Analysis of **World Happiness** Data 2015 - 2017

Delon Saks | Abid Ladhani
W200 | Fall 2018

Berkeley
School of Information

Our Goal

Goal
To conduct an exploratory analysis of the relationships between various factors and world happiness globally using the World Happiness Report for 2015, 2016, & 2017

What is the World Happiness Report?
The World Happiness Report is a landmark survey of the state of global happiness. The first report was published in 2012. The report ranks 150+ countries by their happiness levels, and is released at the United Nations at an event celebrating International Day of Happiness on March 20th

Source: https://www.kaggle.com/unsdsn/world-happiness

How is the Happiness Score calculated?
Scores are gathered from the Gallup World Poll. View the next slide for methodology.

Methodology

Happiness Scores and their components are sourced from the Gallup World Poll

The **World Poll Survey** includes more than 100 global questions as well as region-specific items. Gallup asks residents from Australia to Pakistan the same questions, every time, in the same way[(1)]

Gallup uses:

- Telephone surveys in countries where telephone coverage represents at least 80% of the population
- Area frame design for face-to-face interviewing in randomly selected households in the developing world, including much of Latin America, the former Soviet Union countries, nearly all of Asia, the Middle East, and Africa

Scores are calculated using the **Cantril ladder**[(2)]:

- It asks respondents to think of a ladder, with the best possible life for them being a 10, and the worst possible life being a 0
- The sub-scores show the estimated extent to which each of the factors contribute to making life evaluations higher in each country than in Dystopia

Important Weakness of this analysis: You may notice that some of the values of the various factors (e.g. GDP per capita) are meant to quantify the extent to which the factor contributes to happiness. It is not clear from the report how exactly this was determined by Gallup. If we had more time, we would dig into this methodology a little more.

Data Cleaning / Sanity Checks

The Data: The dataset was presented in the form of three .csv documents that we combined. One of the CSV files had Whisker High and Whisker Low variables that we excluded from the combined dataset. The resulting combined dataset (WorldHappinessReport.csv) had 13 columns, and 471 rows of data.

Max/Min: We ran a maximum and minimum analysis on each of the variables. No negative values were found or values that were greater than the maximum happiness score.

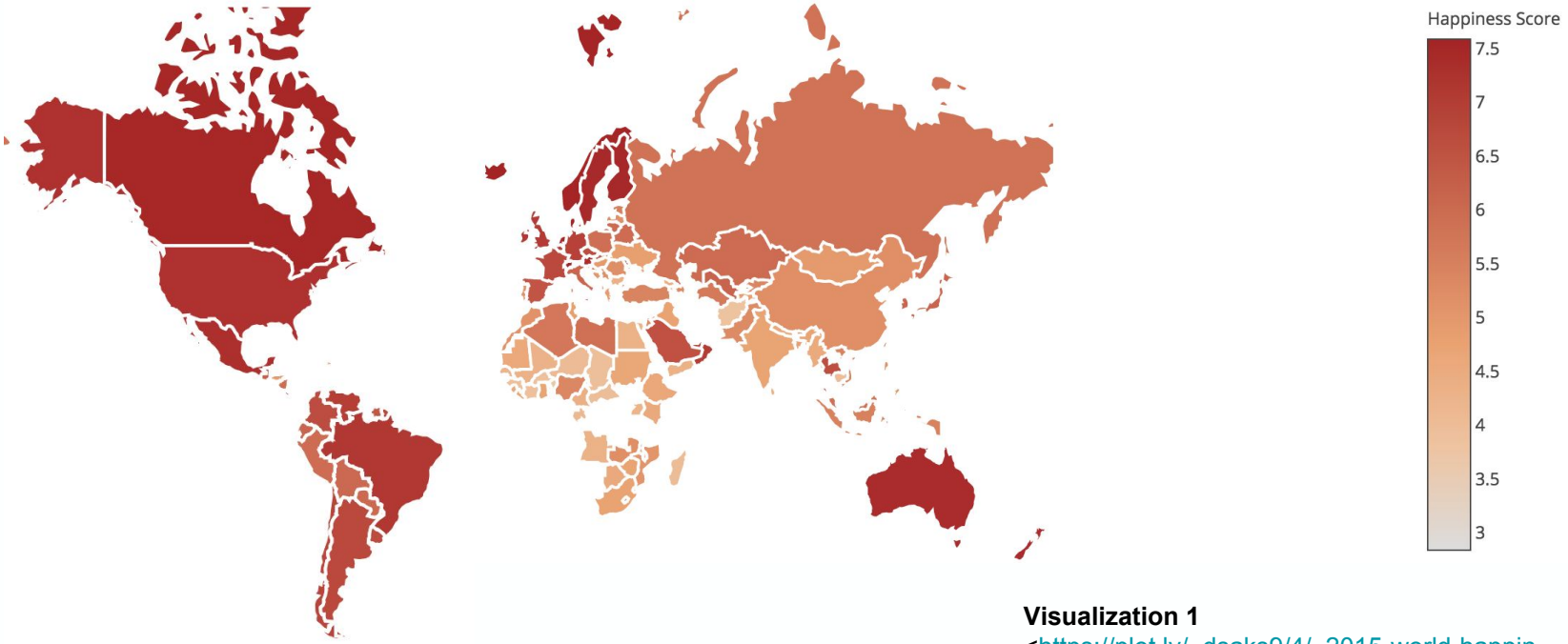
Plotting the data: We plotted the data to ensure the data passed a sanity check (e.g. that countries with higher GDP per capita generally speaking would have higher happiness scores). We also ensured there were no strange outliers.

Double-check against other sources: We checked our findings against findings published in news articles or reports to see if we were off significantly in what we found. (e.g. this Huffington Post article here, and this Bloomberg article here)

Combining Datasets: Data for 2015, 2016, and 2017 were all in separate CSVs. We wanted to combine the data into one CSV for easier data analysis. However, not all of the columns were in the right order, and one of the files contained regional data while others did not. We had to do some clean up in order to consolidate this data.

Adding Datasets: In order to create the required maps and analyze the happiness scores on a continental basis, additional .csv files had to be imported and merged with the main dataset.

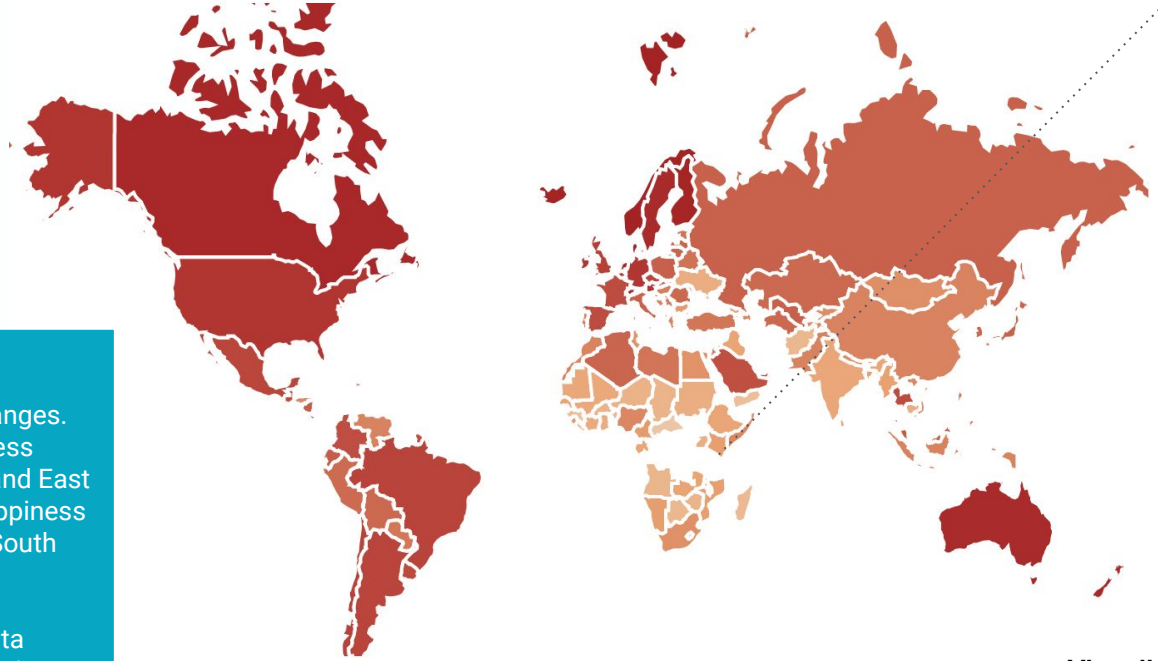
What did World Happiness look like in the world in 2015?

