

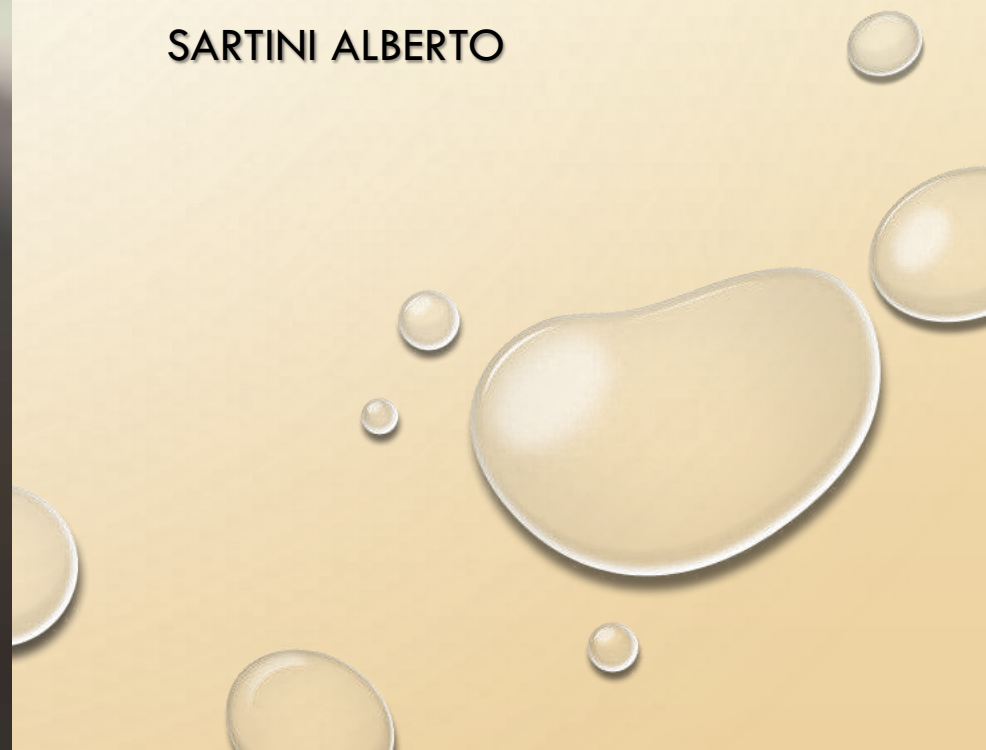


HAPPINESS AND ALCOHOL CONSUMPTION

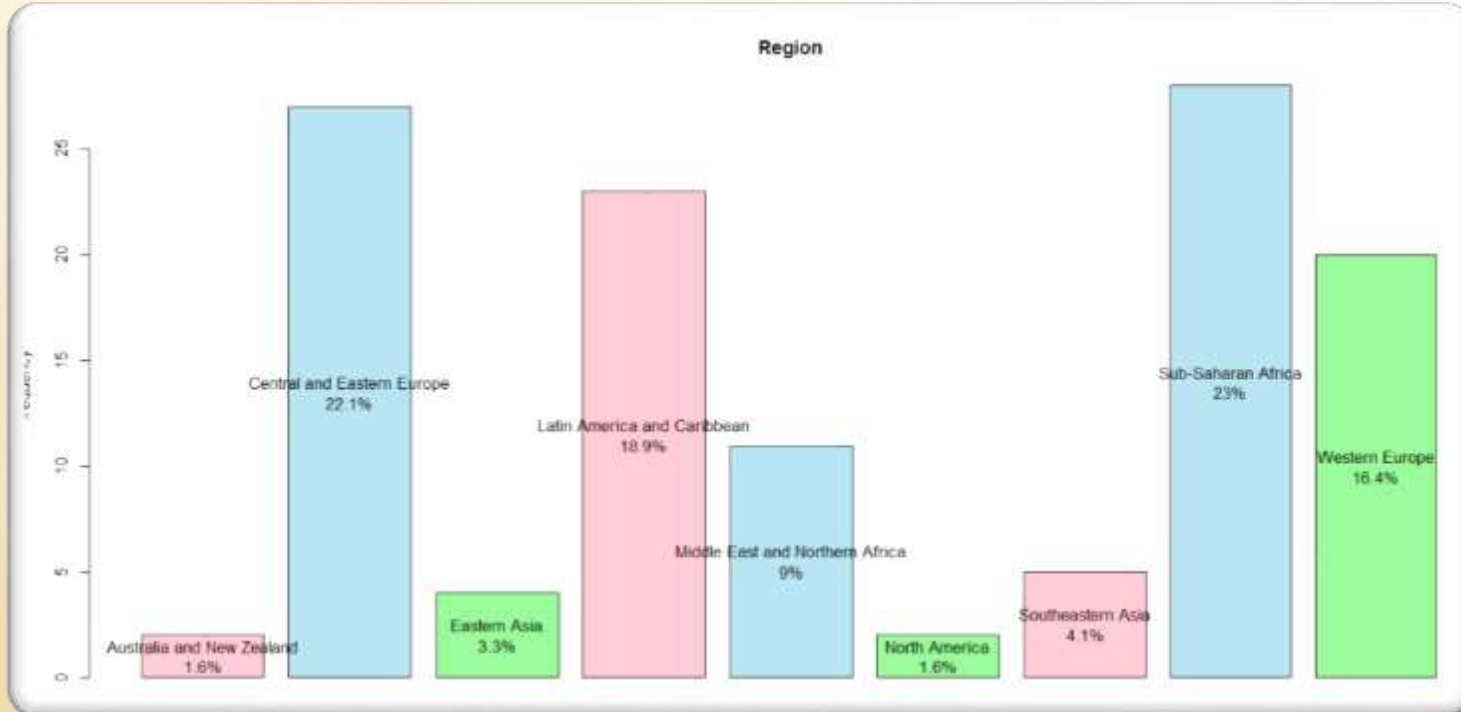
BOLLETTA OSCAR MARIA

GALASSI LEONARDO

SARTINI ALBERTO

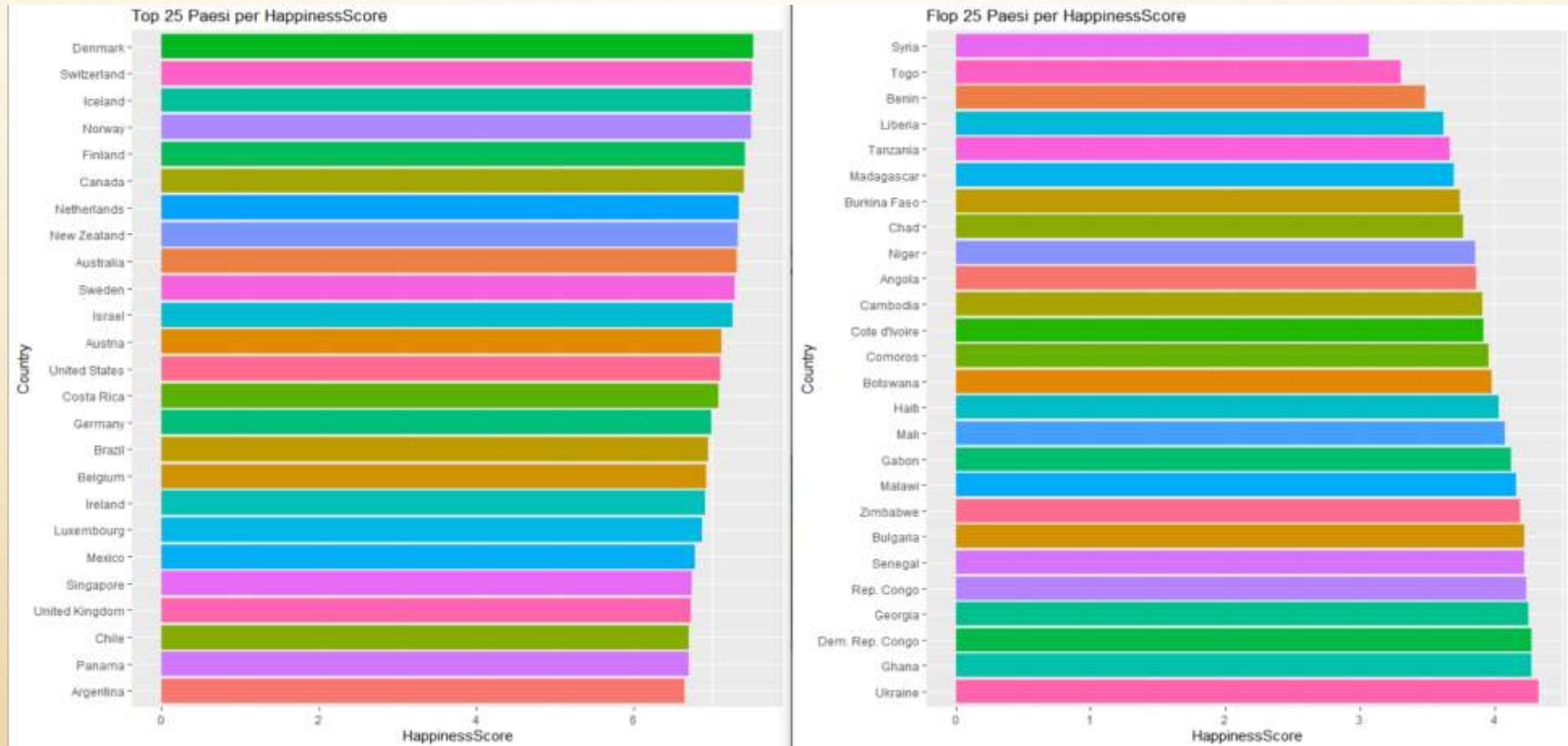


ANALISI ESPLORATIVA

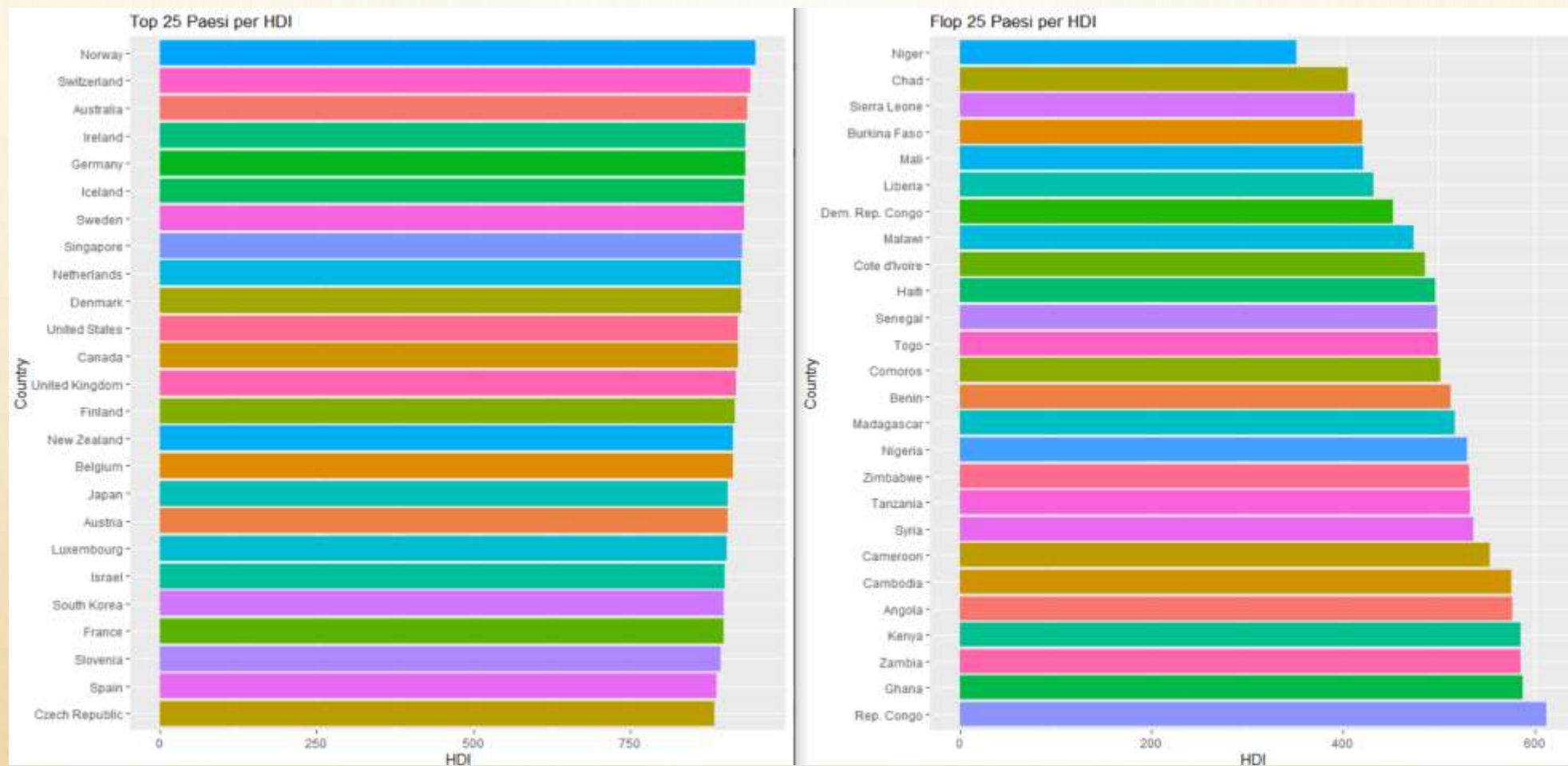


- DESCRIZIONE DATASET
- PULIZIA DEL DATASET
- STUDIO VARIABILI QUANTITATIVE
- BOXPLOT E DENSITY PLOT
- CORRELAZIONE TRA LE VARIABILI

