The Happiness Formula

Beejal Parekh

Feb 2025

Objective

Identify the most important factor(s) influencing happiness.

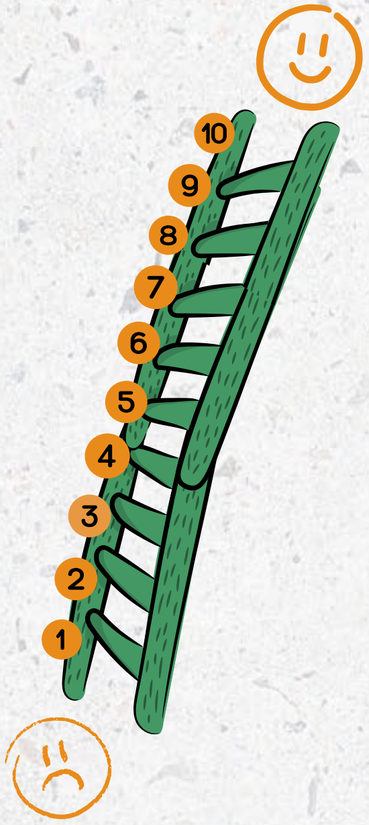
Factors considered

GDP per capita | Social support | Life expectancy | Freedom | Generosity | Corruption perception

Data Source: World Happiness Report (2018 & 2019)

Understanding the data

Survey methodology



- Happiness scores based on Gallup World Poll
- Cantil ladder: nationally representative partipants rate their life between 0 (worst possible life) and 10 (best possible life)
- Sample size approx. 300'000 people in 130 countries over a 3 year rolling average (100'000 a year | 1-3000 per country)
- Variables illustrate correlation not causation. Some variables (e.g. unemployment and inequality) are not included due to insufficient comparable data
- A low perception of corruption is a high value

Cleaning the data

Overview

- Two years: 2018 and 2019
- 160 unique countries (152 in both years, 4 in 2018 only and 4 in 2019 only)
- Rank is by year so there are duplicate values

Deleted columns:

- Healthy life expectancy (97.76% null values)

Null and ‘0’ values

- One null value (UAE perception of corruption) , left in the dataset as other columns provide useful info
- Some 0 in each variable, left these in the dataset for now as unclear if these are null or value is 0

