World Happiness Report – Data Analysis using SPSS and Excel

By – Tomer Itzhakov-Nevo

Happiness define as a "state of well-being and contentment" (Merriam-Webster). In each year the World Happiness Report ranks happiness scores of 156 countries by a verity of variables. The rank is based on the pooled results from Gallup World Poll Surveys from 2015 – 2017. This report tries to understand the happiness form top to bottom through analysis of several questions:

1. What is the difference between the happiness score of 2015, 2016, 2017?
2. Who are the tenth happiest countries in the world? And why?
3. What is the main difference of happiness score by each Region in the world?
4. What are the best variables to predict happiness score?

Q1: what is difference between the happiness score of 2015, 2016, 2017?

To check if there is a difference between the world happiness score between the years 2015, 2016, 2017 it was conducted One-Way ANOVA test. The ANOVA test produce mean to each group and then compare them to each other to find if there is a statistically significant difference between them. The results are in the table below.

Variable Name	Group Name	Mean value	SD value	F value	P value
Happiness score	2015	5.37	1.14	F (2,467) =0.026	P > 0.05
	2016	5.38	1.14		
	2017	5.35	1.13		

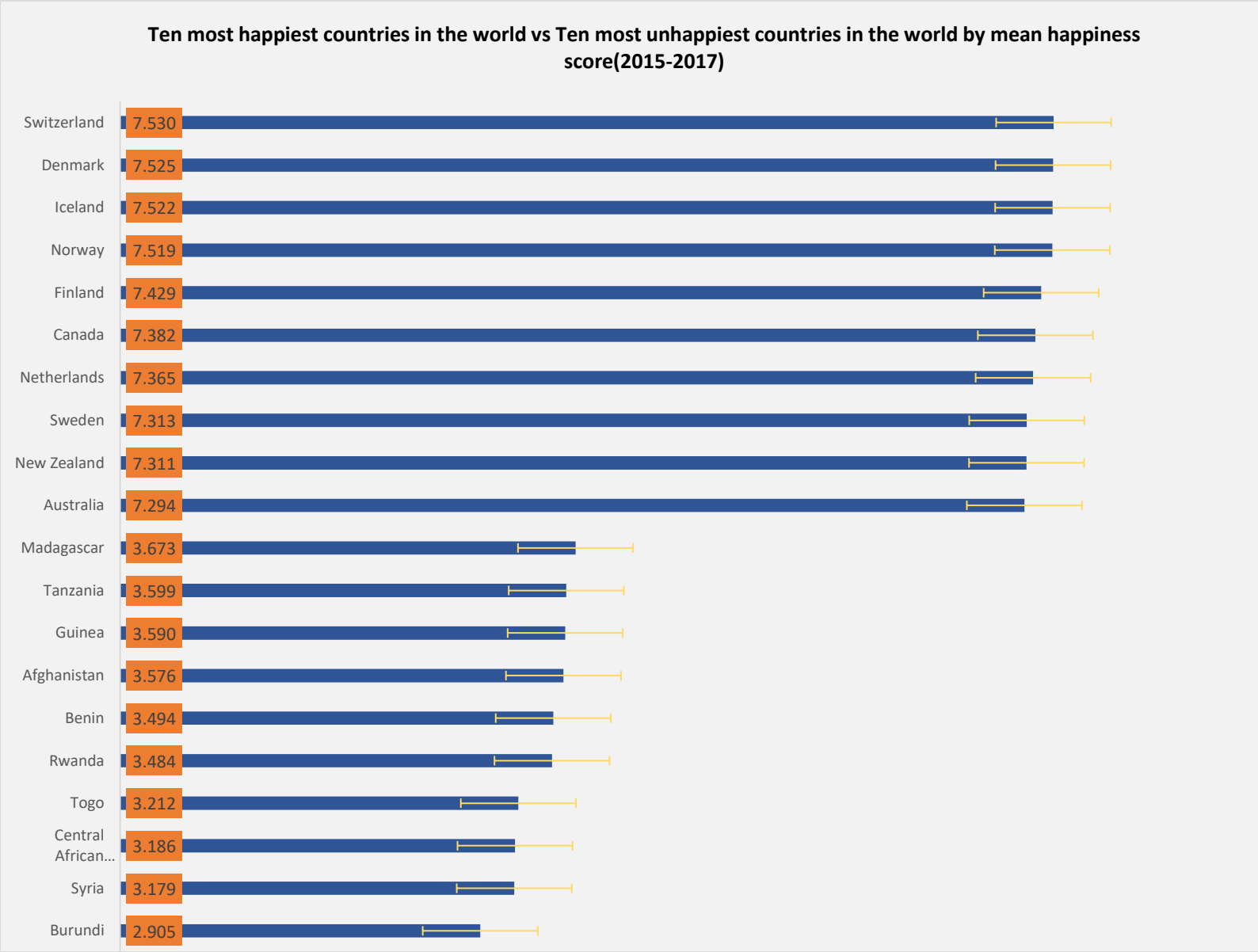
The analysis found that there are no significant differences between each year, F (2,467) =0.026, P > 0.05. In other words, there is no difference in the Happiness score between the year 2015, 2016, 2017. It means that overall the world happiness score is a stable state in a period of three years. But, what about the countries within that state? Who are the countries with the highest happiness score? Who are the countries with the lowest one? And what are the difference between them. Let's find out in the next question.

