A Study of the Science of Happiness

A look at relationships between happiness and other national indicators
W200 Summer 2020 Final Project
Whit Blodgett

Overview

What does it take to be happy? A loving family? Money? A positive outlook? This is a question all people must face at one point in their lives and something this study aims to answer. Utilizing the latest data sets available, we draw observations on data focused from 2015-2020, a time in which much has shifted with globalization, growing inequality, and improving global standards of living among other things. In a world where so many are scrambling to forge a “happy” life, this study aims to shed some light on what factors move the needle and which are unrelated. While these are mainly observational and not inferential, we hope that this study will inspire a seasoned data scientist to investigate the keys to a happy life.

Hypotheses

We connected data from world events aggregated on a yearly level per country and correlated it with yearly country based happiness reports, and GDP measures. Our goal was to explore how world events affect different types of “prosperity” for countries. In our initial exploration and discussion of these data, we formulated the following hypotheses around happiness:

1. Productivity: There is a positive relationship between national productivity (measured using GDP per capita) and national happiness.

2. Personal Variables of Happiness : There is a positive relationship between average national family size, life expectancy, trust, freedom and happiness.

3. Bad News: There is a negative relationship between bad news in a country and happiness.

We then divided these three questions among the team and iteratively performed the exploratory analysis detailed below. Throughout this process we conferred together as our work revealed unexpected limitations or insights, ensuring those were considered and reflected across the questions.

Data

In contemplating these questions, and in search of data that might start to get us forming hypotheses, we happened upon 3 datasets that appeared of particular interest to the topic at hand. In this section we describe these datasets, observations around their provenance, our import procedures, as well as any change or exclusions we made to the core datasets prior to analysis.

World Happiness Report

This dataset[1], sourced from Kaggle, contains happiness scores and associated factors for 2016 - 2020. The data were collected by the Sustainable Development Solutions Network and the Gallup World Poll. The primary measure of interest in this report, which we use as a proxy for national happiness, is the happiness score. The score itself consists of qualitative surveys conducted across 153 countries.

We obtained this dataset programmatically through the Kaggle API. While reported features changed over time, the core features of interest in our analyses required little more than column name re-mapping (ex.: happiness score, GDP per capita, etc.) and merging of separate files. No other material changes

[1] World Happiness Report up to 2020

were made to the raw data prior to those discussed in the hypothesis sections below. Initial exploration of the happiness score reveals an upward trend of the average across 153 nations over the 5 year period - which initially seems noteworthy (see **figure 1**):

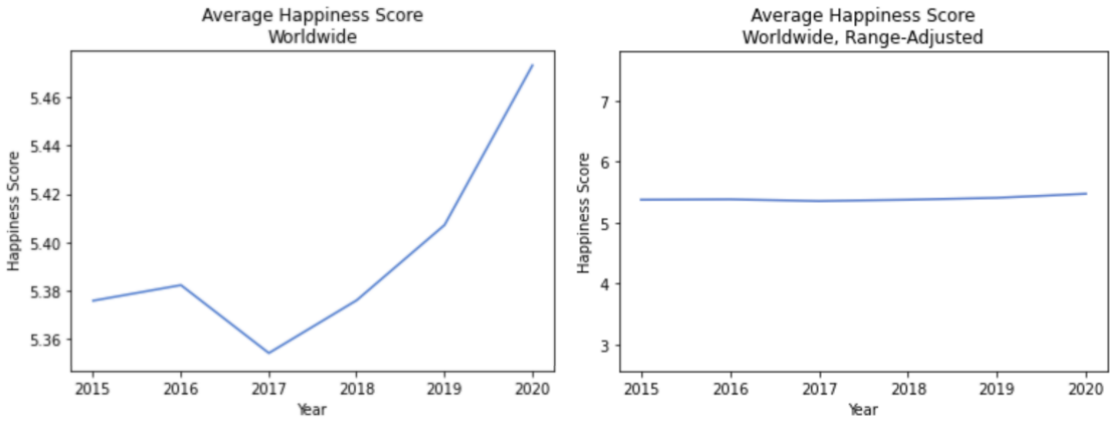

Figure 1: Trends in un-processed source data - happiness.

