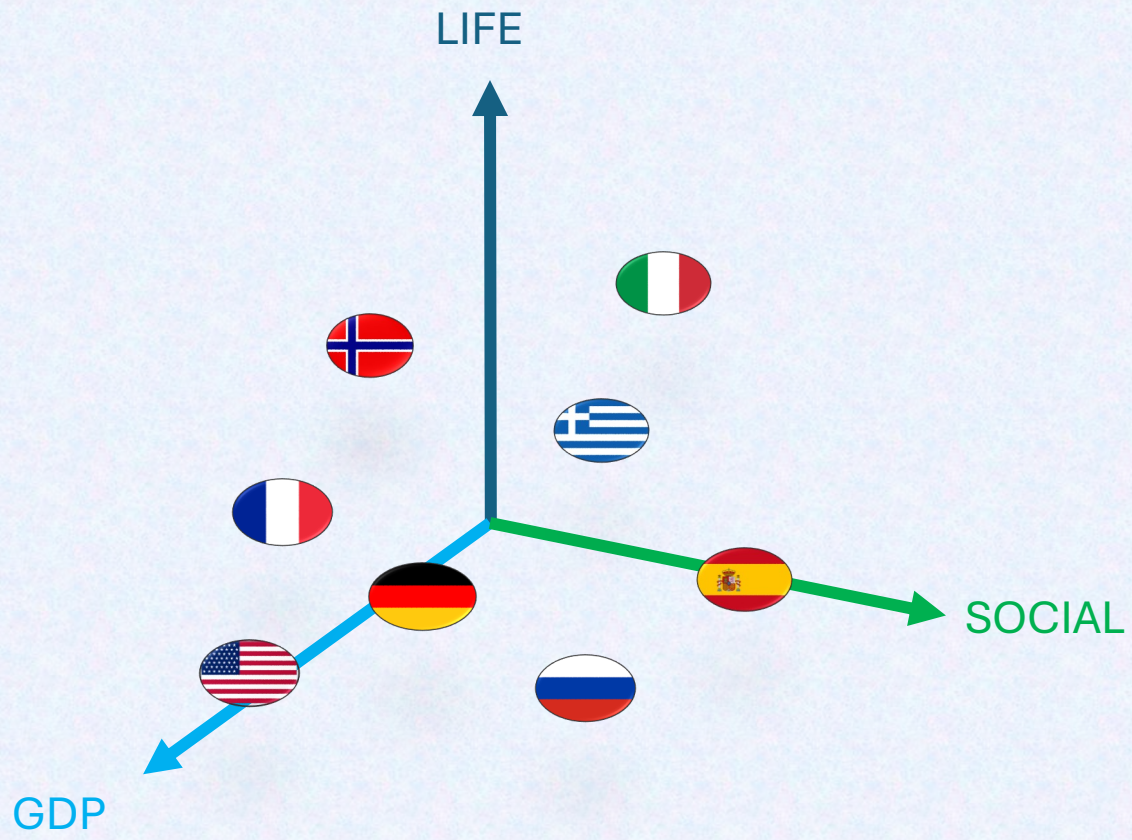
HATLAN

- Honest Albert Tèmu
- Andrea Corradini
- Tommaso Premoli
- Luca Codeluppi
- Antonio De Patto
- Niccolo’ Cibei

WORLD HAPPINESS REPORT

WHAT MAKES A COUNTRY HAPPY?

- Introduction
- Variables
- Unsupervised Learning
- Supervised Learning
- Linear Regression
- Ridge and Lasso Regression



1 - INTRODUCTION

An overview of the ladder score values provided by the World Happiness Report

No.	Country name	Ladder Score	No.	Country name	Ladder Score
1	Finland	7,804	15	United States	6,894
2	Denmark	7,586	25	Singapore	6,587
3	Iceland	7,53	28	Uruguay	6,494
4	Israel	7,473	33	Italy	6,405
5	Netherlands	7,403	59	Mauritius	5,902
6	Sweden	7,395	64	China	5,818
7	Norway	7,315	85	South Africa	5,275
8	Switzerland	7,24	126	India	4,036
9	Luxembourg	7,228	136	Lebanon	2,392
10	New Zealand	7,123	137	Afghanistan	1,859