Q2: Who are the tenth happiest countries in the world? And why?

The chart in the next page demonstrates the difference between ten of the happiest countries and ten of the unhappiest countries by happiness scores. Because happiness is generated form a different type

of variables, the report also gives each country six variables that can tell us the story of happiness from a different angle. These are the variables:

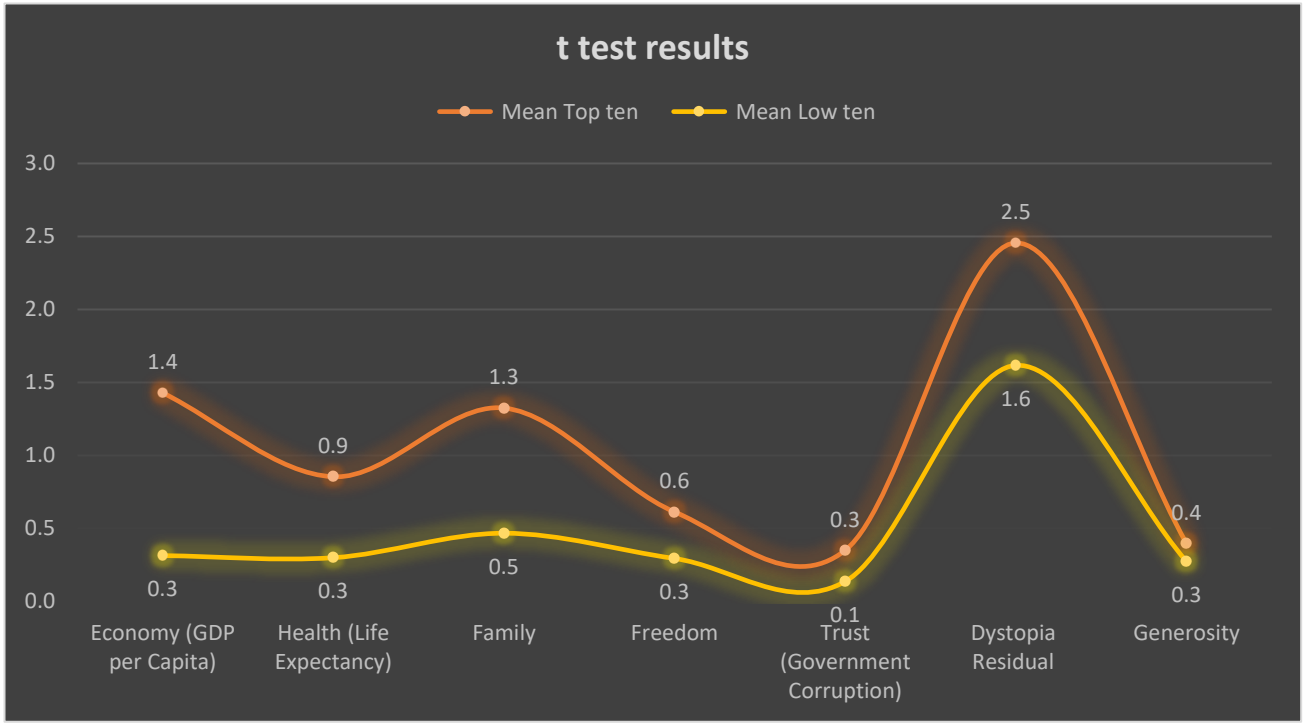
1. Economy (GDP per Capita)
2. Family
3. Health (Life Expectancy)
4. Freedom
5. Trust (Government Corruption)
6. Generosity
7. Dystopia Residual



To check why these countries are happier than other countries it was conducted Two-independent t tests for all the variables in the happiness reports data for 2015-2017. In the analysis it was found that there is a significant statistical difference between all the variables of the report.

Variable Name	Group Name	Mean value	SD value	t test value	p value
Economy	Top ten	1.42800	.089704	t (39.255) = 28.290	p < .001
	Low ten	.31293	.193071		
Family	Top ten	1.32423	.169706	t (43.071) =13.117	p < .001
	Low ten	.46641	.310138		
Health	Top ten	.85340	.046075	t (32.853) =18.675	p < .001
	Low ten	.29862	.153431		
Freedom	Top ten	.61203	.032373	t (30.151) =10.382	p < .001
	Low ten	.29310	.162331		
Government Trust	Top ten	.34973	.085587	t (57) = 7.396	p < .001
	Low ten	.13610	.132125		
Generosity	Top ten	.39627	.082169	t (57) = 5.452	p < .001
	Low ten	.27297	.091425		
Dystopia Residual	Top ten	2.45547	.221519	t (35.340) =7.108	p < .001
	Low ten	1.61645	.597178		