However, re-framing the chart with the minimum and maximum happiness score in the dataset as lower and upper bounds for the x-axis reveals a much more modest picture, with said average moving little over time.

World Events

Sourced from the GDELT Project, this dataset[2] contains metadata and computed signals from worldwide broadcast, print, and online news; identifying actors, locations, event types, the tone of events, etc. This dataset is freely available on the Google Cloud Platform's Marketplace for Datasets (hosted on and available through Google BigQuery). While the full events table is more than 211 GB, we aggregated the data for our use and restricted the columns to those of interest. In order to encourage re-use of our work and avoid dealing with API service keys, we performed this processing in Google BigQuery as follows, and then included the resulting, summarized view as a .csv file in our repo.

```
select
  Year,
  ActionGeo_CountryCode,
  AVG(GoldsteinScale) as avg_GoldsteinScale,
  SUM(NumArticles) as sum_NumArticles,
  AVG(AvgTone) as avg_AvgTone,
  MIN(AvgTone) as min_AvgTone,
  MAX(AvgTone) as max_AvgTone
from gdelt-bq.gdeltv2.events
where Year between 2015 and 2020
group by 1, 2;
```

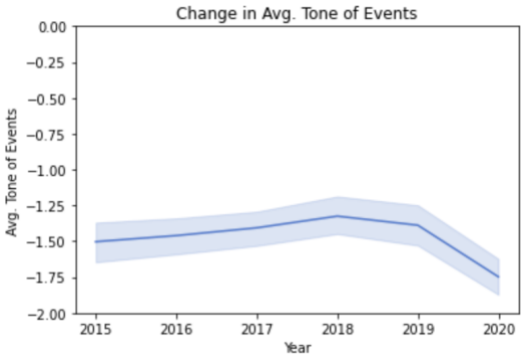

Figure 2: Trends in un-processed source data - events.

As noted above (see **figure 2**), the average tone of world events is slightly negative, with a downward swing starting in 2018 and continuing through 2020 (year-to-date).

[2] GDELT 2.0 Event Database

Country Code Mapping

Because a number of the prior datasets are keyed by either country names, country alpha-2 or alpha-3 codes, we needed a way of mapping either one of the 3 identifiers to the other in order to join datasets that use different taxonomies. For this we used IBAN's list of country codes[3], copied from their website and saved as a .csv file. We made no changes from the source in our import.

Analysis

Hypothesis 1 - Productivity

There is a positive relationship between national productivity (measured using GDP per capita) and national happiness.

When thinking about the things that make us happy, one of the first things to come to mind is wealth. On a national level, wealth can make all the difference in determining the average quality of life for citizens. In fact, the criteria for determining whether a country is "developed" or "develop-ing" is directly tied to Gross-National-Product (GDP) per capita. GDP is the total monetary or market value of all the finished goods and services produced within a country's borders over a period of time. It functions as a comprehensive scorecard of a country's economic health.[4] By dividing this amongst the total population of a given country (per capita) we get an easily comparable metric by which to compare countries.