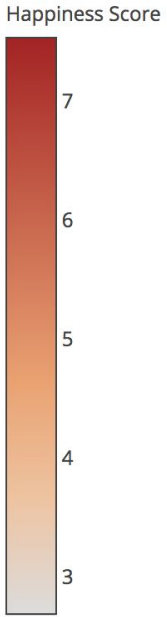


Visualization 1
<https://plot.ly/~dsaks9/4/_2015-world-happiness-score-by-country-hover-for-breakdown/>

What did World Happiness look like in the world in 2017?

Note: Missing data in parts of Africa point to challenges in data collection process.





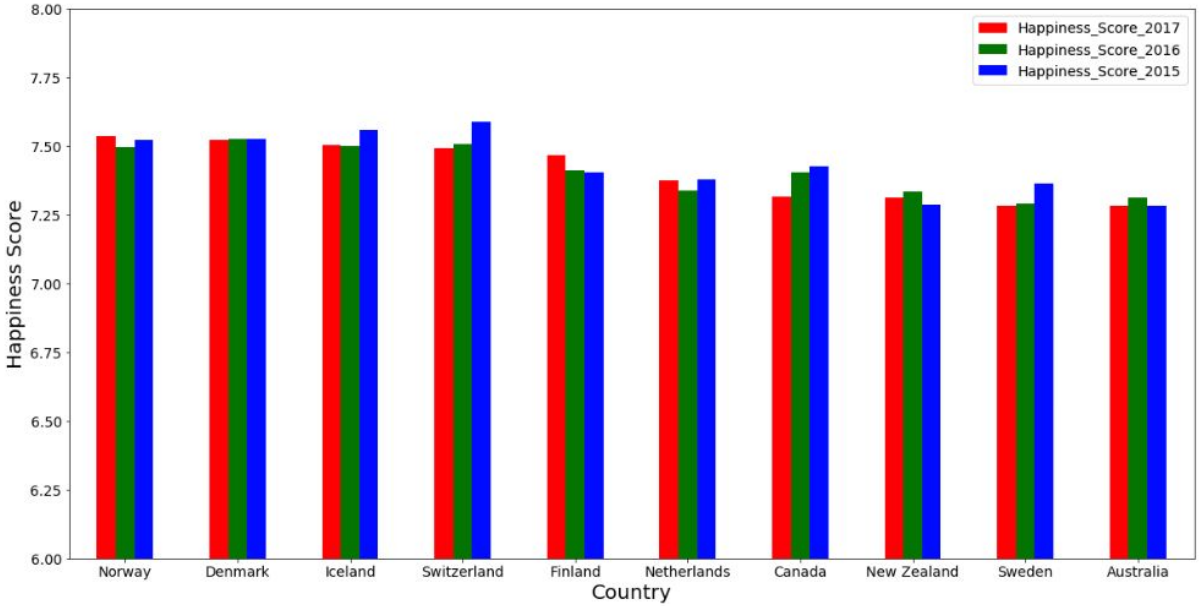
Key Takeaways:

No significant changes. Increased happiness score for Russia and East Asia. Reduced happiness score in parts of South America

Also additional data collection in Africa in 2017 to fill in some gaps

Visualization 2
<https://plot.ly/~dsaks9/8/_2017-world-happiness-score-by-country-hover-for-breakdown/>

Who are the **top 10** happiest countries?



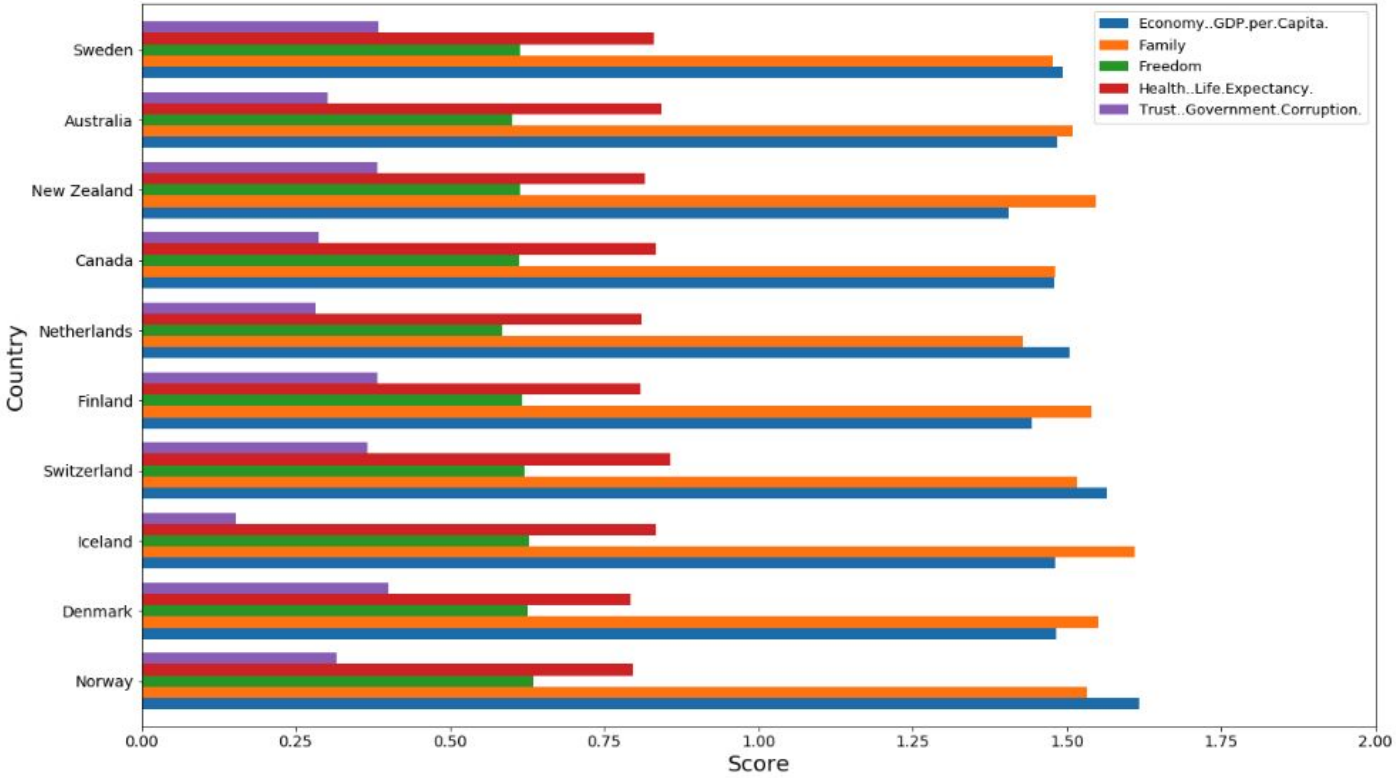
	Rank_2015	Rank_2016	Rank_2017	Rank_Diff	Continent
Norway	4	4	1	3.000	Europe
Denmark	3	1	2	1.000	Europe
Iceland	2	3	3	-1.000	Europe
Switzerland	1	2	4	-3.000	Europe
Finland	6	5	5	1.000	Europe
Netherlands	7	7	6	1.000	Europe
Canada	5	6	7	-2.000	North America
New Zealand	9	8	8	1.000	Oceania
Sweden	8	10	9	-1.000	Europe
Australia	10	9	10	0.000	Oceania

Key Takeaways:

No significant changes in rank. There is a heavy European representation with only one country from North America appearing in the top 10.