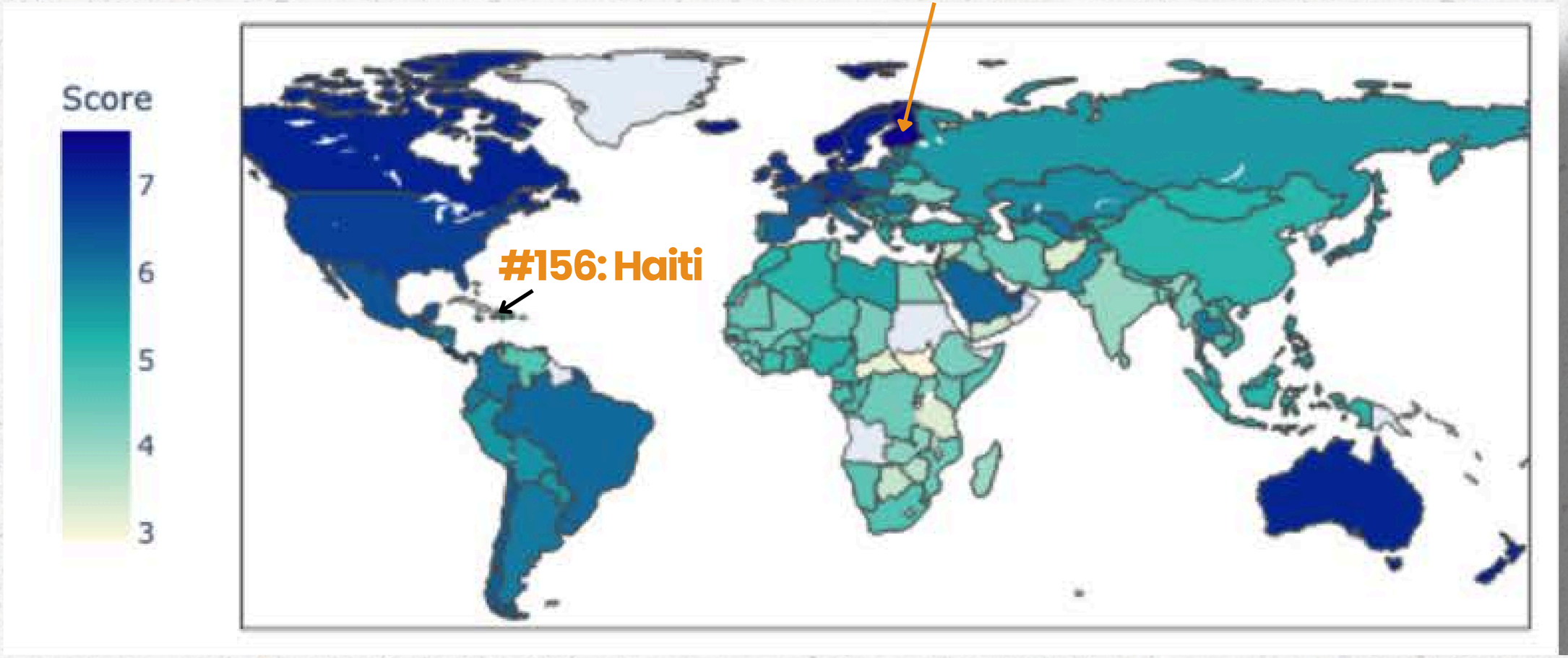
Separated datasets

- Created a separate dataset for each year

Headlines

2019

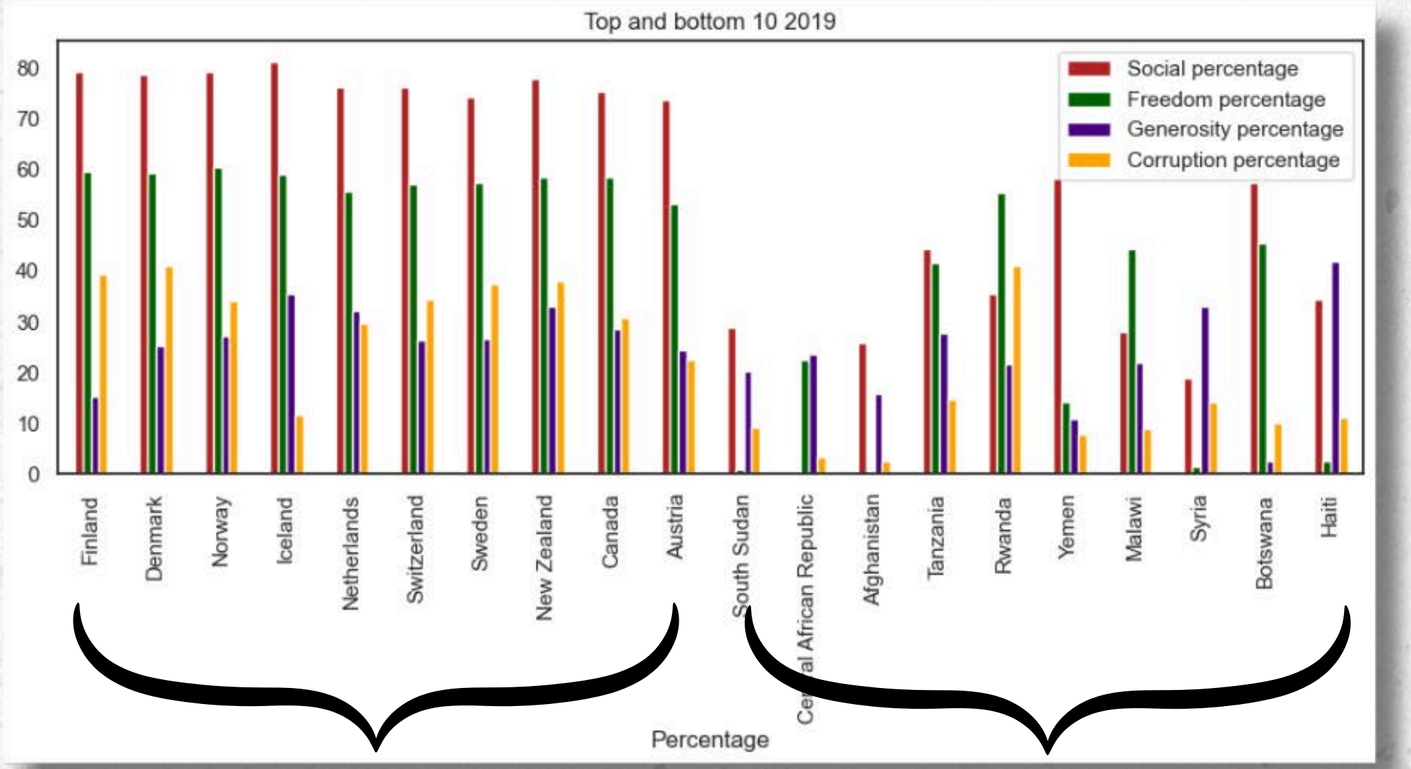
World map of Happiness Score

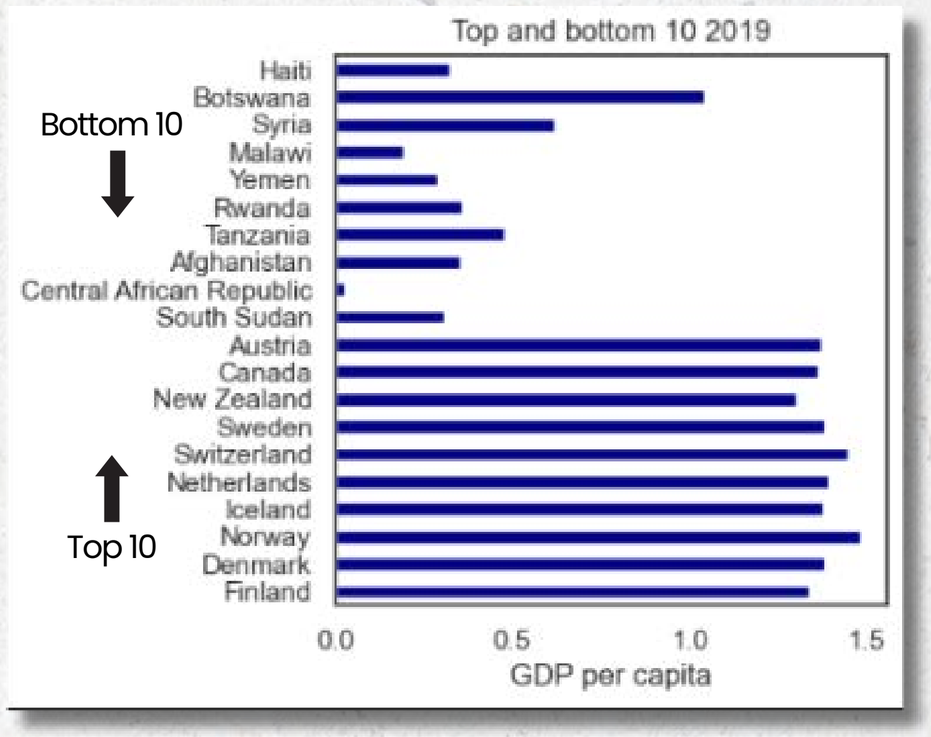


Headlines

2019

Top & Bottom Ranked Countries

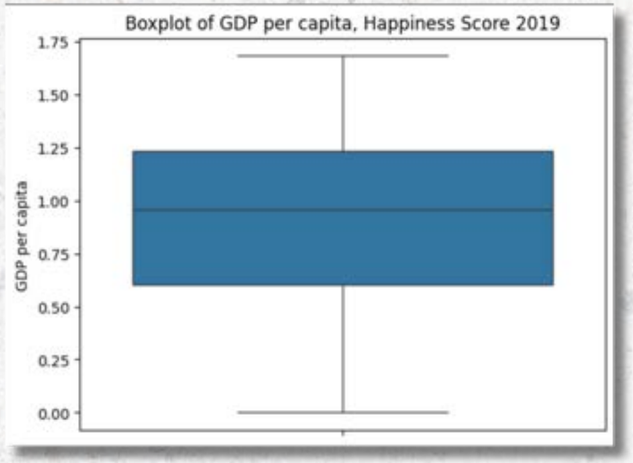


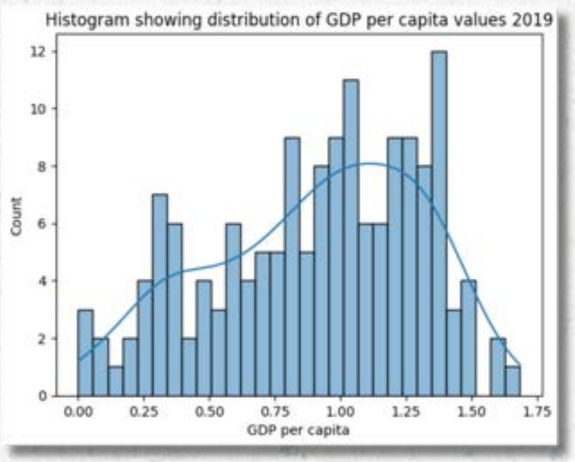


Univariate analysis 2019

Boxplots and histplots showing distribution of data for numerical variables

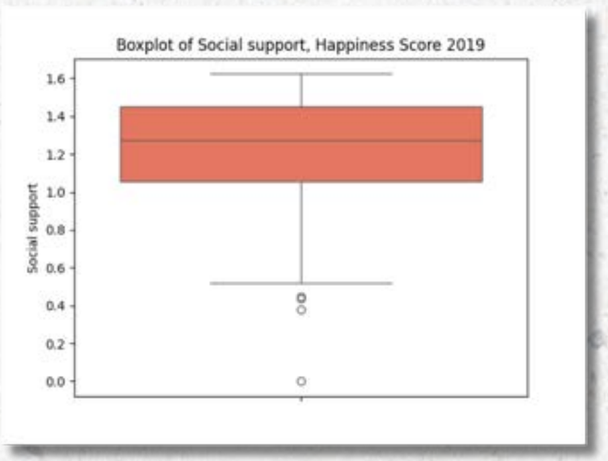
GDP per capita

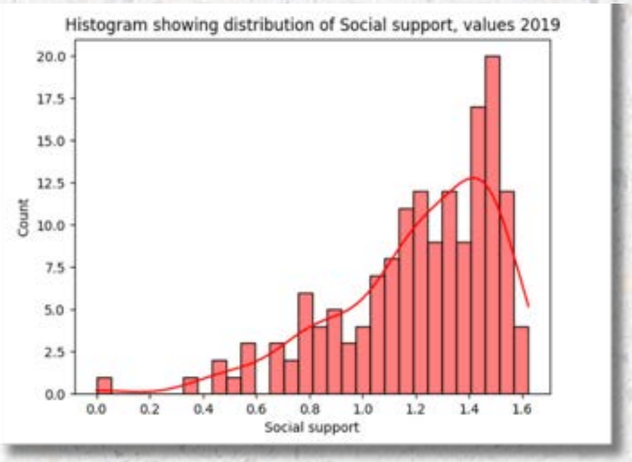




