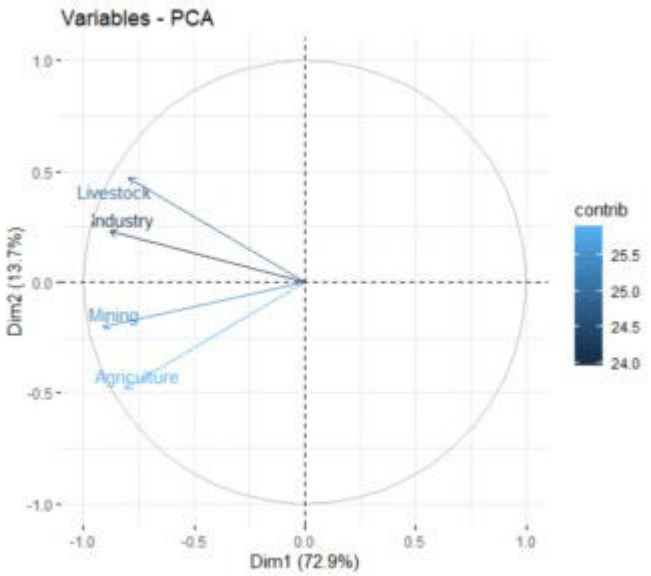

## Graph of Variables

Let's see how we can represent the variables into two dimensions by taking into account their

contribution.

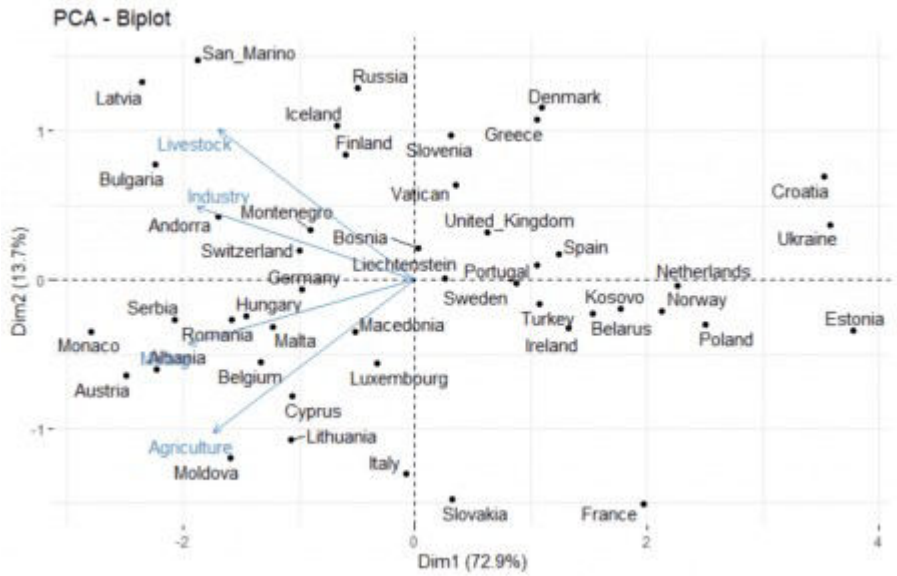
```
#  select.var = list(contrib = 15)

fviz_pca_var(dfPCA, col.var = "contrib", repel = TRUE)
```




## Graph of the Biplot

```
# Graph of the Biplot
fviz_pca_biplot(dfPCA, repel = TRUE)
```




## Eigenvalues, Variables and Individuals

Let's see how we can get the Eigenvalues and statistics for Variables and Individuals such as the **Coordinates**, the **Contributions to the PCs** and the **Quality of representation**

### Eigenvalues

```
# Eigenvalues
eigens_vals <- get_eigenvalue(dfPCA)
eigens_vals
```

```
> eigens_vals
       eigenvalue variance.percent cumulative.variance.percent
Dim.1  2.9178495        72.946236                      72.94624
Dim.2  0.5461993        13.654982                      86.60122
Dim.3  0.4070160        10.175400                      96.77662
Dim.4  0.1289352         3.223381                     100.00000
```

## Variables

```
# By Variable
by_var <- get_pca_var(dfPCA)
by_var$coord
by_var$contrib
by_var$cos2
```

```
> by_var <- get_pca_var(dfPCA)
> by_var$coord
                Dim.1      Dim.2      Dim.3       Dim.4
Agriculture -0.8156361 -0.4812370  0.2920628  0.13359627
Livestock   -0.7972667  0.4716469  0.3700919 -0.07033563
Mining      -0.9159486 -0.2010451 -0.2367584 -0.25409539
Industry    -0.8820380  0.2274651 -0.3587378  0.20390122
> by_var$contrib
               Dim.1     Dim.2    Dim.3     Dim.4
Agriculture 22.79975 42.400095 20.95758 13.842580
Livestock   21.78434 40.727033 33.65174  3.836888
Mining      28.75275  7.400069 13.77207 50.075113
Industry    26.66317  9.472804 31.61861 32.245418
> by_var$cos2
                Dim.1      Dim.2      Dim.3       Dim.4
Agriculture 0.6652623 0.23158902 0.08530070 0.017847963
Livestock   0.6356342 0.22245077 0.13696798 0.004947101
Mining      0.8389619 0.04041912 0.05605453 0.064564465
Industry    0.7779911 0.05174039 0.12869281 0.041575706
```

### Individuals

```
# By ndividual
by_ind <- get_pca_ind(dfPCA)
by_ind$coord
by_ind$contrib
by_ind$cos2  …
```