The chart below presents the t test results from the most statistically significant t test to the lowest statistically significant t test. The top 3 variables are GDP, Health and Family while the lowest 3 variables are Trust, Dystopia-Residual and Generosity. In conclusion strong economy, better health



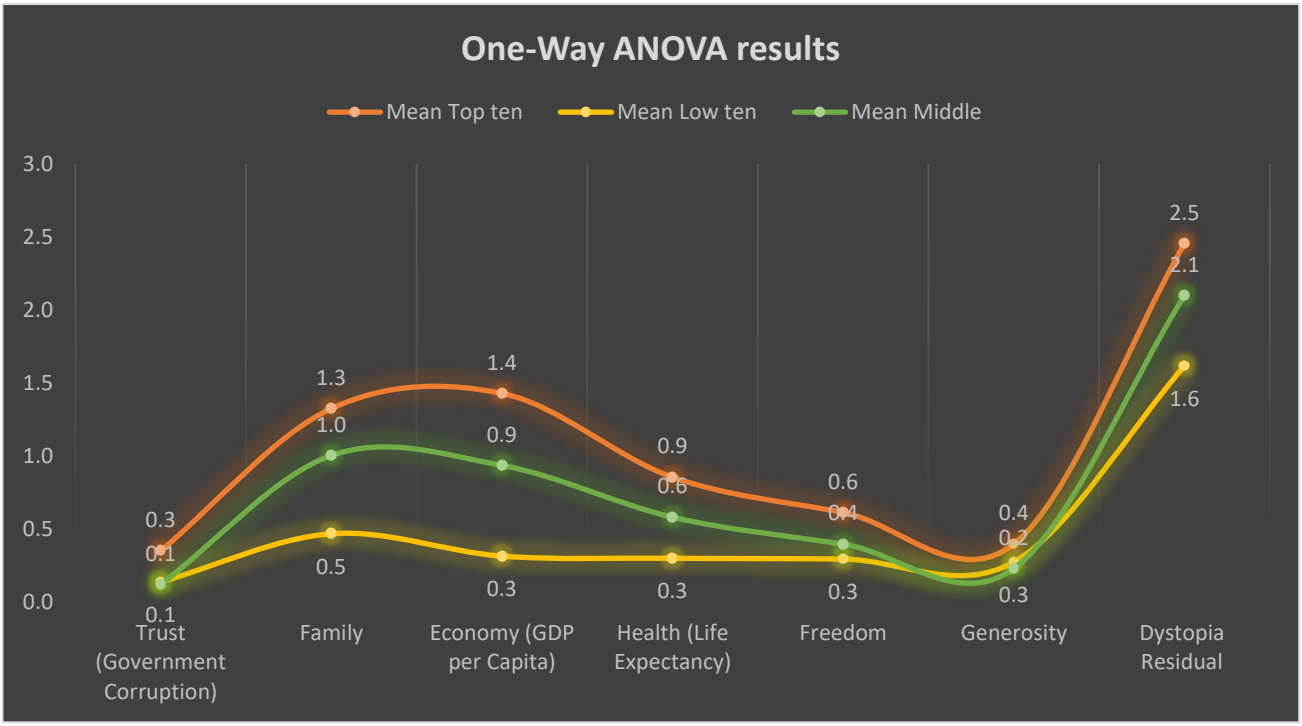
and good family communication are connected variables to happier life. The lowest variables are interesting, generosity is present at both of the top and lowest countries, civilians in the top ten are less Dystopia believers but they have better quality of life while the low ten have harder life. In addition, both of the civilians among the two groups don't' trust their government reliability.

What will happen to the results if we enter another group – the middle group. To answer this question, first I created the third group and then One-Way ANOVA test was conducted. The results are presented in the table below just for the statistically significant results of the ANOVA.

Variable Name	**Group Name**	**Mean**	**SD**	**F value**	**p value**
Economy	Low ten	.31293	.193071	F (2, 467) = 69.052	p < .001
	Top ten	1.42800	.089704		
	Middle	.93469	.386495		
Family	Low ten	.46641	.310138	F (2, 467) =73.214	p < .001
	Top ten	1.32423	.169706		
	Middle	1.00295	.282562		
Health	Low ten	.29862	.153431	F (2, 467) =47.083	p < .001
	Top ten	.85340	.046075		
	Middle	.57986	.230521		
Freedom	Low ten	.29310	.162331	F (2, 467) = 44.122	p < .001
	Top ten	.61203	.032373		
	Middle	.39530	.141002		
Government Trust	Low ten	.13610	.132125	F (2, 467) = 80.434	p < .001
	Top ten	.34973	.085587		
	Middle	.11899	.093969		
Generosity	Low ten	.27297	.091425	F (2, 467) = 25.984	p < .001
	Top ten	.39627	.082169		
	Middle	.22884	.129497		
Dystopia Residual	Low ten	1.61645	.597178	F (2, 467) = 17.643	p < .001
	Top ten	2.45547	.221519		
	Middle	2.09982	.559123		

The chart below presents the one-way ANOVA test results from the most statistically significant ANOVA test to the lowest statistically significant ANOVA test. The top 3 variables are Trust, Family and GDP while the lowest 3 variables are Freedom, Generosity and Dystopia-Residual. In conclusion all the variables were statistically significant, and most of the variables remain the same for the top 3

variable and the low 3 variables. The family, GDP and Health are the most contribute variables to happiness among countries. Additionally, I was assuming that there will be no statistically significant difference between the top group and the middle group. But, the significant of the results teach us about the inequality of the world in terms of happiness and the other variables.