One clear main peak, with a slight secondary peak slightly higher than the surrounding values. Suggesting there **might be two subsets of data.**

Social support



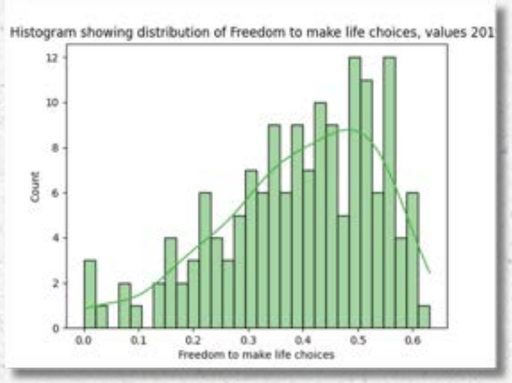


Right/positively-skewed distribution, most of the values on the higher end.

Univariate analysis

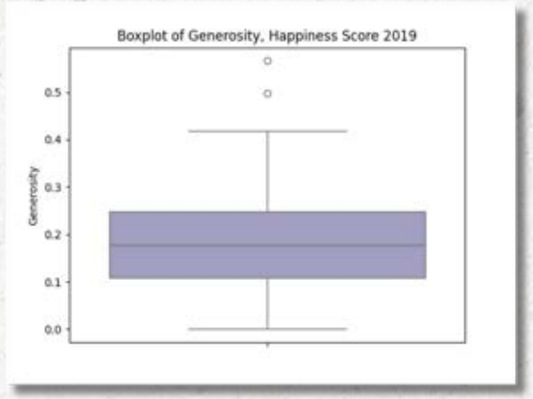
2019

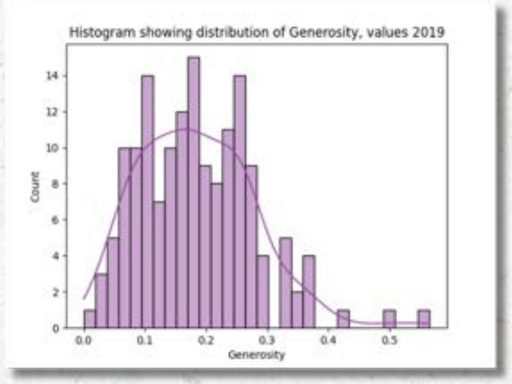
Freedom to make life choices



Right / positively skewed, most of the values on the higher end.

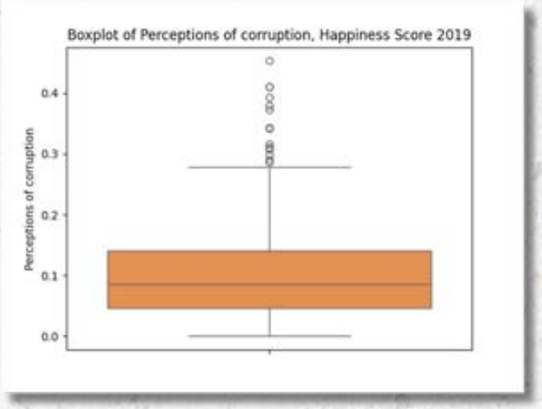
Generosity

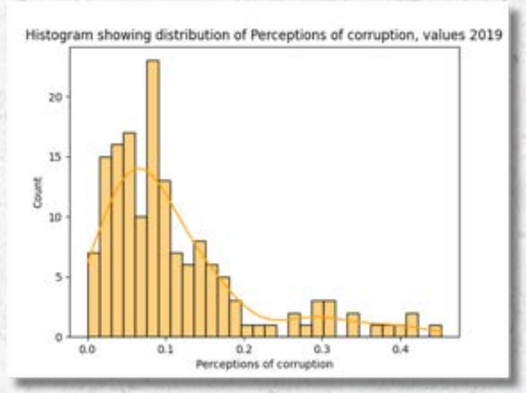




Left/negatively-skewed distribution, most of the values are on the lower end.

