What Affects Happiness

Yunzhe Li, Nathan Sanders

University of California, Santa Cruz

yli566@uscs.edu nqsander@ucsc.edu

December 8, 2018

Background

- Gallup Conducts a survey called the Gallup World Poll for around 155 countries.
- In this Poll they ask random individuals questions that relate to their trust in government, family, health, freedom, economy and their happiness.
- This data is often used by governments, organizations to inform their policy decision making.

Background

- Sustainable Development Solutions Network summarizes this data.
- They use a question called the "Cantril ladder" from the poll.
- They use some of the questions asked in the poll and condense them into 6 different factors.

Data Presentation

Country	Region	Happiness Rank	Happiness Score	Standard Error	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedom	Trust (Government Corruption)	Generosity	Dystopia Residual
Switzerland	Western Europe	1	7.587	0.03411	1.39651	1.34951	0.94143	0.66557	0.41978	0.29678	2.51738
Iceland	Western Europe	2	7.561	0.04884	1.30232	1.40223	0.94784	0.62877	0.14145	0.4363	2.70201
Denmark	Western Europe	3	7.527	0.03328	1.32548	1.36058	0.87464	0.64938	0.48357	0.34139	2.49204
Norway	Western Europe	4	7.522	0.0388	1.459	1.33095	0.88521	0.66973	0.36503	0.34699	2.46531
Canada	North America	5	7.427	0.03553	1.32629	1.32261	0.90563	0.63297	0.32957	0.45811	2.45176

Table: Head of World Happiness Data of 2015.

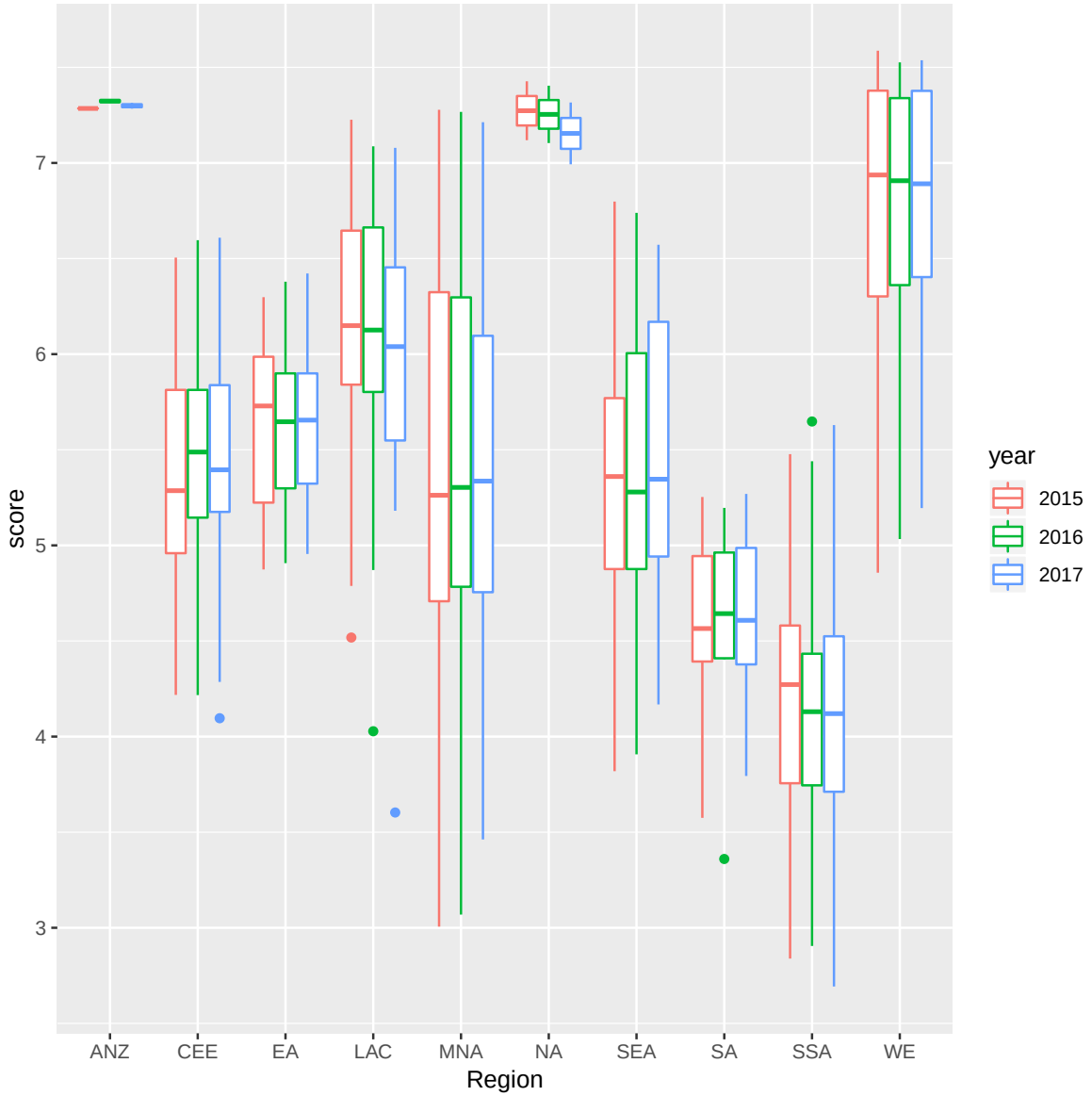
Country	Region	Happiness Rank	Happiness Score	Lower Confidence Interval	Upper Confidence Interval	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedom	Trust (Government Corruption)	Generosity	Dystopia Residual
Denmark	Western Europe	1	7.526	7.46	7.592	1.44178	1.16374	0.79504	0.57941	0.44453	0.36171	2.73939
Switzerland	Western Europe	2	7.509	7.428	7.59	1.52733	1.14524	0.86303	0.58557	0.41203	0.28083	2.69463
Iceland	Western Europe	3	7.501	7.333	7.669	1.42666	1.18326	0.86733	0.56624	0.14975	0.47678	2.83137
Norway	Western Europe	4	7.498	7.421	7.575	1.57744	1.1269	0.79579	0.59609	0.35776	0.37895	2.66465
Finland	Western Europe	5	7.413	7.351	7.475	1.40598	1.13464	0.81091	0.57104	0.41004	0.25492	2.82596