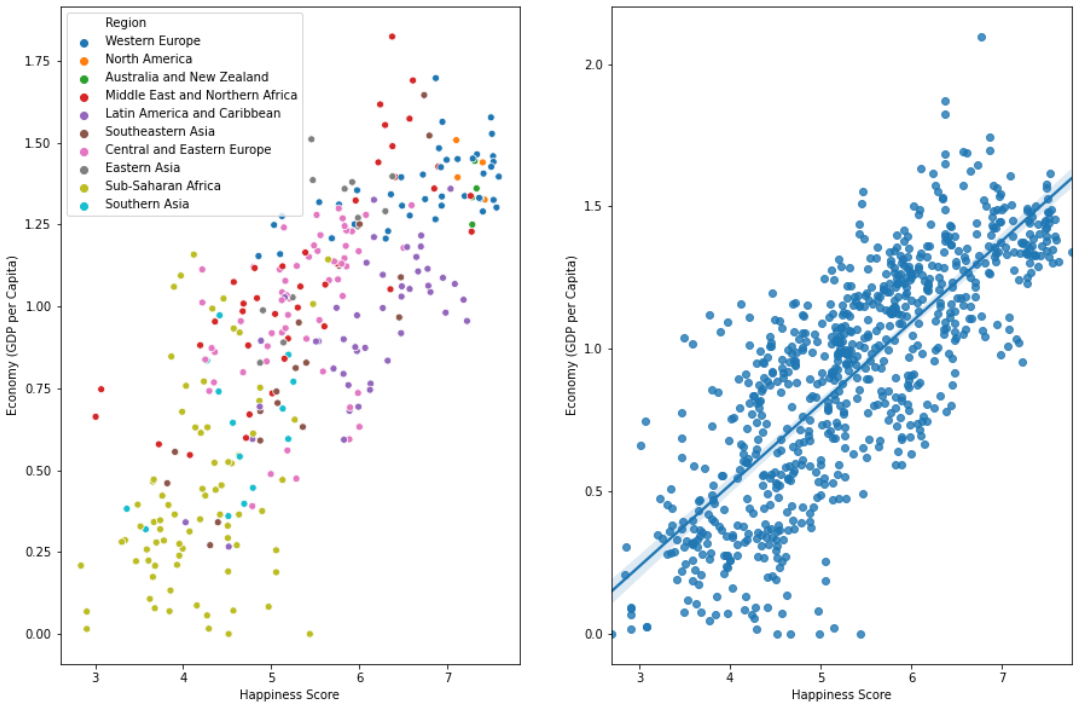

Figure 3: Regional overview of correlation between GDP per capita and happiness

Off the bat, our analysis showed promising results. Looking at **figure 3** above, we can see the strong correlation between happiness and GDP per capita. As we move from countries with lower GDP to ones with higher GDP, they tend to be happier in general. This is unsurprising given high GDP feeds directly

[3] Country Codes Alpha-2 and Alpha-3
[4] https://www.investopedia.com/terms/g/gdp.asp

into activities that benefit the individual citizens' happiness. For example, with more GDP per capita, a government has a larger tax base (or whatever the equivalent taxation system) and more revenue which it can dedicate to public works such as power plants, water plants, sewage disposal and many other areas of public development which subsequently impact the happiness of its citizens.

The breakdown of country by region as shown by the colors of the left graph in **figure 3** reveals some interesting clustering. Specifically, we can see the sub-saharan African region make up the majority of the left most points, however despite varying levels of GDPpc, happiness seems to be solidly rooted in the 3-5 range, stacked on top rather than moving left with higher GDPs. This is where cracks in the hypothesis started to form. By diving into the per country view, we quickly see a different story than the one highlighted above.

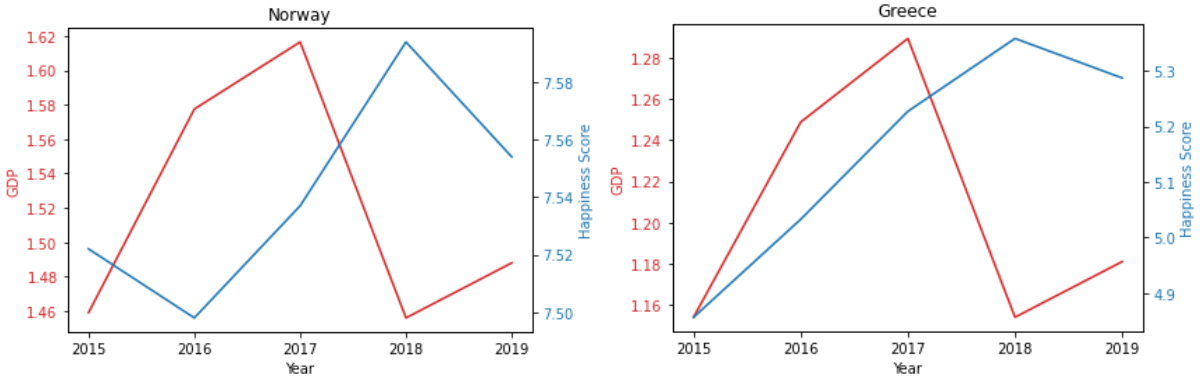

Figure 4: Norweigan and Greek happiness rates dropped in response to drops in GDP per capita