Perception of corruption



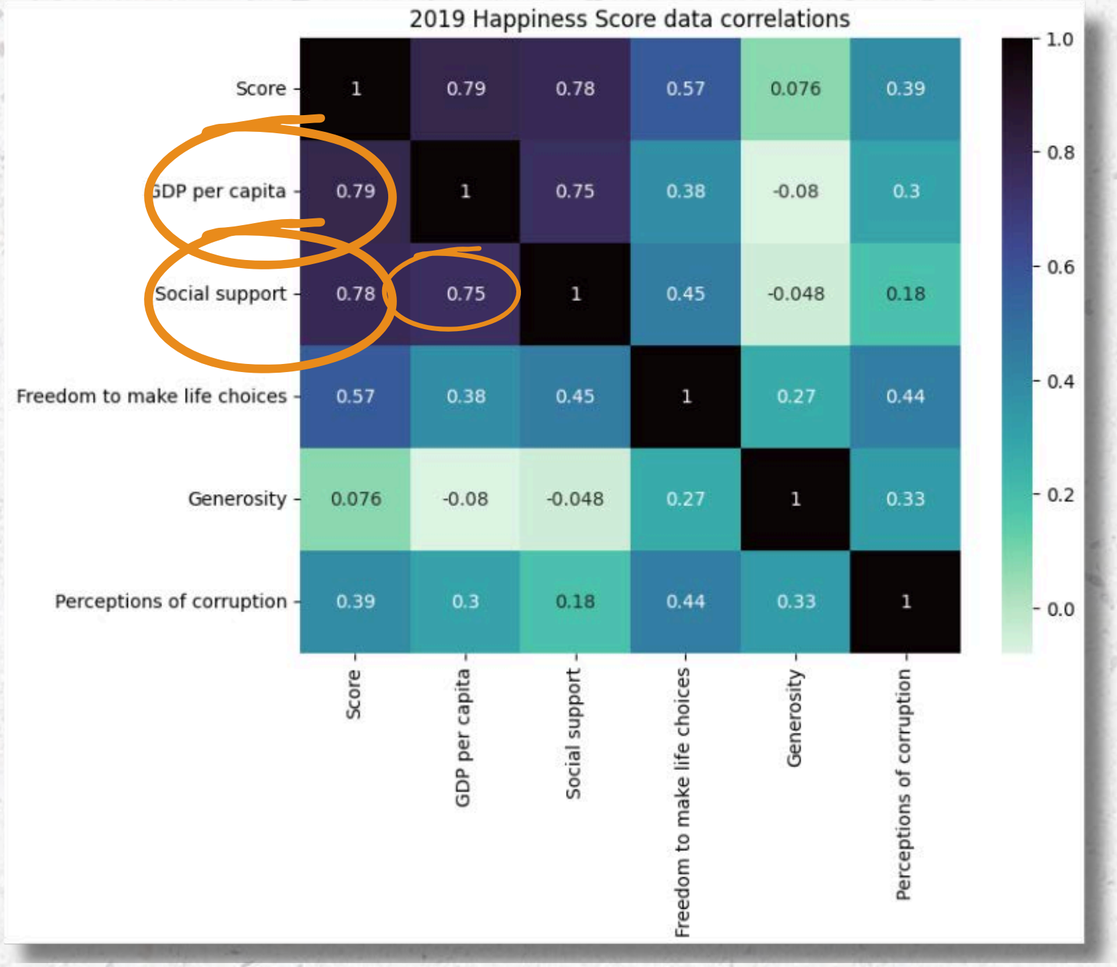


Left/negatively-skewed distribution, most of the values are on the lower end. **Lots of outliers (most high outliers are amongst the highest ranked countries)**

What matters most?

2019

Heatmaps to show which factors have the strongest correlation with happiness.



Strong correlation

Moderate correlation

Weak correlation

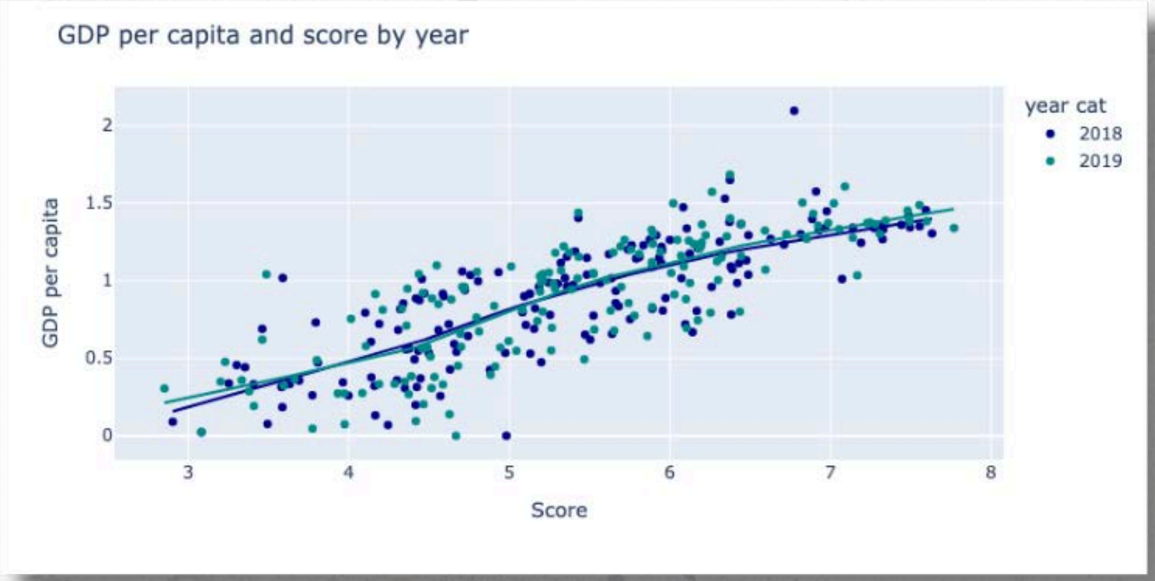
NOTE: We choose to look at variables against score rather than rank to preserve variability, capture subtle distinctions and allow for more sophisticated statistical analysis.