Table: Head of World Happiness Data of 2016.

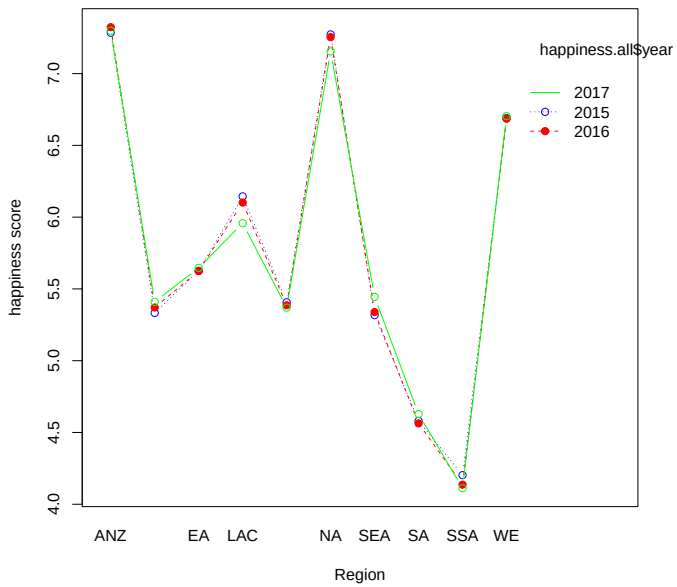
"Country"	"Happiness.Rank"	"Happiness.Score"	"Whisker.high"	"Whisker.low"	"Economy..GDP.per.Capita."	"Family"	"Health..Life.Expectancy."	"Freedom"	"Generosity"	"Trust..Government.Corruption."	"Dystopia.Residual"
"Norway"	1	7.53700017929077	7.59444482058287	7.47955553799868	1.61646318435669	1.53352355957031	0.796666502952576	0.635422587394714	0.36201223731041	0.315963834524155	2.27702665328979
"Denmark"	2	7.52199983596802	7.58172806486487	7.46227160707116	1.48238301277161	1.55112159252167	0.792565524578094	0.626006722450256	0.355280488729477	0.40077006816864	2.31370735168457
"Iceland"	3	7.50400018692017	7.62203047305346	7.38596990078688	1.480633020401	1.6105740070343	0.833552122116089	0.627162635326385	0.475540220737457	0.153526559472084	2.32271528244019
"Switzerland"	4	7.49399995803833	7.56177242040634	7.42622749567032	1.56497955322266	1.51691174507141	0.858131289482117	0.620070576667786	0.290549278259277	0.367007285356522	2.2767162322998
"Finland"	5	7.4689998626709	7.52754207581282	7.41045764952898	1.44357192516327	1.5402467250824	0.80915766954422	0.617950856685638	0.24548277258873	0.38261154294014	2.4301815032959

