Northwestern University Data Science Boot Camp Project

Team Members
Nii Odateifio
Ellen White
Andrew Kim
Lizzy Nguyen

Project Overview
During the stresses and impact of Covid-19, we as a team wanted to see if there were any data sets with various elements that could help us determine the change in happiness levels across the world. We refer to a World Health Organization report that looked at a variety of metrics by country, including happiness.
We first asked, what are these country metrics? Which countries are the happiest, and which are the saddest? Is that due to these metrics?
Following that, we looked at WHO's Covid Mortality rates. We asked, with the hypothesis being "yes, of course"; did COVID mortality affect happiness? Analyzing the correlation between country happiness and country mortality rates, we were able to come to a surprising conclusion.

Findings
After merging the data files that we had collected, we cleaned up the information by limiting our reports to reflect information across 80+ countries (13 sub-regions) from a time span of 2011 to 2020, except Covid-19 mortality rates which were only used to reflect 2019-2020. We reviewed the Statistical Correlations and Regressions on the Happiness Survey vs 5 different factors:

- Log GDP
- Perception of Correlation
- Social Support
- Generocity
- Mortality Rates due to to Covid-19 from 2019-2020

The results of the YOY (2019-2020) change did not follow the statistical results. The 2019-2020 Change in Happiness study shows that the country characteristics that our regressions suggest influence a country's happiness, in fact, do not. A sub-region can have changes to its countries GDP, corruption, support, and generosity without an overall change in its happiness.

Not only that, resilience to COVID is evident globally. Besides South-East Asia, sub-regions did not significantly change average happiness between 2019 and 2020 (average between all subregions is an increase in happiness by 0.06).

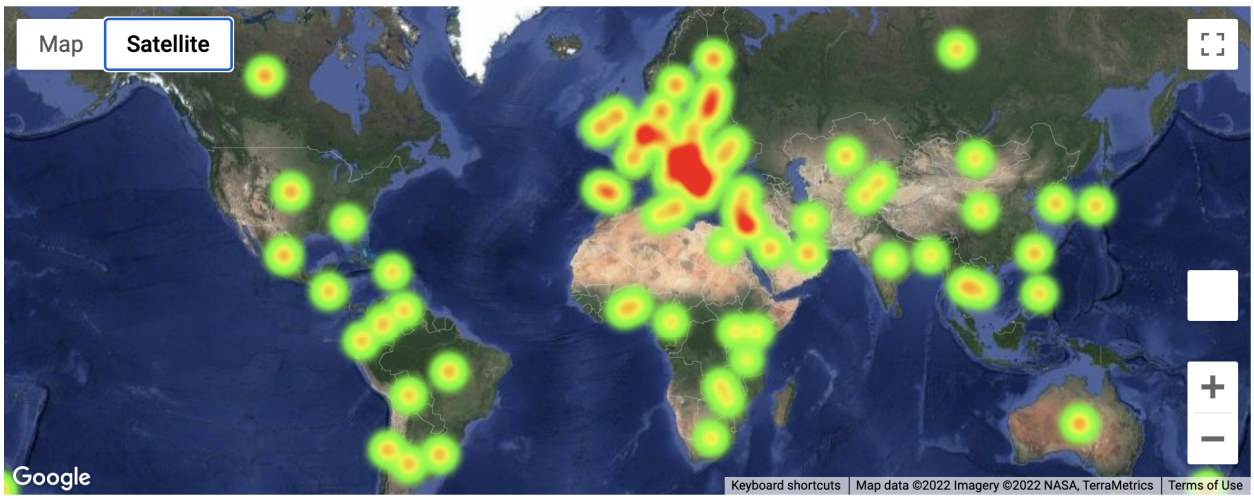
Data Overview
WHO Happiness report, Mortality Rate due to Covid, SubRegions, and Lat/Lng were used to derive a clean data set. Due to the varying countries and inconsistent countries in the report we needed to ensure that we had data for every country accounted for.