Bivariate analysis

Key factors driving happiness: scatterplots showing correlation between variables and happiness score

GDP per capita

Does wealth = happiness?



Strong liner correlation.

Social Support

The power of relationships & strong communities.

Strong liner correlation.

Bivariate analysis

Freedom to Make Life Choices

The role of personal autonomy

Generosity

Do giving and kindness lead to happiness?

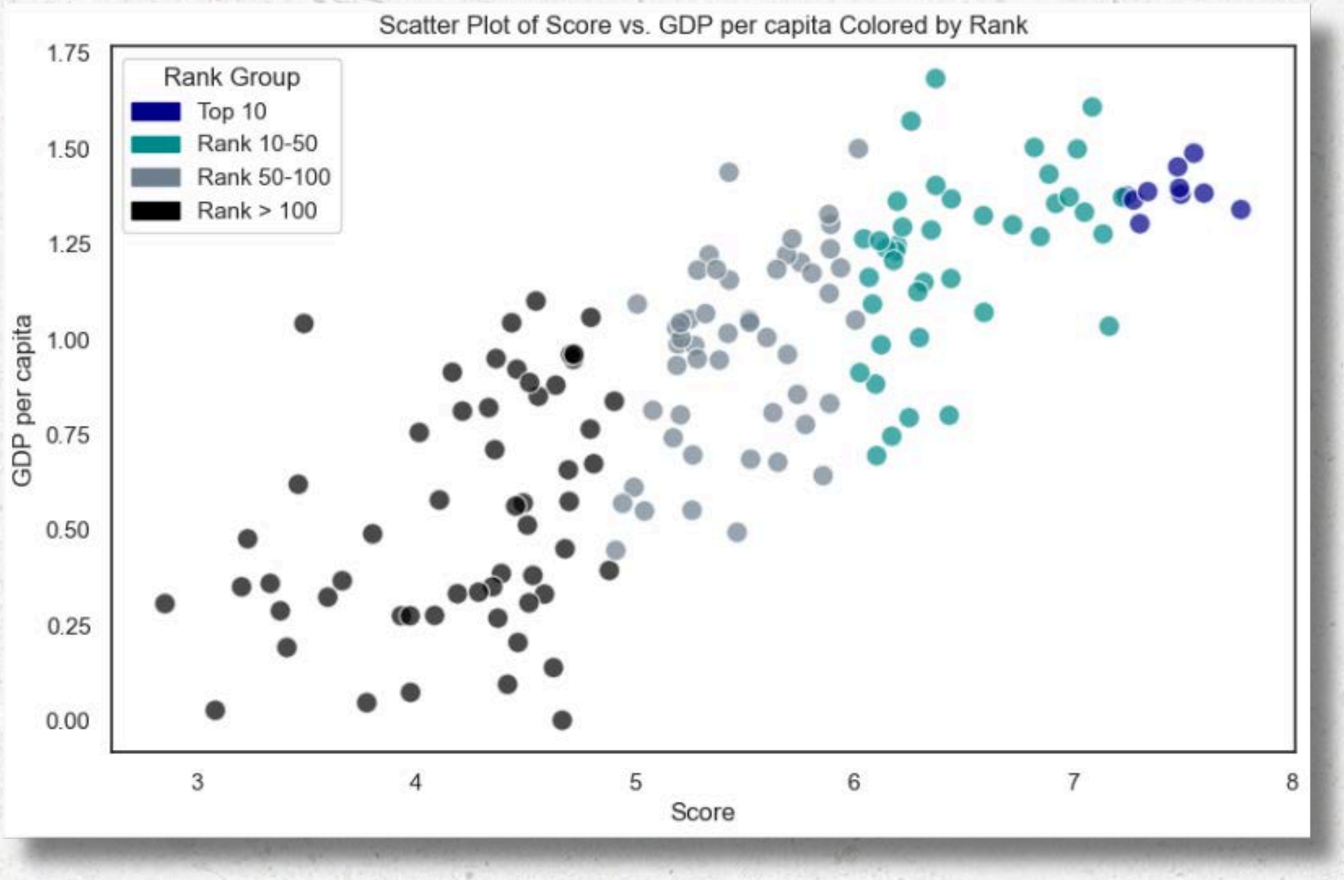
Corruption Perception

Trust in government & institutions

Weak overall correlation, but it seems there may be a non-linear correlation where these factors, especially generosity and perceptions of corruption make **more difference amongst higher scoring countries.**

What matters most?