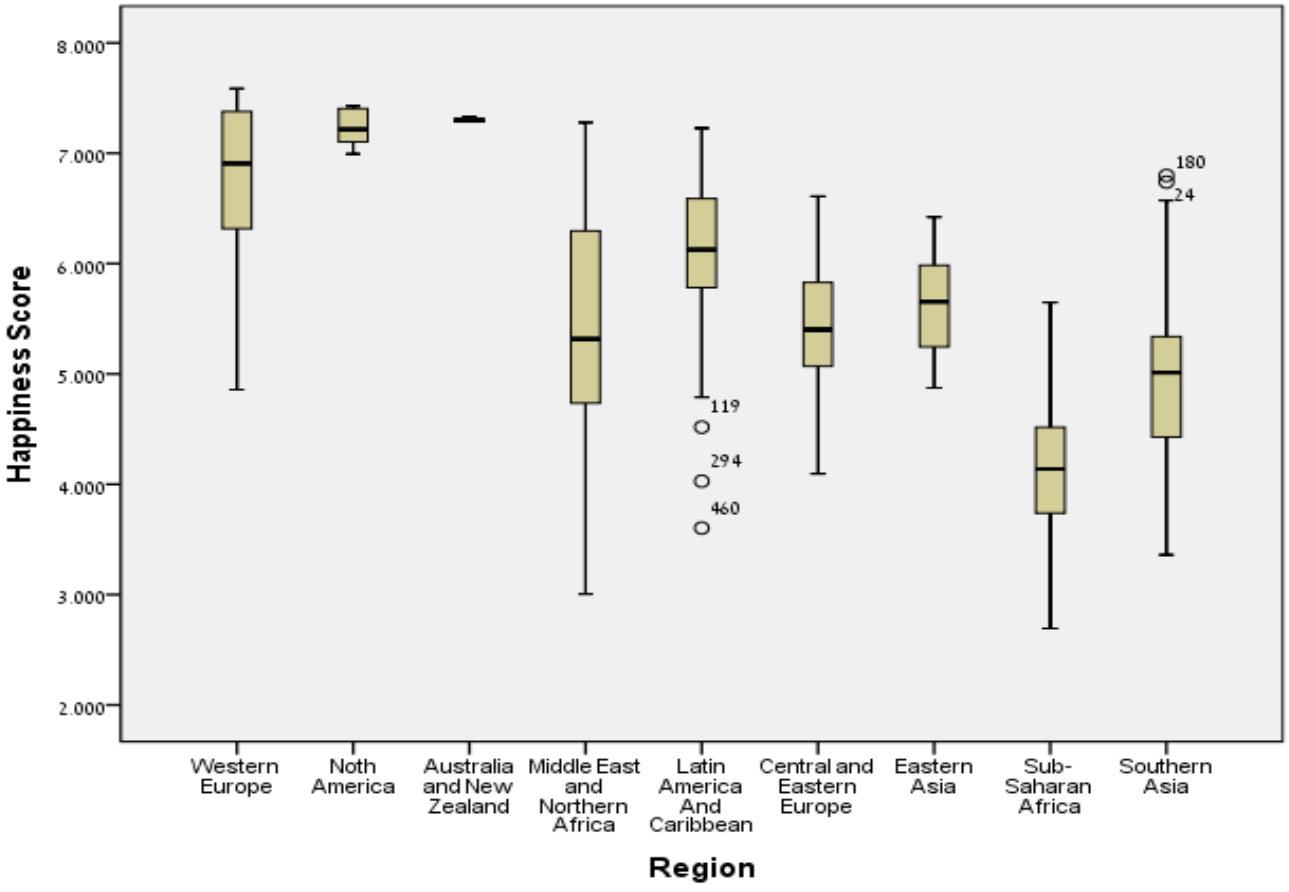
Q3: What is the main difference of happiness score by each Region in the world?

The report of happiness divides all the countries into a Region area. In total there is 10 regions in the report: Western Europe, North America, Australia and New Zealand, Middle East and Northern Africa, Latin America And Caribbean, Central and Eastern Europe, Eastern Asia, Sub-Saharan Africa and Southern Asia. To answer this question, it was conducted one-way ANOVA test. Here are the mean and SD for each region.