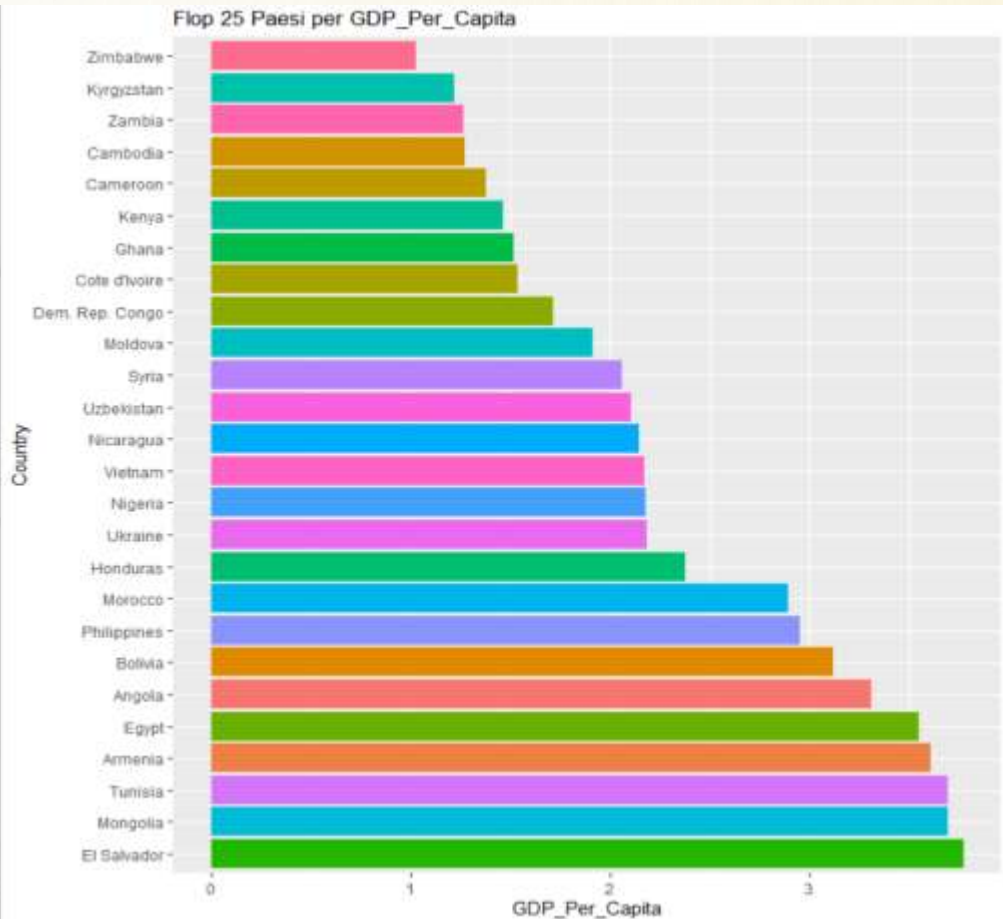
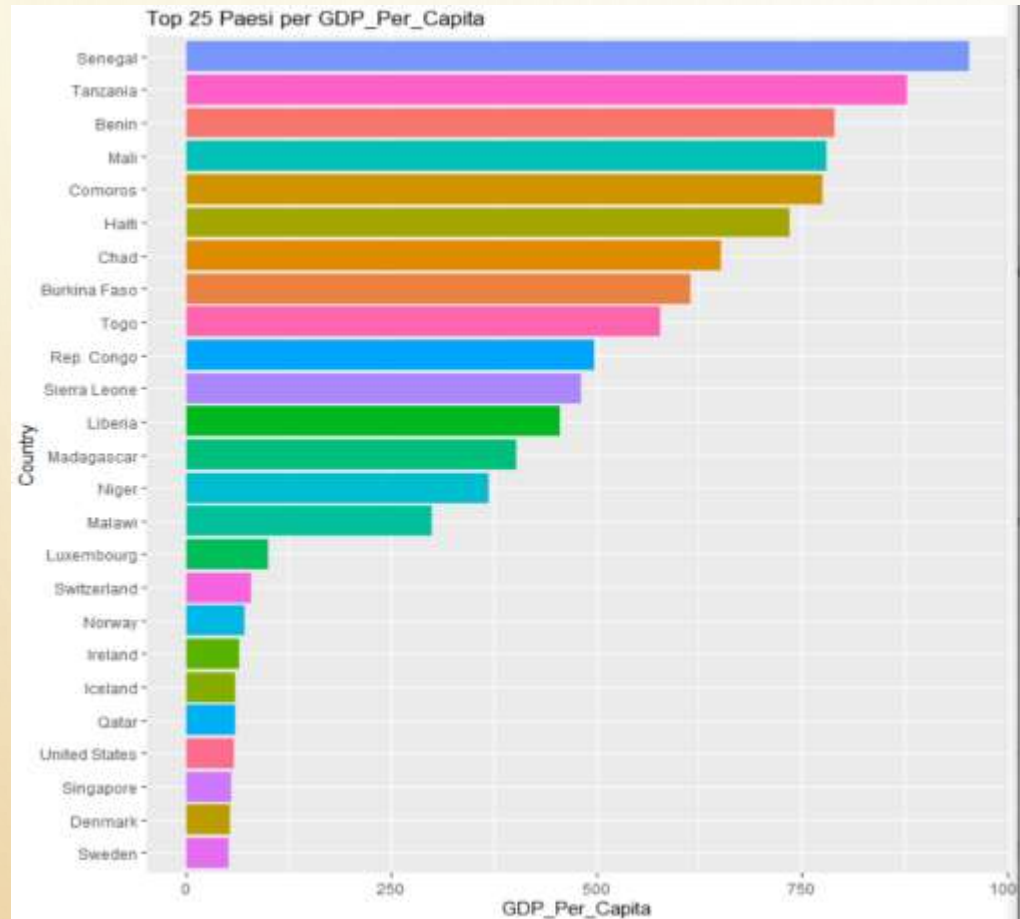
HAPPINESS SCORE