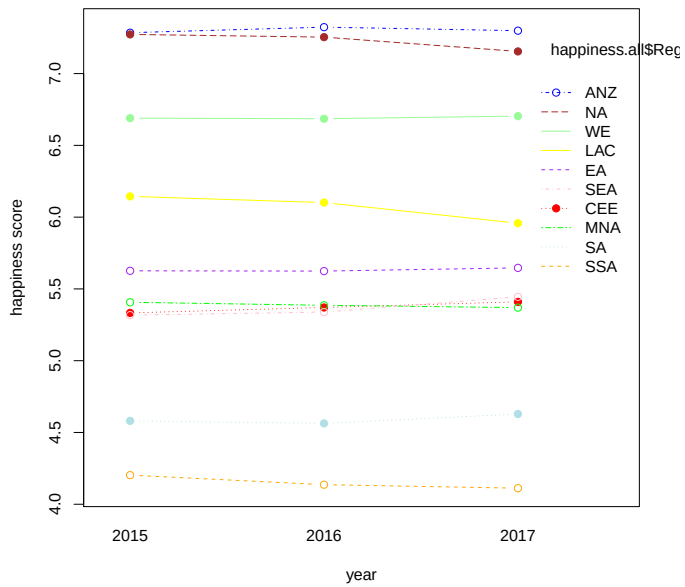
Table: Head of World Happiness Data of 2017.