For the WHO Happiness Report we reviewed the past 10 years of data, and cleaned the data set to the countries that had data in each country which reduced the data set to 80+ countries.

Mortality report from Covid Deaths was added to the data set which reduced our data set to 60+ countries.

To be able to look at subregions we utilized the ISO-3166 Country and Dependent Territories Lists with UN Regional Codes to ensure the ability to summarize by regions.

Lastly we used Geopy and Nominatim packages to gain access to Latitude and Longitude coordinates to enable the gmaps API and created a heat map of Happiness by country. One is able to see that Europe was a concentrated area of happiness.



This is an API map that was created using the datasets mentioned above for the year of 2020, where weight is used to display the happiness levels that are being reported across 80 countries. The redder the region, the higher their levels of happiness are on a scale from 1-10.

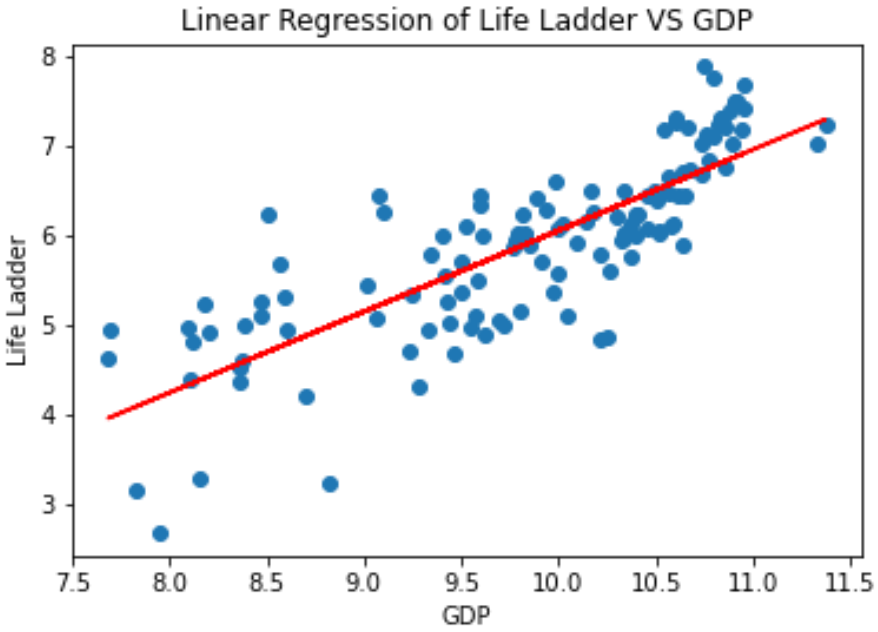
Linear Regressions and Correlation

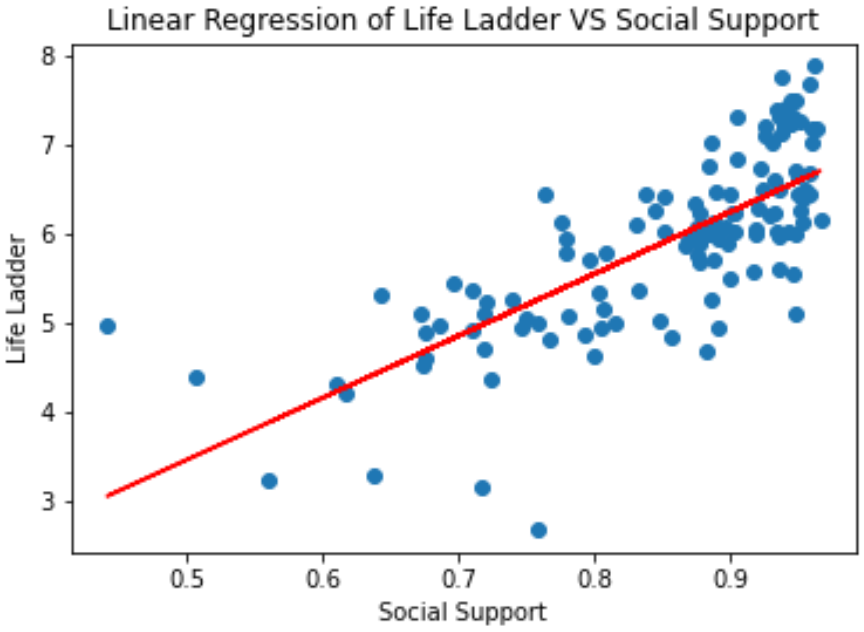
As we analyzed the relationship between overall country happiness and the country variables, we wanted to visually confirm whether or not our countries followed those regression trends between 2019 and 2020, or in other words, is it just some correlation, or is there causation as well. If not, we wondered if the changes in happiness are then due to COVID?