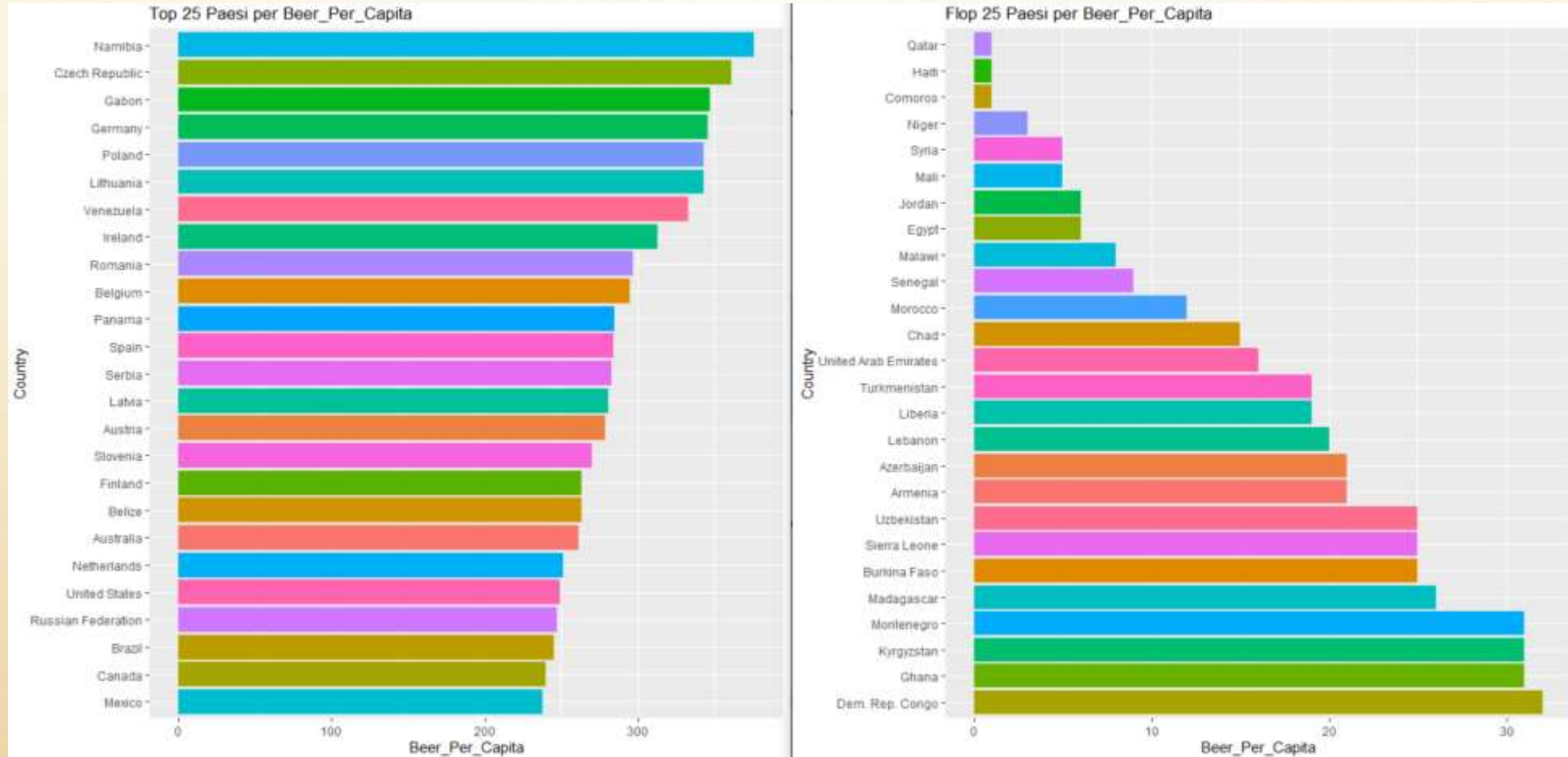
HUMAN DEVELOPMENT INDEX



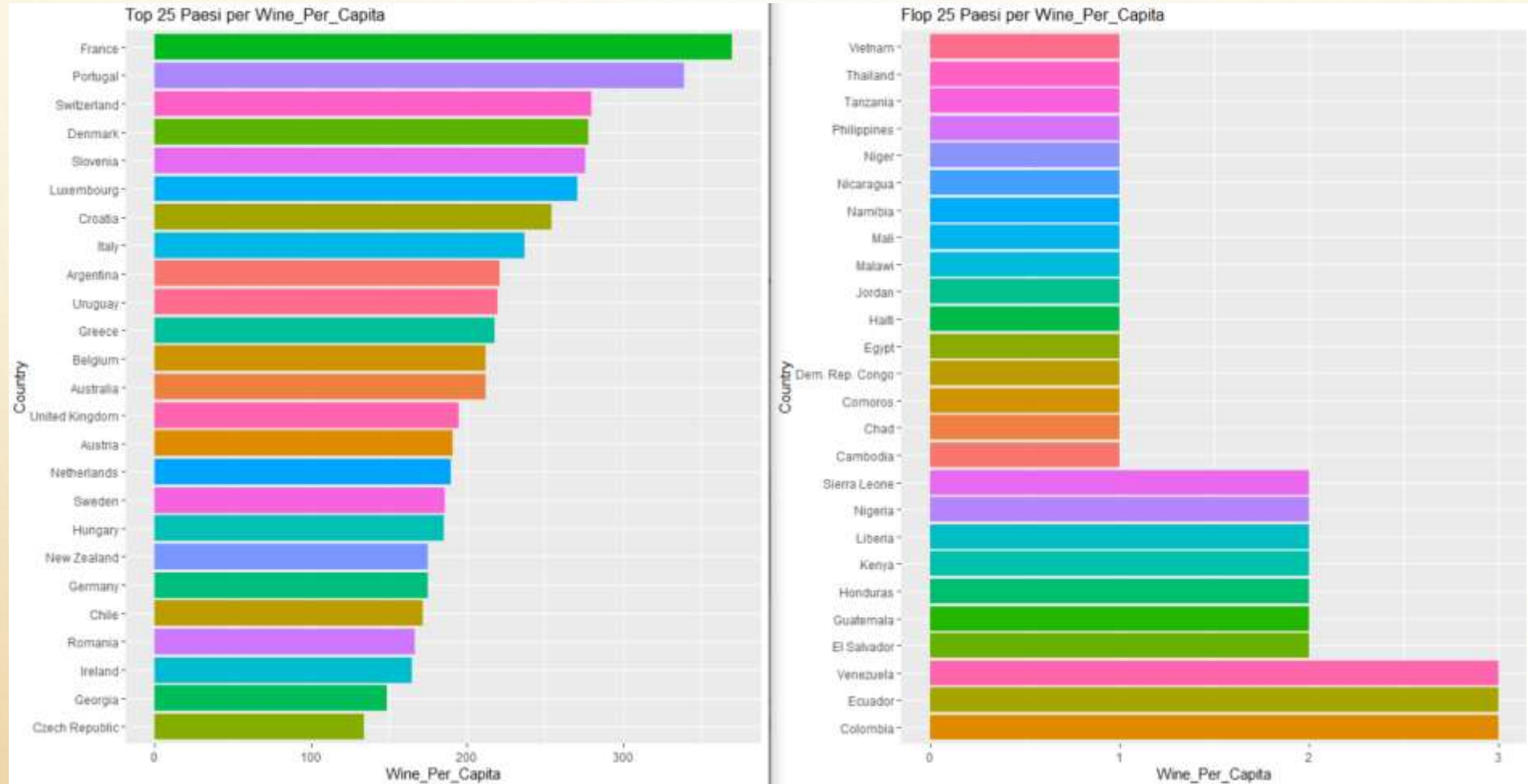
GDP PER CAPITA



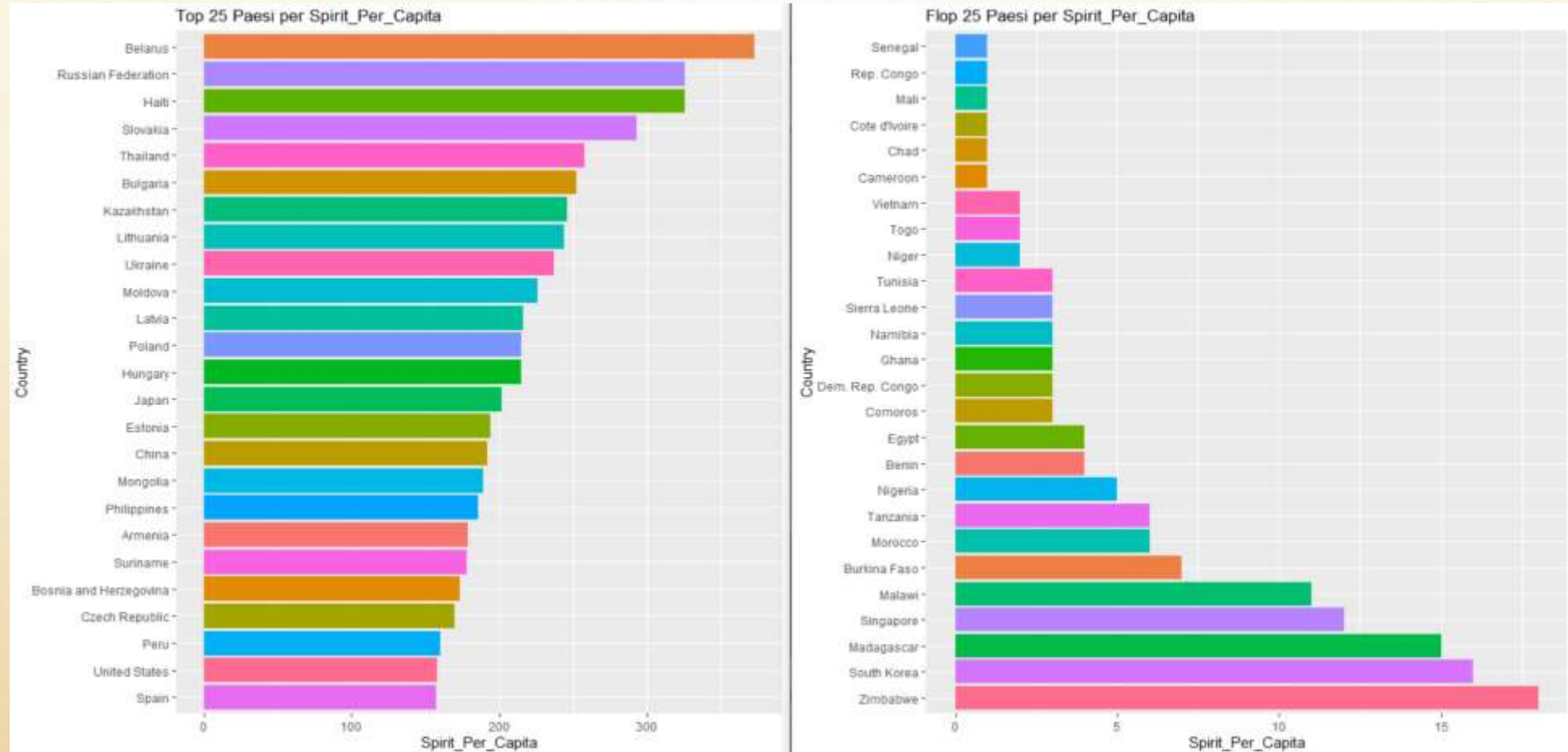
BEER PER CAPITA

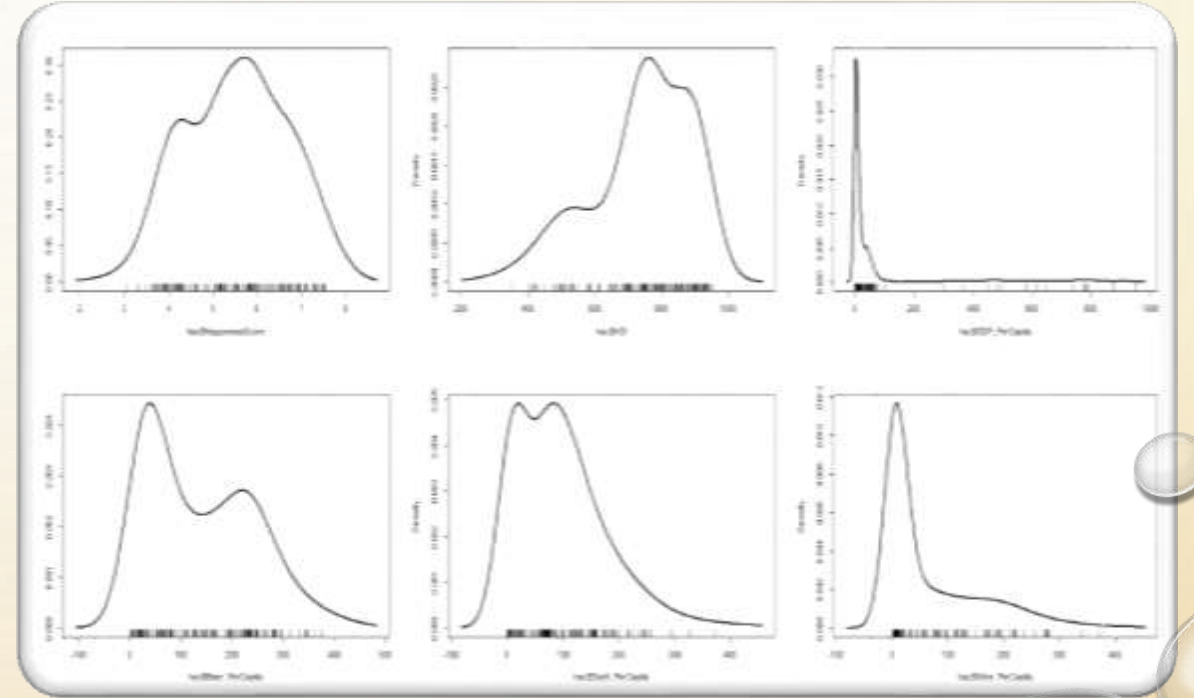
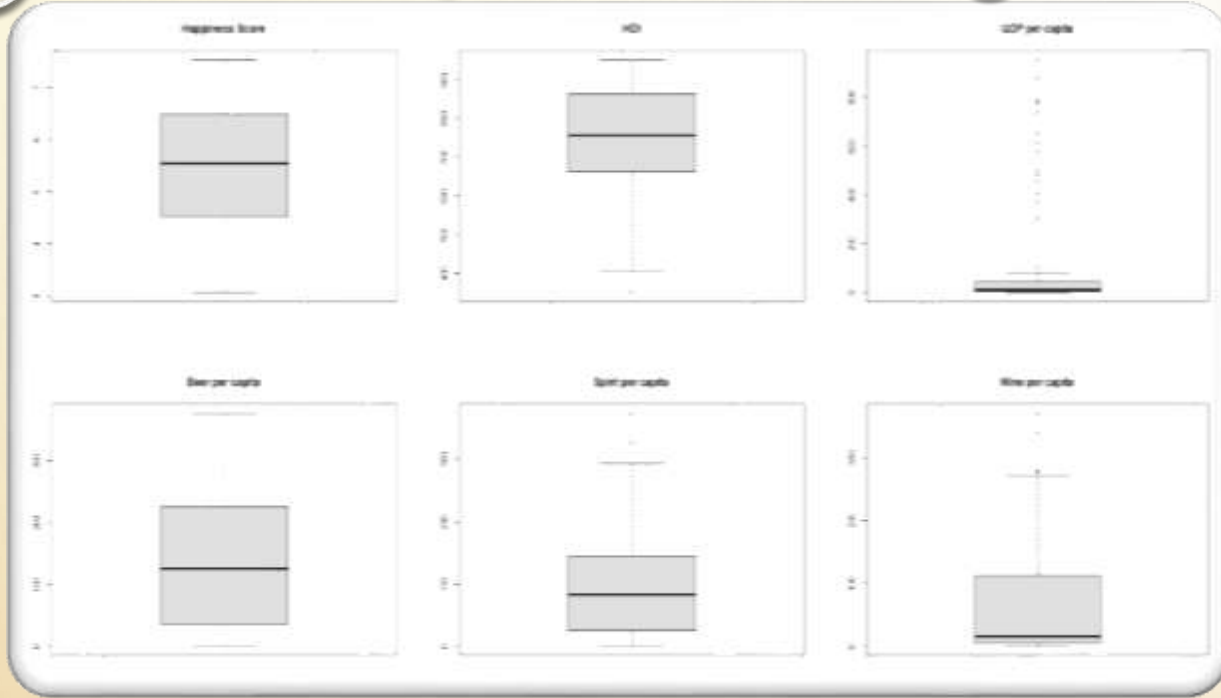