Stepping into a per country view of the data, we can get at the root of the inconsistencies we see in the overview. In **figure 4** we see 2 countries that experienced a recession in the past 5 years. In each case, there is a sharp drop in GDP, followed by a subsequent drop in happiness. This makes intuitive sense - whenever a country enters recession, there's a surge in unemployment, lower wages and lost opportunities generally that runs counter to people's happiness.[5]

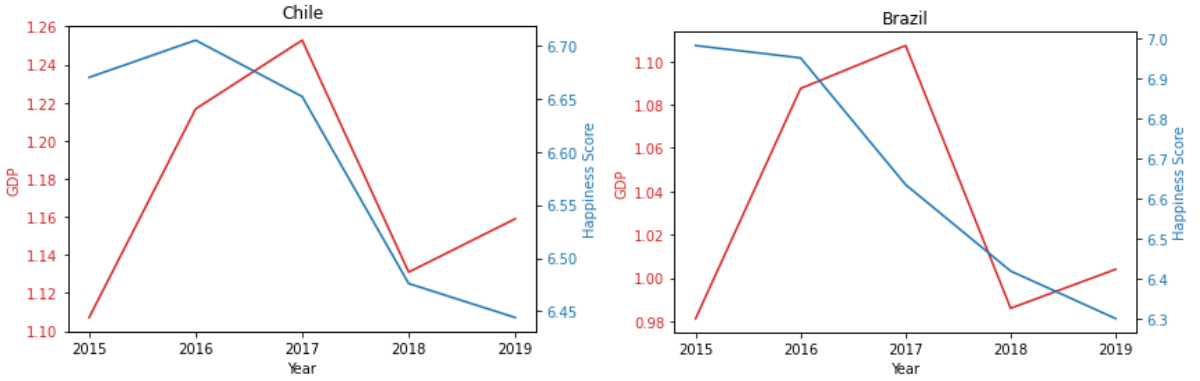

Figure 5: Chilean and Bazillian happiness rates preempted a drop in GDP per capita

In **figure 5**, we see a similar story, where happiness and GDPpc largely track together. However in this case, we see that drops in happiness actually preempt the drop in GDPpc. This makes sense given that at the end of the day, a population drives GDP. If citizens are unhappy, it is unlikely they will operate at peak performance, ultimately resulting in a contracting economy. This raises the interesting question of whether economists should include happiness data in their studies to more accurately predict oncoming recessions or crises.

[5] https://www.epi.org/publication/bp243

Despite this nice fit for some countries, others show very counterintuitive reactions to drops in GDPpc.

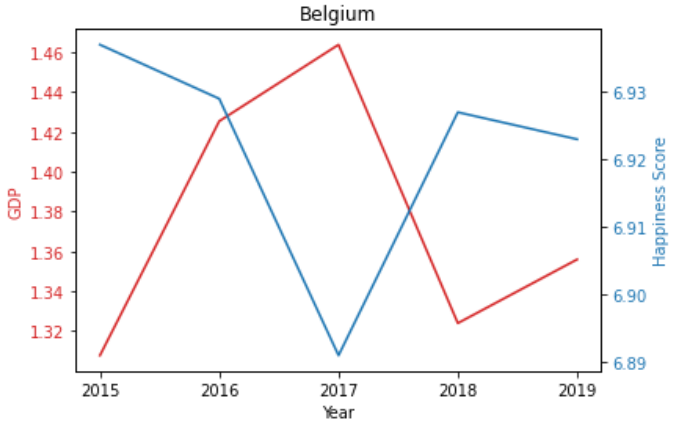

Figure 6: Belgium experienced a drop in GDP per capita from 2017-2019

For example, in **figure 6** we can see this inverse relationship between Belgium's GDP and happiness rates. Further investigation suggests external factors were likely at play, for example the small country came in 3rd in the 2018 FIFA World Cup which may have boosted national happiness.[6] All in all, while there certainly seems to be strong correlation between national happiness and GDP per capita, external emotional factors seem to quickly throw the trend in one direction or another, despite a strong or weak economy.