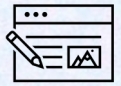
THE VARIABLES

GDP

Social support

Healthy life expectancy

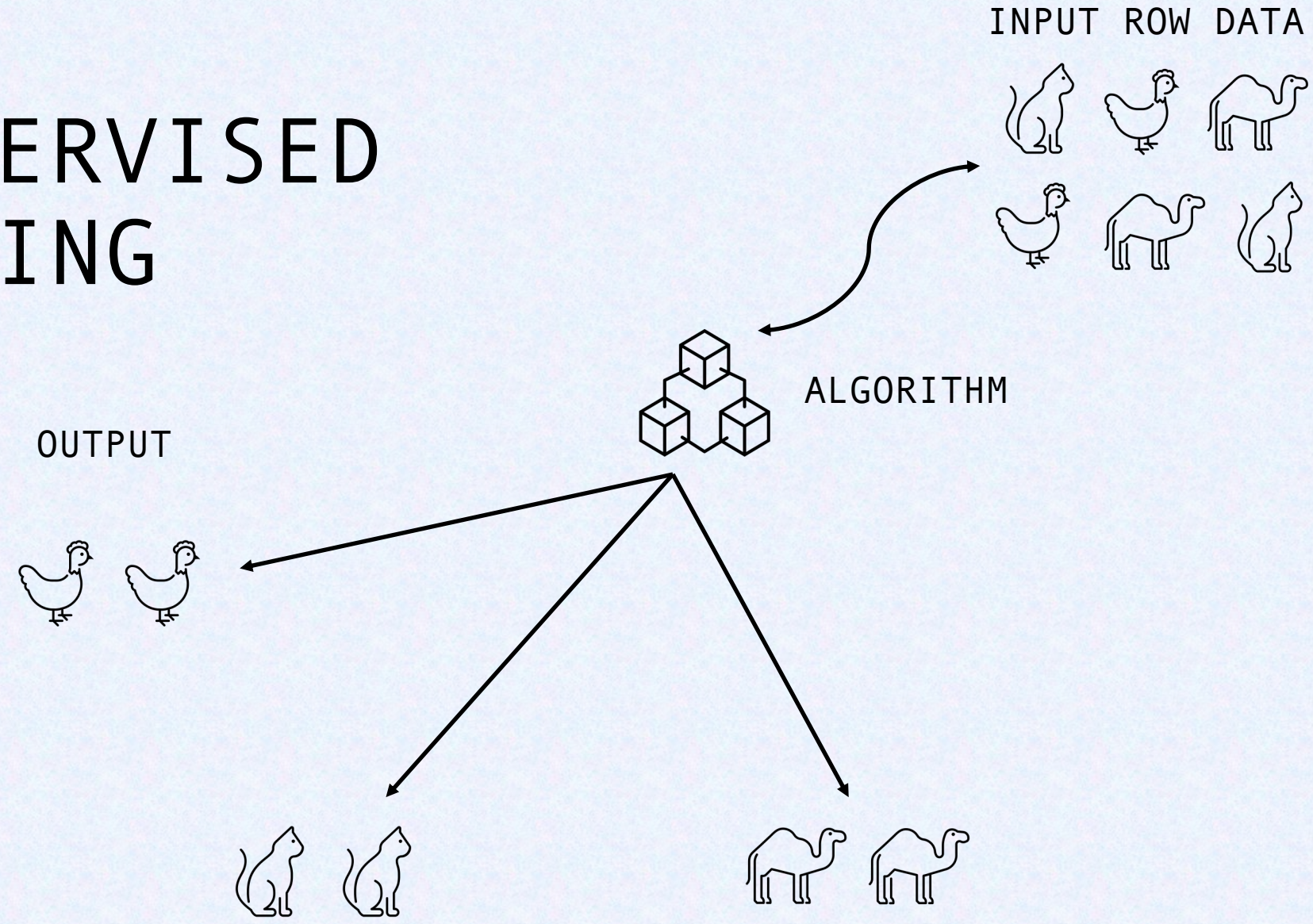
Generosity



Corruption perception

Freedom to make life choices

UNSUPERVISED LEARNING



2- PCA-ANALYSIS

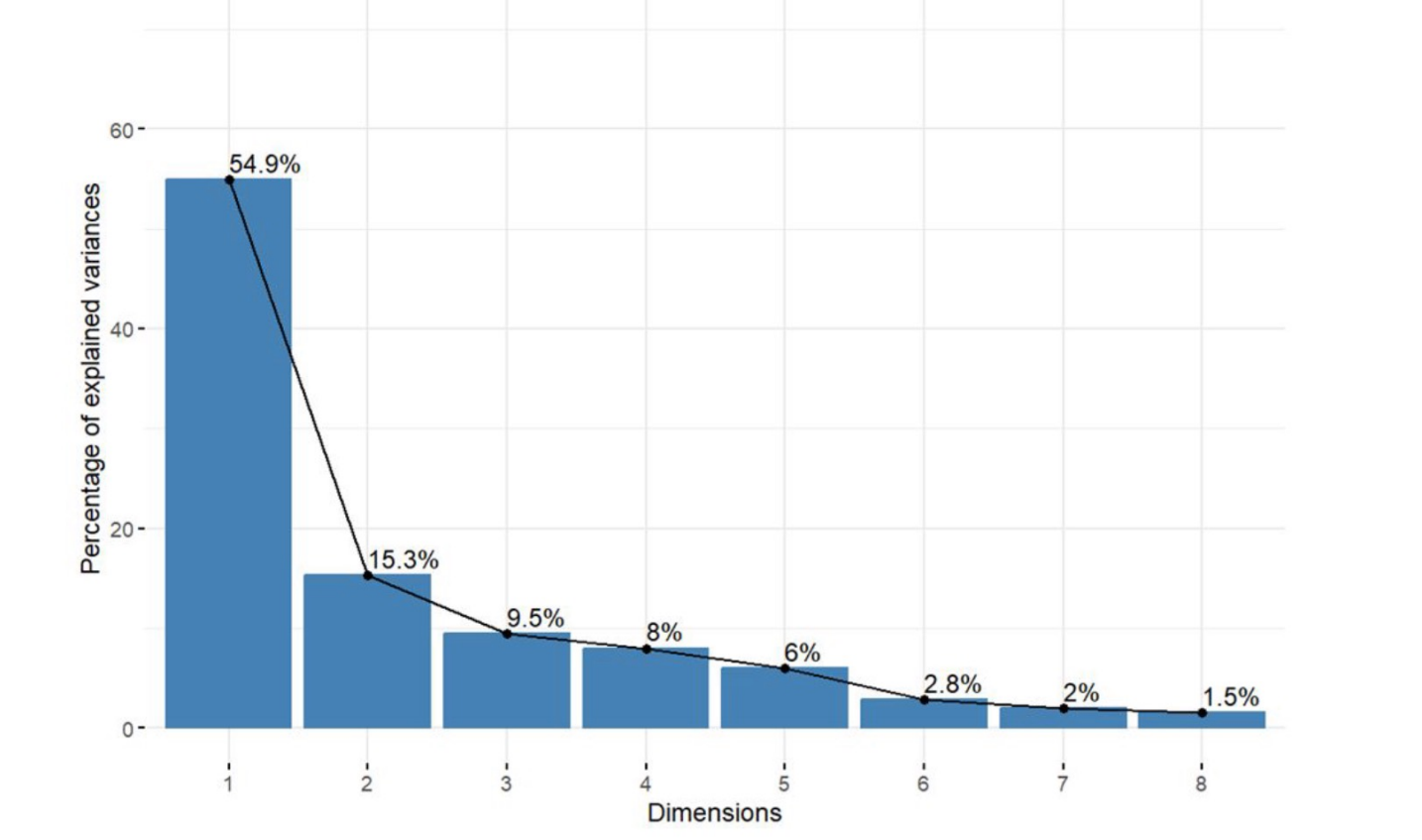
- REMOVE VARIABLE WITH LESS SIGNIFICANCE

- COMPUTE THE PCA GENERATING COMPONENTS princomp()

```
> # Summary of PCA
> summary(pca_result)
Importance of components:
                          PC1    PC2     PC3     PC4     PC5     PC6     PC7     PC8
Standard deviation     2.0958 1.1074 0.86972 0.79876 0.69285 0.47620 0.39663 0.35047
Proportion of Variance 0.5491 0.1533 0.09455 0.07975 0.06001 0.02835 0.01966 0.01535
Cumulative Proportion  0.5491 0.7023 0.79688 0.87663 0.93664 0.96498 0.98465 1.00000
```