Year-Region two way boxplot



Year-Region Interaction



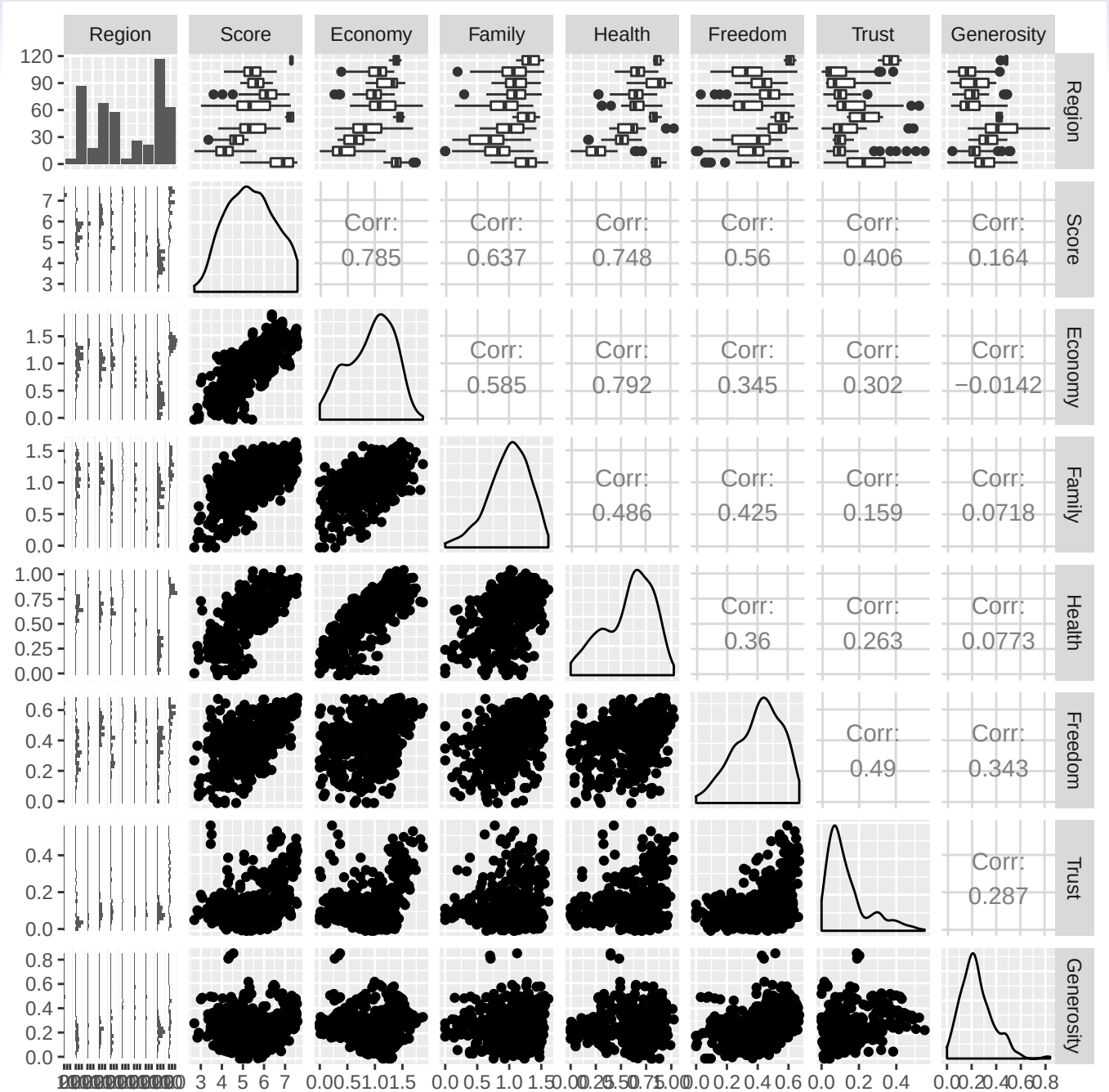


Anova output

```
Analysis of Variance Table

Response: score
             Df Sum Sq Mean Sq F value Pr(>F)
year          1   0.04   0.037  0.0711 0.7898
Region        9 374.51  41.612 81.0327 <2e-16 ***
year:Region   9   0.67   0.075  0.1455 0.9983
Residuals   450 231.09   0.514
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Analysis of Variance Table

Response: score
            Df Sum Sq Mean Sq F value Pr(>F)
year         1   0.04   0.037  0.0723 0.7881
Region       9 374.51  41.612 82.4135 <2e-16 ***
Residuals  459 231.76   0.505
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Combined Data

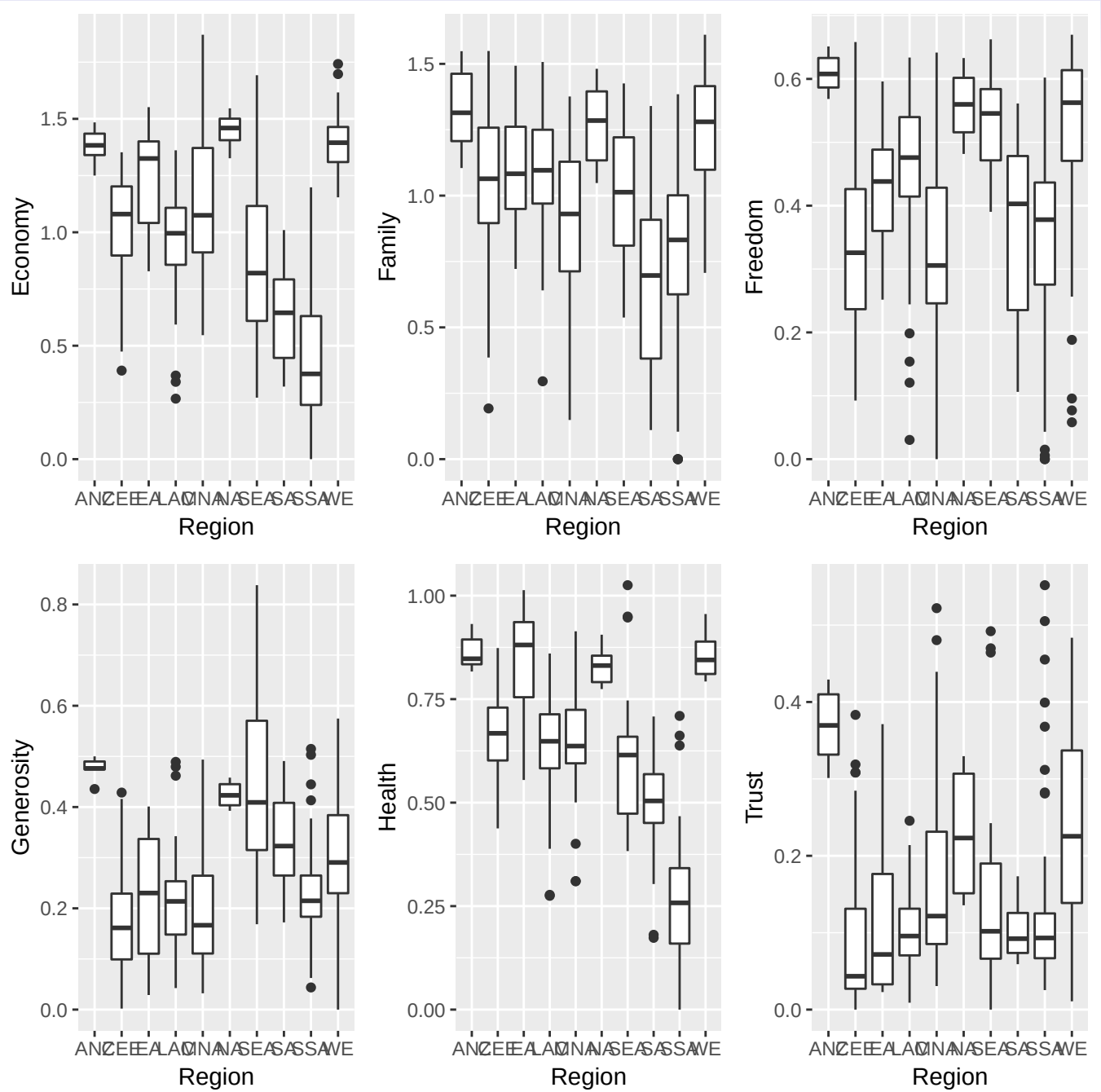
```
'data.frame': 470 obs. of  11 variables:
 $ Country                       : Factor w/ 166 levels "Afghanistan",..:
                                   136 59 38 106 25 46 100 135 101 7 ...
 $ Region                        : Factor w/ 10 levels "ANZ","CEE","EA",..:
                                   10 10 10 10 6 10 10 10 1 1 ...