Ultimately, through this analysis we found the hypothesis to be true on a macro level - higher GDPs tend to correlate with happier citizens. However when evaluating the per country performance of this hypothesis, it is inaccurate to say if a country has a strong economy they will be happier and vice versa. There simply are too many variables at play to isolate and judge happiness by GDP per capita.

Hypothesis 2 - Family, Life Expectancy, Freedom, Government Trust & Happiness

Is there a positive relationship between family size, life expectancy, freedom, government trust and happiness? The answer, based on our study, is that: it depends on the region. Looking at data from 2015-2019 across the four variables and happiness for the whole world, we see relatively strong correlations between Life Expectancy, Family, Freedom and Happiness, while there seems to be a less strong correlation between Government trust and Happiness. Looking at the worldwide data per region, we saw some clustering happening, so we decided to take a deeper dive on a regional level.

[6] https://en.wikipedia.org/wiki/Belgium_at_the_FIFA_World_Cup#2018_FIFA_World_Cup

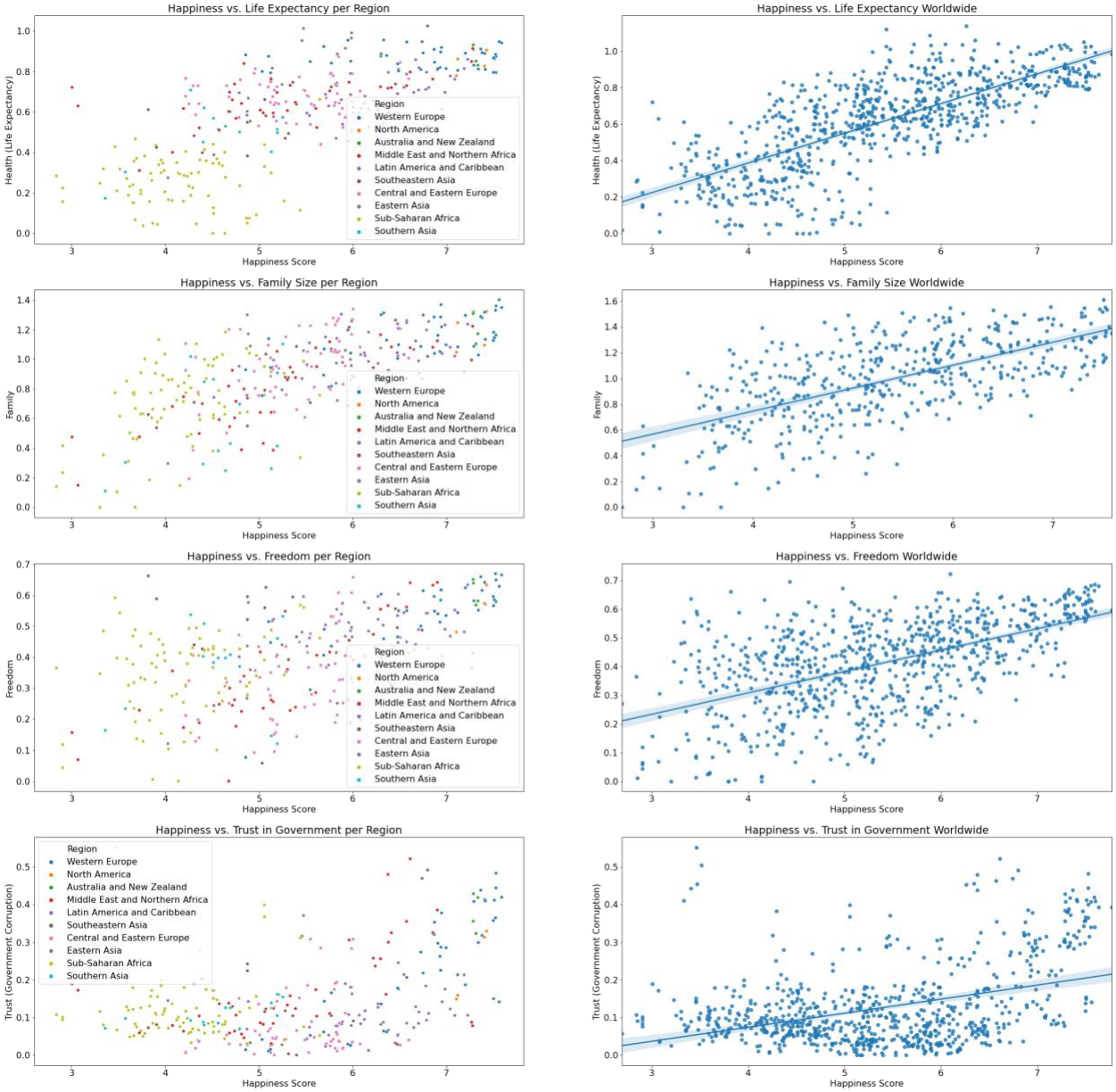

Figure 7: Positive worldwide correlations with happiness for 2015-2019

Taking a deeper look into the data on a regional level in **figure 7**, most of the regions did not have strong positive correlations with the variables (though most graphs trended upwards) and happiness and some regions did not have enough data points to make any conclusions. But two regions did stick out in having stronger correlations with the variables: Western Europe and the Middle East and North Africa.

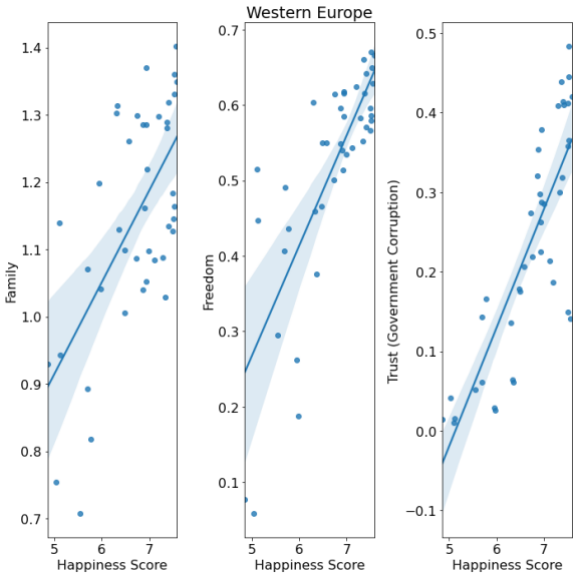

Figure 8: Strong positive correlations and high ranges for happiness in Western Europe

Looking at the scatter plots and best fit lines for Western Europe in **figure 8**, they showed particularly strong positive correlations for Government Trust and Happiness, strong positive correlations for Freedom and Happiness when in the upper ranges, and some positive correlation for family size and happiness.

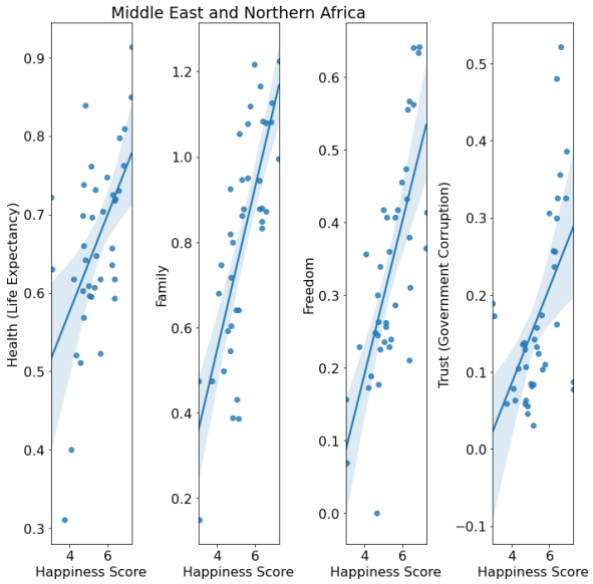

Figure 9: Very strong positive correlations but wider and lower ranges for happiness in the ME