THE SCREE PLOT

fviz_eig() function

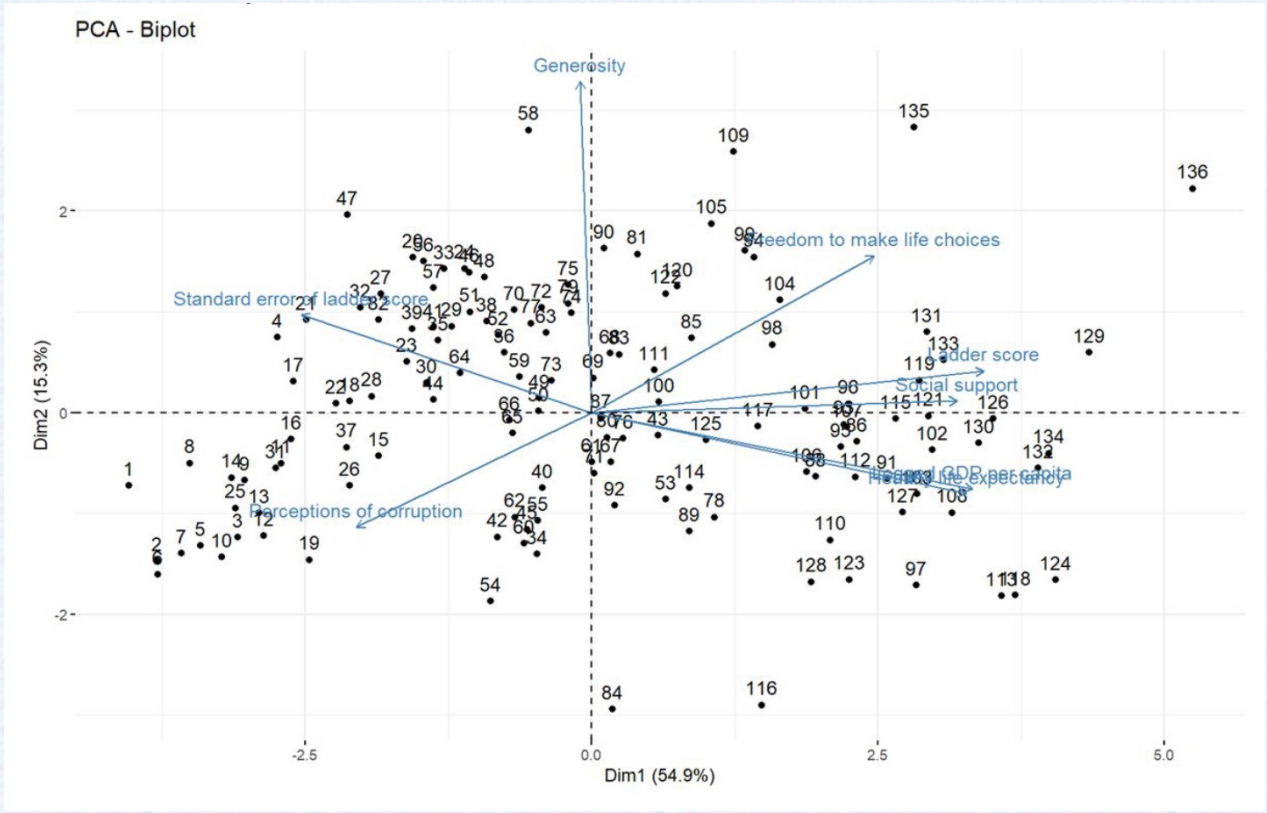


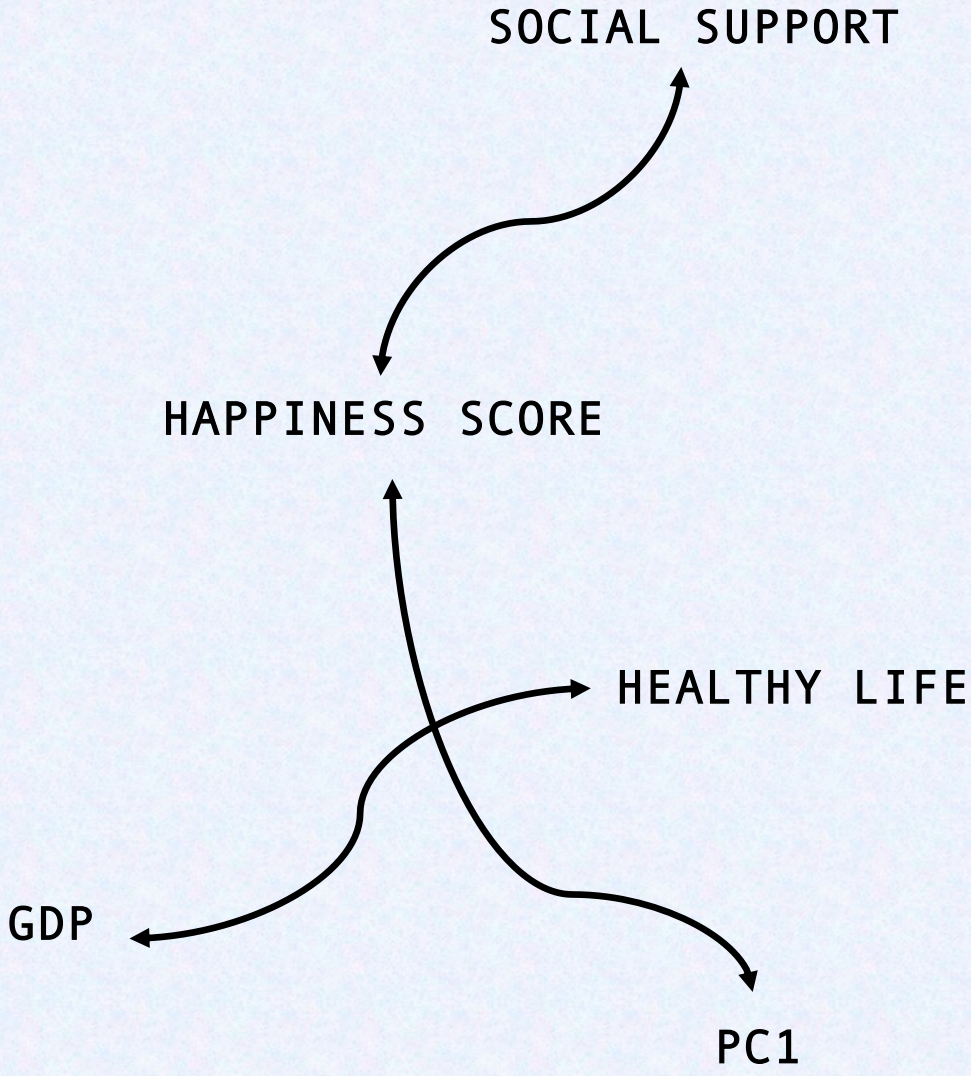
54,9%

of the entire data-set explained by the first component

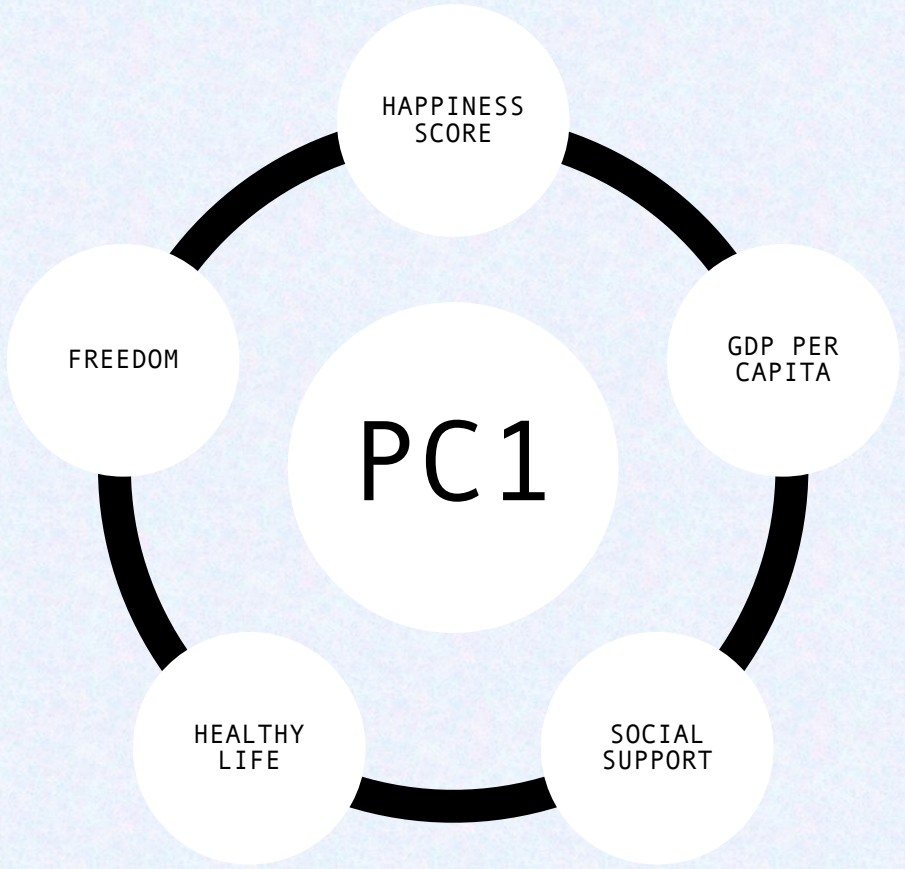
THE BIPLOT

USEFUL TO SEE CORRELATION