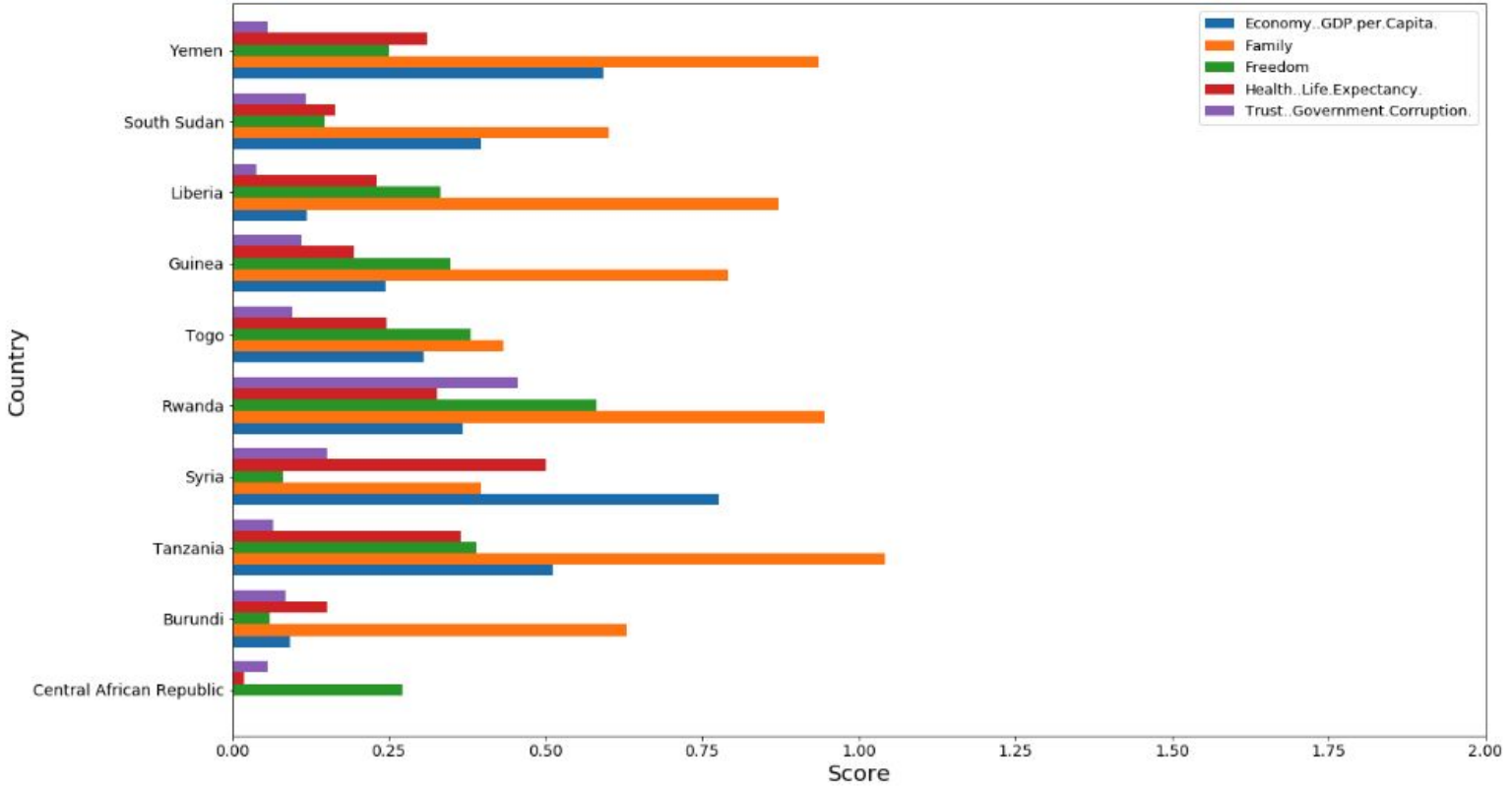
Visualization 3 & 4

Factors in the **top 10** happiest countries (2017)



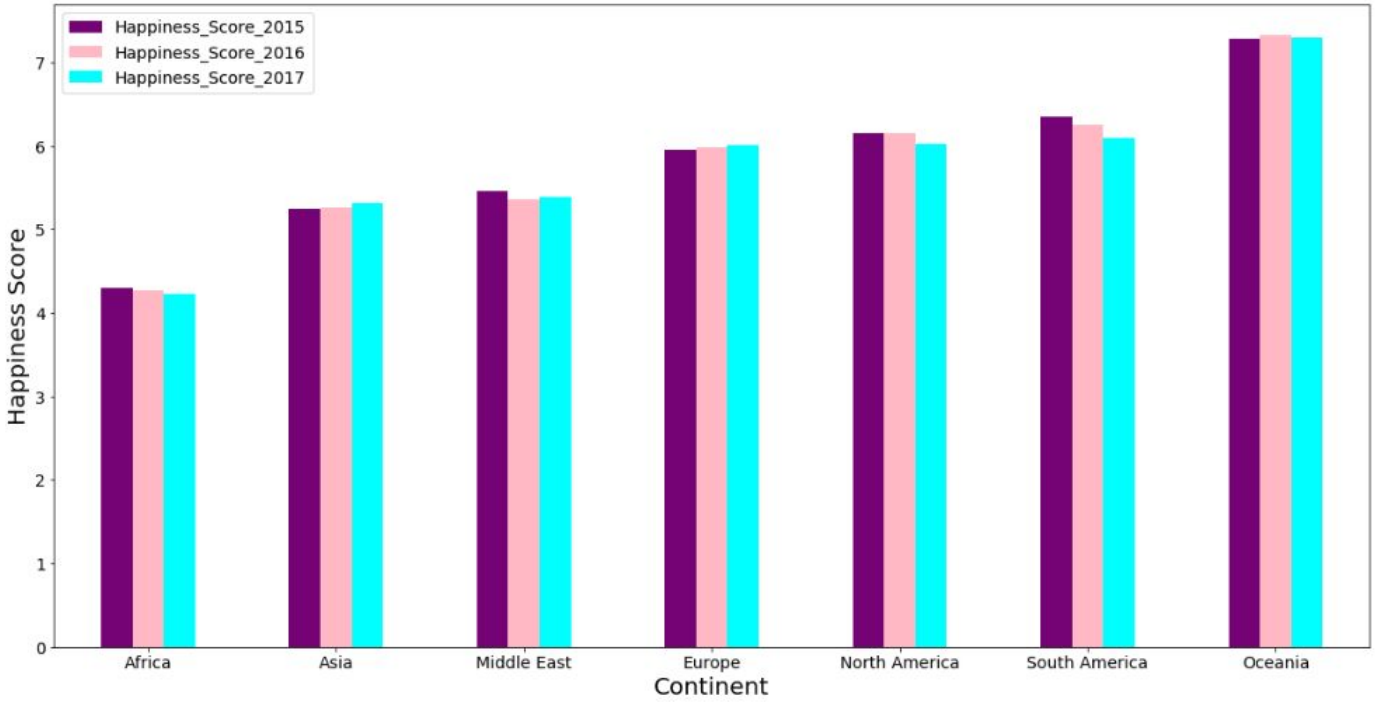
Visualization 5

Factors in the **bottom 10** happiest countries (2017)



Visualization 6

What have been the major **regional changes**?

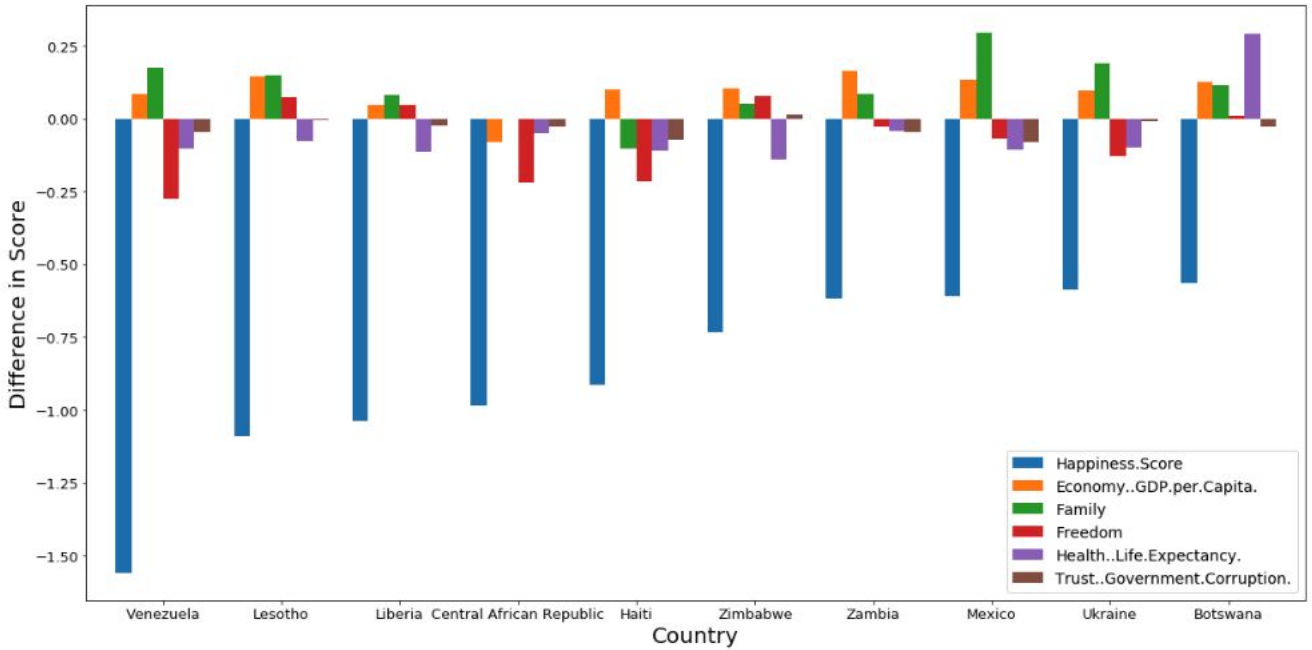


Key Takeaways:

Declining happiness scores in Africa, South America, and North America

Increasing happiness scores in Asia & Europe

What countries have driven **declining** Happiness Scores?