WINE PER CAPITA

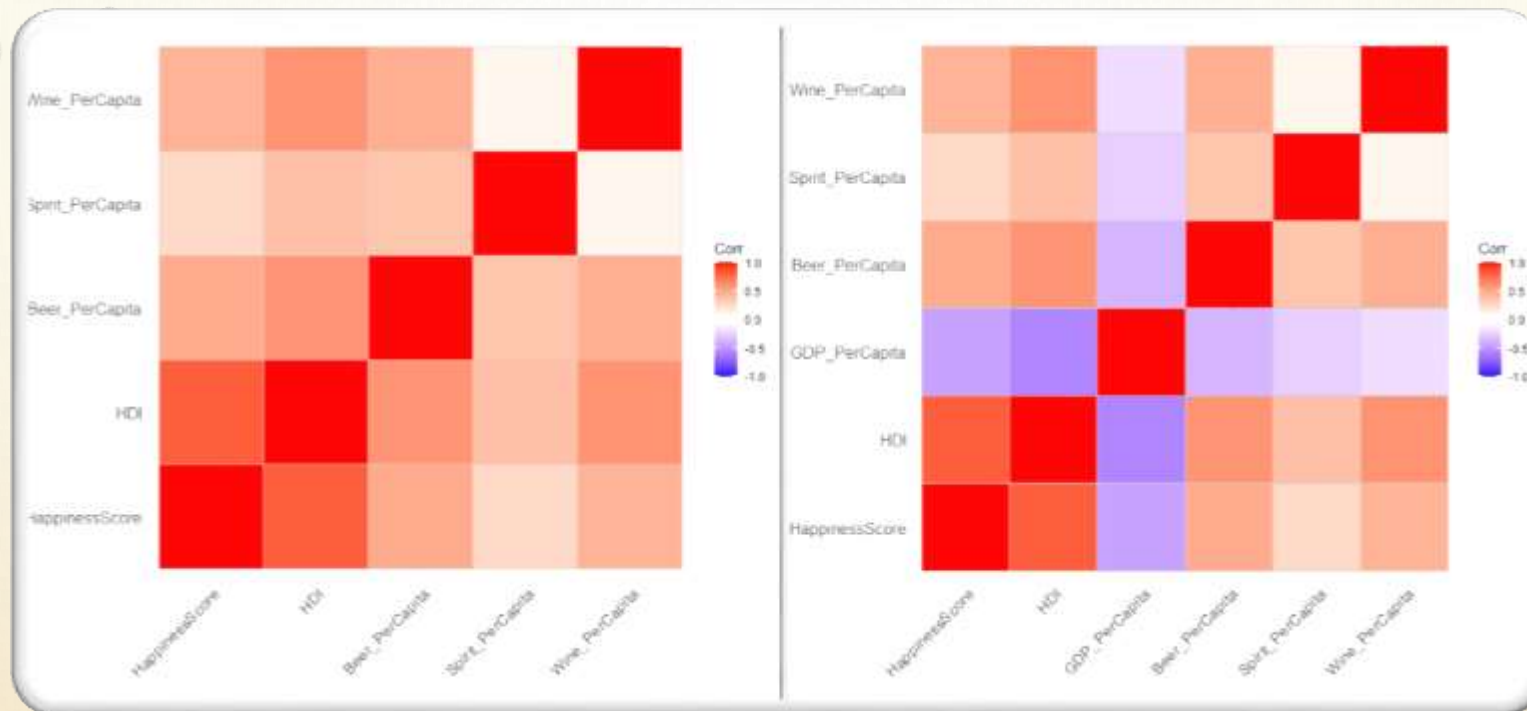


SPIRIT PER CAPITA





BOXPLOT E DENSITY PLOT

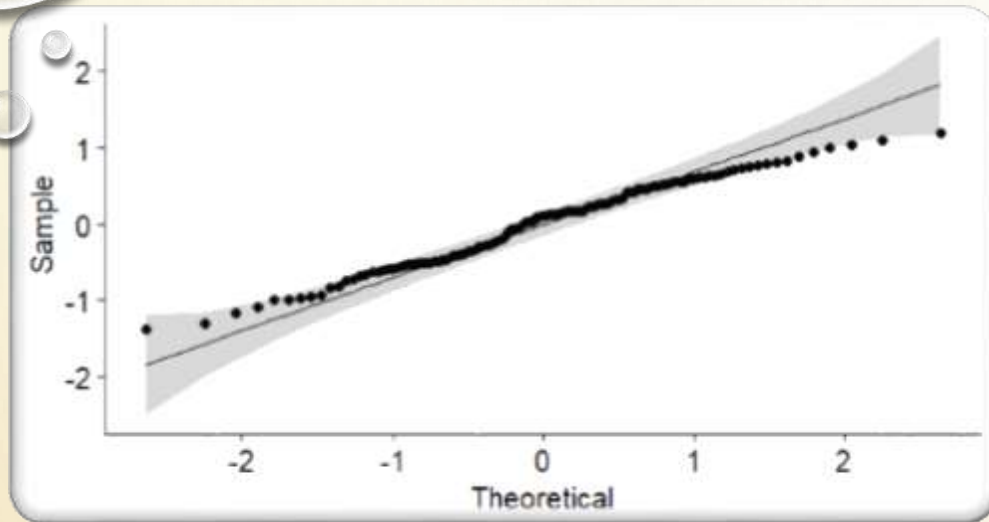


MATRICE DELLE CORRELAZIONI

REGRESSIONE LINEARE

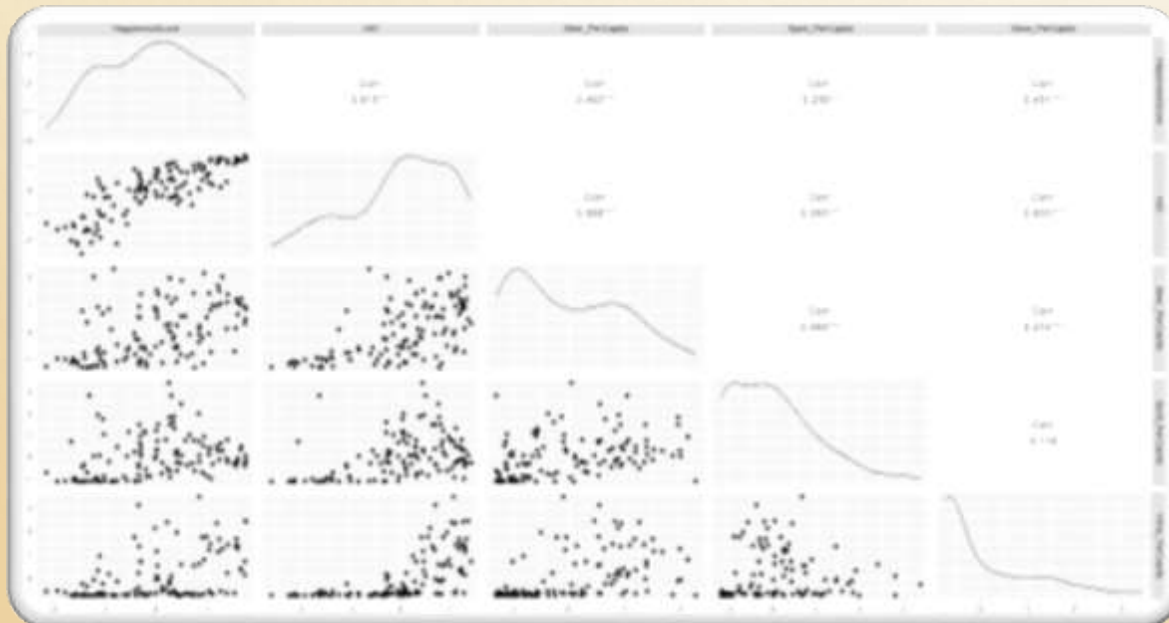
| | Estimate | Std. Error | t value | Pr(> t) | |
|------------------|------------|---------------|---------|-----------|-----|
| (Intercept) | -2.097e-16 | 5.246e-02 | 0.000 | 1.000 | |
| HDI | 8.756e-01 | 7.669e-02 | 11.418 | <2e-16 | *** |
| Beer_PerCapita | 5.806e-02 | 6.790e-02 | 0.855 | 0.394 | |
| Spirit_PerCapita | -9.824e-02 | 5.946e-02 | -1.652 | 0.101 | |
| Wine_PerCapita | -9.290e-02 | 6.857e-02 | -1.355 | 0.178 | |

- HAPPINESS SCORE VARIABLE Y
- HDI VARIABLE SIGNIFICATIVA
- R-SQUARED = 0.6754
- F-STATISTIC = 60.85
- P-VALUE: < 2.2E-16



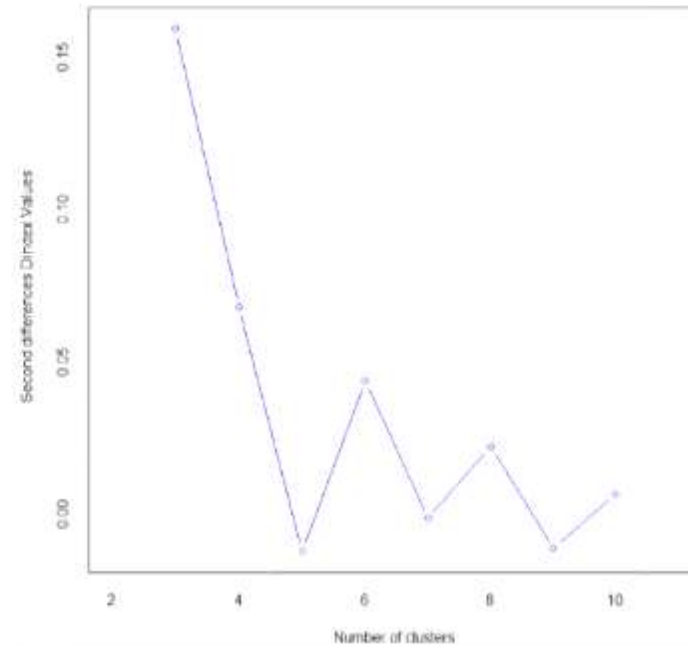
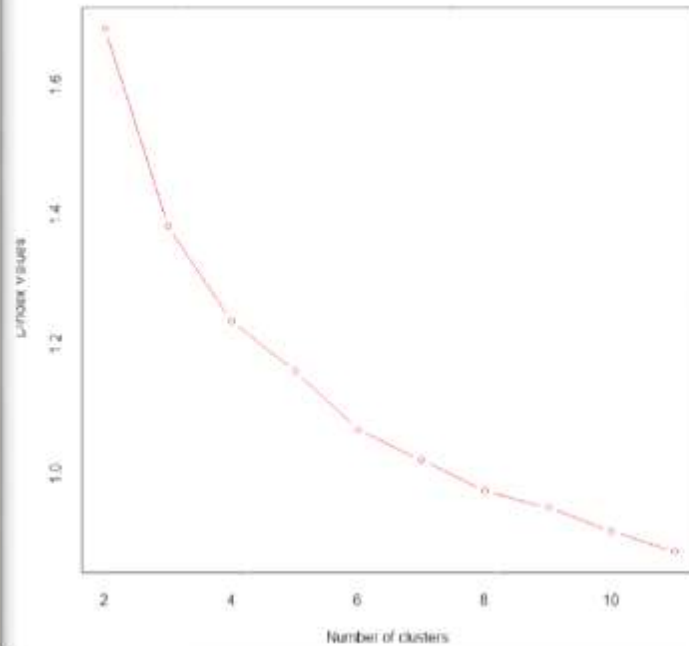
TEST

- LINEARITÀ TRA LE VARIABILI
- ANOVA TEST
- VIF TEST
- ISTOGRAMMA E GGQQPLOT
- NCV TEST E BP TEST



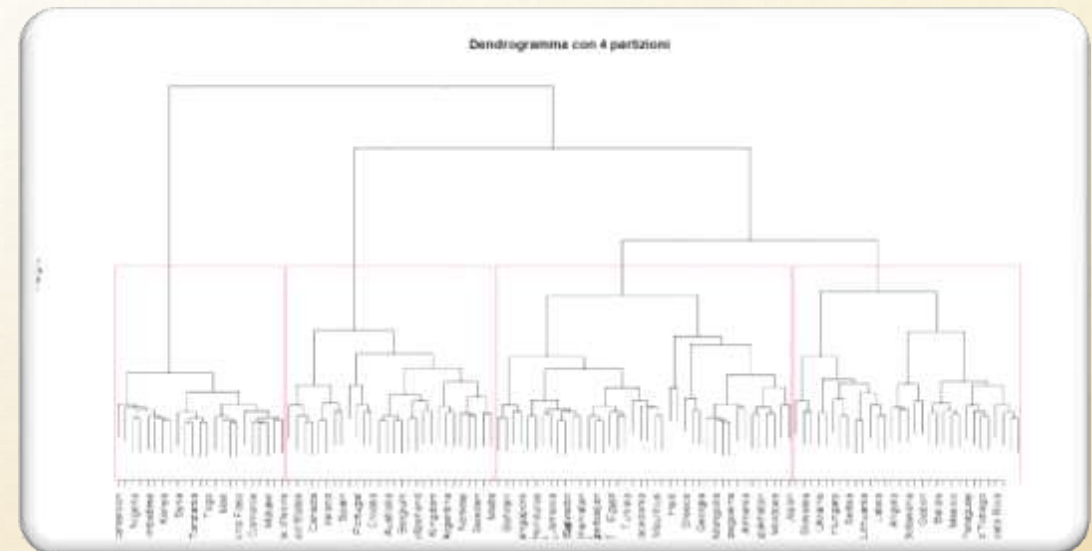
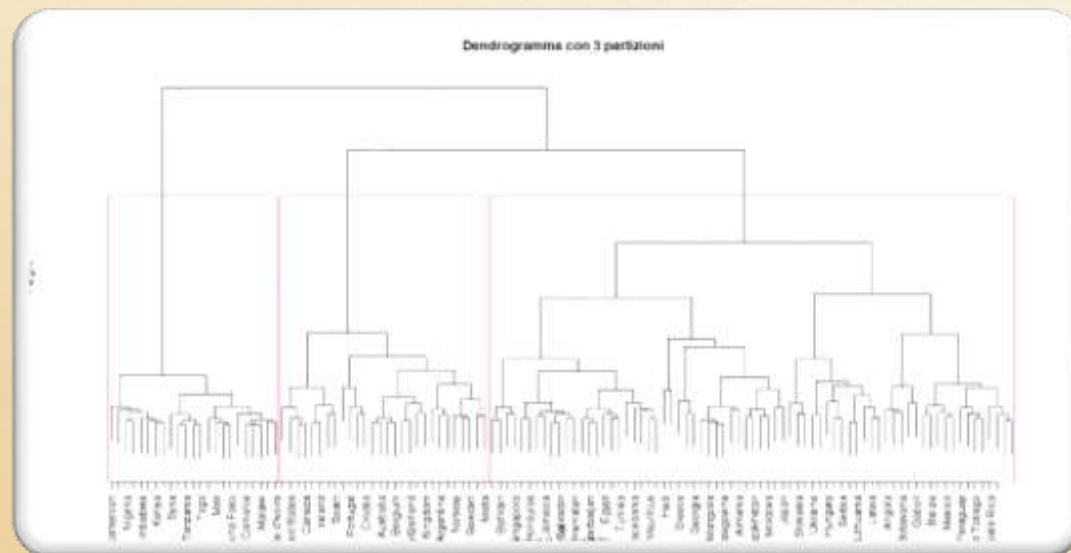
CLUSTER GERARCHICO

- DISTANZA EUCLIDEA
- METODO DI WARD
- HUBERT INDEX E D-INDEX
- 3 O 4 GRUPPI?