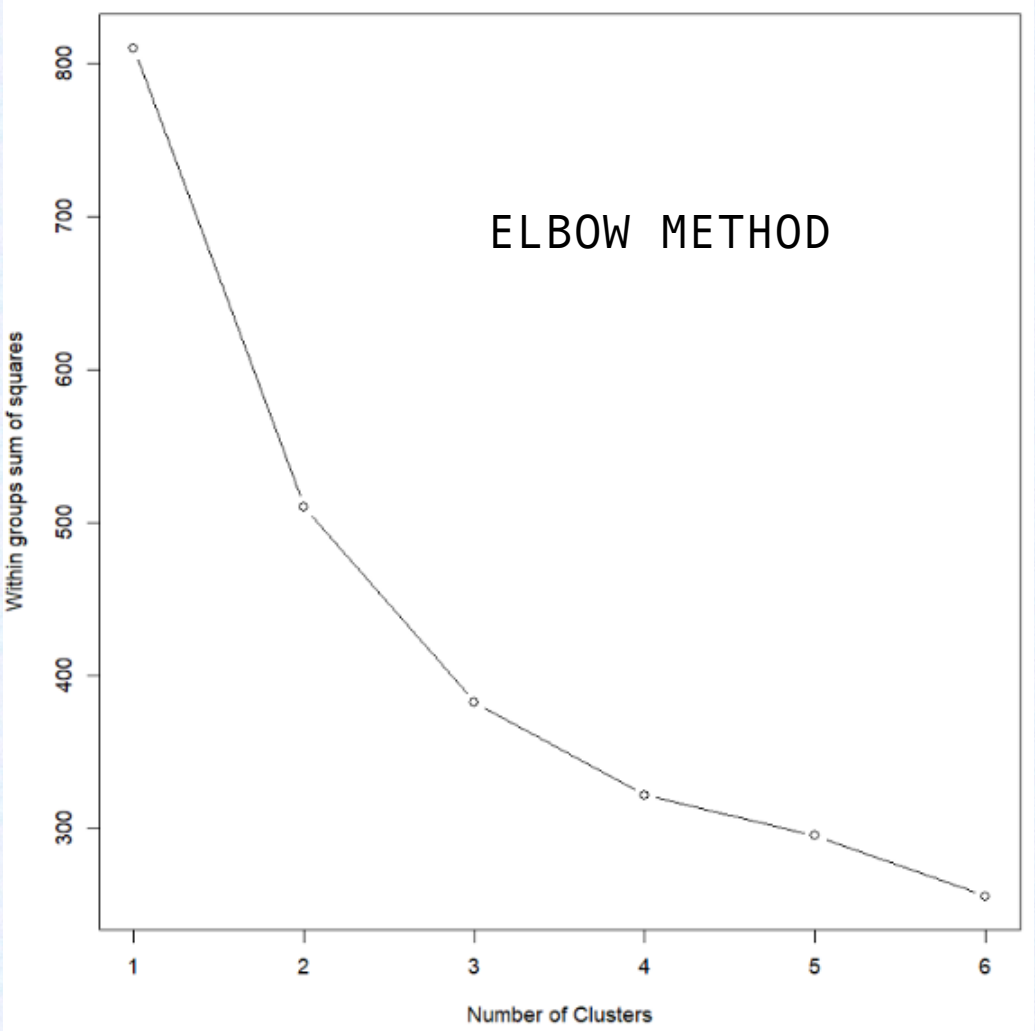


CORRELATION BETWEEN P.C. AND OTHER VARIABLES



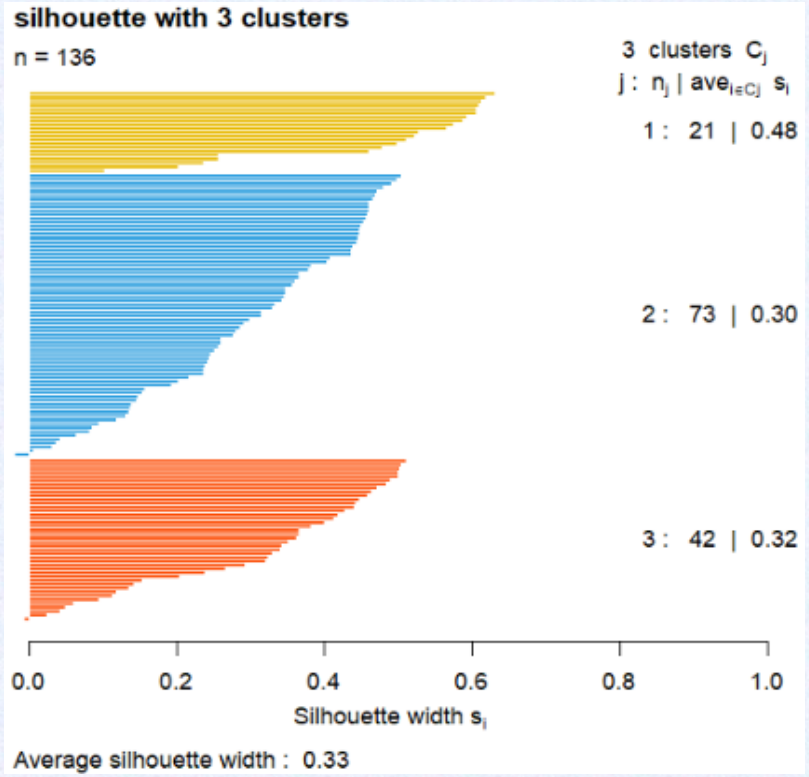
3- K MEANS CLUSTERING

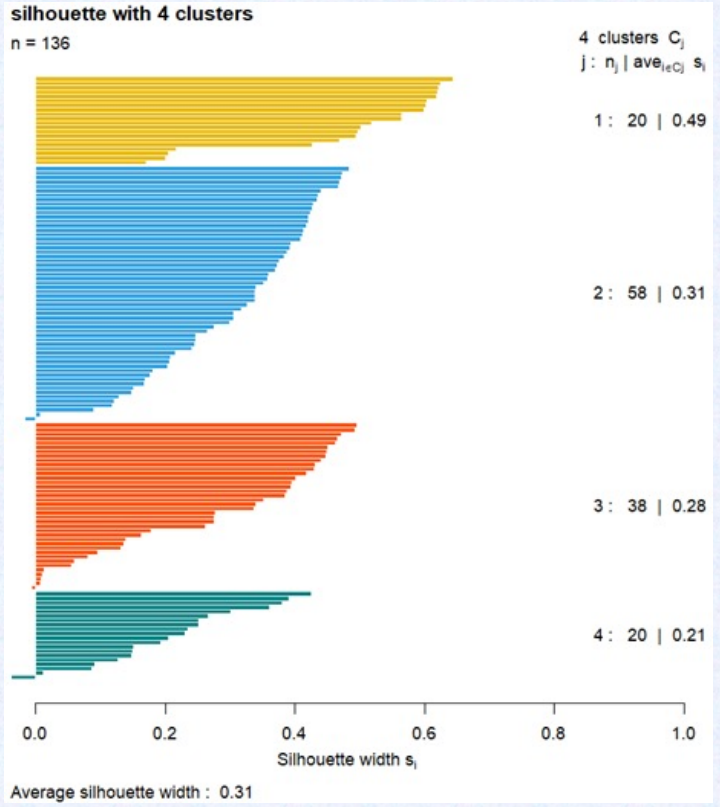
CALCULATE THE PERFECT NUMBER OF CLUSTERS