Key Takeaways:

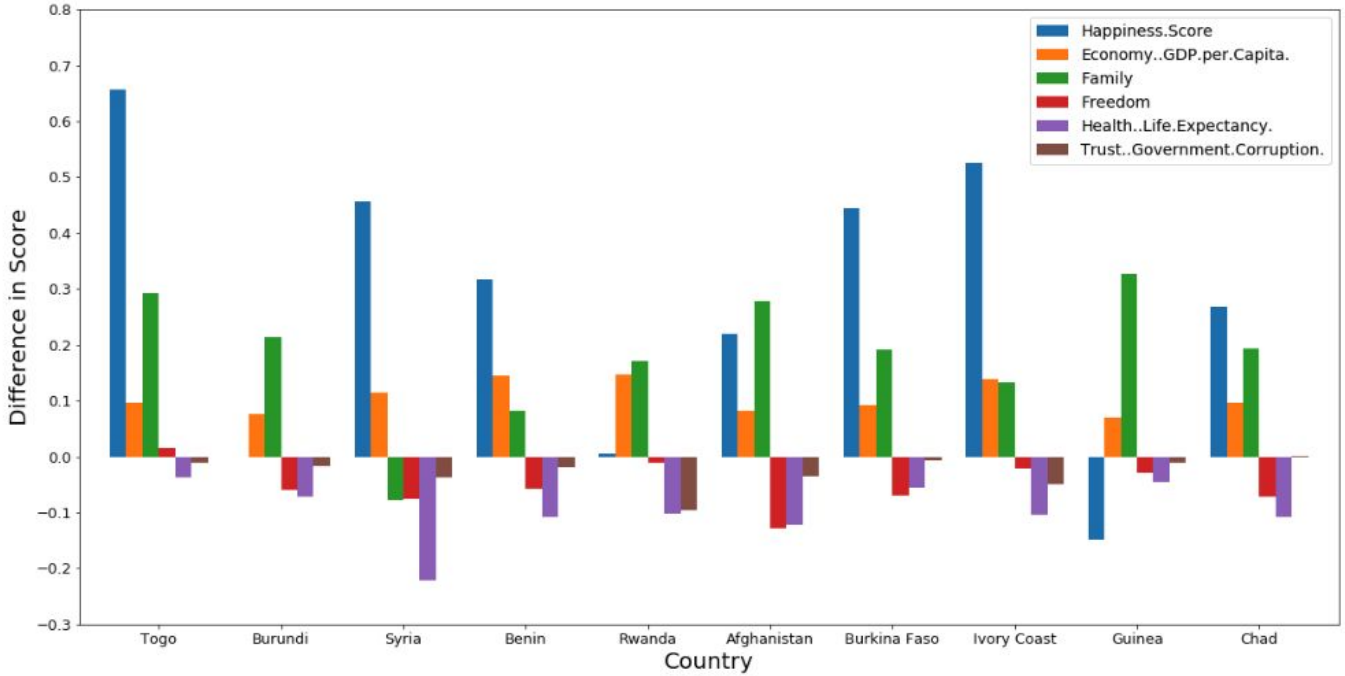
These are the 10 countries that have experienced the largest decline in happiness score.

Many of these are African countries, and a couple of Latin American countries.

Given the political turmoil in Venezuela over the past few years, the decline in happiness score is understandable, and is coupled with a decline in Freedom.

Many of these countries have also experienced a corresponding decline in Freedom and/or Life Expectancy.

What countries have **improved** the most?



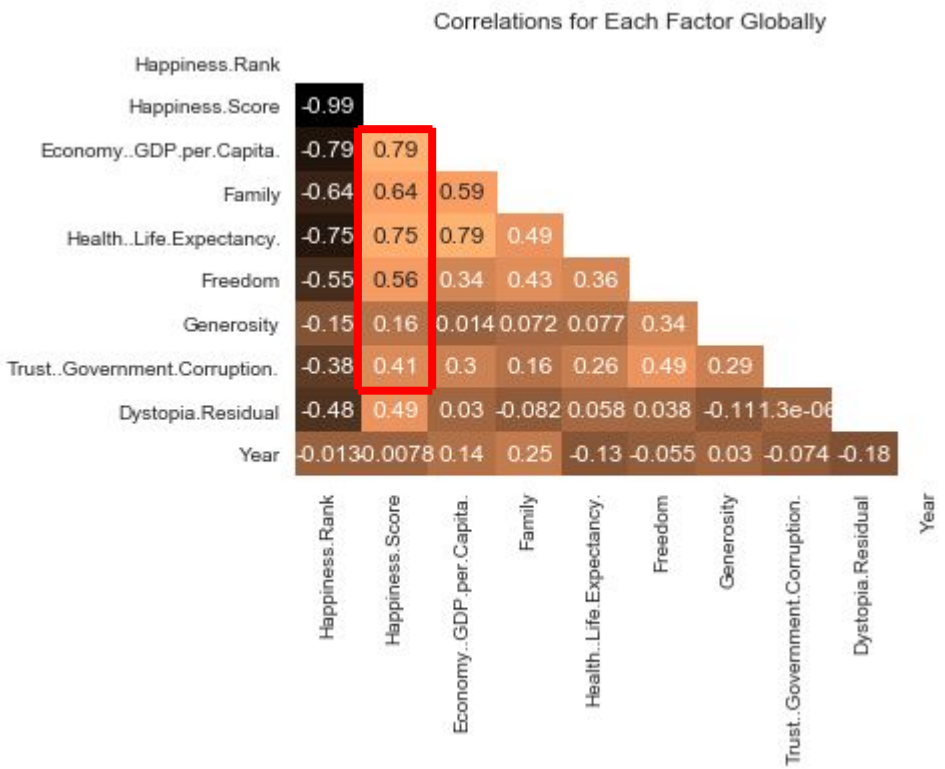
Key Takeaways:

Most improvements have been coupled with large improvements in family and some improvements in GDP per capita

Syria is the exception, where only an improvement in GDP per capita drove dramatic improvement in happiness score

It's clear that there are a number of different factors that influence overall happiness score. Let's dig into each of those factors a little bit more in the following slides.

What is the **relationship** between Happiness & various factors?



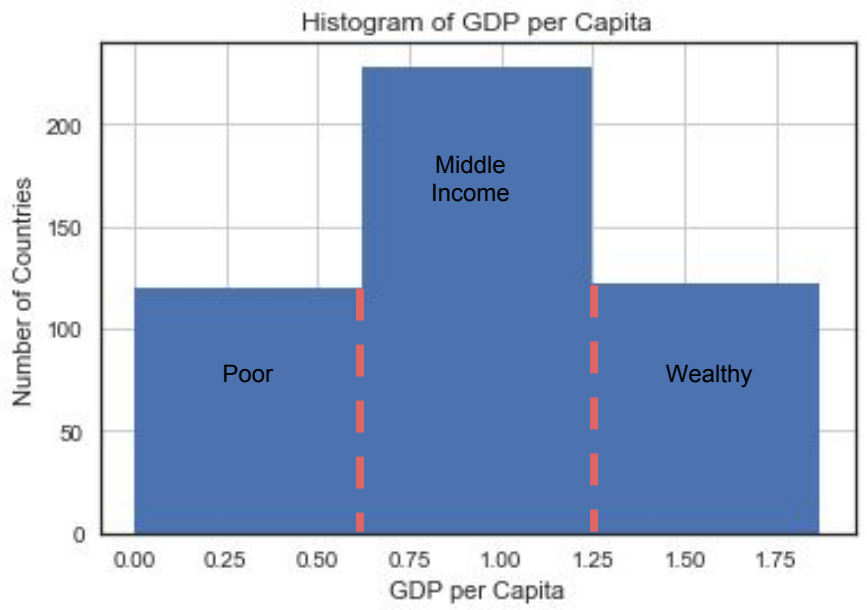
Key Takeaways:

GDP per Capita has the largest overall correlation with happiness score

Generosity seems to play the smallest role in driving happiness score for countries

Hypothesis: The role that each factor plays may vary for countries at different income levels.

How can we separate the countries into Income Levels?



