COGS9: Introduction to Data Science
Assignment #1: Project Proposal
Due date: Friday 2019 November 01 23:59:59
Grading: 5% of overall course grade. 20 points total.
Completed as a group. One submission per group on Gradescope.

Group Member Information:

First Name	Last Name	PID
██████████	██████████	██████████

Question

How does a country's freedom to make life choices, perceptions of corruption, and social support correlate with their citizens' levels of happiness using the number of national measures of the quality of government and the national average happiness?

We will be measuring all the above factors in the form of an online survey in which we ask people to rate the following factors on a scale of 1 to 10 (1 being no freedom and 10 being full freedom, 1 being extremely corrupt and 10 being no corruption, 1 being no social support and 10 being full support).

Hypothesis

Because freedom is often the thing that most people desires the most, it is hypothesized that Freedom Index (freedom to make life choices) positively correlates with citizens' levels of happiness. The connection between governments and happiness function in both ways, what government do to influence happiness and in exchange, the citizens' happiness determines what kind of government they support and to what measure they substantiate it.

In addition, it is believed that the Corruption Index negatively correlates with citizens' levels of happiness as corruption often negatively impacts residents' lives. The index is the average of

binary answers to two questions including "Is corruption widespread throughout the government or not?" and "Is corruption widespread within businesses or not?"

Background Information

There are many variables that contribute to the calculation of Happiness Index, such as Corruption Index, GDP, Freedom to make life choices, Confidence in government, etc. According to the World Happiness Report, the locals and immigrants within a country share "similar happiness levels". This suggests that immigrants adapt to the local psychological climate. The role of government can be an essential factor in determining the psychological climate of a country (Sachs). As a result, it is likely that this Happiness Index , which is based on local residents' response to the survey, is greatly impacted by the country's current state and government's effectiveness.

The corruption index in the data set is determined by survey answers provided by residents in a variety of countries. It is the average of the binary (Yes = 1, No= 0) answers to two questions in the survey: " Is corruption widespread among the government?" and "Is corruption widespread in the business?" It is used in this data set to measure a country's corruption index in the public sector (Helliwell). Higher index indicates that the country is corrupted while low index indicates the lack of corruption reported by local residents.

Because residents of some countries may not always have the freedom to make their life changing decisions as they wish, freedom to make life choices is often a variable depending on which country you are from. As a result, freedom to make life choices is included when considering the inhabitants happiness index. It is calculated by averaging the is binary responses to the GWP question "Are you satisfied or dissatisfied with your freedom to choose what you do with your life?", with 1 representing Yes and 0 equivalent to No (Helliwell) in the dataset.

Data

For this question, the ideal dataset would be one that covers all the variables we discussed in our question, democratic quality, perceptions of corruption, and social support, and at least one observation for each country. More ideally, we would use a dataset with multiple observations for each country, perhaps several years per country, to observe changes over time.

Online, the World Happiness Report has an easily accessible dataset to use, that includes several variables, like the country's life satisfaction, GDP per capita, and life expectancy. Most of the variables measure a country's population and their feelings towards things like their trust in the government, their cost of living, and freedom to make choices. It also includes measurements for

corruption, democracy, and social services, all of which we wish to examine. The weakness of a dataset like this is that for some countries certain measurements couldn't be taken, whether they be due to circumstances in that country or otherwise. Furthermore, due to the limited scope of surveys, they could only sample a certain portion of the population, not its entirety. There could be biases in the survey that would be more or less significant on the whole population. Moreover, since some data is measured by surveying the population, there are portions of the data that couldn't be measured directly, so it had to be taken by survey. This is less powerful than if there was an actual government record of corruption, but since it is unlikely that such a document exists, there had to be a compromise.

Ethical Considerations

Data Collection:
We wish to collect data from the citizens of each country through an online survey. Anyone who has access to the internet will be able to participate in our research. We will obtain informed consent from all participants. The data that we collect from them will be anonymous. It will also include questions about people's gender, age, citizenship and race to ensure the diversity of our data. If we decided to use data from APIs, we need to make sure that we have permissions.

Data Storage:
To protect our data, we will require a two-step login process to access the data set, using the app Duo Mobile. We will send a confirmation email with partipant's survey response to their email that they provided in the survey so that people are able to verify what they submitted and contact us if there is any concern. To delete the data we no longer need, we will access the data set and delete the corresponding row of data.

Analysis:
After we collected our data, we will use codes to generate several different plots to ensure that our data is well represented. For example, we will have a venn diagram for each country showing the composition of the percentage of people that chose each number(1-10). Also, we will have plots comparing the percentages of the population that chose each number across different countries. The plot will not display any personal identity information since it should provide trends about countries not individuals. Then, we will check our data trends (2017-2018) with the trends from years before to ensure that the differences in data are reasonable. We will also check with credible sources online such as the United Nations' report to make sure our analysis is in the right direction. All the plots and code will be put in files and stored in GitHub to ensure reproducibility. We will only include countries where at least 20% of the population participated. There might be a bias in data since not all countries have equal access of internet for everyone.

Modeling:
The variables we chose to measure are not biased because they are not related to personal traits and behaviors. However, since our data collecting method is through an online survey, we might have less data in countries' with people who do not have internet access on a daily basis or people who do not obtain any kind of social media. This can cause biases in our model. Also, since we are rating the government based on people's feelings and judgment, it might not be an accurate measure of the true performance of the government.

Deployment:
The transnational corporations might be able to use the data to make money since the happiness level of a country relates closely to their purchasing behaviors. Moreover, politicians can use our data to better target their campaigns.

In case the participants emotionally get harmed by the result, we will make sure to leave our contact information in the confirmation email and ask them to contact us if they ever feel uncomfortable that their data are being used. We will also provide them with personal psychiatrist if they need it. Also, we have full access to all the information, graphs, and models we publish. Therefore, we can recall all results if necessary. The change in the development of a country will probably change how people feel about their current government, thus it is important to recollect the data every year to ensure the most accurate and reliable results are displayed. We will give a clear explanation of the meaning of what we published. Our graphs and models will have watermarks on them so people know where to find reference.

In addition, we will present the data in as unbiased a manner as possible, as to not give the perception that we have any particular biases towards any one country. The point of this data is to present what countries perceive themselves to be unhappy, not the opinions of anyone else, including this group. Moreover, the conclusions drawn by the "happiness index" must be presented in such a way not to intentionally on our part impose stereotypes onto any country. The "happiness index" was built using a wide variety of variables, any combination of which could have lead to any country's particular score. The point of this data was to examine what variables led to a satisfactory population, as opposed to which country was "happiest." Presenting our data from this point of view will minimize the level of stereotyping that will naturally occur.

Work Cited

Sachs, Jeffrey D., et al. "World Happiness Report Finds That People Are Feeling Worse."
    Greater Good,
    https://greatergood.berkeley.edu/article/item/world_happiness_report_finds_that_people_
    are_feeling_worse.

Helliwell, John F., et al. *World Happiness Report*.
    https://s3.amazonaws.com/happiness-report/2019/Chapter2OnlineData.xls.