3- K MEANS CLUSTERING

CALCULATE THE PERFECT NUMBER OF CLUSTERS

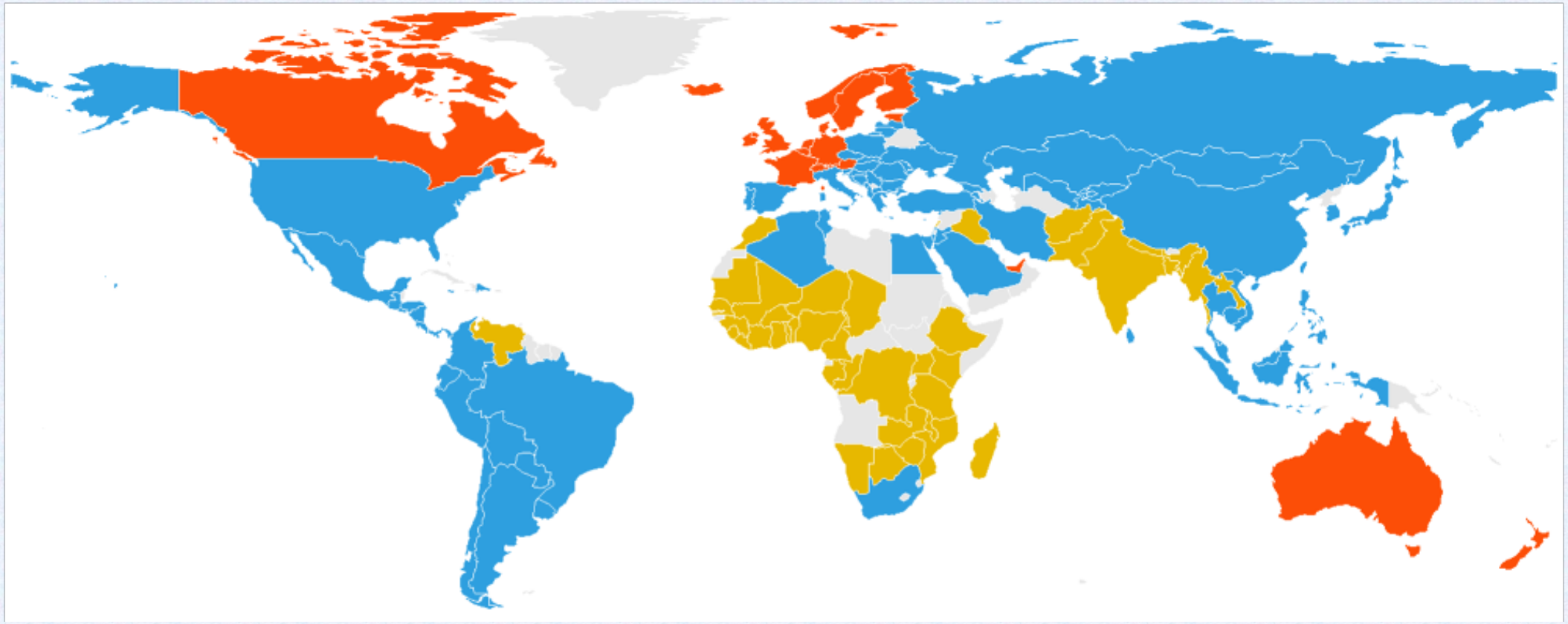




$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, \text{ if } |C_i| > 1$$

SILHOUETTE

K – MEANS CLUSTERING WITH 3 CLUSTERS



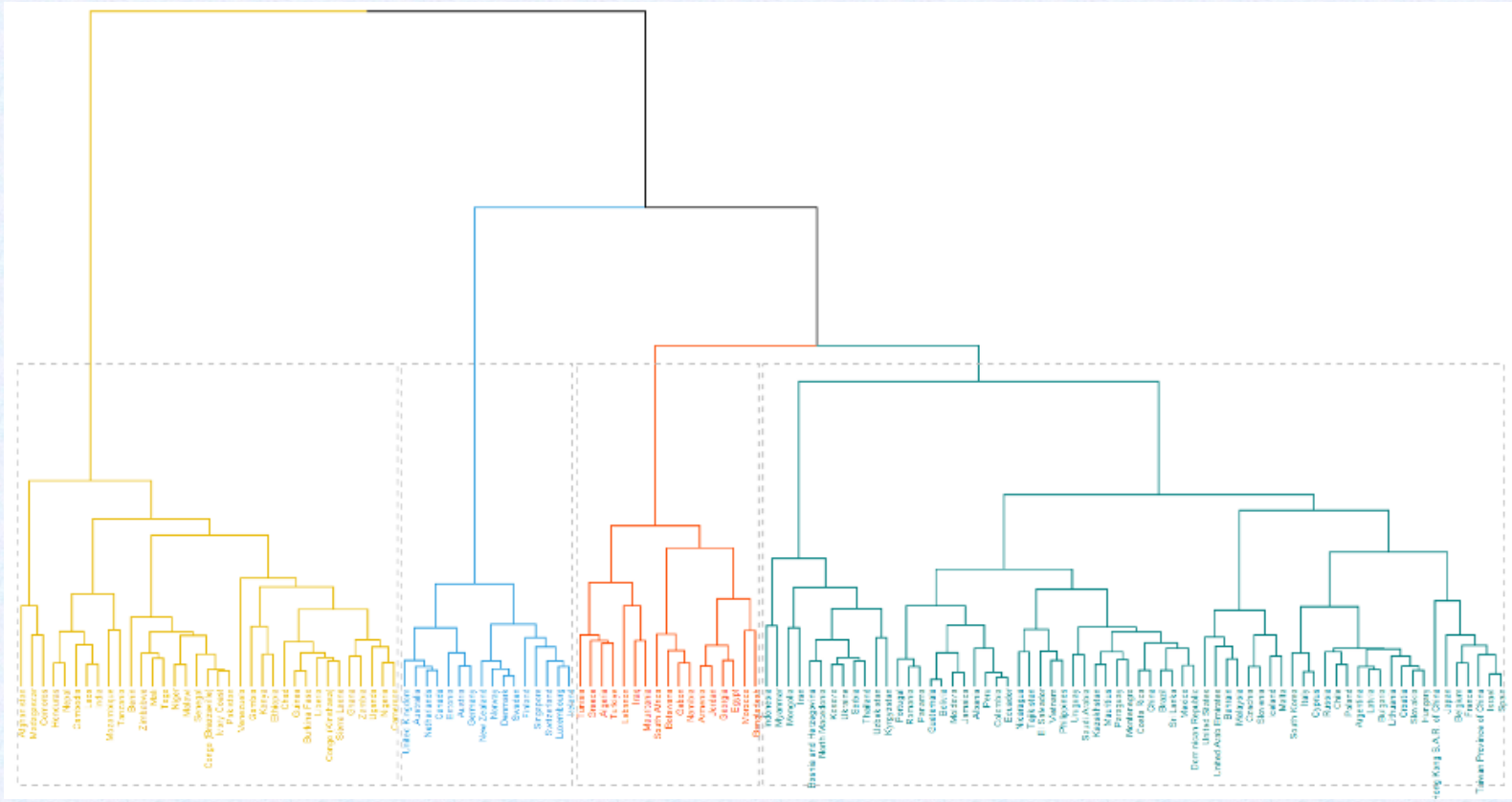
gdp	social_supp	life_exp	freedom	generosity	corruption
Min. :10.54	Min. :0.817	Min. :66.24	Min. :0.6870	Min. :-0.10000	Min. :0.146
1st Qu.:10.79	1st Qu.:0.888	1st Qu.:71.05	1st Qu.:0.8550	1st Qu.: 0.02700	1st Qu.:0.271
Median :10.90	Median :0.920	Median :71.30	Median :0.8870	Median : 0.09600	Median :0.385
Mean :10.97	Mean :0.914	Mean :71.49	Mean :0.8846	Mean : 0.09357	Mean :0.384
3rd Qu.:11.09	3rd Qu.:0.943	3rd Qu.:72.05	3rd Qu.:0.9340	3rd Qu.: 0.16500	3rd Qu.:0.496
Max. :11.66	Max. :0.983	Max. :77.28	Max. :0.9610	Max. : 0.25300	Max. :0.668

gdp	social_supp	life_exp	freedom	generosity	corruption
Min. : 8.237	Min. :0.7160	Min. :56.99	Min. :0.4750	Min. :-0.25400	Min. :0.5220
1st Qu.: 9.367	1st Qu.:0.8110	1st Qu.:65.30	1st Qu.:0.7690	1st Qu.:-0.09900	1st Qu.:0.7210
Median : 9.811	Median :0.8670	Median :67.00	Median :0.8090	Median :-0.05700	Median :0.8080
Mean : 9.816	Mean :0.8546	Mean :67.06	Mean :0.8085	Mean :-0.01501	Mean :0.7889
3rd Qu.:10.353	3rd Qu.:0.9060	3rd Qu.:69.00	3rd Qu.:0.8770	3rd Qu.: 0.04000	3rd Qu.:0.8660
Max. :11.048	Max. :0.9530	Max. :74.35	Max. :0.9580	Max. : 0.53100	Max. :0.9290

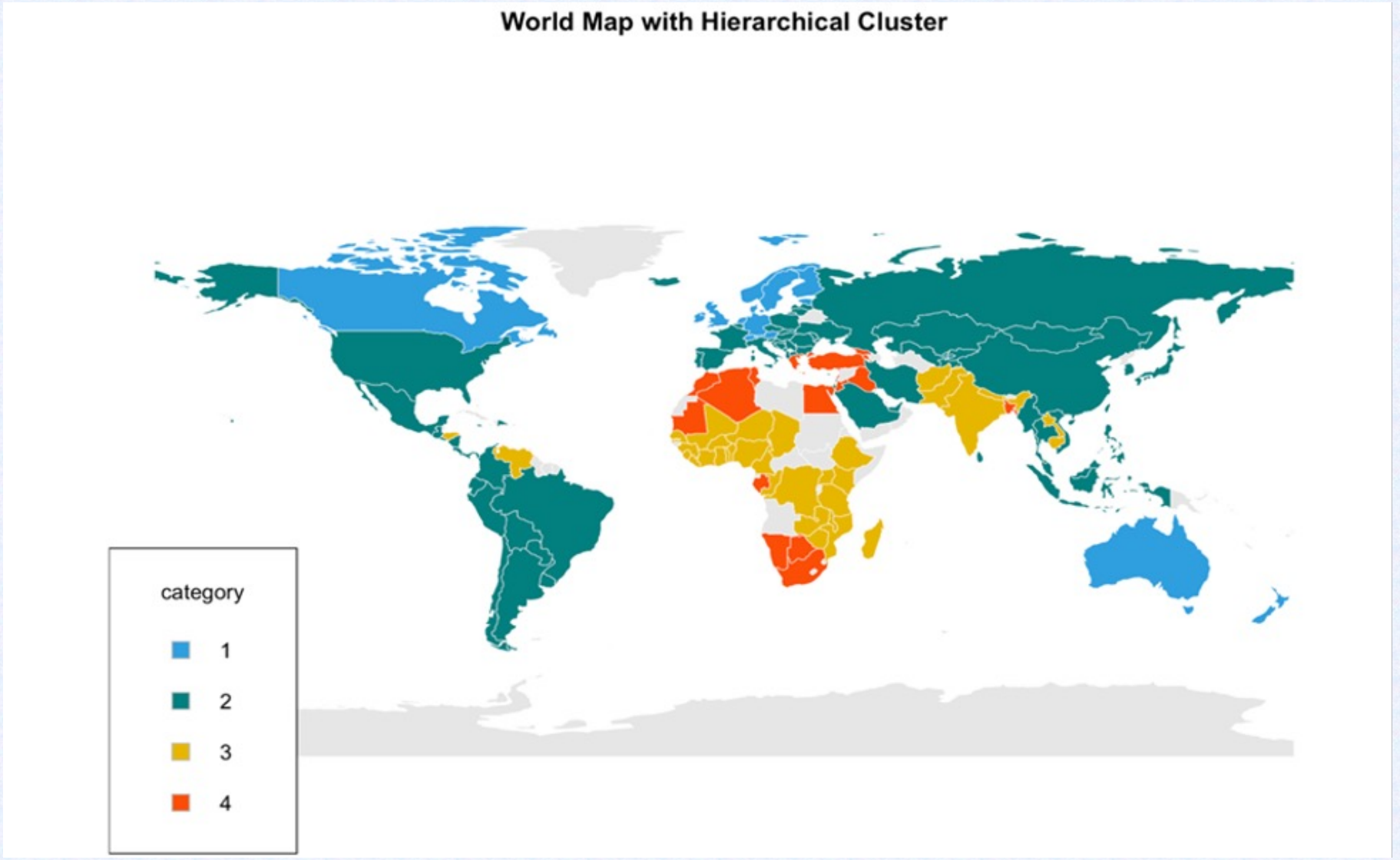
gdp	social_supp	life_exp	freedom	generosity	corruption
Min. :5.527	Min. :0.3410	Min. :51.53	Min. :0.3820	Min. :-0.23100	Min. :0.5540
1st Qu.:7.643	1st Qu.:0.5907	1st Qu.:55.42	1st Qu.:0.6590	1st Qu.:-0.01100	1st Qu.:0.7405
Median :8.085	Median :0.6495	Median :57.67	Median :0.7155	Median : 0.03950	Median :0.7880
Mean :8.070	Mean :0.6437	Mean :58.07	Mean :0.7043	Mean : 0.05562	Mean :0.7831
3rd Qu.:8.587	3rd Qu.:0.7120	3rd Qu.:60.36	3rd Qu.:0.7705	3rd Qu.: 0.12225	3rd Qu.:0.8460
Max. :9.629	Max. :0.8390	Max. :66.15	Max. :0.9190	Max. : 0.49100	Max. :0.9110