We divided the data into three bins:
(1) Wealthy, (2) Middle Income, (3) Poor

Definitions (GDP per capita):

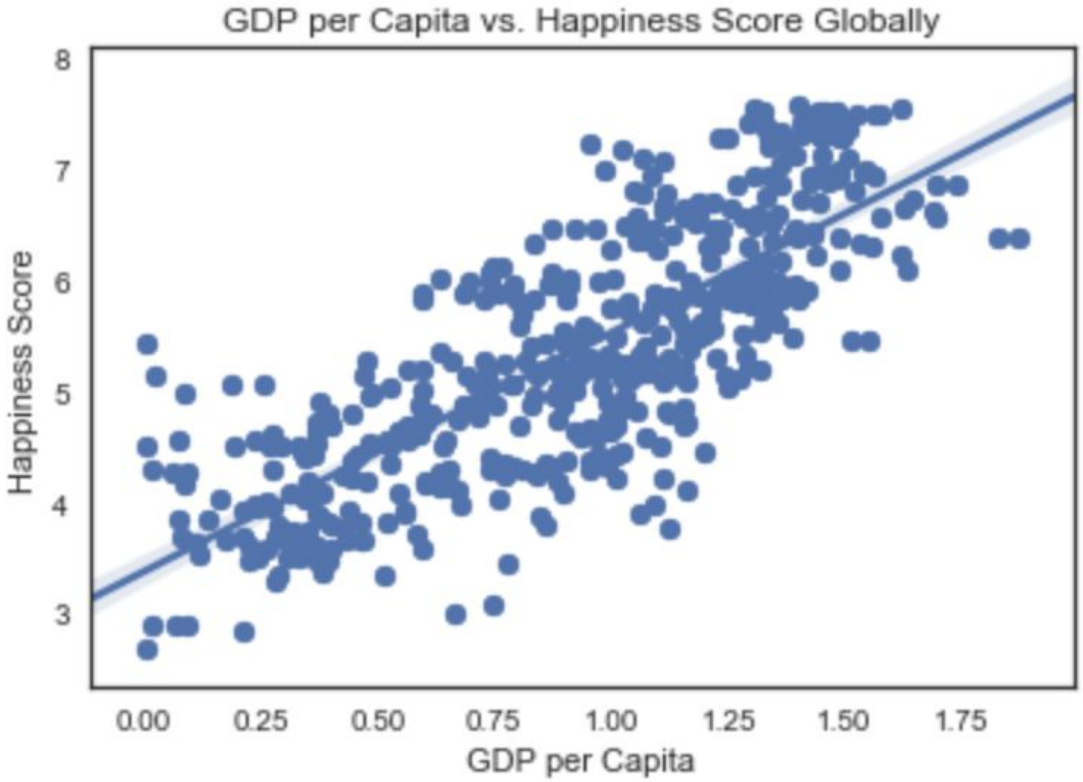
1. Wealthy >= 1.25
2. 1.24 >= Middle Income > 0.6
3. Poor =< 0.6

Count of Countries

	index	2015	2016	2017
0	Middle Income	82	79	72
1	Poor	47	35	34
2	Wealthy	29	43	49

In the next slides, we look into the role that each factor plays both globally and within each of the three Income Levels.

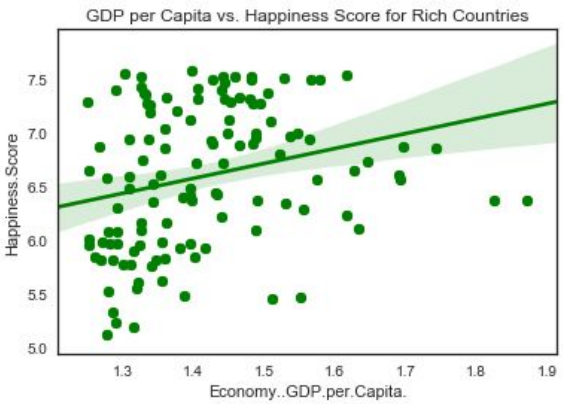
What role does **GDP per capita** play?



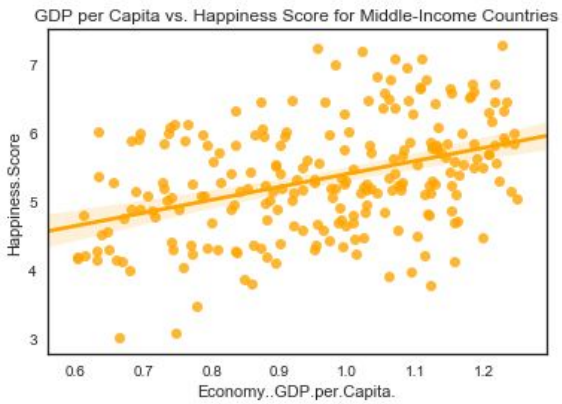
Correlation = 0.79

GDP per capita is among the highest correlated factors globally between 2015 and 2017

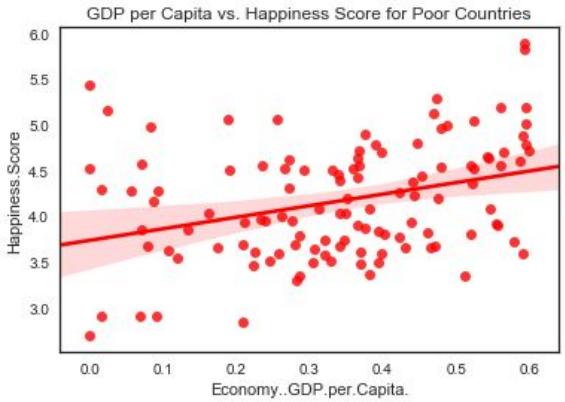
What role does **GDP per capita** play?



Correlation = 0.26



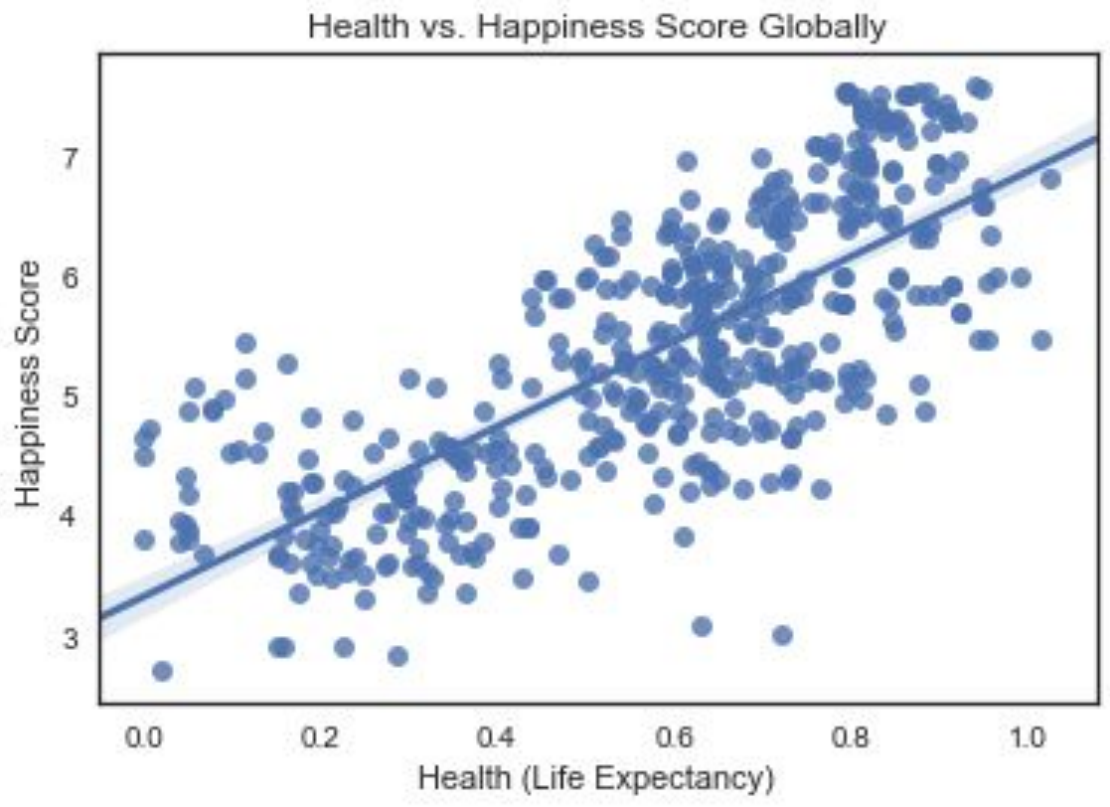
Correlation = 0.4



Correlation = 0.33

Key Takeaway:
GDP per Capita seems to have diminishing returns for richer countries, and play a larger role in influencing Happiness Score for Middle Income countries.

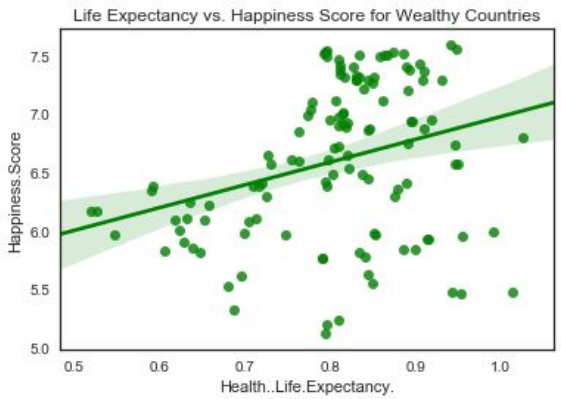
What role does **life expectancy** play?



