

Ryan Dancoes, Sydney Gallo, Sandhya Sangappa

Professor Eddie

MA 346 Section 1

October 12, 2022

Midterm Project Proposal

The data we want to use for our midterm project is the World Happiness Report from 2005 to 2020 [found on Kaggle](#). Gallup began conducting its poll in 2005 to obtain a “happiness” score for 160 countries worldwide.

There are several research questions we wish to investigate:

- Which countries and regions are the most happy and unhappy? Have these rankings changed overtime? We are interested in gaining a starting point of what the happiest and unhappiest countries are and the characteristics they tend to have (I.e., are countries more generous if they have a higher happiness score).
 - Which countries and regions are the most and least free to make their own lifestyle choices? Have these rankings changed over time?
 - Which countries and regions are the most and least generous? Have these rankings changed over time?
 - What similarities and differences exist between the countries on these lists?
- What is the relationship between social support and perceptions of corruption overall, between countries, and between broader regions? We are curious to see if a lack of social support in a part of the world is associated with higher perceptions of corruption within the respective government, as well as businesses.
 - We want to include Generosity in the previous question to discern if countries and regions can be consistently generous in the face of poor social conditions.
- What is the relationship between Happiness (Life Ladder) and Social Support, Healthy Life Expectancy at Birth, Freedom to Make Life Choices, Generosity, and Perceptions of Corruption? We are curious to see which of these features have relationships with Life Ladder (and to what extent) and to see what matters most for people to be “happy.”
- Has the freedom to make life choices improved overall over time? For which countries and regions have this freedom improved or regressed? We are interested to see if the state of people’s lives (I.e., “average of laughter and enjoyment”) in different countries has improved.
 - If so, is this improvement consistent?
 - What economic and/or political events may have influenced changes in lifestyle freedom? Are future generations likely to experience higher levels of lifestyle freedom?
 - What parts of the world, if any, have lost this freedom and if so, for what periods of time?

- What is the mean and median positive and negative affect for a country in 2020? A region in 2020? We are interested in seeing the correlation between positive and negative affects for a country and how it relates to overall happiness.
 - How have these ‘affects’ changed over time over locations? In which countries and regions has the ‘affect’ increased and/or decreased?

Regarding methods, we will need to create a categorical variable “Region” to determine how broader portions of the world perform in our analysis. We will explore using plotly dash to create a 3D plot of social support, corruption, and generosity. Both seaborn and matplotlib will be used in this analysis, and we will take advantage of the knowledge learned through Datacamp to include visualizations such as swarm plots and heat maps. Most features in the dataset contain null values, which we will fill in with the median or the closest non-null values since we do not want to drop rows with null values and lose data that can impact the results of our analysis. We will also create a data dictionary based on the official World Happiness Report website to include in the report.