4- HIERARCHICAL CLUSTERING

FOR A BETTER VISUALIZATION OF THE PROCESS WE HAVE CHOSEN WARD'S METHOD



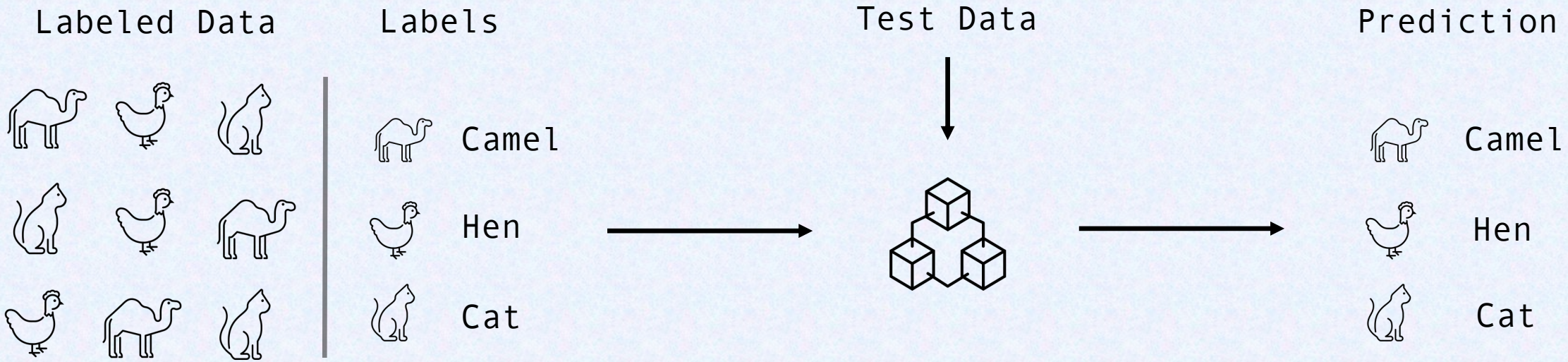
DENDROGRAM USING WARD LINKAGE



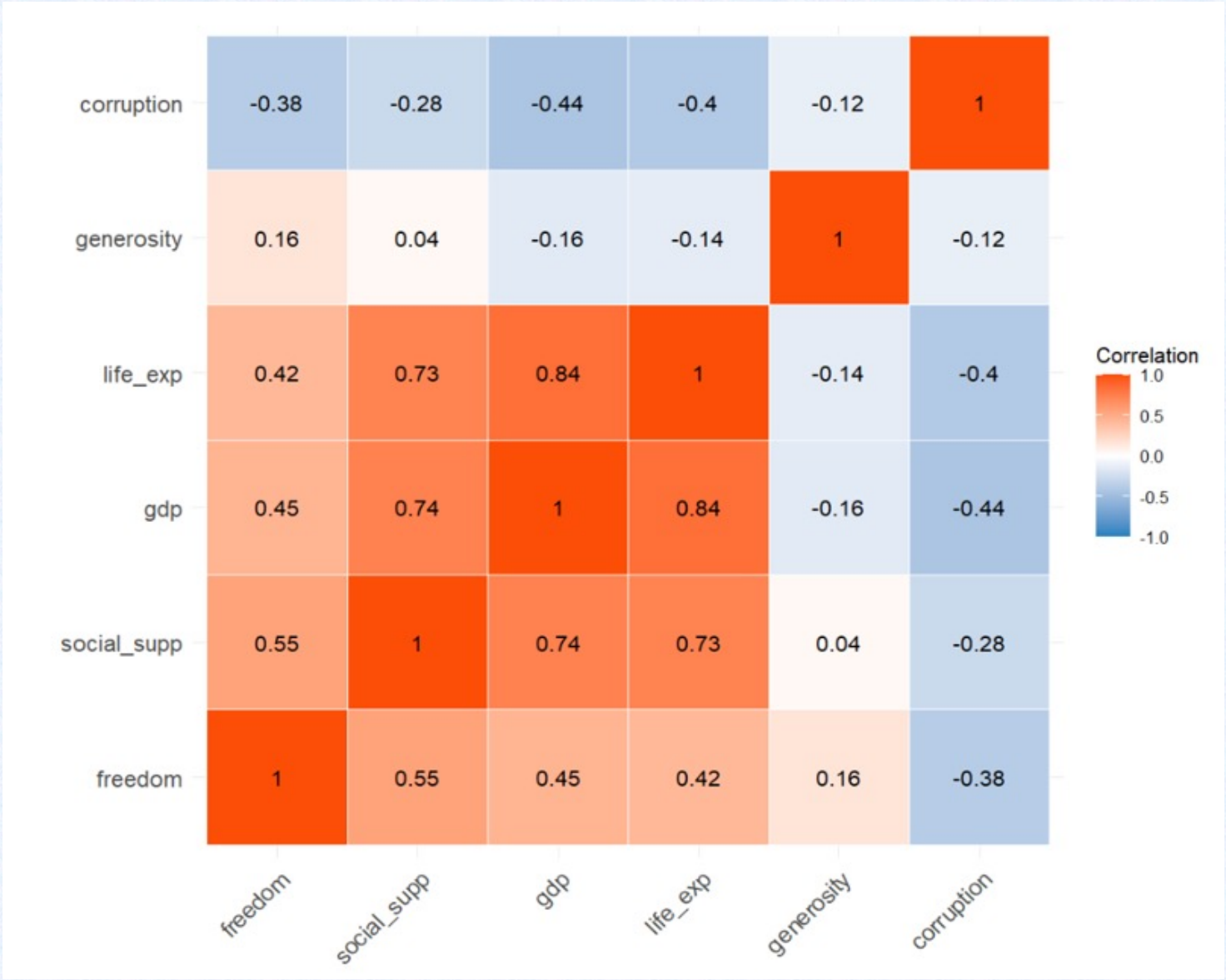
GEOGRAPHICAL MAP WITH THE COUNTRIES DIVIDED BY CLUSTER

Group.1	gdp	social_supp	life_exp	freedom	generosity	corruption
1	10.991750	0.9215000	71.40469	0.8980000	0.11143750	0.3327500
2	9.942706	0.8680294	67.58069	0.8257500	0.01298529	0.7728235
3	7.841343	0.6424286	57.57506	0.7208286	0.08365714	0.7872571
4	9.381588	0.7270000	63.67700	0.6724118	-0.14052941	0.7714118

SUPERVISED LEARNING



LINEAR REGRESSION



LINEAR REGRESSION

```
Call:
lm(formula = happiness ~ gdp + life_exp + social_supp + freedom +
    generosity + corruption + cluster, data = happiness_data)

Residuals:
     Min       1Q   Median       3Q      Max
-1.30510 -0.19680  0.01897  0.26772  0.95222

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.34921    1.15239  -2.039  0.04359 *
gdp          0.24241    0.07554   3.209  0.00169 **
life_exp     0.02139    0.01449   1.477  0.14219
social_supp  3.98488    0.58413   6.822 3.36e-10 ***
freedom      1.99937    0.48606   4.113 6.99e-05 ***
generosity  -0.33621    0.34960  -0.962  0.33805
corruption  -0.46592    0.40837  -1.141  0.25607
cluster2    -0.23392    0.23178  -1.009  0.31480
cluster3    -0.04205    0.33169  -0.127  0.89932
cluster4    -0.58541    0.30160  -1.941  0.05449 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4706 on 126 degrees of freedom
Multiple R-squared:  0.8417,    Adjusted R-squared:  0.8304
F-statistic: 74.45 on 9 and 126 DF,  p-value: < 2.2e-16
```