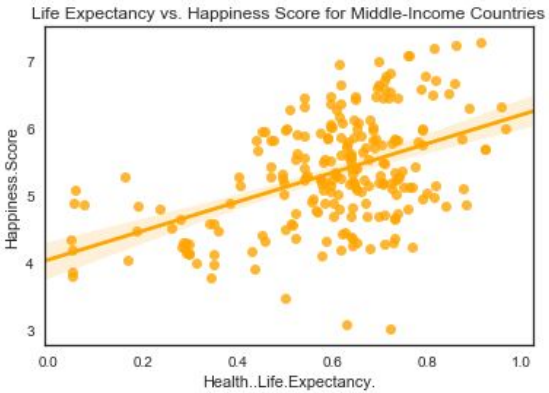
Correlation = 0.75

Health (life expectancy) is the second most correlated factor globally

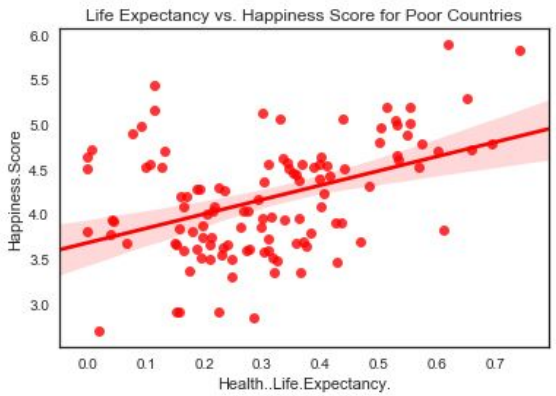
What role does **life expectancy** play?



Correlation = 0.3



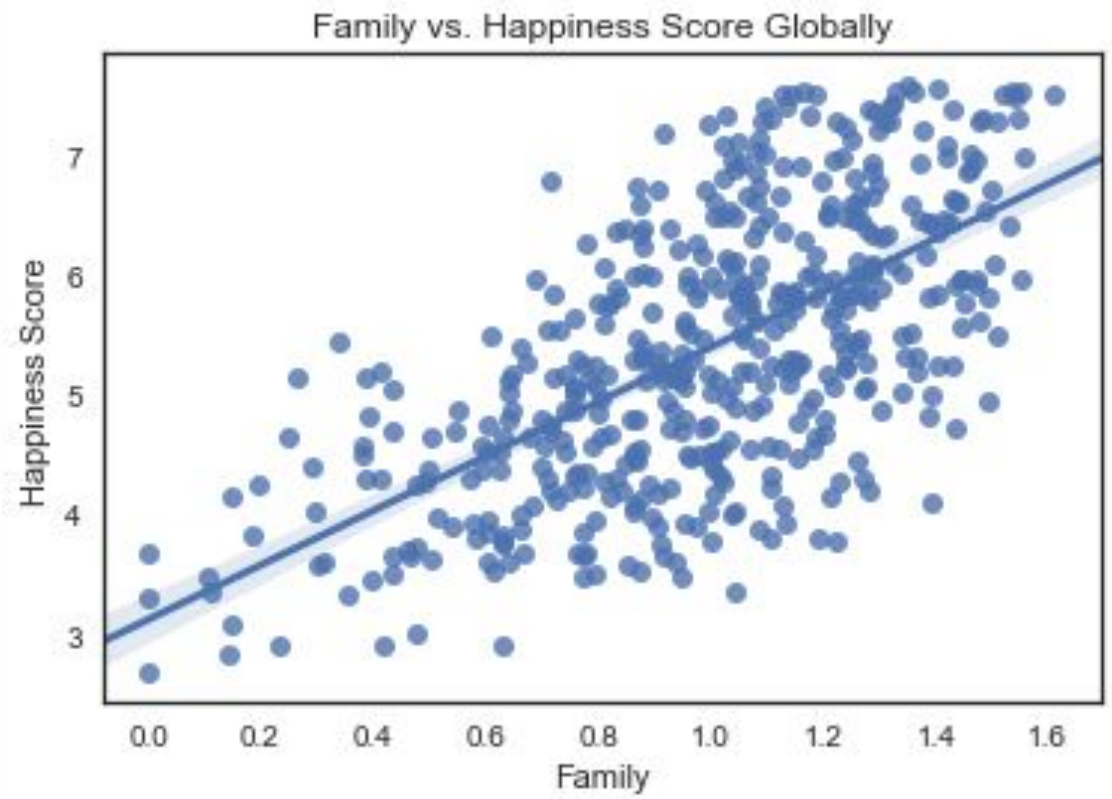
Correlation = 0.47



Correlation = 0.43

Key Takeaway:
Changes in life expectancy play much more of an important role on happiness score for Middle Income and Poor countries.

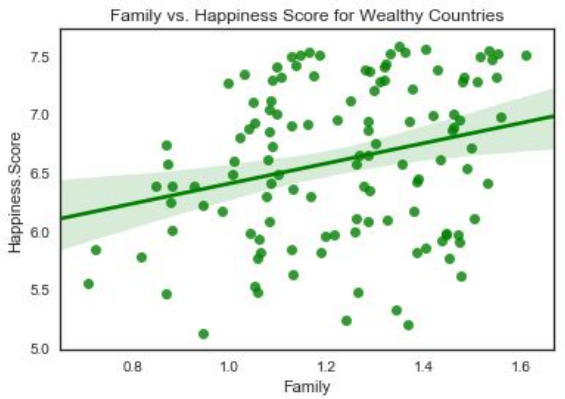
What role does **family** play?

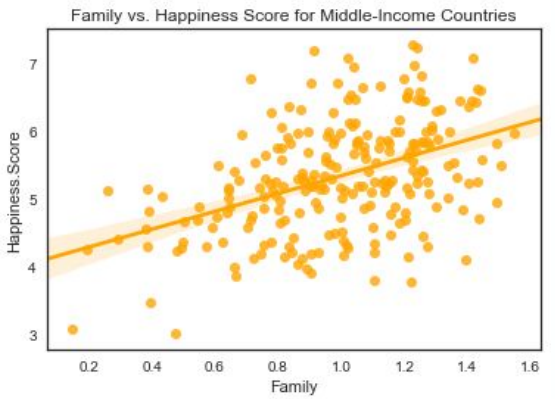


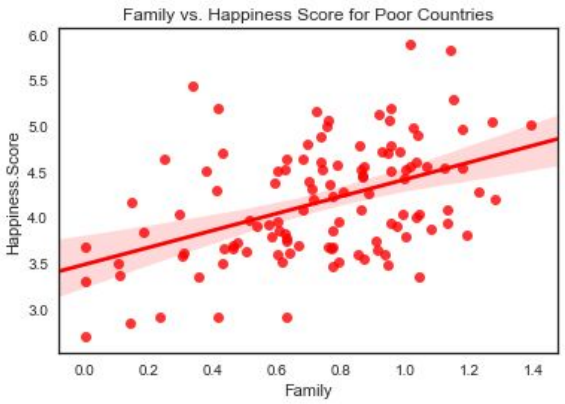
Correlation = 0.64

Family is the third most correlated factor globally

What role does **family** play?