DENDROGRAMMA

- CLUSTERIZZAZIONE A 4 GRUPPI PIÙ OMOGENEA
- COME SONO CLASSIFICATI I PAESI?



```
> print(gruppi[[1]])
```

| | | | |
|---------|---------------|-----------------|------------|
| Nigeria | Zambia | Sierra Leone | Cameroon |
| 86 | 88 | 90 | 92 |
| Kenya | Ghana | Dem. Rep. Congo | Rep. Congo |
| 96 | 98 | 99 | 101 |
| Senegal | Zimbabwe | Malawi | Mali |
| 102 | 104 | 105 | 107 |
| Comoros | Cote d'Ivoire | Cambodia | Niger |
| 110 | 111 | 112 | 114 |
| Chad | Burkina Faso | Madagascar | Tanzania |
| 115 | 116 | 117 | 118 |
| Benin | Togo | Syria | |
| 120 | 121 | 122 | |

```
> print(gruppi[[2]])
```

| | | | | |
|------------|----------------|-------------|-----------|----------------|
| Denmark | Switzerland | Iceland | Norway | Finland |
| 1 | 2 | 3 | 4 | 5 |
| Canada | Netherlands | New Zealand | Australia | Sweden |
| 6 | 7 | 8 | 9 | 10 |
| Austria | United States | Germany | Belgium | Ireland |
| 12 | 13 | 15 | 17 | 18 |
| Luxembourg | United Kingdom | Chile | Argentina | Czech Republic |
| 19 | 22 | 23 | 25 | 26 |
| Uruguay | Malta | France | Spain | Italy |
| 28 | 29 | 31 | 34 | 44 |
| Slovenia | Croatia | Portugal | | |
| 56 | 66 | 80 | | |

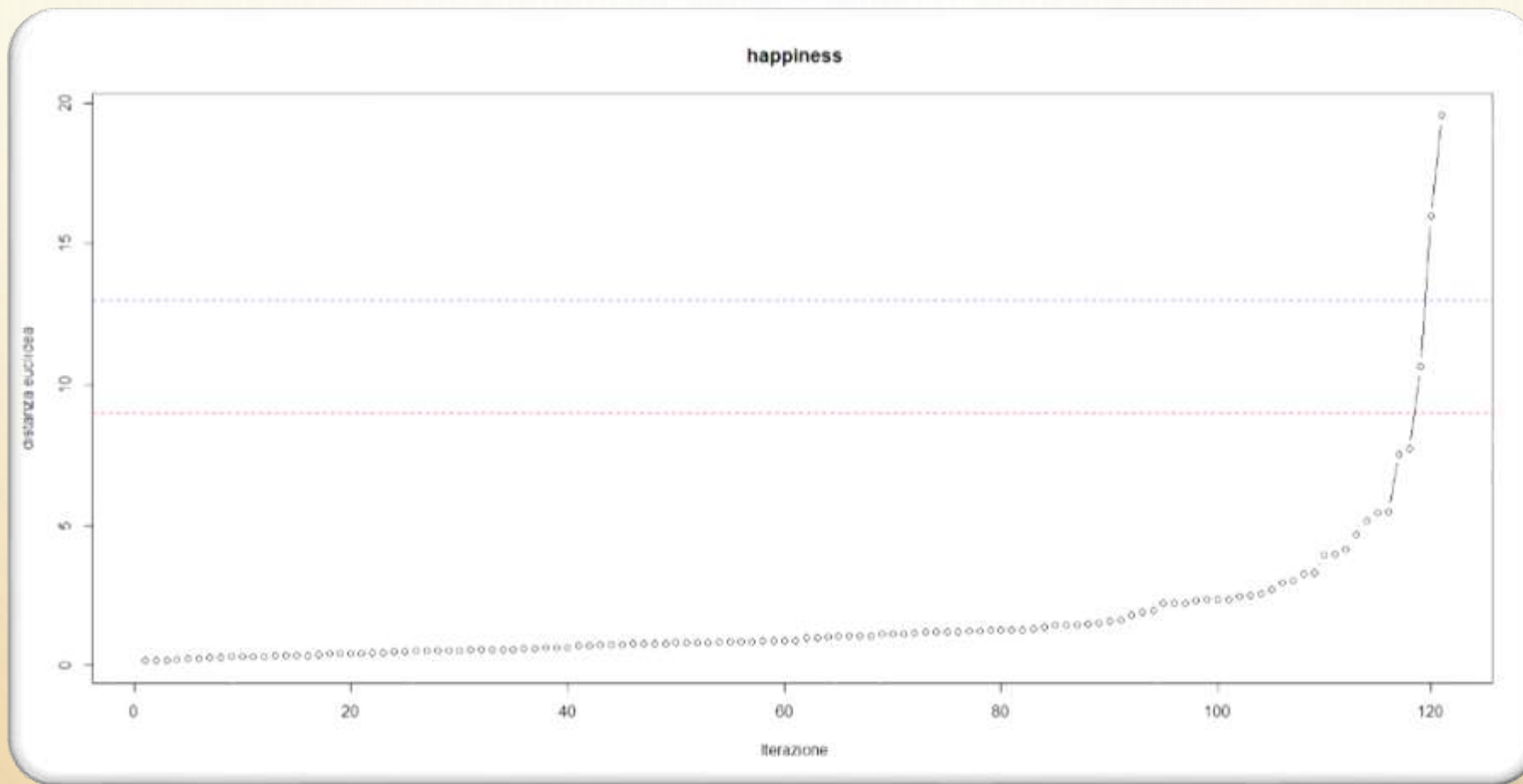
```
> print(gruppi[[3]])
```

| | | |
|------------------------|------------|----------------------|
| Israel | Singapore | United Arab Emirates |
| 11 | 21 | 27 |
| Thailand | Qatar | Guatemala |
| 32 | 33 | 35 |
| Suriname | Bahrain | El Salvador |
| 36 | 37 | 41 |
| Nicaragua | Uzbekistan | Japan |
| 42 | 43 | 47 |
| Kazakhstan | Moldova | South Korea |
| 48 | 49 | 52 |
| Turkmenistan | Mauritius | Jamaica |
| 58 | 59 | 65 |
| Turkey | Jordan | Azerbaijan |
| 67 | 68 | 69 |
| Philippines | China | Kyrgyzstan |
| 70 | 71 | 72 |
| Bosnia and Herzegovina | Montenegro | Morocco |
| 74 | 75 | 77 |
| Lebanon | Macedonia | Vietnam |
| 79 | 81 | 82 |
| Tunisia | Greece | Mongolia |
| 83 | 84 | 85 |
| Honduras | Albania | Egypt |
| 87 | 89 | 94 |
| Armenia | Georgia | Haiti |
| 95 | 100 | 108 |
| Liberia | | |
| 110 | | |

```
> print(gruppi[[4]])
```

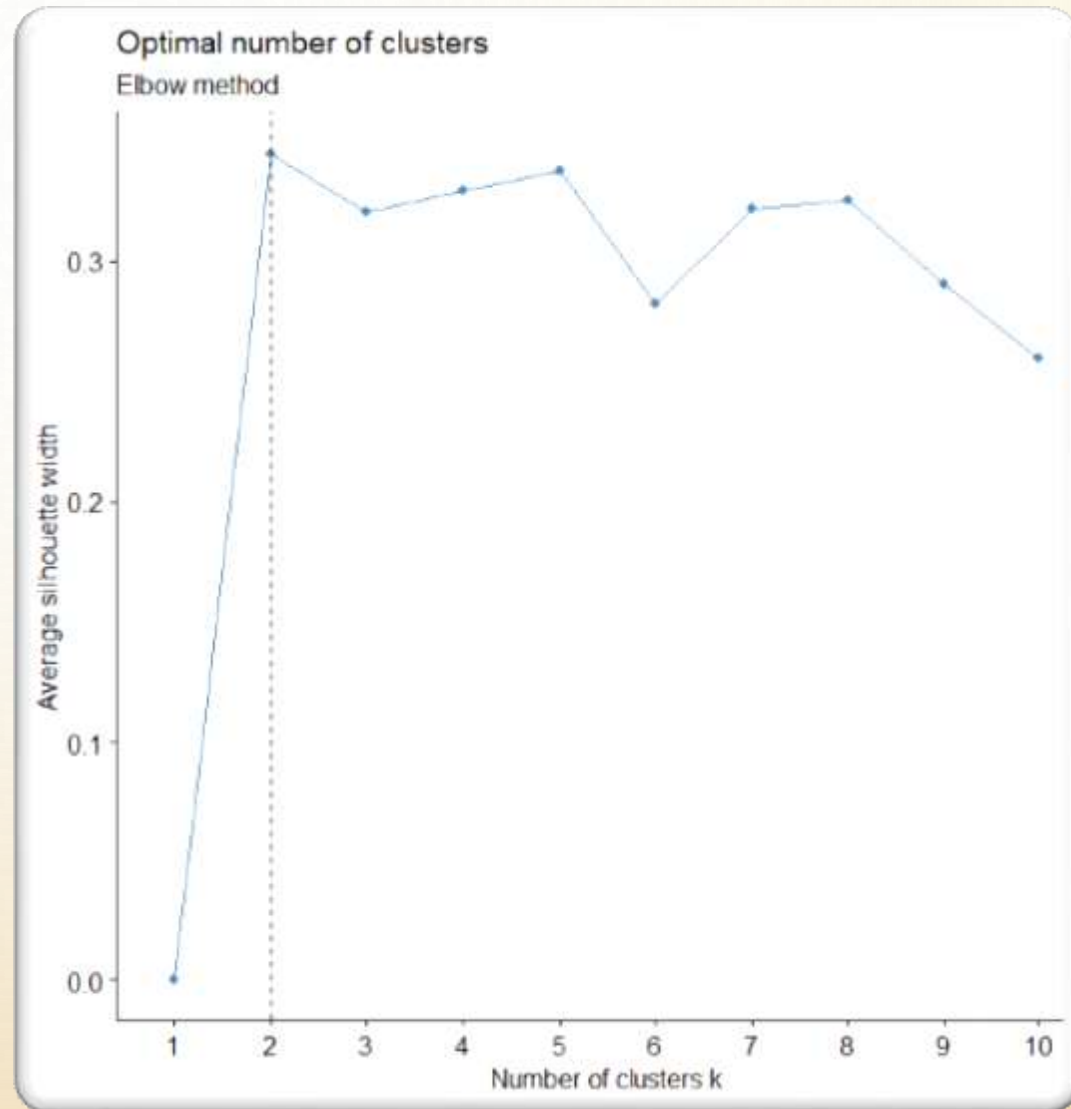
| | | |
|------------|--------------------|---------------------|
| Costa Rica | Brazil | Mexico |
| 14 | 16 | 20 |
| Panama | Colombia | Trinidad and Tobago |
| 24 | 30 | 38 |
| Venezuela | Slovakia | Ecuador |
| 39 | 40 | 45 |
| Belize | Russian Federation | Poland |
| 46 | 50 | 51 |
| Bolivia | Lithuania | Belarus |
| 53 | 54 | 55 |
| Peru | Latvia | Cyprus |
| 57 | 60 | 61 |
| Paraguay | Romania | Estonia |
| 62 | 63 | 64 |
| Serbia | Dominican Republic | Hungary |
| 73 | 76 | 78 |
| Namibia | South Africa | Ukraine |
| 91 | 93 | 97 |
| Bulgaria | Gabon | Botswana |
| 103 | 106 | 109 |
| Angola | | |
| 113 | | |

GRAFICO DELLE DISTANZE

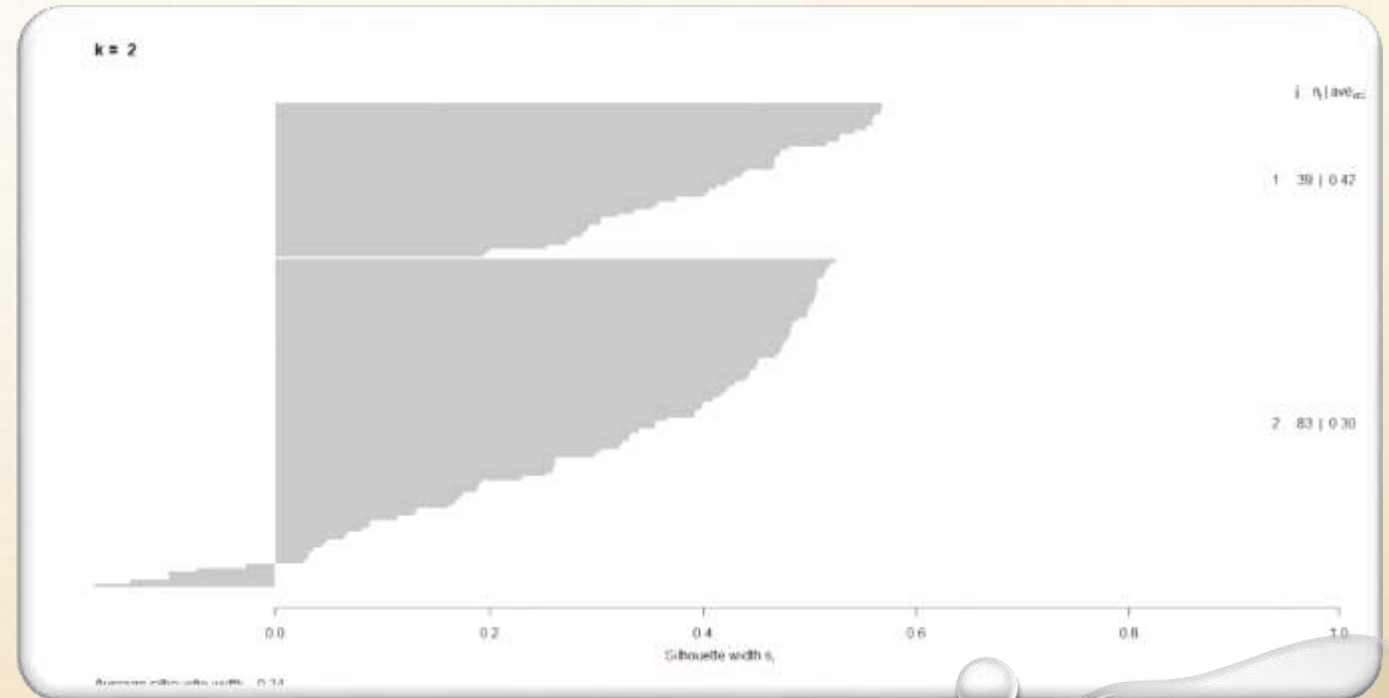
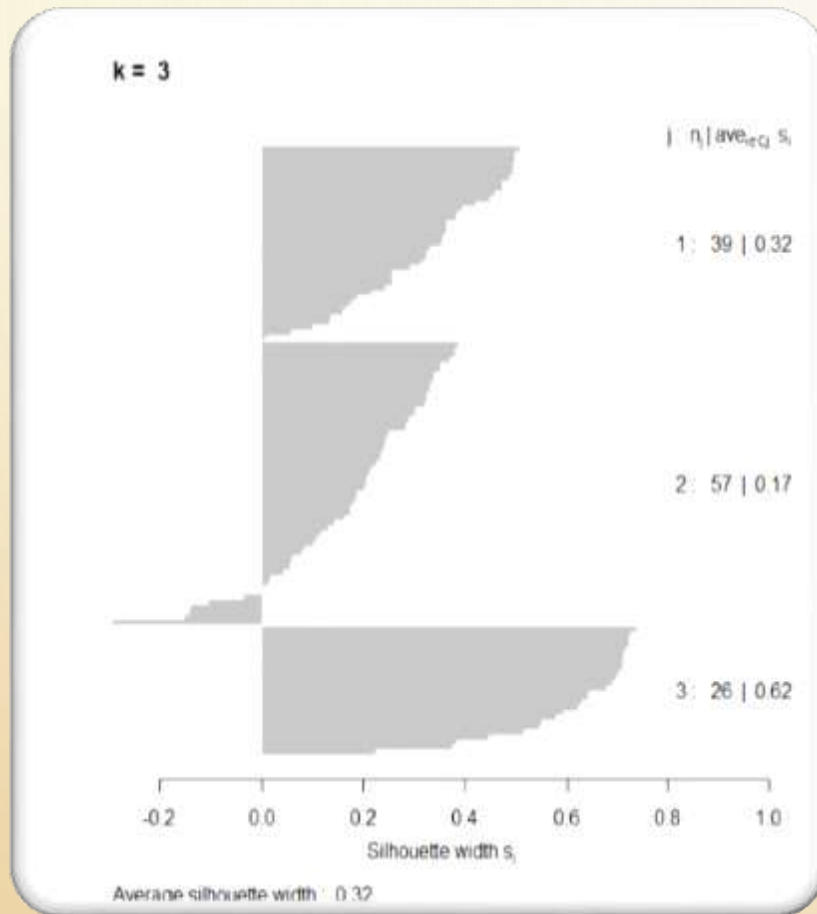