Region Name	**Mean**	**Std. Deviation**
Western Europe	6.69	0.78
North America	7.23	0.18
Australia and New Zealand	7.30	0.02
Middle East and Northern Africa	5.39	1.03
Latin America And Caribbean	6.07	0.73
Central and Eastern Europe	5.37	0.58
Eastern Asia	5.63	0.50
Sub-Saharan Africa	4.15	0.58
Southern Asia	5.02	0.84

The analysis of the ANOVA test revels a statistically significant difference of happiness scores between the region groups, F (8, 461) = 88.7, p < 0.001. In a post hoc tests (Scheffe) it was found that Australia and New-Zealand (M=7.30, SD =0.02), North America (M= 7.23, SD =0.18) and Western Europe (M= 6.69, SD =0.78) was statistically higher than all the other regions, but not statistically significant among themselves. Another post hoc test reviles that Latin America And Caribbean (M=6.07, SD =0.73) was statistically significantly higher than Middle East and Northern Africa (M=5.39, SD =1.03), Central and Eastern Europe (M=5.37, SD =0.58), Sub-Saharan Africa (M=4.15, SD =0.58) and Southern Asia (M=5.02, SD =0.84). Another post hoc test reviles that the Middle East Northern Africa (M=5.39, SD =1.03), Central and Eastern Europe (M=5.37, SD =0.58), and Southern Asia (M=5.02, SD =0.84) was statistically significantly higher than Sub-Saharan Africa (M=4.15, SD =0.58) but not statistically significant among themselves. In conclusion the happiest regions in the world are Australia and New-Zealand, North America and Western Europe and the unhappiest region in the world is Sub-Saharan in Africa.

the chart below demonstrates the difference between the regions by happiness scores. The black line in each part is the mean happiness score of each region. The lines that starches outside the box can teach us about the variance of each region. Among all the regions, the variance of the middle east is the biggest one with an SD of 1.03.



To dive and understand the variance of the middle east it was conducted one-way ANOVA among the countries in the middle east by happiness score and all the other variables. In conclusion there is a statistically significant deference between all the countries among all the variables. This is not the place for introducing all the post hoc test, but if you wish to search inside the results, look in the output file of the SPSS.

Variable Name	**F value**	**p value**
Happiness Score	F (18, 38) = 112.011	p < .001
Economy (GDP per Capita)	F (18, 38) =71.953	p < .001
Family	F (18, 38) = 4.238	p < .001
Health (Life Expectancy)	F (18, 38) = 9.957	p < .001
Freedom	F (18, 38) = 16.203	p < .001
Trust (Government Corruption)	F (18, 38) = 97.283	p < .001
Generosity	F (18, 38) = 90.789	p < .001
Dystopia Residual	F (18, 38) = 8.171	p < .001

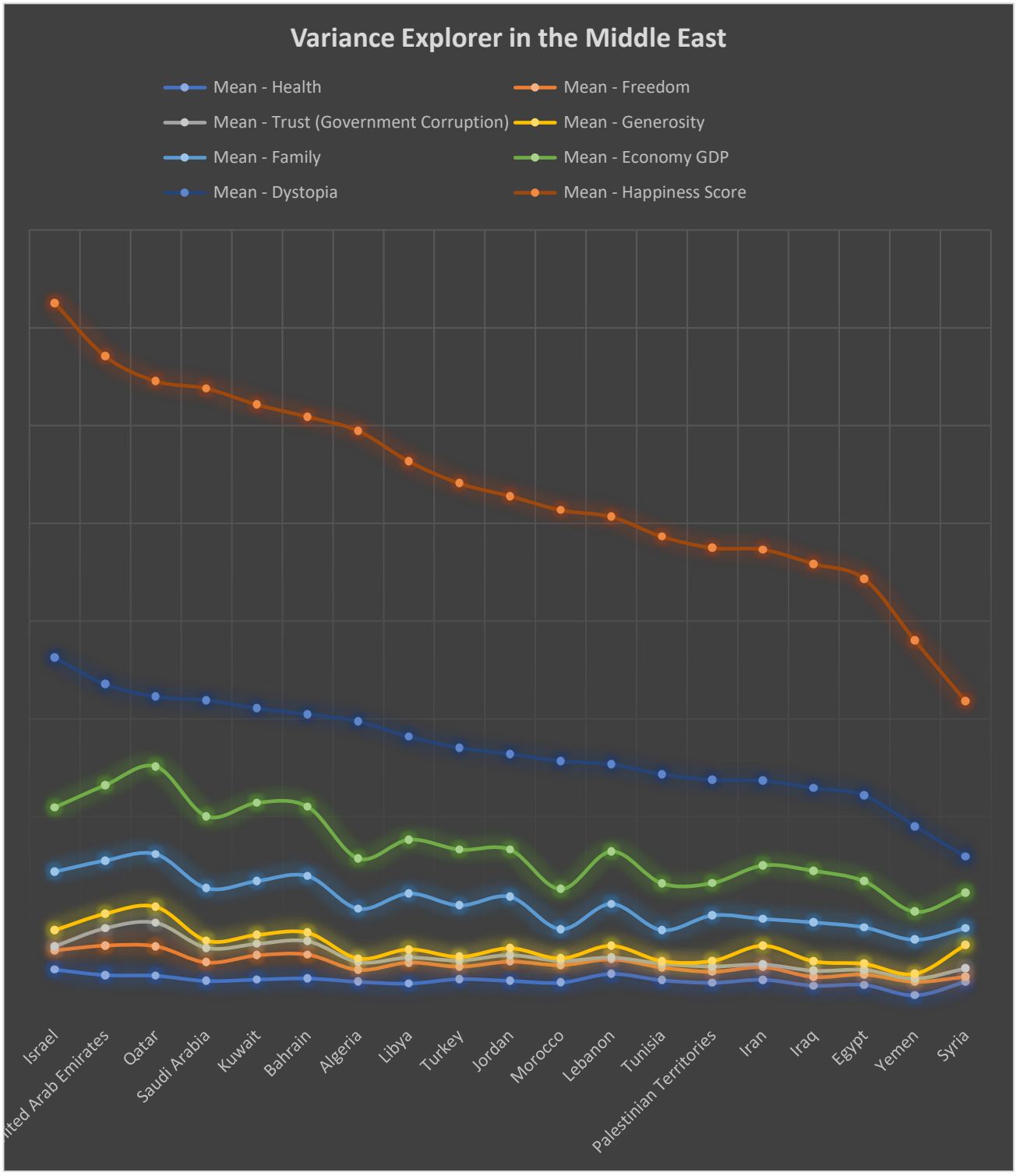
The chart in the next page demonstrates the variance in the Middle east by happiness score and all the other variables. The chart is order by the happiness score from top to bottom. The variance explorer can be conducted in two ways: the first way is to look on the rows for the difference (variance) between the countries in all the variables. The second way is to look on the columns to explorer all the different variables in one country.