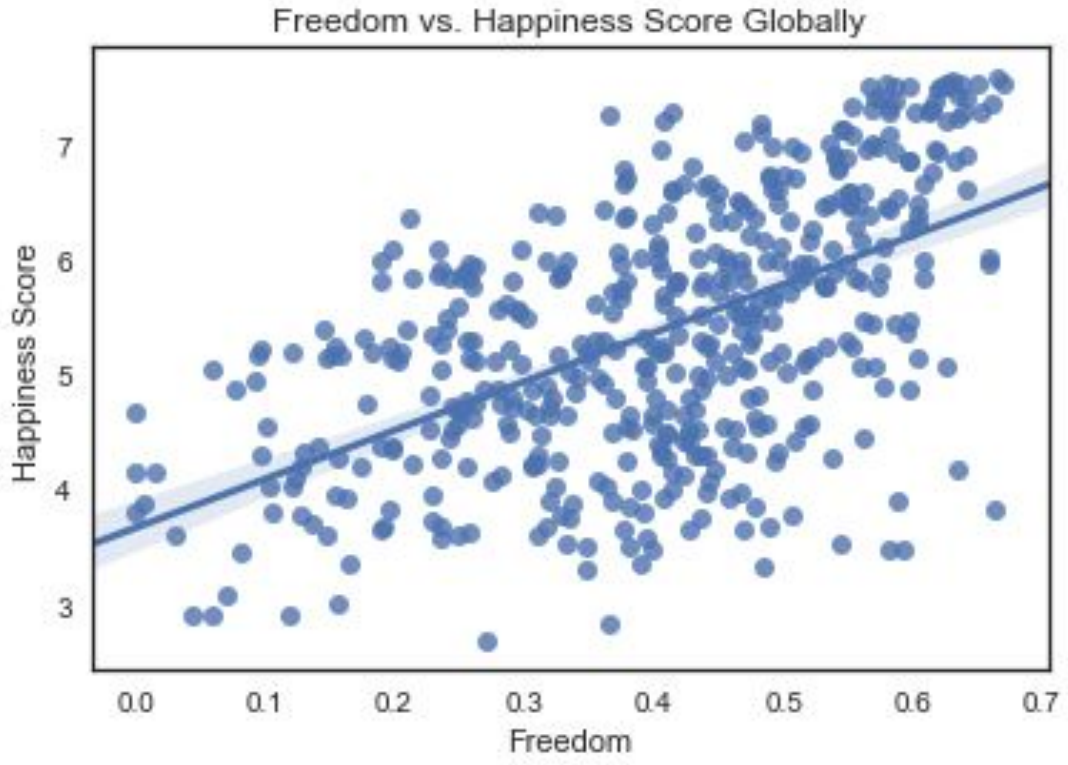
Correlation = 0.27

Correlation = 0.44

Correlation = 0.45

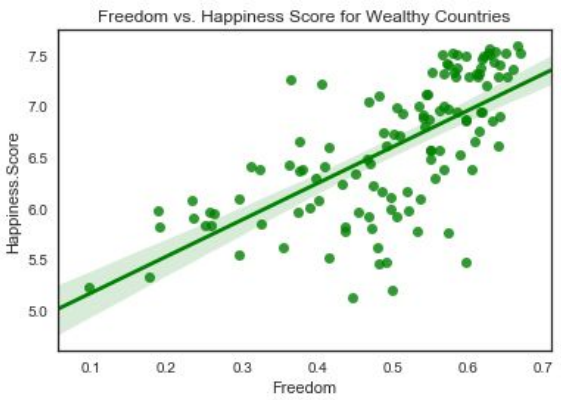
Key Takeaway:
The role of family seems to be much more important in Middle Income and Poor countries where family is much more necessary for support (i.e. potentially less economic independence)

What role does **freedom** play?

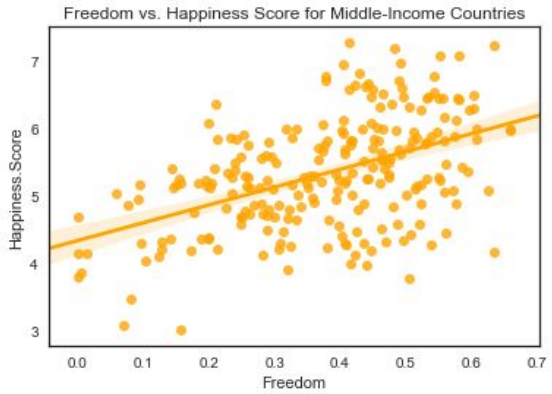


Correlation = 0.56

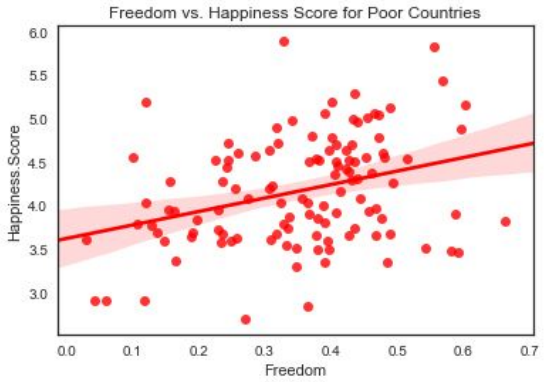
What role does **freedom** play?



Correlation = 0.67



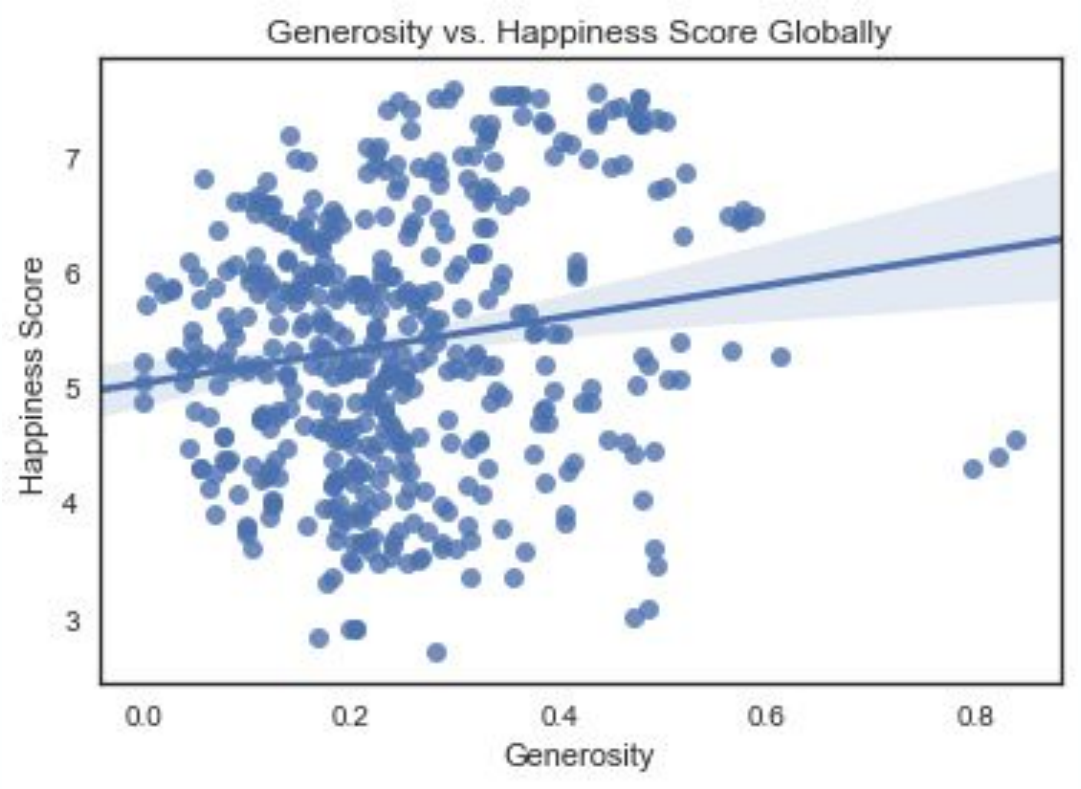
Correlation = 0.48



Correlation = 0.33

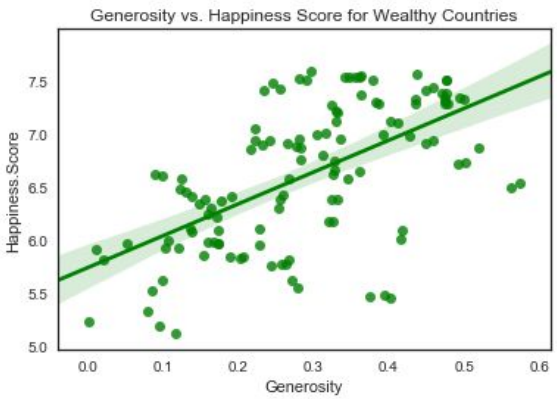
Key Takeaway:
Interestingly enough, freedom plays a very important role for Wealthy countries and Middle Income countries, and less so for Poor countries.

What role does **generosity** play?

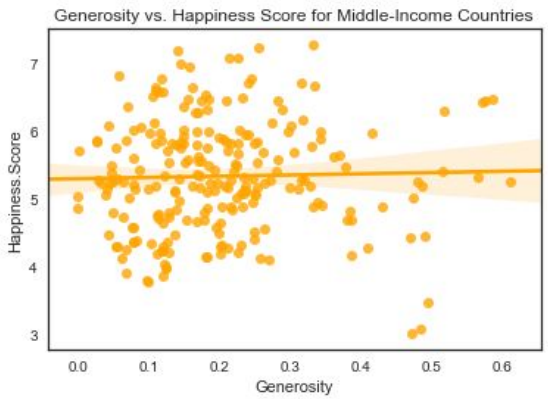


Correlation = 0.16

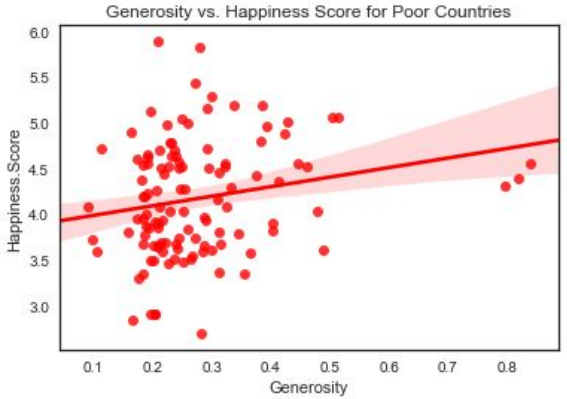
What role does **generosity** play?