CLUSTER NON GERARCHICO

- IL NUMERO OTTIMALE DI GRUPPI È 2
- ULTERIORE ANALISI CON 3 GRUPPI
- CONFRONTO CON GERARCHICO

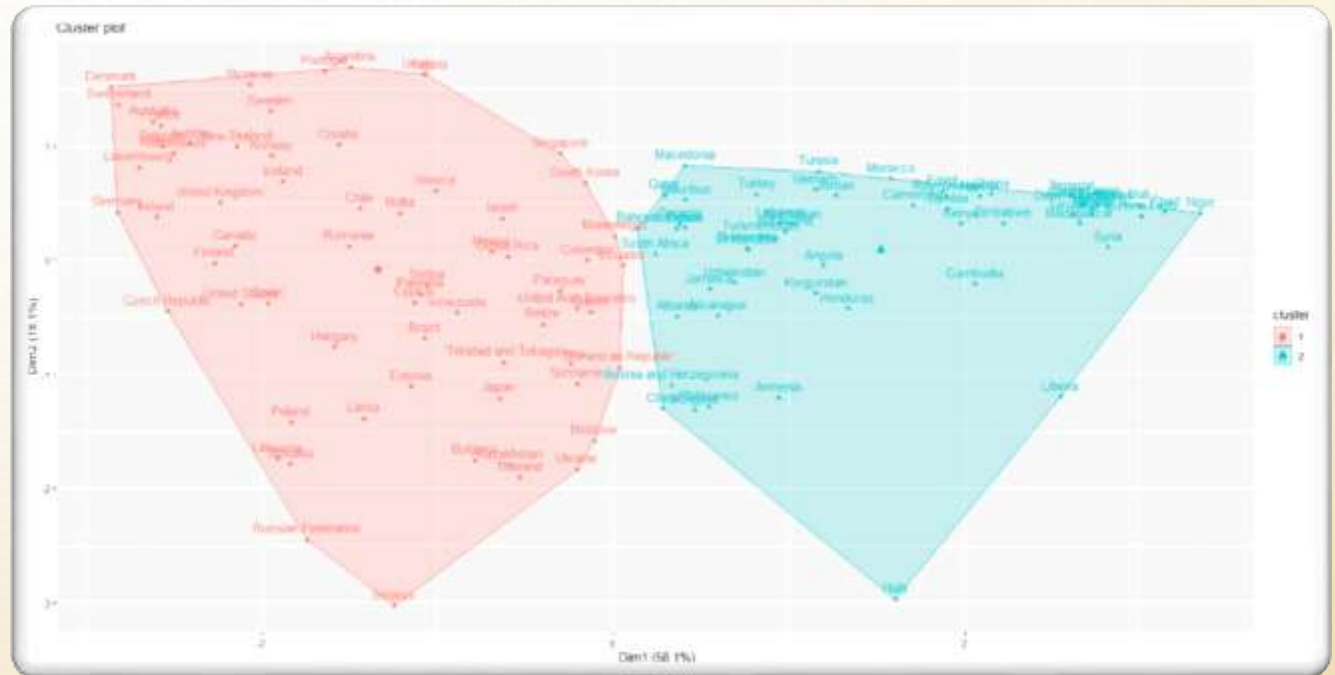


SILHOUETTE



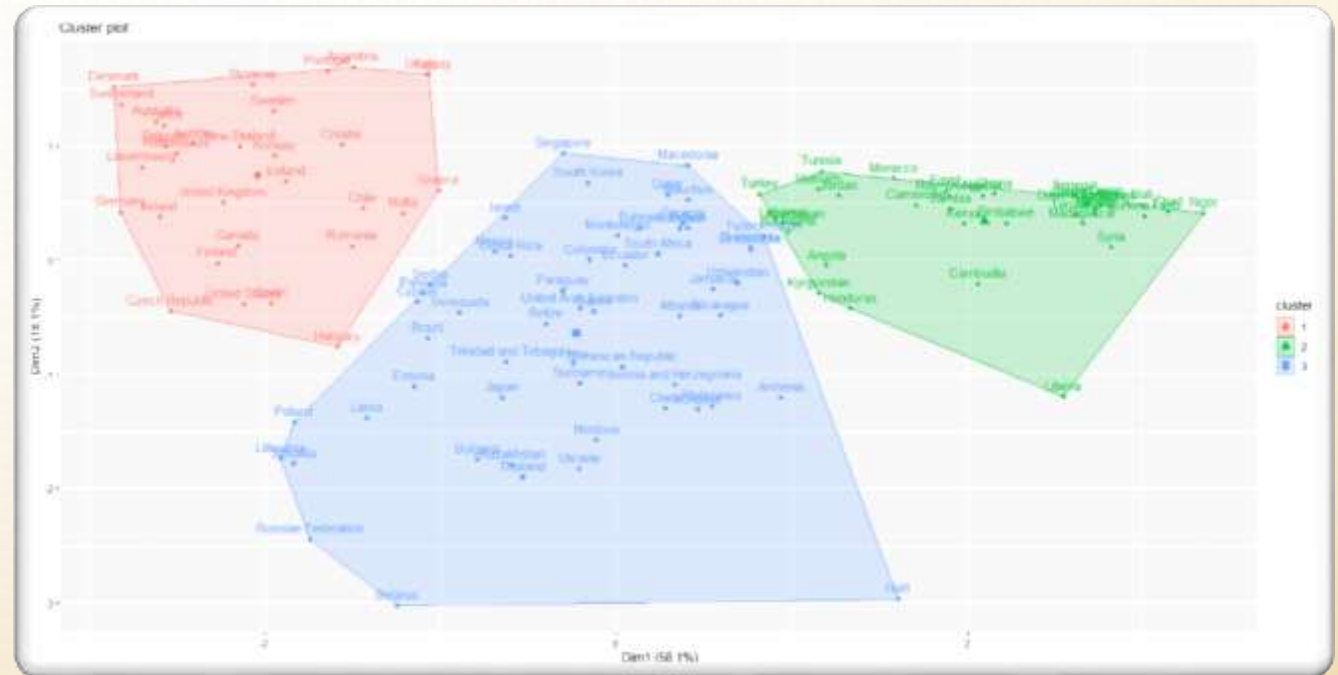
K-MEANS CON 2 CLUSTER

- DEVIANZA BETWEEN = 249.56
- DEVIANZA WITHIN = 221.06 – 134.37
- UNITÀ IN OGNI CLUSTER: 65 E 57



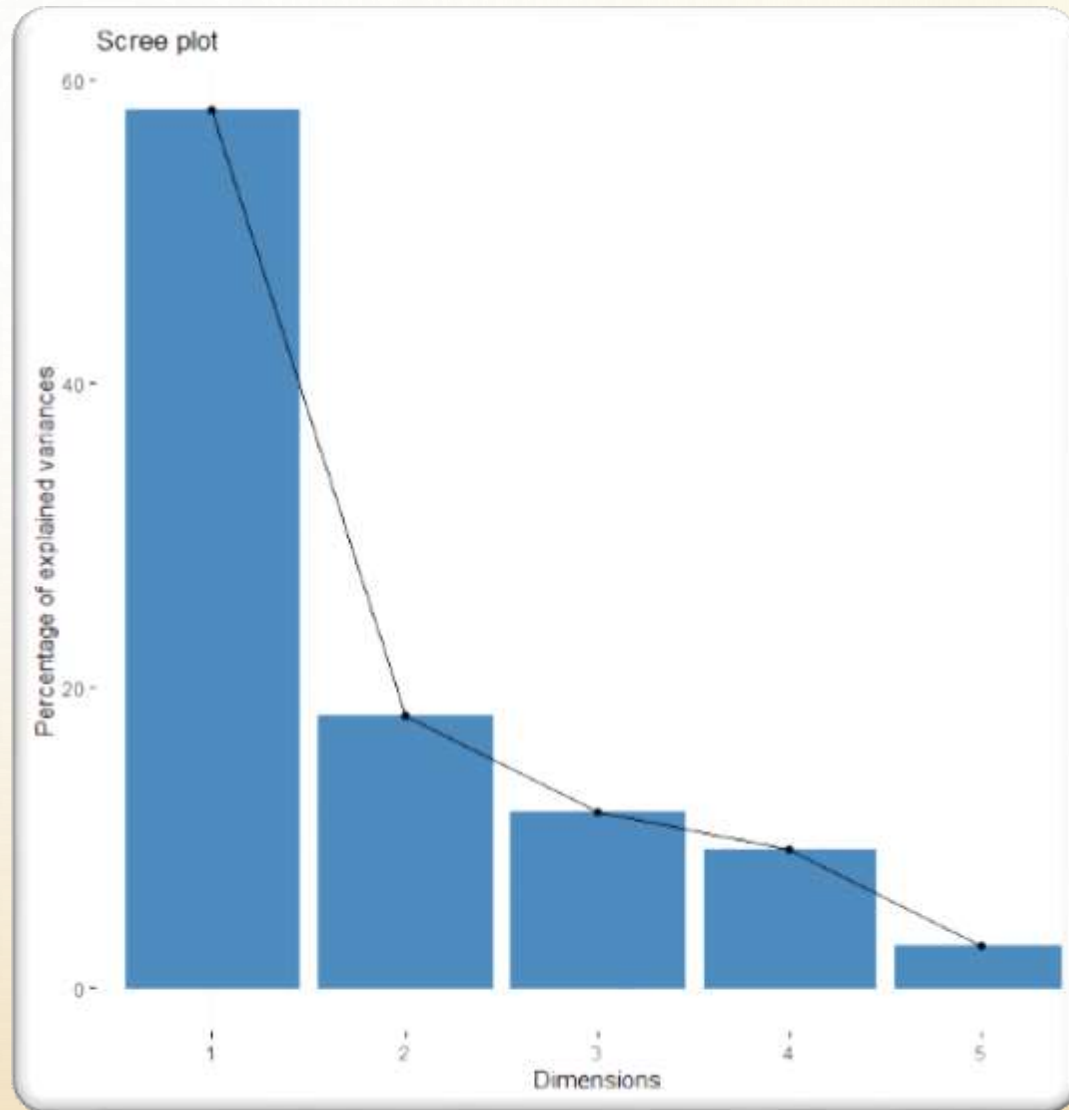
K-MEANS CON 3 CLUSTER

- DEVIANZA BETWEEN = 341.46
- DEVIANZA WITHIN = 57.07 – 44.93 – 161.52
- UNITÀ IN OGNI CLUSTER: 31, 36 E 55
 - ELEPHANT CLUSTER?