GDP is the most strongly correlated variable with score and rank



Hypothesis

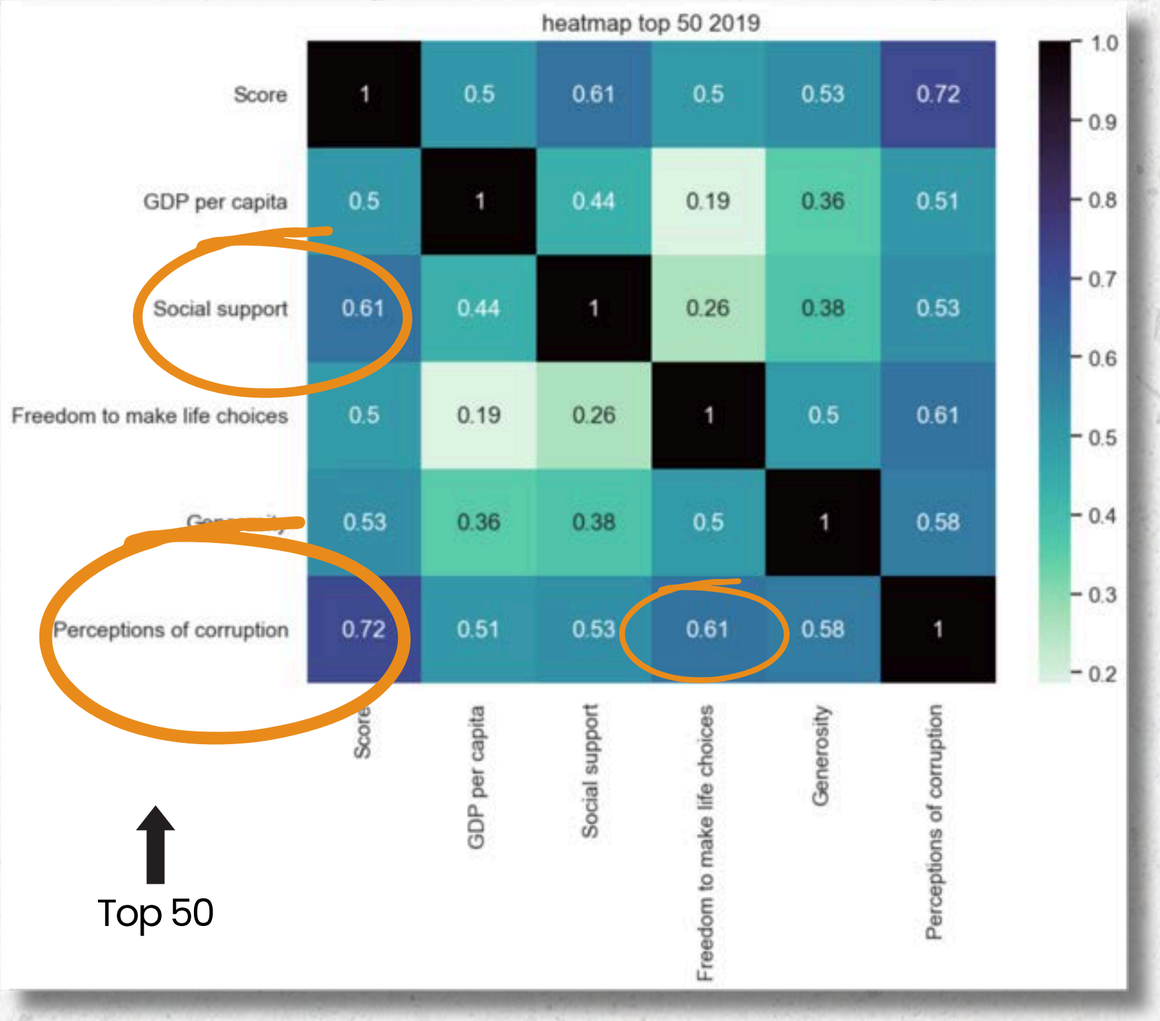
GDP matters but only up to a point, amongst higher ranked countries, does this show something like the Maslow's hierarchy?

How to get from 50 to 1

Top 50

2019

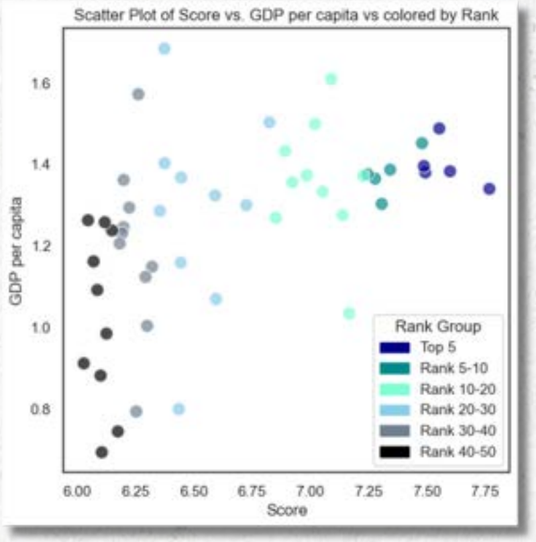
Correlations between the top 50 ranked countries only....