Correlation = 0.6



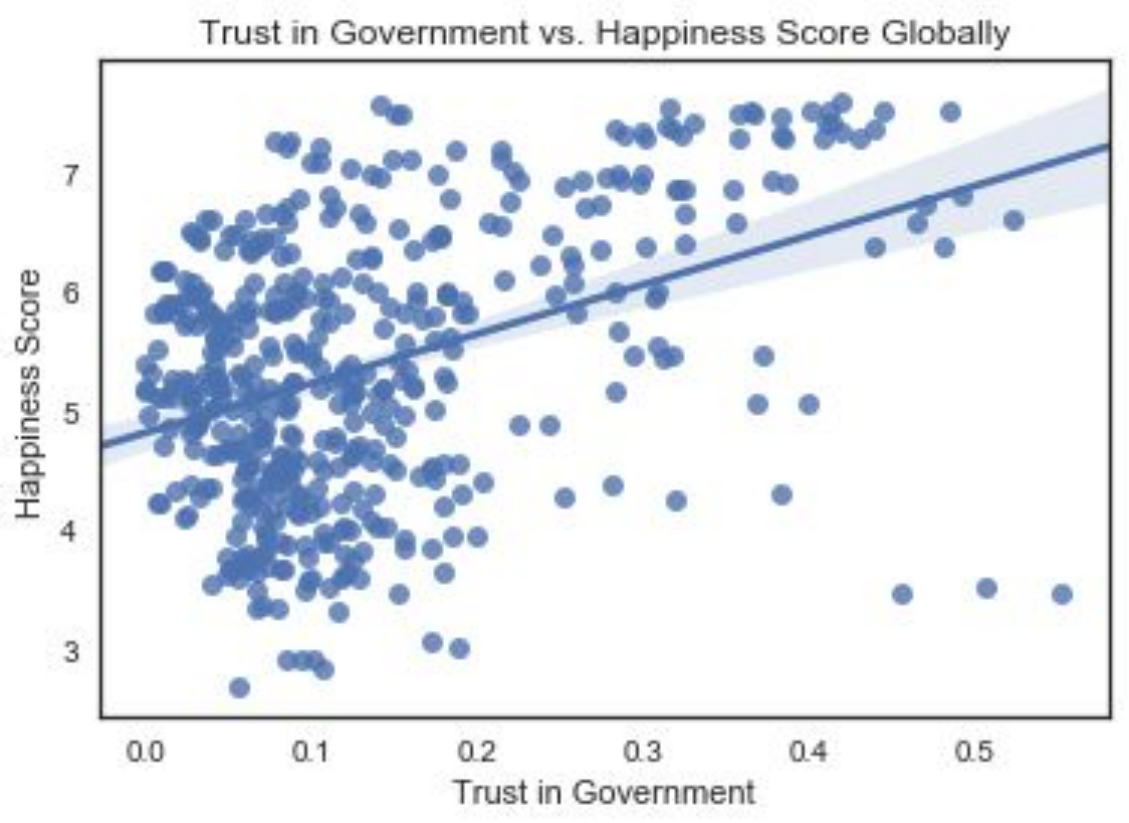
Correlation = 0.027



Correlation = 0.21

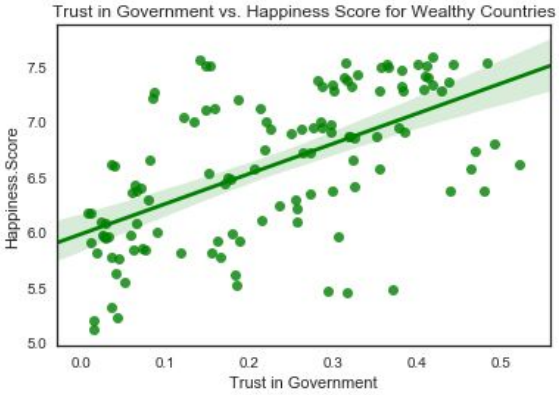
Key Takeaway:
Generosity plays a much more important role for Wealthy countries and a negligible role for Middle Income and Poor countries. It seems like a luxury used to find fulfillment. This finding supports supports Maslow's Hierarchy of Needs.

What role does **trust in government** play?

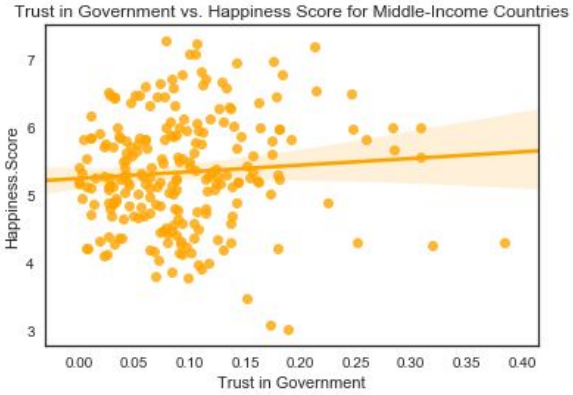


Correlation = 0.41

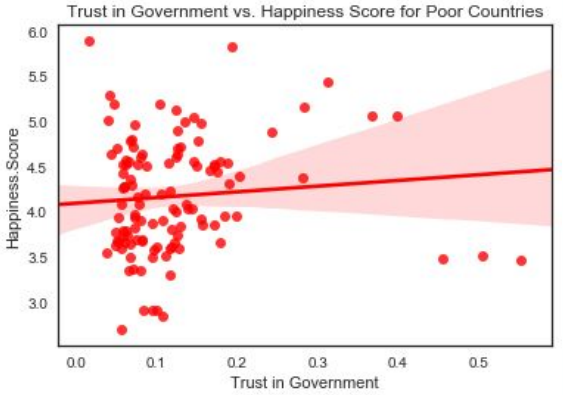
What role does **trust in government** play?



Correlation = 0.59



Correlation = 0.078



Correlation = 0.091

Key Takeaway:
Trust in government plays a negligible role in Middle Income and Poor countries relative to Wealthy countries where it appears important for happiness score.

Summary of **correlations** by Income Levels

	index	Wealthy Country Correlations	Middle Income Country Correlations	Poor Country Correlations	Global Correlations
0	Happiness.Rank	-0.985051	-0.993522	-0.969276	-0.993268
1	Happiness.Score	1.000000	1.000000	1.000000	1.000000
2	Economy..GDP.per.Capita.	0.262909	0.404775	0.333741	0.785450
3	Family	0.267657	0.442185	0.452901	0.636532
4	Health..Life.Expectancy.	0.299946	0.474024	0.426604	0.748040
5	Freedom	0.671377	0.479378	0.327824	0.560353
6	Generosity	0.595333	0.027164	0.206392	0.163562
7	Trust..Government.Corruption.	0.585222	0.077693	0.090567	0.406340
8	Dystopia.Residual	0.631690	0.749523	0.529603	0.489747
9	Year	-0.233873	-0.121457	-0.173777	-0.007761

Key Takeaways

Observations

1. **Top 10:** heavy European representation in the top 10 happiest countries
2. **Top 10 most improved countries:** driven primarily by increase in family score and GDP per capita, many are middle income
3. **Top 10 most declined countries:** driven primarily by reduced freedom and health (life expectancy)
4. **GDP per capita:** highest global correlation, but particular influential for the happiness score of middle income and poor countries - diminishing returns in wealthy countries
5. **Health (life expectancy):** second highest global correlation, but particularly influential for happiness in middle income and poor
6. **Family:** plays much more of a role for middle income and poor countries, potentially because it's a necessary source of economic support
7. **Freedom**: much more important in wealthy and middle income countries than poor countries
8. **Generosity:** very important in rich countries, almost negligible for middle income and poor countries
9. **Trust in Government:** important in rich countries, negligible for middle income and poor countries

Further questions / analysis
(if we had more time)

1. Dig deeper into how the Happiness Score is calculated because it is not totally clear from the explanation in the report.
2. Compare these results against the World Bank's Human Development Index (HDI) data to see if there is any relationship between the happiness scores and HDI.
3. Use other datasets on government services offered in these countries. The hypothesis is that trust in government may be negligible in middle income and poor countries in part because not as much government services are available.
4. Run more rigorous statistical analysis on the data that can lead to more conclusive findings.
5. Find ways to better define Wealthy, Middle Income, & Poor countries. There may be a standardized definition of different Income levels.

References

1. https://www.gallup.com/178667/gallup-world-poll-work.aspx
2. http://worldhappiness.report/faq/