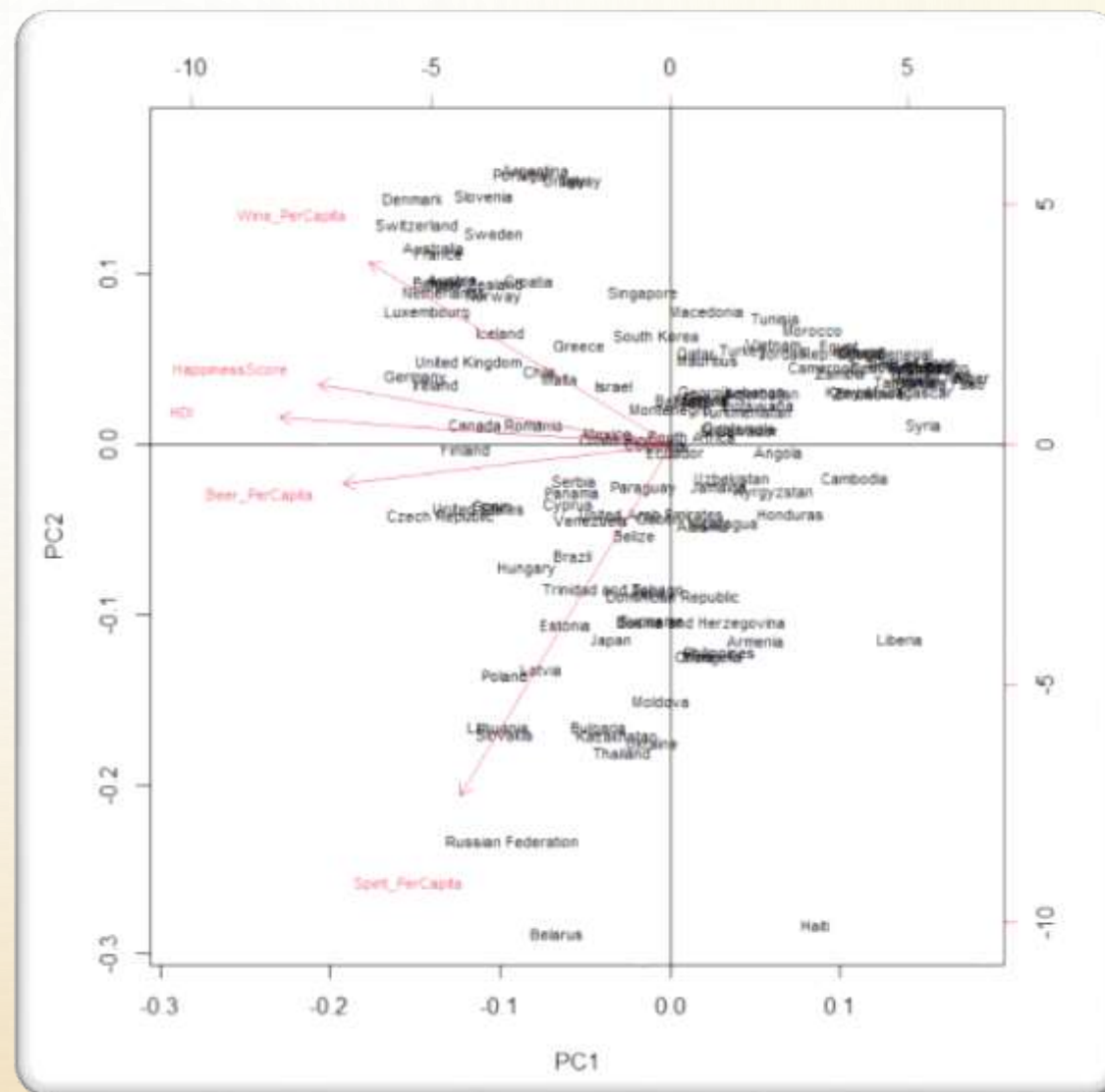
PCA

- LE PRIME 2 COMPONENT SPIEGANO QUASI L'80% DELLA VARIANZA
- LA PRIMA COMPONENTE SPIEGA IL 58%

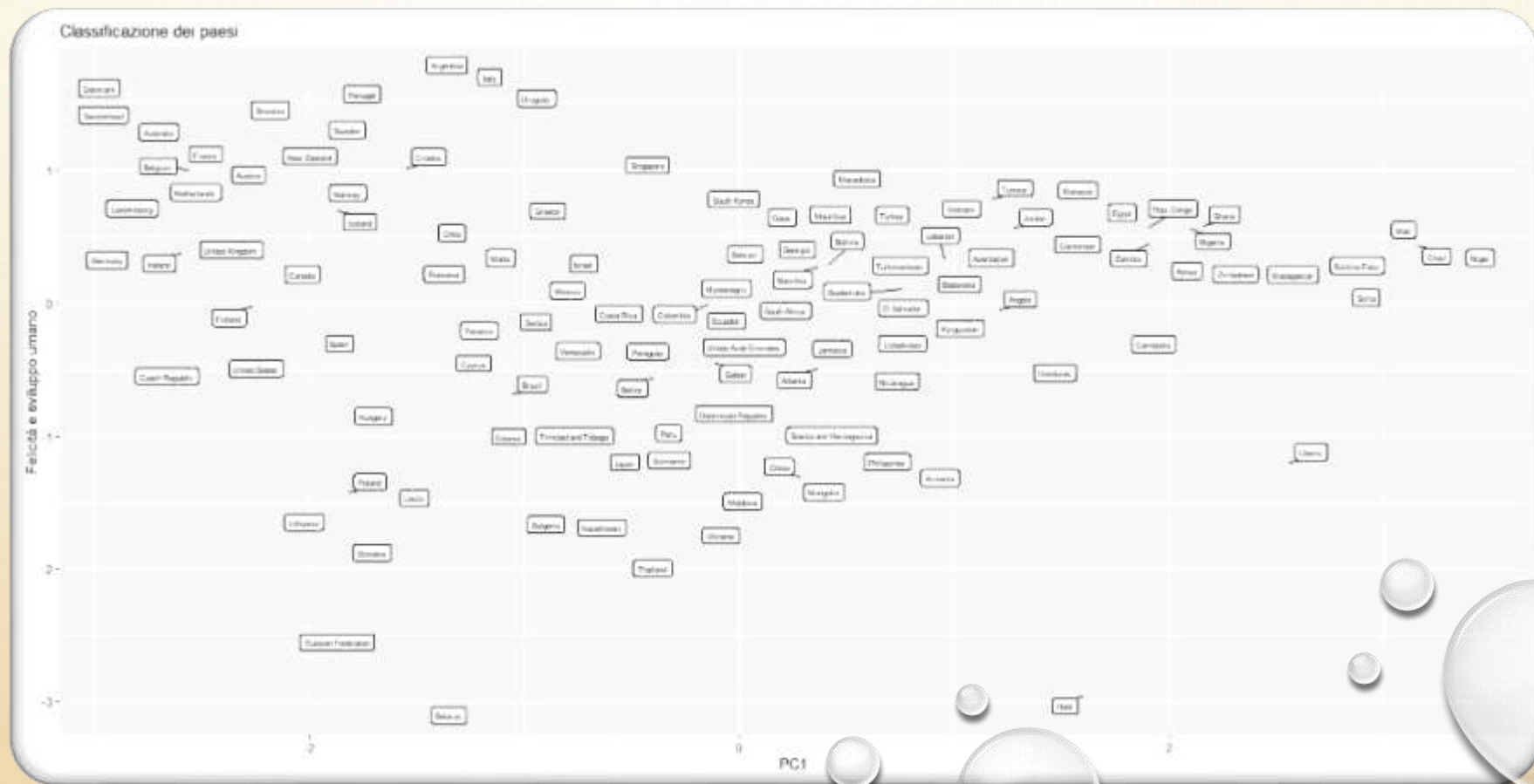


BIPLOT

- COME INTERPRETARE LE COMPONENTI PRINCIPALI?
 - PC1 = ??
- PC2 = FELICITÀ E SVILUPPO UMANO



CLASSIFICAZIONE DEI PAESI




```
Call:
glm(formula = Hemisphere ~ Beer_PerCapita + Spirit_PerCapita +
    HDI, family = "binomial", data = train)
```

Deviance Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|--------|--------|
| -2.4225 | 0.1163 | 0.4091 | 0.7030 | 1.2873 |

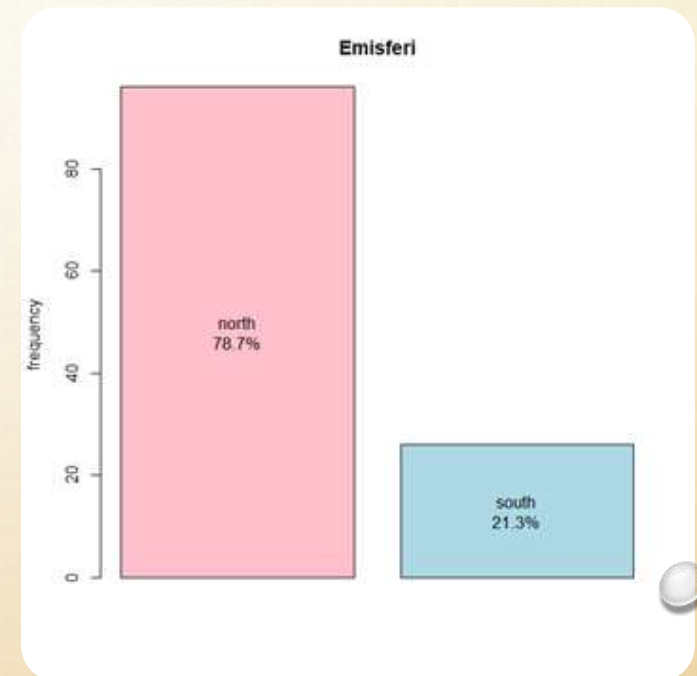
Coefficients:

| | Estimate | Std. Error | z value | Pr(> z) |
|------------------|-----------|------------|---------|----------|
| (Intercept) | -2.196316 | 1.552185 | -1.415 | 0.1571 |
| Beer_PerCapita | -0.009859 | 0.004352 | -2.266 | 0.0235 * |
| Spirit_PerCapita | 0.016088 | 0.006872 | 2.341 | 0.0192 * |
| HDI | 0.005181 | 0.002768 | 1.872 | 0.0612 . |

```
> varImp(logit2)
```

| | Overall |
|------------------|----------|
| Beer_PerCapita | 2.265563 |
| Spirit_PerCapita | 2.341091 |
| HDI | 1.871687 |

REGRESSIONE LOGISTICA



TEST

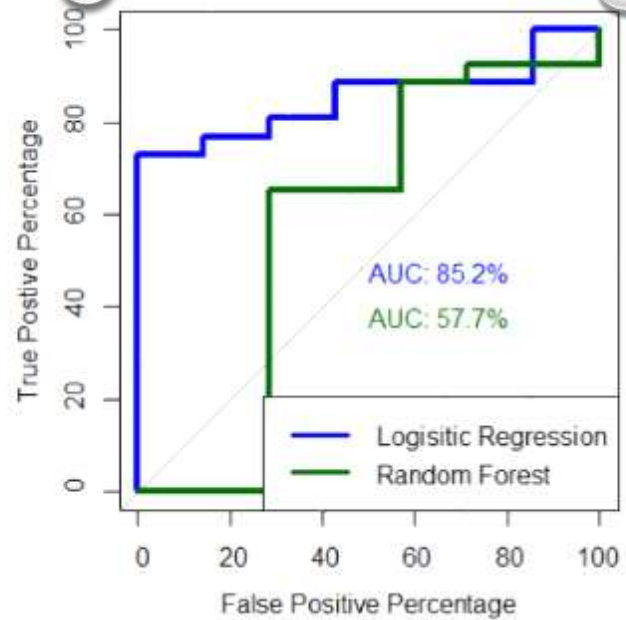
- WALD TEST
- MCFADDEN
- VIF

Wald test:

Chi-squared test:

$X^2 = 8.9$, $df = 3$, $P(> X^2) = 0.03$

```
> pR2(logit2)["McFadden"]
fitting null model for pseudo-r2
McFadden
0.2011446
> vif(logit2)
Beer_PerCapita Spirit_PerCapita HDI
2.746580 1.807144 2.439175
```

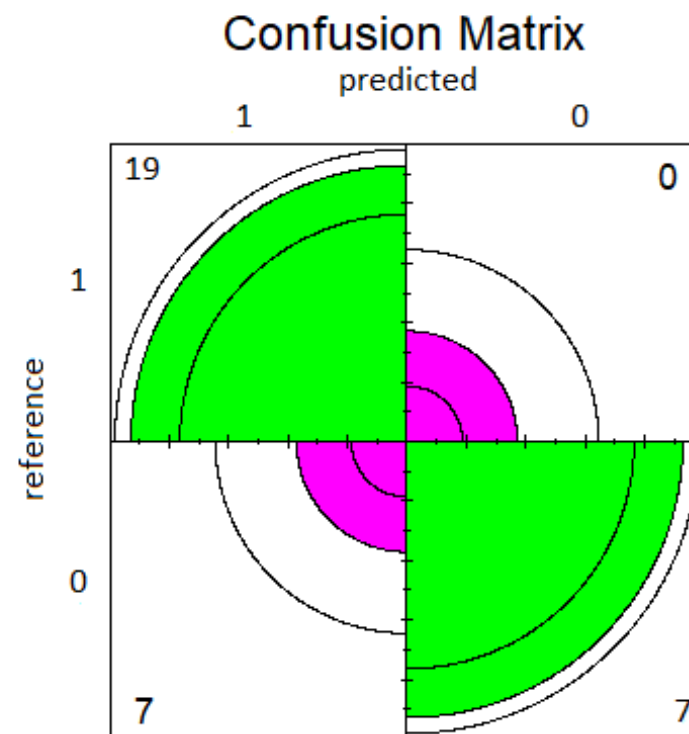


```
> roc.df[roc.df$tpp > 65 & roc.df$tpp < 76,]  
      tpp      fpp thresholds  
14 73.07692 14.28571  0.8547989  
15 73.07692  0.00000  0.8638175  
16 69.23077  0.00000  0.8682961  
17 65.38462  0.00000  0.8802020  
> |
```