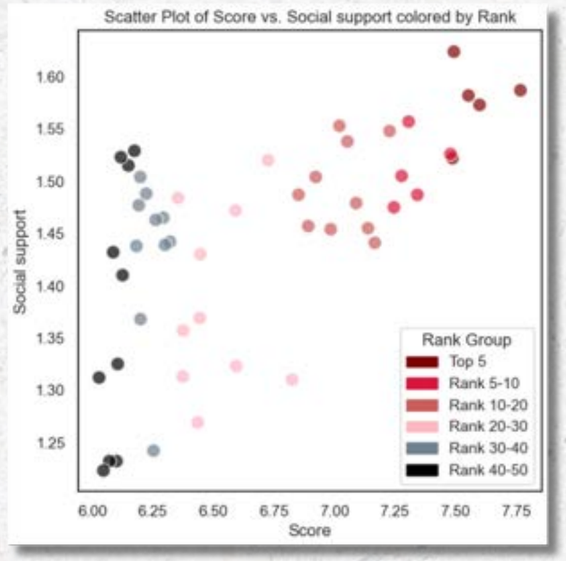
How to get from 50 to 1

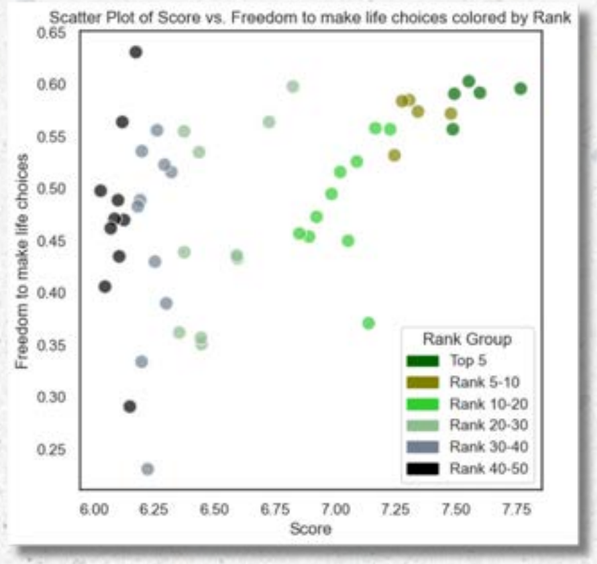
Top 50

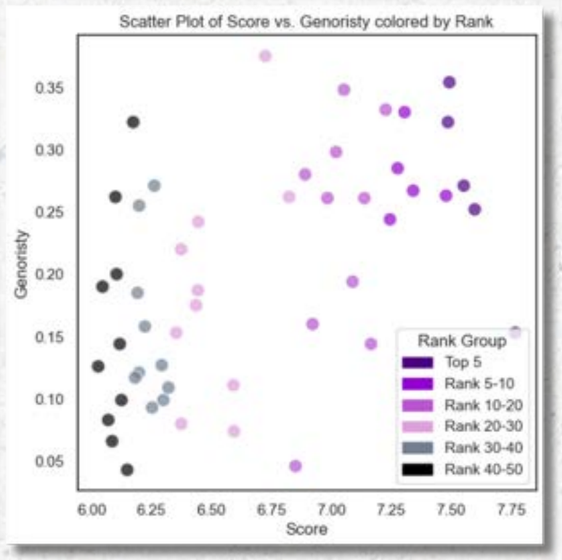
2019

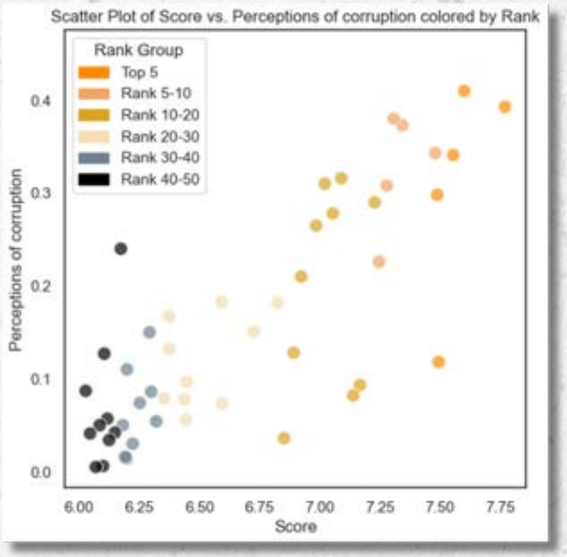


Look at the **horizontal distribution** of the different colour groups to see how significant that variable is.