LINEAR REGRESSION

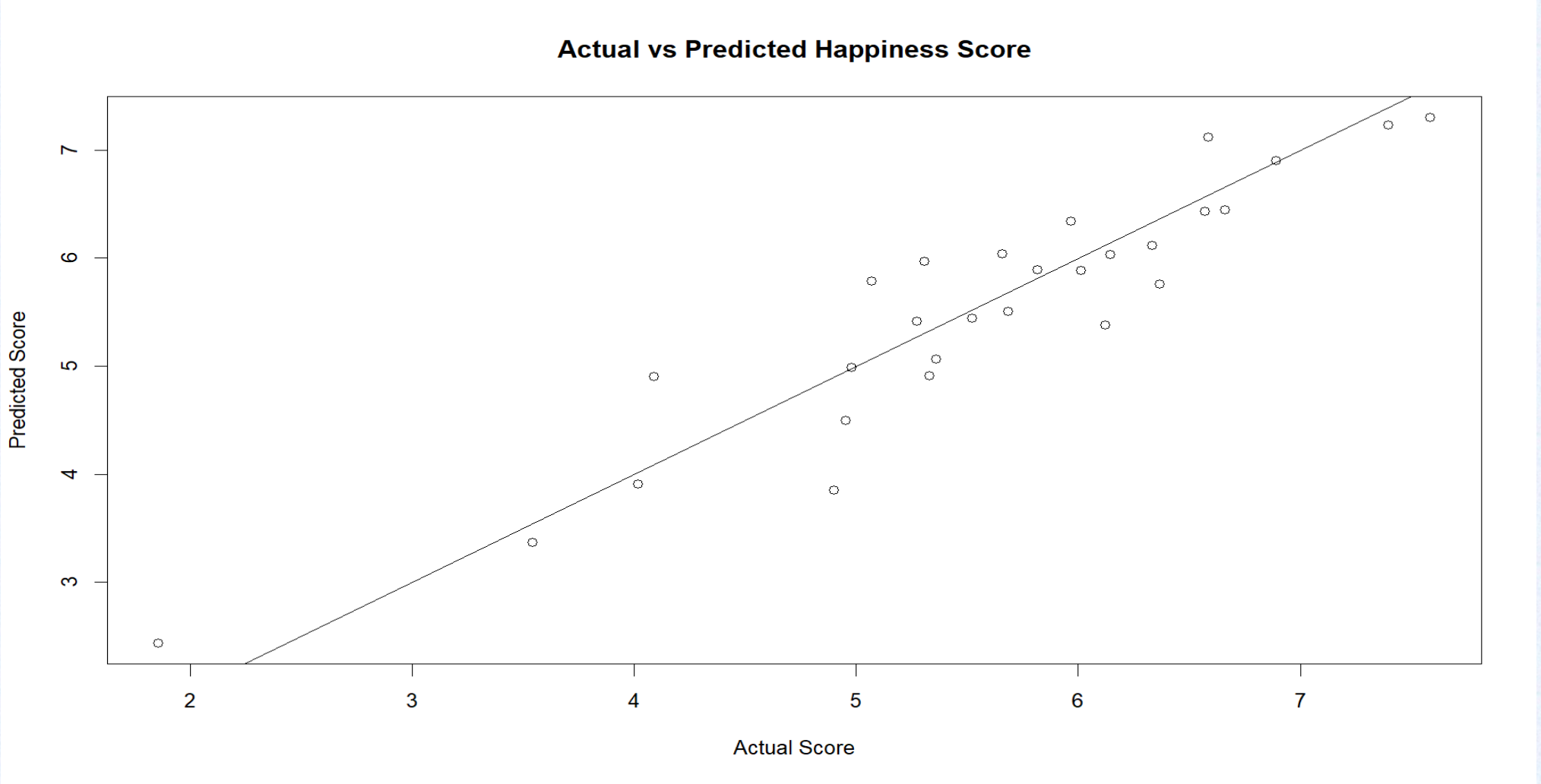
```
Call:
lm(formula = happiness ~ life_exp + social_supp + freedom + generosity +
    corruption + cluster, data = happiness_data)

Residuals:
     Min       1Q   Median       3Q      Max
-1.30251 -0.21926  0.00313  0.29014  1.01767

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.29531    1.14432  -1.132  0.25979
life_exp     0.03939    0.01384   2.847  0.00515 **
social_supp  4.33856    0.59427   7.301 2.77e-11 ***
freedom      2.05031    0.50327   4.074 8.09e-05 ***
generosity  -0.54808    0.35566  -1.541  0.12580
corruption  -0.53406    0.42248  -1.264  0.20851
cluster2    -0.38767    0.23493  -1.650  0.10138
cluster3    -0.42400    0.32073  -1.322  0.18854
cluster4    -0.77985    0.30607  -2.548  0.01203 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4876 on 127 degrees of freedom
Multiple R-squared:  0.8288,    Adjusted R-squared:  0.818
F-statistic: 76.84 on 8 and 127 DF,  p-value: < 2.2e-16
```

LINEAR REGRESSION

```
Call:
lm(formula = happiness ~ social_supp + freedom + generosity +
    corruption + cluster, data = happiness_data)

Residuals:
    Min      1Q  Median      3Q     Max
-1.4627 -0.2332  0.0491  0.3005  1.0558

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.3870     0.6673   2.078  0.03966 *
social_supp   4.7417     0.5930   7.997 6.57e-13 ***
freedom       1.8825     0.5135   3.666  0.00036 ***
generosity   -0.7460     0.3583  -2.082  0.03935 *
corruption   -0.7390     0.4277  -1.728  0.08644 .
cluster2     -0.4582     0.2400  -1.909  0.05850 .
cluster3     -0.7984     0.3005  -2.656  0.00890 **
cluster4     -1.0037     0.3039  -3.303  0.00124 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5009 on 128 degrees of freedom
Multiple R-squared:  0.8179,    Adjusted R-squared:  0.8079
F-statistic:  82.1 on 7 and 128 DF,  p-value: < 2.2e-16
```