ROC CURVE

CONFUSION MATRIX

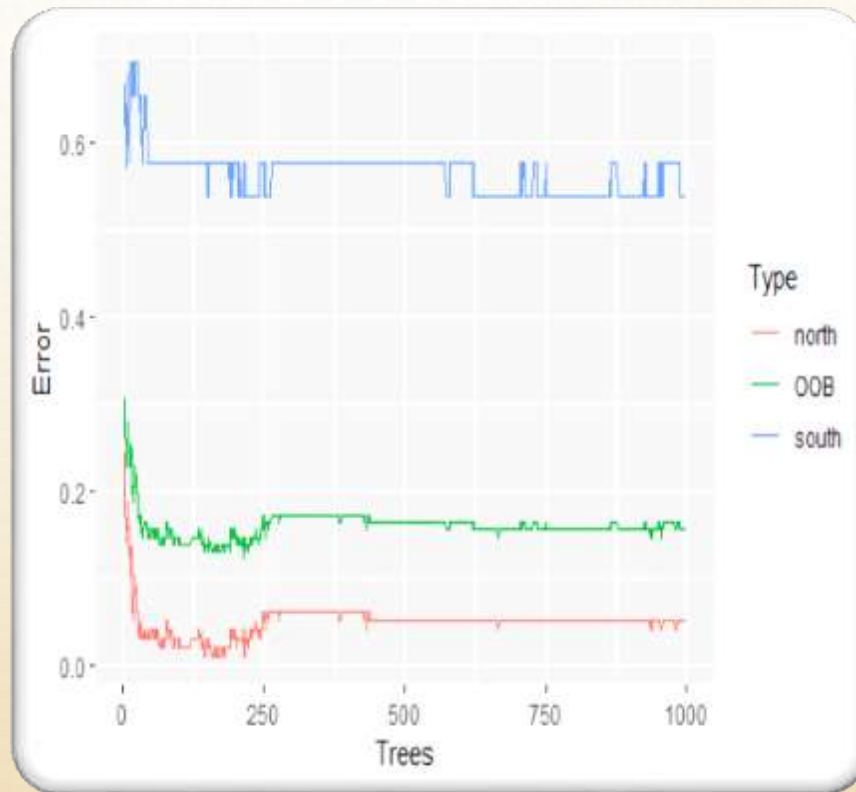
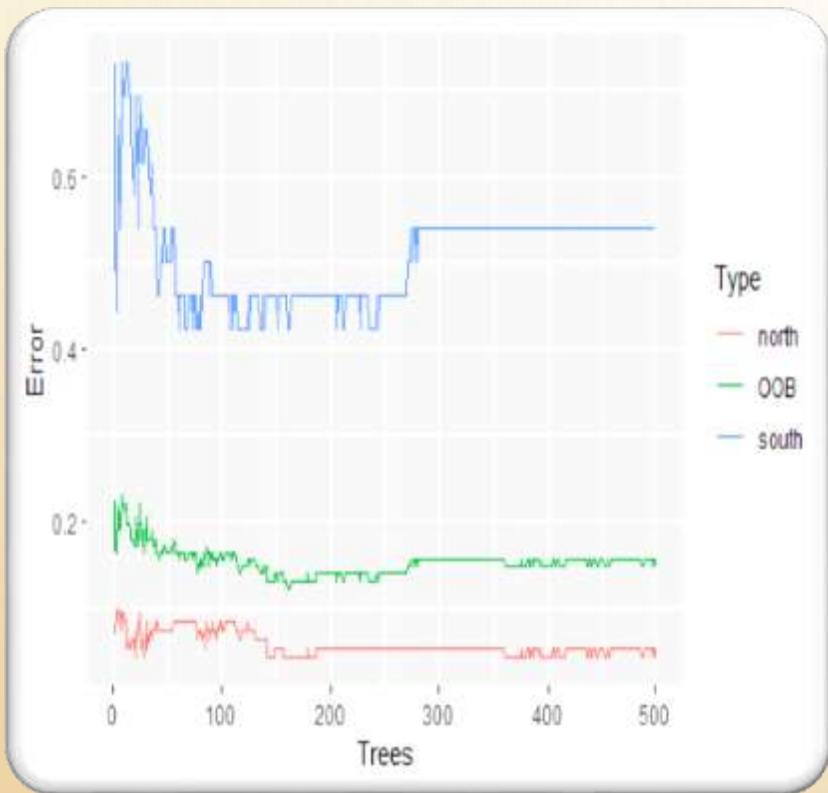
| | |
|--------------------|--------------|
| Accuracy | 0.789 |
| Precision | 0.731 |
| Sensitivity | 1.000 |
| Specificity | 0.500 |



The background is a light beige color. It features a faint, repeating pattern of binary code (0s and 1s) in a light greenish-yellow hue. Overlaid on this is a thin, white line graph that fluctuates across the upper half of the image. Scattered throughout the image are several realistic water droplets of various sizes, some with highlights and shadows, giving them a three-dimensional appearance.

RANDOM FOREST

OOB ERROR RATE

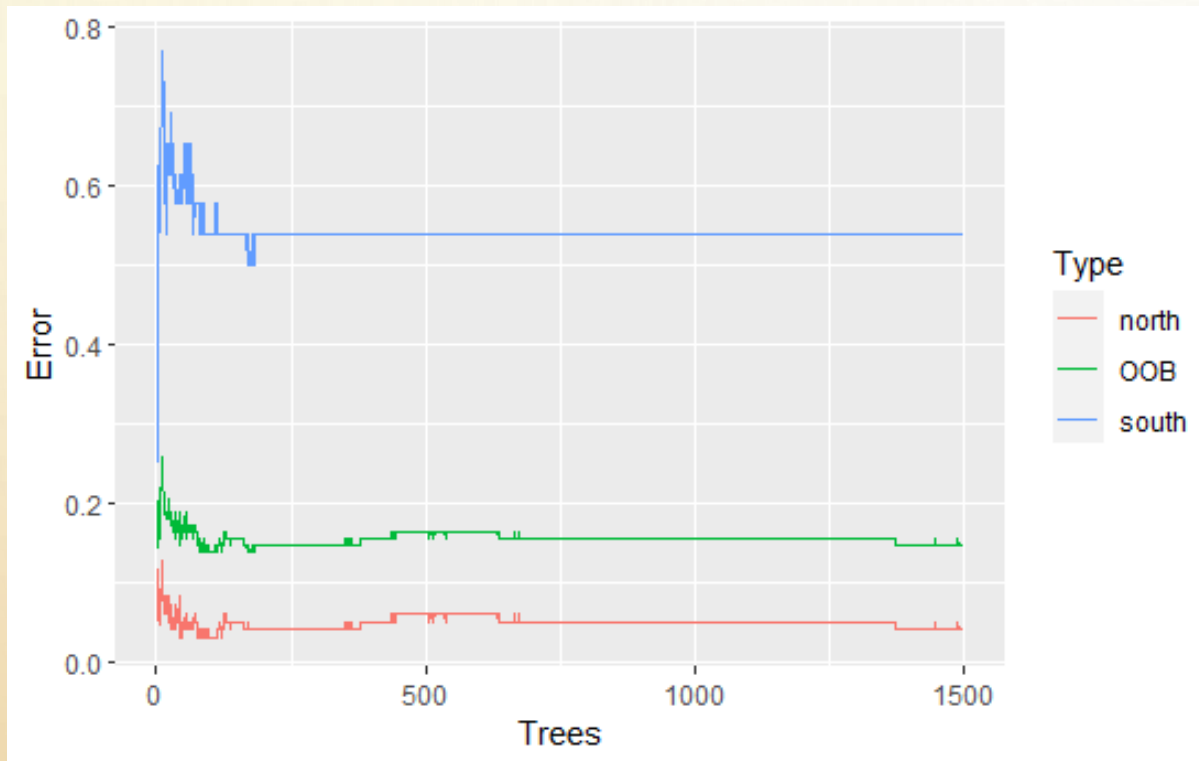


OOB estimate of error rate: 15.57%

Confusion matrix:

| | north | south | class.error |
|-------|-------|-------|-------------|
| north | 91 | 5 | 0.05208333 |
| south | 14 | 12 | 0.53846154 |

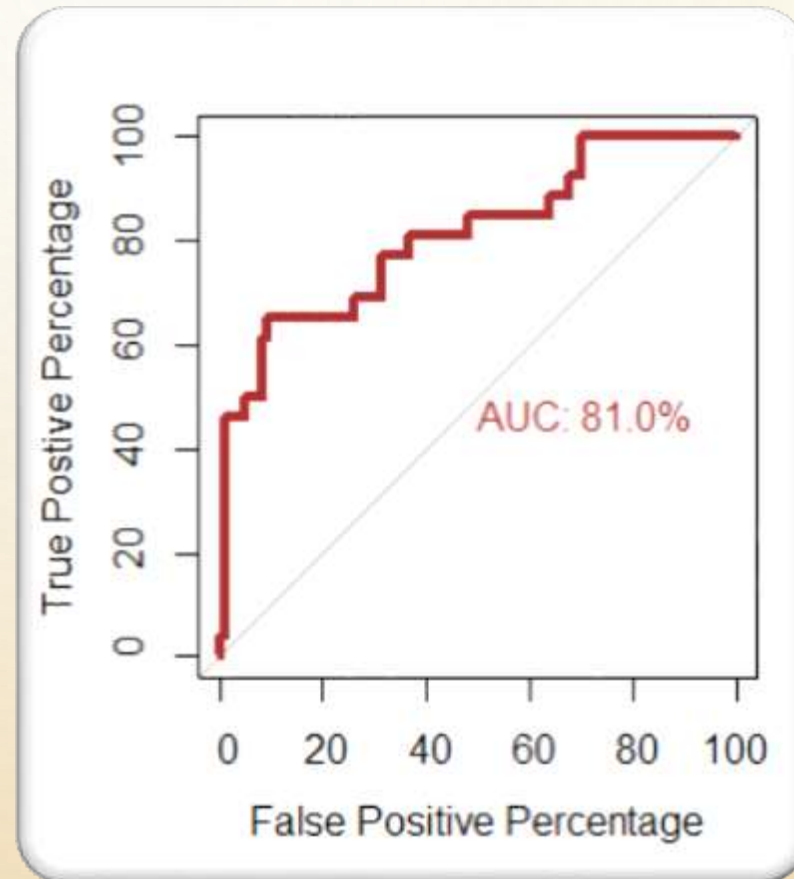
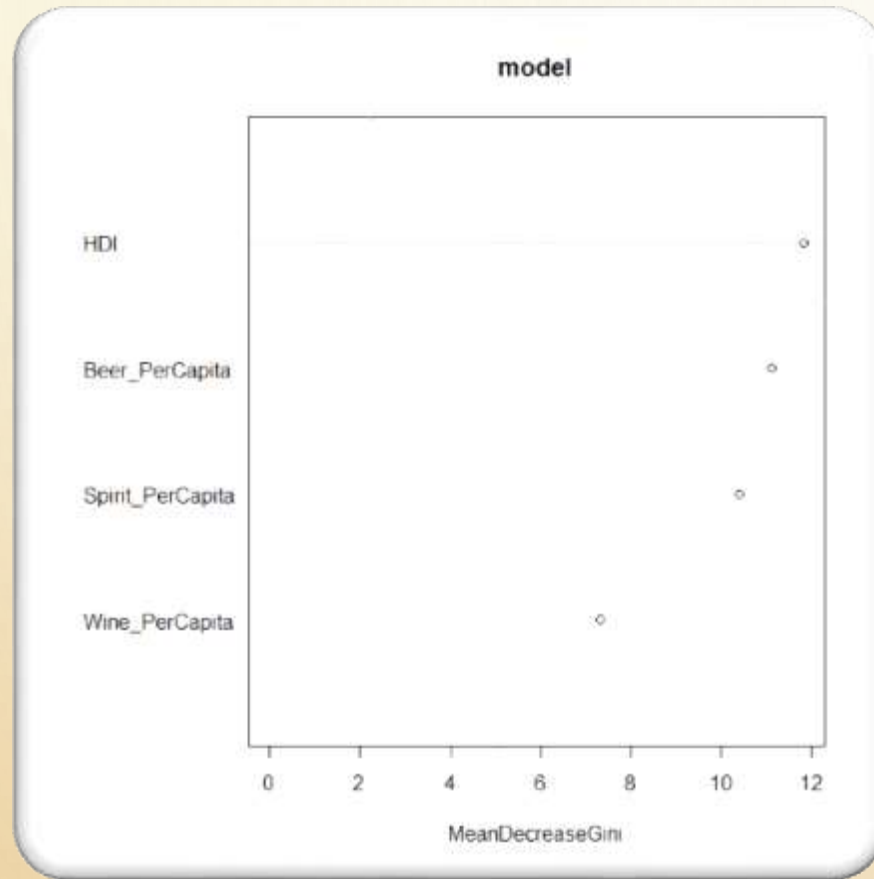
OOB ERROR RATE



| | |
|-------------|-------|
| Accuracy | 0.852 |
| Precision | 0.867 |
| Sensitivity | 0.958 |
| Specificity | 0.461 |

```
      OOB estimate of  error rate: 14.75%
Confusion matrix:
      north south class.error
north   92     4  0.04166667
south   14    12  0.53846154
```

VARIABLE IMPORTANCE E ROC CURVE







GRAZIE PER L'ATTENZIONE!!!