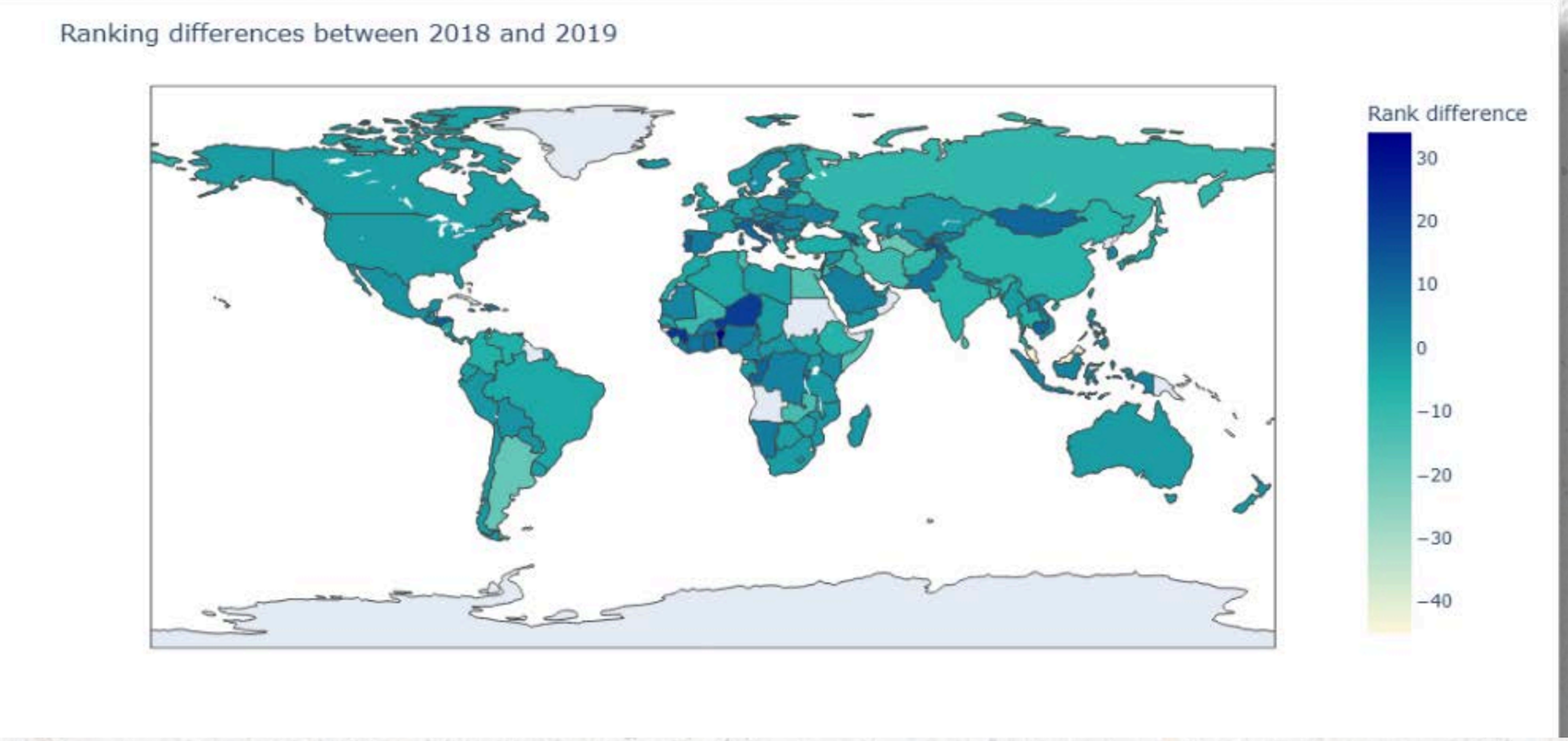








Change of over time

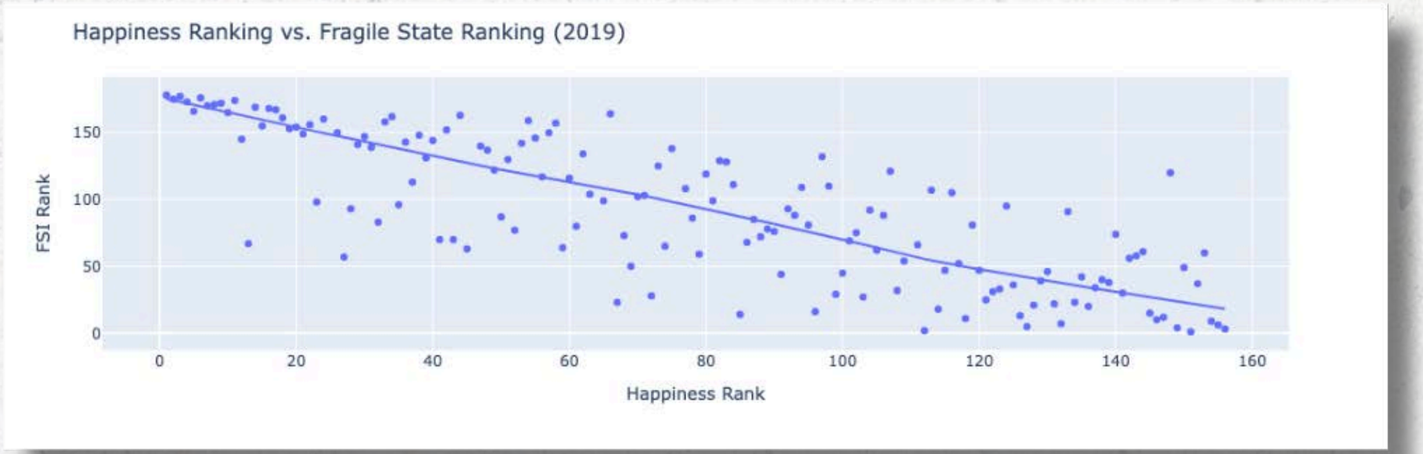


Malasya and Turkmenistan went significantly down the ranking.

Kosovo, Niger, Guinea and Benin went significantly up the ranking.

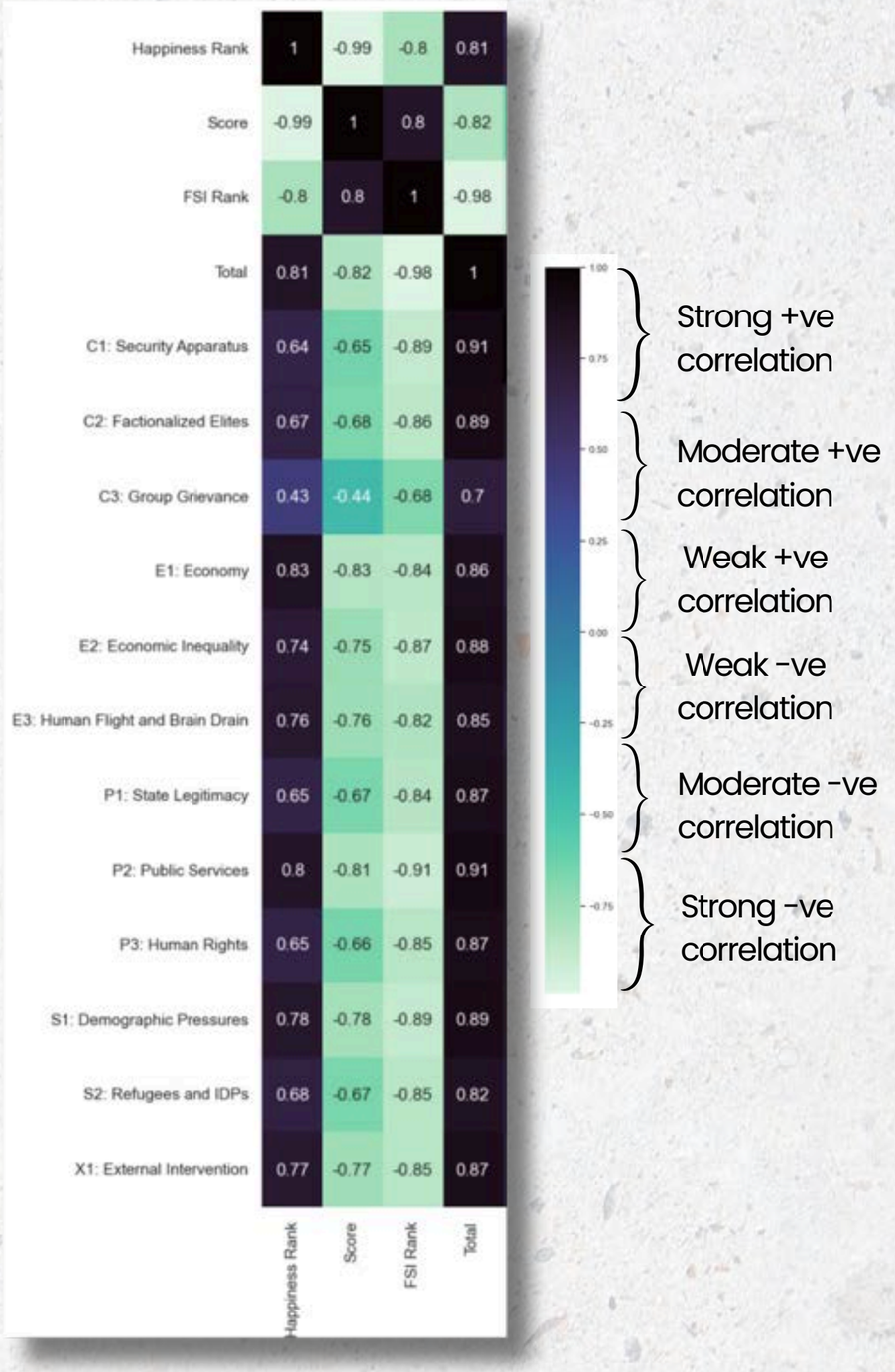
Other factors

Comparison to rank from the fragile states index



High liner correlation (negative correlation) between these two rankings, as we might expect:

- A high fragile state rank is strongly correlated to a low happiness rank
- A low fragile state rank (i.e. a stable country) is correlated to a high happiness rank



Data Sources:

World Happiness Report (2019) & Fragile States Index (2019)

Next steps

What else to consider

- Other data sets:
  - unemployment and inequality as these are identified as important factors
  - healthy life expectancy as this column was omitted due to insufficient data
  - other country statistics e.g. population size
  - more world happiness score years to see changes over time

- Group variables in different ways to look for patterns e.g. by continent, population size etc.

Conculsions

What else to consider

- GDP is important overall but other factors can be more significant when looking at smaller subsets in the ranking or score.

- The stability of a country seems to be highly correlated with happiness rank, and may contribute not only to the rank but also explain significant differences between years i.e. a war or environmental disaster seems to have a big localised impact on ranking.

- This may also explain why there is more change in rank and score between years amongst lower ranking countries (more likely to have volatile situation), than higher ranking countries (more likely to remain stable across years)