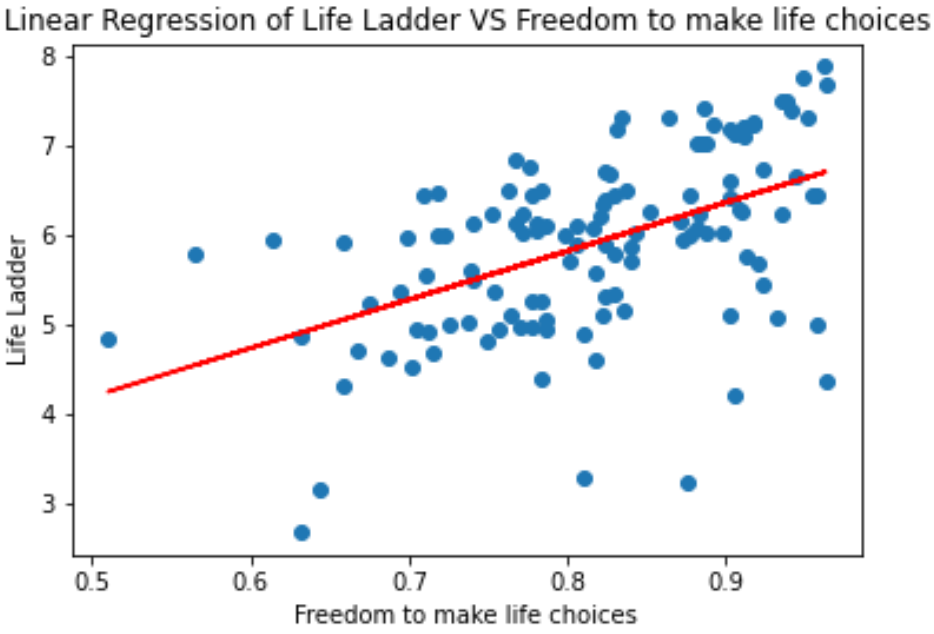
We merged data from another data set to bin all of our countries into sub-regions. We then plotted average happiness in both 2019 and 2020 for each sub-region to see who changed and by how much; only about 5 sub-regions changed significantly enough to explore why.