interesting thing is to compare between the top 5 countries and the bottom 5 countries. While the top 5 countries (Israel, United Arab Emirates, Qatar, Saudi-Arabia and Kuwait) have stability in their leadership the 5 bottom countries (Iran, Iraq, Egypt, Yemen and Syria) during the creation of the report have instability leadership and most of them are a war zone area.

If we explorer the Economy (GDP - Green line) and Happiness score we can see that Israel has the happiest citizens in the middle east but not the strongest Economy like in United Arab Emirates and Qatar. This example shows that Money can influence on happiness but not absolutely. Extreme example for this insight we can see in the case of Morocco - the citizens are quite happy but the Economy is weak.

Another interesting insight is found in the Generosity (yellow line) variable. In a post hoc test (Scheffe) it was found that the most unhappy country (Syria) has statistically significant more higher mean rank of Generosity (M=0.48, SD=0.011) than the other countries [Israel (M=0.32, SD=0.005), United Arab Emirates (M=0.29, SD=0.055), Qatar (M=0.32, SD=0.003), Saudi-Arabia (M=0.14, SD=0.009), Kuwait (M=0.18, SD=0.039),Bahrain (M=0.17, SD=0.002), Libya (M=0.16, SD=0.016), Algeria

(M=,0.07 SD=0.005), Turkey (M=0.07, SD=0.044),Jordan(M=0.14, SD=0.031), Morocco (M=0.04, SD=0.021), Lebanon (M=0.24, SD=0.023), Tunisia (M=0.05, SD=0.013), Palestinian Territories (M=0.11, SD=0.001), Iraq (M=0.19, SD=0.011), Egypt (M=0.12, SD=0.009), Yemen (M=0.09, SD=0.007)] except Iran (M=0.38, SD=0.003). we can explain the result by trying to understanding the human behavior – civilians during war suffer form many types of reasons like hunger, injury many more. Maybe Because of the suffering people are more generous with each other's.



Q4: What are the best variables to predict happiness score?

Two steps were made to answer this question:

1. Pearson Correlations test to see if there is a connection between the happiness score and the other variables.
2. If there is strong connection, I will conduct a stepwise regression to predict happiness scores of countries in the future by tree of the most correlated variables among all the variables in the report.

Pearson Correlation Test Results:

To check if there is a statistically significant correlation between the variables of the happiness report it was conducted Pearson Correlation Test. In the analysis it was found that among all the variables there is a statistically significant strong and positive connection between them. In other words, when the mean in every variable is higher thus the mean in other variables is getting higher.

Correlations

		Happiness Score	Economy	Family	Health	Freedom	Government Trust	Generosity	Dystopia Residual
Happiness Score	Pearson Correlation	1	.785**	.637**	.748**	.560**	.406**	.164**	.490**
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000	.000
	N	470	470	470	470	470	470	470	470

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

In the analysis it was found that the three variables who are the most connected to Happiness scores are Economy (0.785), Health (0.748) and Family (0.637). in other words when the Economy stronger the Citizens happier, and when the they are Healthier and have good Family membership, they are happier.

Stepwise regression Test Results:

To check who from the three variables (Economy, Health and Family) is contribute more in the explanation of the variance of the happiness score it was conducted Stepwise Regression test. In the first step the analysis show that the mean of Economy and the explained variance is R^2 =61.7%, p<.001. The second step of the analysis shows the mean of the Family and the explained variance is R^2 = 66.4%, p<.001. In the third step the analysis shows the mean of the Health and the explained variance

R^2 = 70.3%, p<.001. in other words, the most predictable variable of happiness is Economy then Family and last Health. In total, all of them together can explain 70.3% of the happiness score variance and also predict 70.3% of the happiness score in the future. This point gets also to be given strength from the first ANOVA of this report that found no statistically significant difference in happiness scores by year. It means that as long as the Economy, Family structure and Health remine stable by year the model of the regression can be more accurate.