In **figure 9**, for the Middle East and North Africa we see strong positive correlations with Happiness across all of the four variables, showing especially strong correlations with Family and Freedom. As compared to Western Europe in figure 8, however, we see a lower and wider range for the Happiness Scores. Western European countries scored between 4.8 and 8 for happiness, while the Middle East/North African countries scored between 3 and 7. In **figure 8** we also see more Western European Countries clustering on the higher ends of the variables and happiness, while for the Middle East/North Africa, the points are more evenly spread out, except for Government Trust which is shown to cluster on the lower end with low happiness scores.

Exploring our results, it intuitively makes sense that the Middle East/North Africa would have high correlations (as compared to Western Europe) to the variables of Family, Freedom, and Trust due to having less developed nations than Western Europe. We hypothesize that developed nations may have stronger relations with the economy/GDP for happiness than less developed nations, which would depend more on other facets like family and government.

Looking at the high correlation of freedom and government trust and happiness in Western Europe led us to explore why this might be the case. What we found is that Western European positive sentiment towards voting has increased over the last 30 years. Furthermore, educated Europeans tend to be satisfied with their country's democracies, even while there still remains skepticism with the government across Europe.[7]

On the contrary, for the Middle East and Northern Africa, we see high correlation with government and trust and happiness but more clustering on the lower end over the last 5 years. This makes sense due to the political upheavals in the region like the 2011 "Arab spring" and 2013 Syrian civil war. Studies are showing that the chaos in the Middle East is even decreasing life expectancy in some countries. In 2016, the Institute for Health Metrics and Evaluation reported that "Yemen, Tunisia and Egypt — all sites of regime change or collapse -- also lost around three months in life expectancy."[8]

Lastly, we wanted to explore the stronger correlation of Family size and happiness in the Middle East as opposed to the rest of the world/Western Europe. What we found was that while 36% of people in Europe live alone, in a single parent or coupled household, only 6% do in the Middle East. 56% of people live in a two-parent household in the Middle East, while only 26% do in Europe.[9] Tied closely to family is culture and we believe that the secular culture of Europe probably allow for higher divorce and separated families as compared to Middle Eastern/Islamic culture, which also allows for polygamy. Thus, it would make sense that strong family ties may be more important in Middle Eastern cultures for happiness.

Hypothesis 3 - Bad News

There is a negative relationship between bad news in a country and happiness.

Another reasonable question one might ask is whether the tone of events in a country has any meaningful relationship to the happiness its citizens experience? To study this question we computed the average happiness and average tone of events on a per-country basis for the 5-year study period. At first glance, when plotted together, there appears to be no discernable pattern or trend between the two variables (see **figure 10**):

[7] https://www.pewresearch.org/global/2019/10/14/democratic-satisfaction/
[8] https://www.washingtonpost.com/news/worldviews/wp/2016/08/26/middle-east-chaos-leads-to-drop-in-life-expectancy-in-some-arab-countries-study-says/
[9] https://www.pewresearch.org/fact-tank/2020/03/31/with-billions-confined-to-their-homes-worldwide-which-living-arrangements-are-most-common/

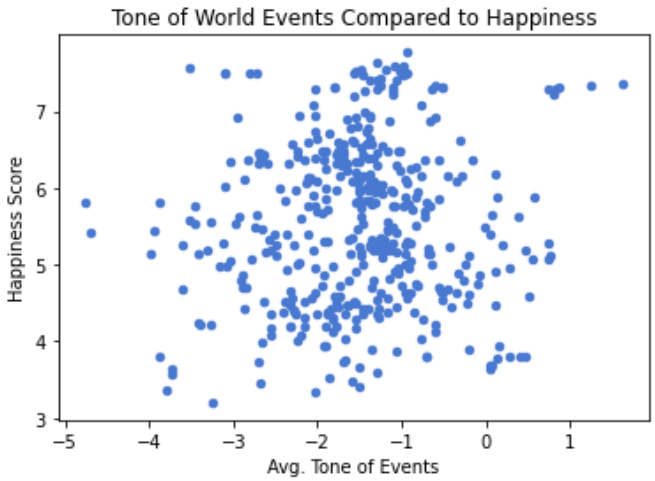

Figure 10: No clear relationship between macro happiness scores and event tones.