 $ Happiness.Rank                : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Happiness.Score               : num  7.59 7.56 7.53 7.52 7.43 ...
 $ Economy..GDP.per.Capita.      : num  1.4 1.3 1.33 1.46 1.33 ...
 $ Family                        : num  1.35 1.4 1.36 1.33 1.32 ...
 $ Health..Life.Expectancy.      : num  0.941 0.948 0.875 0.885 0.906 ...
 $ Freedom                       : num  0.666 0.629 0.649 0.67 0.633 ...
 $ Trust..Government.Corruption.: num  0.42 0.141 0.484 0.365 0.33 ...
 $ Generosity                    : num  0.297 0.436 0.341 0.347 0.458 ...
 $ Dystopia.Residual             : num  2.52 2.7 2.49 2.47 2.45 ...
```


Region
Score
Economy
Family
Health
Freedom
Trust
Generosity
Corr: 0.785
Corr: 0.637
Corr: 0.748
Corr: 0.56
Corr: 0.406
Corr: 0.164
Corr: 0.585
Corr: 0.792
Corr: 0.345
Corr: 0.302
Corr: −0.0142
Corr: 0.486
Corr: 0.425
Corr: 0.159
Corr: 0.0718
Corr: 0.36
Corr: 0.263
Corr: 0.0773
Corr: 0.49
Corr: 0.343
Corr: 0.287

Model 1 (Full Model)

```r
model1 = lm(Score~Economy+Family+Health+
                  Freedom+Trust+Generosity,
                  data=selected.table)
drop1(model1)
```