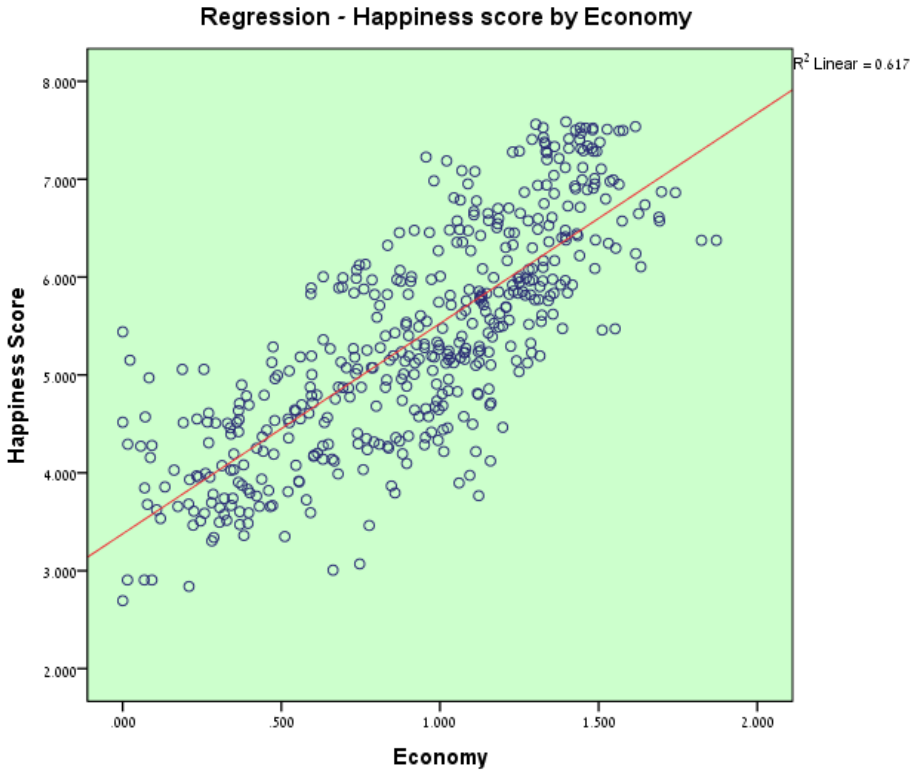
Steps	**Variables**	**B**	**SE B**	**β**	**R^2**	**Δ R^2**
Step 1	Economy	1.037	0.099	0.379***	0.617***	0.617***
Step 2	Economy	1.037	0.099	0.379***	0.617***	0.617***
	Family	0.920	0.111	0.258***	0.664***	0.047***
Step 3	Economy	1.037	0.099	0.379***	0.617***	0.617***
	Family	0.920	0.111	0.258***	0.664***	0.047***
	Health	1.526	0.196	0.322***	0.703***	0.039***
B = the x value (mean) of each variable SE B = std. Error of x value β = Pearson correlation strength R^2 = The explained variance Δ R^2 = The change in the contribution of each variable to the model					*p<.05 **p<.01 ***p<.001	

To visualize the steps of the model I conducted 4 scatterplot maps. Three of the scatterplots are representative of a single variable to predict happiness score. The last one is combining all of them into one and show a prediction of 70.3% of happiness scores. The regression will be presented from the most positive correlated variable to the least positive correlated variable:

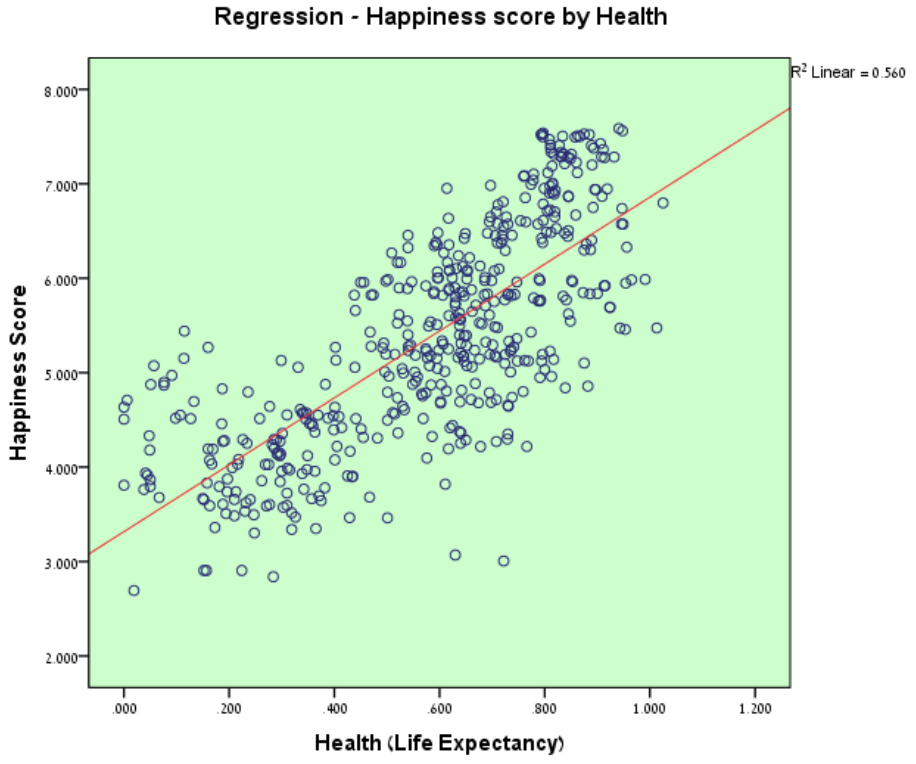
1. The first scatterplot regression is happiness score by Economy (R^2 = 0.617)
2. The second scatterplot regression is happiness score by Family (R^2 = 0.560)
3. The third scatterplot regression is happiness score by Health (R^2 = 0.405)
4. The fourth scatterplot regression is happiness score by the combined value of all the variables together (R^2 = 0.703).