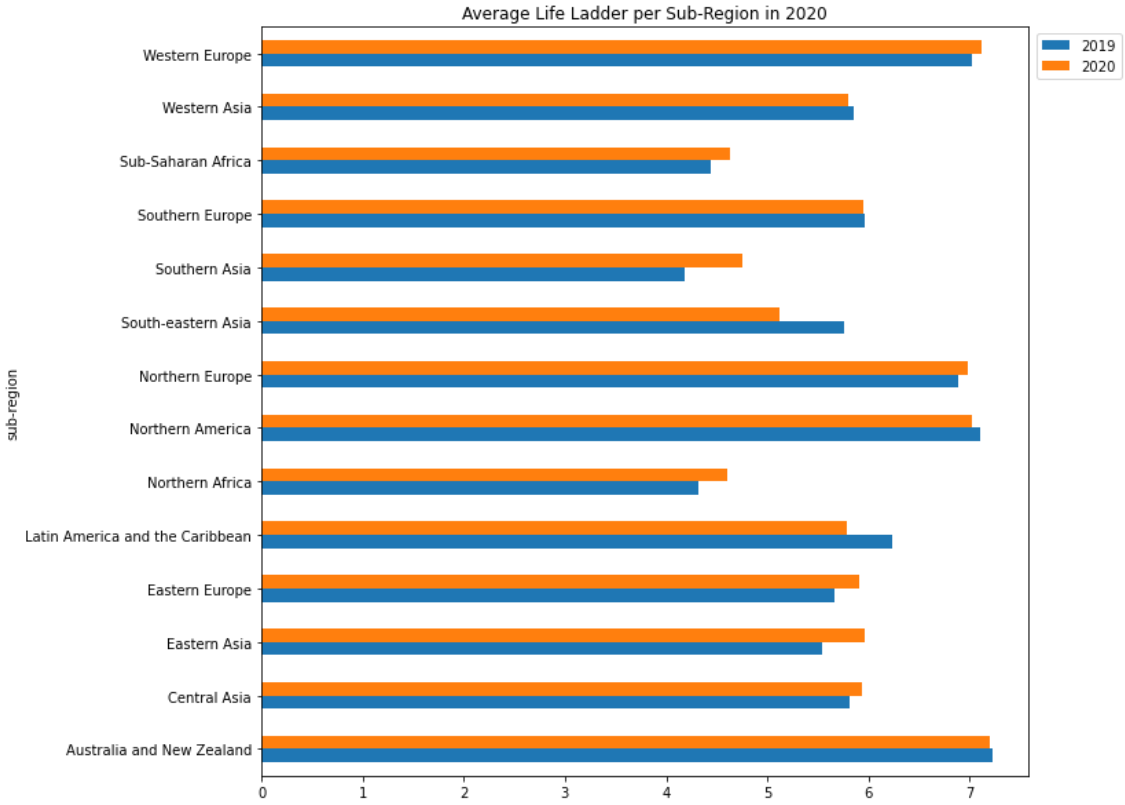
These metrics measuring certain characteristics of a country do not explain change in overall happiness. While some, like social support and Log GDP, tend to be higher in countries with higher happiness and lower in countries with lower happiness, ie a strong correlation, they are not the cause for change. The other variables don't even change enough to affect anything

between 2019-2020. After seeing this, we decided to look further into COVID deaths for a possible cause for the change in happiness, expecting at least a correlation between low happiness and high COVID death rate.






Linear Regression of Life Ladder VS Freedom to make life choices
Life Ladder
Freedom to make life choices
8
7
6
5
4
3
0.5
0.6
0.7
0.8
0.9


Average Life Ladder per Sub-Region in 2020
sub-region
2019
2020
Western Europe
Western Asia
Sub-Saharan Africa
Southern Europe
Southern Asia
South-eastern Asia
Northern Europe
Northern America
Northern Africa
Latin America and the Caribbean
Eastern Europe
Eastern Asia
Central Asia
Australia and New Zealand
0
1
2
3
4
5
6
7

Conclusion

Running the numbers together, based on our determining factors of happiness for this project, the “Happiest” country was Finland and the least “Happiest” was Zimbabwe. However, tying all of our findings mentioned above in our line regression charts and further analysis, we came to a conclusion that the matter of happiness is subjective and that there are too many factors that come into play. Each country has a different perception and reaction to the varying levels of the elements chosen above, therefore change in one element does not necessarily mean a large change in the scale of happiness as a whole.

Factors to consider:

We must note that a subjective matter such as Happiness cannot be derived statistically as there are variables that need to be controlled. We as Data Scientists must vet the validity of the data and understand that a subjective matter such as this may be more a phenomenon than statistical conclusions. Things we can further research are to create a control. If we can dive deeper by country we may be able to find more statistical significance, however, more variables must be taken into consideration. There are social and economical differences person to person, family to family, country to country that impact the scores. Governmental influence may drive polarizing results that are not a good reflection of “Happiness”. Examining the landscape of the variables and further fine tuning the data is crucial in getting a better understanding of Happiness.