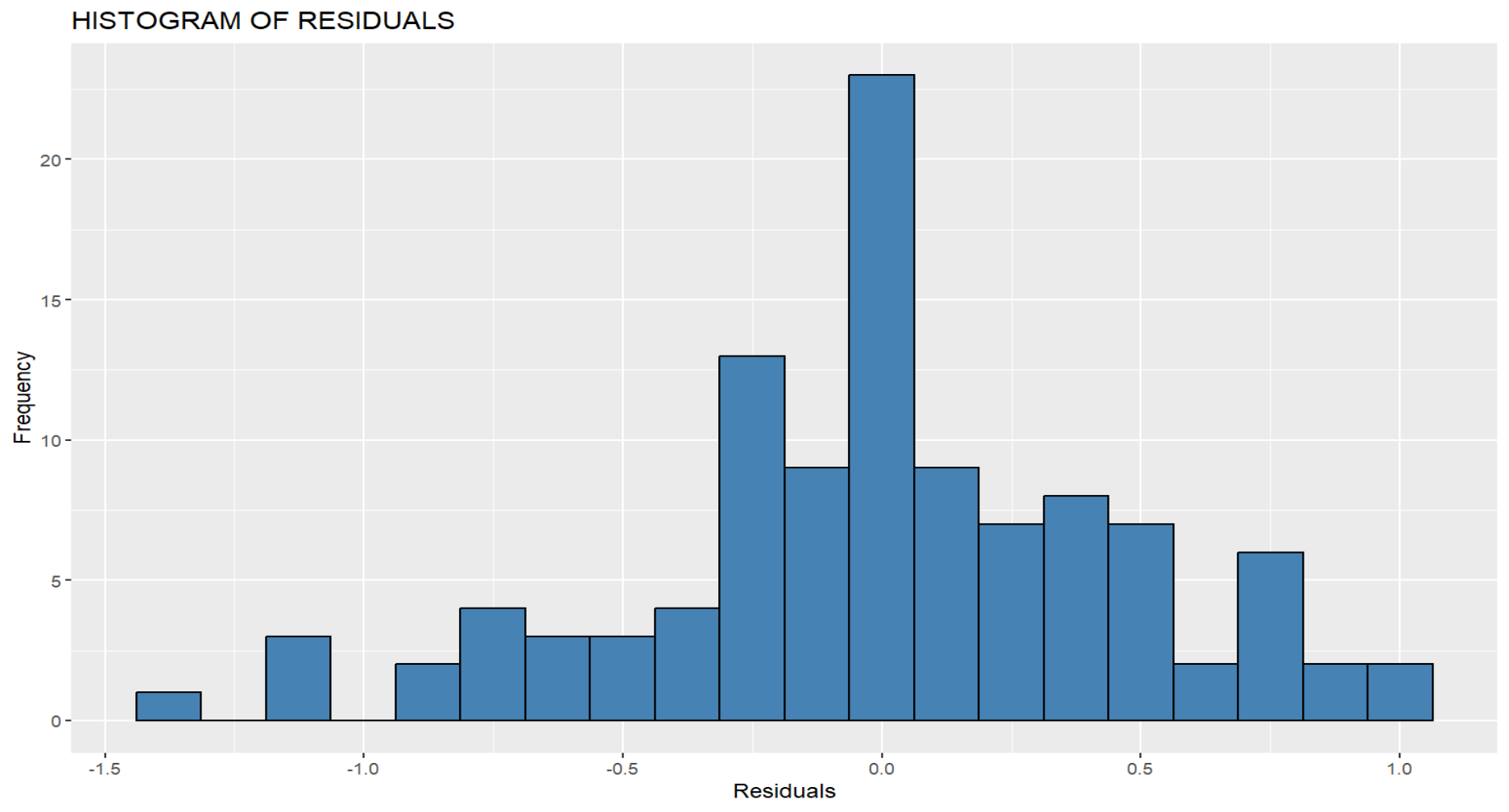
LINEAR REGRESSION

GRAPH BETWEEN PREDICTED VS ACTUAL SCORE



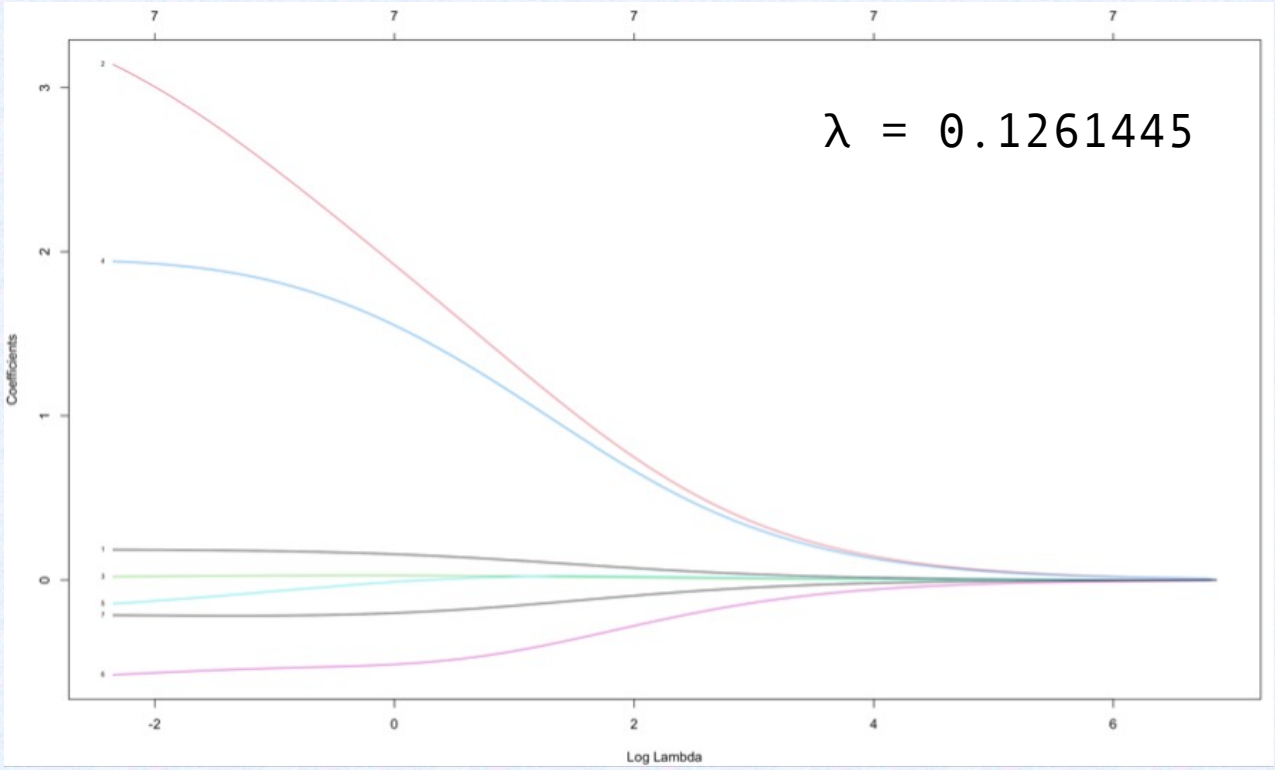
LINEAR REGRESSION

RESIDUALS HISTOGRAM



RIDGE AND LASSO REGRESSION

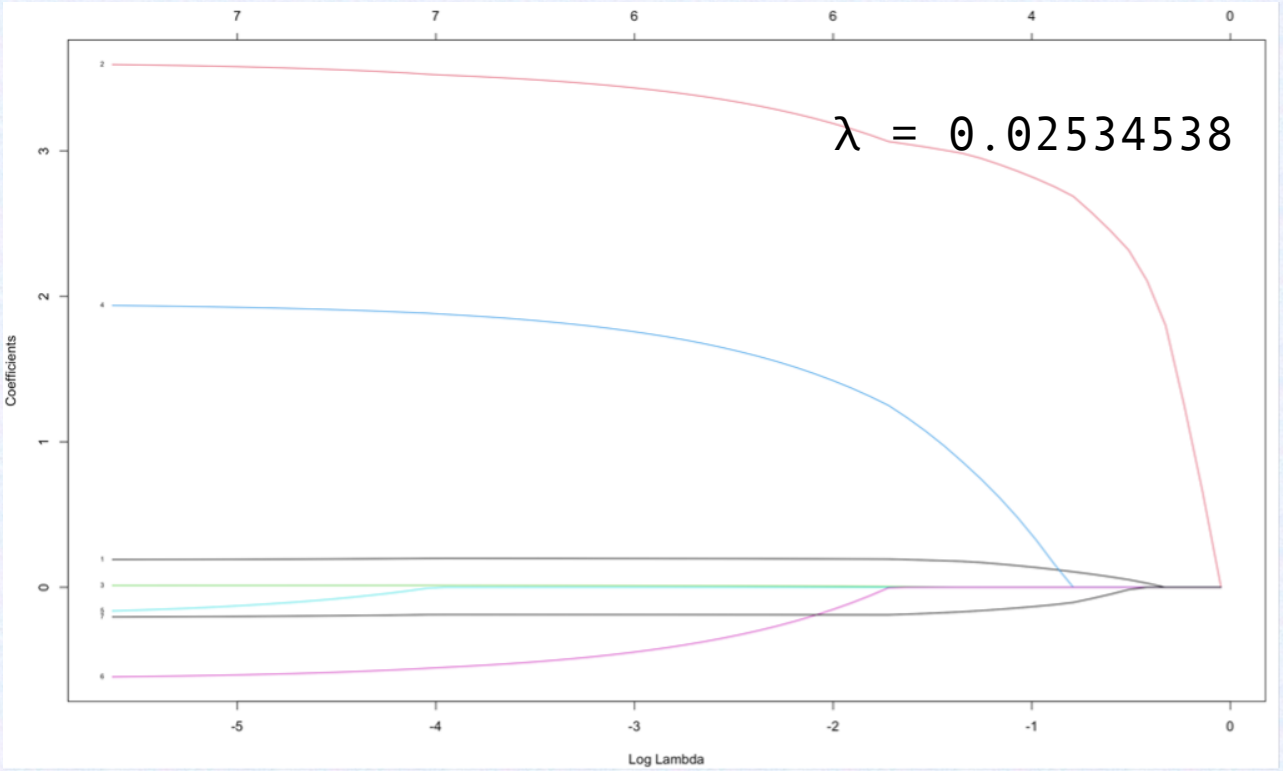
RIDGE REGRESSION



8 x 1 sparse Matrix of c	
	s0
(Intercept)	-0.58601553
gdp	0.18336608
social_supp	3.03212164
life_exp	0.02132172
freedom	1.93052120
generosity	-0.13211830
corruption	-0.56810017
cluster	-0.21623226

RIDGE AND LASSO REGRESSION

LASSO REGRESSION




```
8 x 1 sparse Matrix of c
                       s0
(Intercept) -0.59956374
gdp          0.19871121
social_supp  3.50201462
life_exp     0.01296036
freedom      1.85286613
generosity   .
corruption  -0.52939785
cluster     -0.18842764
```

COMPARISON

```
> mse_ridge
[1] 0.2154341
```

```
> mse_lasso
[1] 0.2406527
```

DECISION TREE