Adding in a time dimension doesn't add clarity to any discernible pattern, either (see **figure 11**).

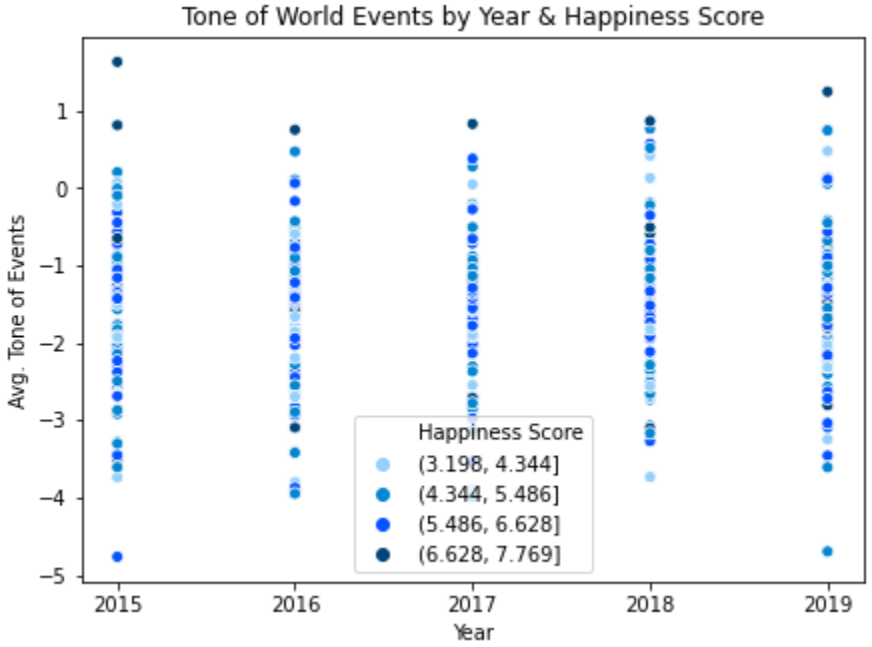

Figure 11: Similarly, no clear relationship over study tears.

Lastly, we looked at the correlation between happiness score and the average tone of events, per country, across all study years and concluded that there were 48 countries where this relationship was positive (i.e. happiness increased along with tone) and 39 where it was negative.

As a result we rejected our hypothesis, having failed to show any descriptive relationship between happiness and the tone of events within a country. It is likely that a variety of confounding factors, including perhaps GDP, family, life expectancy (all discussed previously) and others are more directly related to happiness.

Conclusion

Overall, we failed to show any clear relationship between happiness and GDPpc or the tone of events within a country. We showed weak global correlation for life expectancy, family size, government trust, and freedom with happiness, with some strong regional support for Western Europe and the Middle East and North Africa. We speculate that while individual effects may be too small to measure with descriptive statistics, it is possible that they have a nonlinear relationship to happiness as part of a function that may include, but is not limited to the studied variables. We propose further research attempting to model national happiness scores based on the features included in this study, as well as other demographic and economic variables. It may also be helpful to evaluate the correlation between study variables to identify proxies and other latent relationships.

Moreover, we also learned that merging data sets from the same source, but spread over time, as well as from disparate sources can be challenging due to formatting, inconsistency of features and degrees of granularity. We learned that the appropriate way to map such datasets may differ depending on the question being asked, and that tertiary datasets may be required to connect two primary sources - introducing its own challenges.