In the singles regression there is a big variance compere to the stepwise regression. You can see it in the scattering dots in each graph. When the scattering dots are close to each other the variance is low and the predicted values are more accurate.

single regression - happiness score by Economy (R^2 = 0.617)

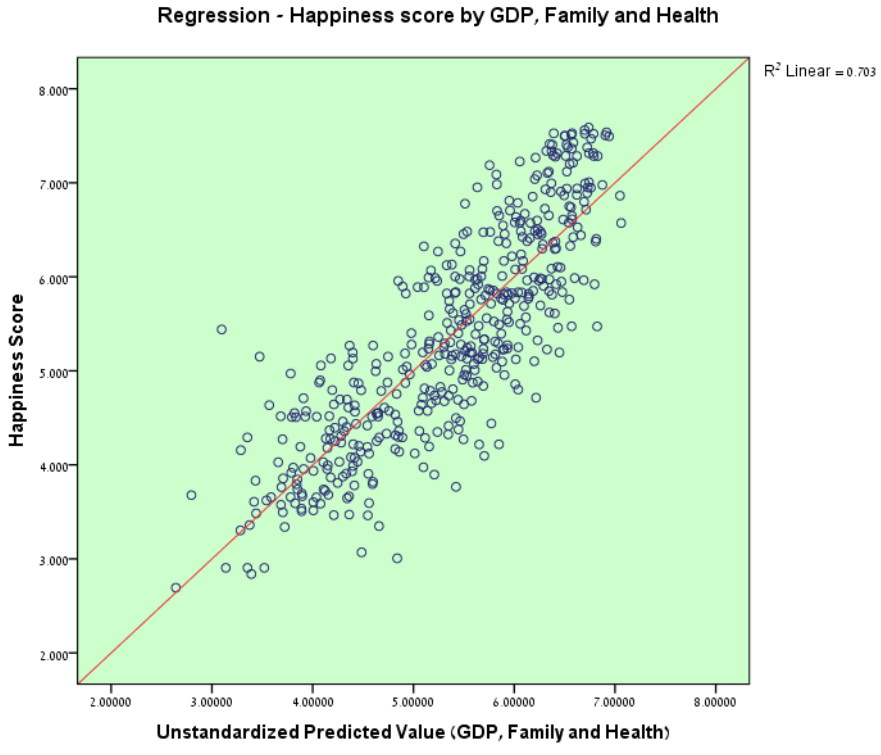


single regression - happiness score by Health (R^2 = 0.560)



single regression - happiness score by Family (R^2 = 0.560)

Stepwise regression - happiness score by Predicted Combine Value (R^2 = 0.703)



Summery

In sum, this report tries to answer four questions:

1. What is the difference between the happiness score of 2015, 2016, 2017?
2. Who are the tenth happiest countries in the world? And why?
3. What is the main difference of happiness score by each Region in the world?
4. What are the best variables to predict happiness score?

It was found that there is no big and statistically significant difference between the happiness scores of the world between 2015 until 2017. But if we look more deeply, we saw that there is a vary big variance between countries in a variety of variables. The most interesting variables that contribute to happiness scores were Economy, Family relationship and Health. We saw the same picture in the analysis of the Regions – Wile New-Zealand and Australia, North America and Western Europe are the happiest regions in the world the Sub-Saharan region is the unhappiest one. in addition, we explorer the region with the biggest variance in the world – The Middle East and analyze within the region why the variance is so big. In the last part of the report we looked if there is a possibility to predict the happiness score by the different variables of the report. It was found that the same variables that was contribute to happiness scores (Economy, Family and Health) are also the most positive correlated variables to happiness scores, and by combining all of their values into single value – we can predict happiness score with an accuracy of 70.3%. I hope you enjoy reading and learned something new.