```
Single term deletions

Model:
Score ~ Economy + Family + Health + Freedom + Trust + Generosity
           Df Sum of Sq    RSS     AIC F value    Pr(>F)
<none>                  143.51 -543.56
Economy     1   24.9520 168.47 -470.22 80.4990 < 2.2e-16 ***
Family      1   12.5422 156.06 -506.18 40.4631 4.817e-10 ***
Health      1   15.2539 158.77 -498.08 49.2114 8.198e-12 ***
Freedom     1   14.4376 157.95 -500.51 46.5779 2.769e-11 ***
Trust       1    2.9714 146.49 -535.93  9.5862  0.002079 **
Generosity  1    0.8624 144.38 -542.74  2.7822  0.095991 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```


Economy
Family
Freedom
Generosity
Health
Trust
Region
ANZCEEEALACMNAMNASEASASSASSAWE

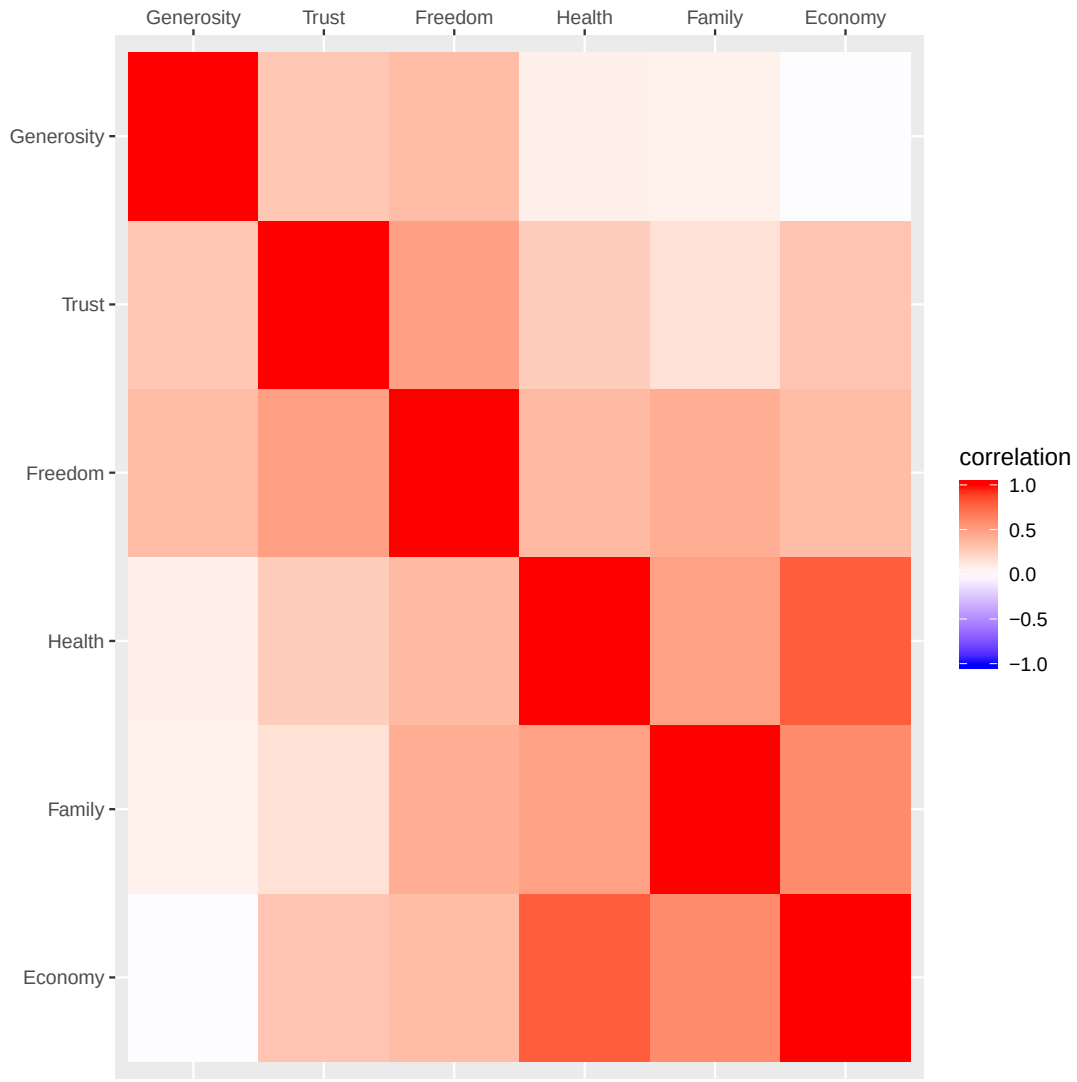
Model 2 W/ Region Factor

```
model2 = lm(Score~(Economy + Health) * Region
            + Family + Freedom + Trust
            + Generosity, data=selected.table)
summary(model2)
```

Part of Summary with only p-values

	Pr(>|t|)	
(Intercept)	0.506	
Health	0.931	
Economy	0.966	
RegionCentral and Eastern Europe	0.801	
RegionEastern Asia	0.703	
RegionLatin America and Caribbean	0.694	
RegionMiddle East and Northern Africa	0.616	
RegionNorth America	0.643	
RegionSoutheastern Asia	0.688	
RegionSouthern Asia	0.798	
RegionSub-Saharan Africa	0.745	
RegionWestern Europe	0.393	
Family	2.86e-07	***
Trust	0.219	
Freedom	1.42e-09	***
Health:RegionCentral and Eastern Europe	0.999	
Health:RegionEastern Asia	0.893	

Correlations



Model 3 PCA

```
Importance of components:
                            PC1     PC2      PC3      PC4      PC5         PC6
Standard deviation     0.0782 0.02447 0.007521 0.005253 0.004173 3.071e-19
Proportion of Variance 0.8972 0.08789 0.008300 0.004050 0.002550 0.000e+00
Cumulative Proportion  0.8972 0.98510 0.993400 0.997450 1.000000 1.000e+00
```

```
1 pca.lm.model = lm(Score ~ PC1 + PC2, data=pca.table)
```

Model 3 PCA

```
Call:
lm(formula = Score ~ PC1 + PC2, data = pca.table)

Residuals:
    Min      1Q  Median      3Q     Max
-2.1214 -0.5489 -0.1044  0.5886  2.3613

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.71567    0.09822  37.829   <2e-16 ***
PC1         -12.91904    0.58502 -22.083   <2e-16 ***
PC2          -2.78116    2.89214  -0.962    0.337
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.7969 on 467 degrees of freedom
Multiple R-squared:  0.5108,Adjusted R-squared:  0.5087
F-statistic: 243.8 on 2 and 467 DF,  p-value: < 2.2e-16
```

Model 4 PCA

```
pca.lm.model = lm(Score ~ PC1 * Region, data=pca.table)
```

First Component:

Economy	Family	Health	Freedom	Trust	Generosity
-0.12536761	-0.04213843	-0.02279518	0.04786830	0.06326022	0.07917271

Model Results

Models	AIC Values	BIC Values
six variables model	759.243	825.465
partial region interaction model	663.873	809.218
two PCs model	1125.423	1142.033
one PC with region model	817.551	904.758

Future Plan

- Check model assumptions (independence, normality).
- Transformations on some variables.
- Other methods to reduce collinearity.
- Adding another variable (for example, Development).

Developed Countries Variable

- The statistical annex contains a set of data that the World Economic Situation and Prospects (WESP) employs to delineate trends in various dimensions of the world economy.

Developed economies

Europe			Other countries	Major developed economies (G7)
European Union	New EU member States	Other Europe		
EU-15	Bulgaria	Iceland	Australia	Canada
Austria	Croatia	Norway	Canada	Japan
Belgium	Cyprus	Switzerland	Japan	France
Denmark	Czech Republic		New Zealand	Germany
Finland	Estonia		United States	Italy
France	Hungary			United Kingdom
Germany	Latvia			United States
Greece	Lithuania			
Ireland	Malta			
Italy	Poland			
Luxembourg	Romania			
Netherlands	Slovakia			
Portugal	Slovenia			
Spain				
Sweden				
United Kingdom				

Other possible explorations

- Clutering countries not according physical locations but based on their happiness, economy, etc.
- Gathering Data from 2018 and check if our model still fits.

References

https://www.kaggle.com/unsdsn/world-happiness/activity
https://www.gallup.com/178667/gallup-world-poll-work.aspx
https://github.com/Jacksoften/WorldHappiness

Thank You!!!