**Country Happiness: Testing Different Measurements for Happiness Against the World Happiness Report**
**CSE 519 Final Project: Progress Report**

**Objective:** To test different metrics for measuring a country's happiness by experimenting with multiple scoring functions. To incorporate new variables not included in the WHR's Happiness Indicators and attempt to create a better measurement for happiness. Secondarily, to test different variables and data sources of self-reported happiness and understand how our new measurements affect happiness rankings.

## Approach

The World Happiness Report (WHR) is an annual report designed to rank every country in the world according to how happy its citizens are. Each country is assigned a happiness index based on a number of objective and self-reported indicators of well-being. Rankings are determined by such factors as GDP per capita, life expectancy at birth, perceived social support, perceived freedom to make life choices, generosity, perceptions of corruption, and positive and negative affect. We attempt to create new happiness rankings using additional indicators that we feel make a difference to the happiness of a given country's citizens.

Our baseline model integrates ten additional variables with the eight indicators that make up the WHR Indices. We believe that these additional variables help provide a more complete picture of factors that may influence well-being within a nation, and these indicators are discussed in detail in the next section. Our second model, which we are currently working on the building, is a bit more experimental. In this model, we attempt to use a different data source for self-reported happiness indicators: The World Values Survey. We utilize variables from this survey which we believe may affect happiness within a country and may even provide better measurements than the Gallup World Poll data used by the World Happiness Report. More details on the datasets and variables used are discussed in the next section.

## Data Sources & Cleaning

***Datasets:***

Multiple data sources provide the needed metric by countries for our analysis. We considered the latest year for which data was available and for how many different countries. The analysis is shown in the table, which helped us to finalize on the World Bank and UN Human Development as our primary data sources.

| Data Source | Latest Year | Countries |
|---|---|---|
| WHR20 | 2019 | 133 |
| World Bank | 2019 | 192 |
| UN Human Development | 2018 | 190 |
| UN SGD Indicators | 2017 | 174 |
| WHO | 2017 | 93 |

***Baseline Dataset:***

In addition to the eight indicators that make up the WHR indices, we incorporated the following additional variables into our baseline model dataset (sources listed next to each data source):

- Maternal Mortality rate (per 100K) [World Bank] (Negative affect)
- Adolescent Fertility Rate (per 100K) [World Bank] (Negative affect)
- % of Parliamentary Seats held by women [World Bank] (Positive affect)
- Secondary School Enrollment- Gender Parity Index [World Bank]
- Adult Literacy Rate [World Bank] (Positive affect)
- Natural Resource Depletion (% of GNI) [UN Human Development Data] (Negative affect)
- Life Expectancy at Birth [UN Human Development Data] (Positive affect)
- Income Inequality (GINI coefficient) [UN Human Development Data] (Negative affect)
- Female and Male suicide rates [UN Human Development Data] (Negative affect)

The inclusion of these variables into our baseline model was motivated by both prior research and (what we felt were) gaps in the WHR's happiness measurements. We felt that the initial eight indicators focused too heavily on self-reported data and overlooked some more objective indicators that may influence well-being.

***Model 2 Dataset:***

In addition to our baseline model dataset, we created an additional dataset for use in our second model. This dataset focuses on a subset of 48 countries and includes data from the World Values Survey (WVS) rather than the WHR's Gallup Poll data. In our second model, we are trying to test how country happiness rankings will change based on the use of different self-reported well-being metrics. We use a subset of countries in this model due to the lack of availability of a worldwide dataset. Gallup does not make its World Poll data publicly available without paying a lot of money, and no similar self-report dataset exists that is publicly accessible. An additional advantage of utilizing the World Values Survey is that, while Gallup uses a binary measurement for many of their indicators, the WVS uses Likert Scale measurements instead. This proves advantageous when compiling indices because instead of relying on an average of binary responses (as calculated by the WHR), a Likert Scale measure can provide a more representative estimate for that metric.
The variables we measure from the WVS are as follows:

- **Freedom to make life choices:** Some people feel they have completely free choice and control over their lives, while other people feel that what they do has no real effect on what happens to them. Please use this scale where 1 means "no choice at all" and 10 means "a great deal of choice" to indicate how much freedom of choice and control you feel you have over the way your life turns out.
- **Life Satisfaction:** All things considered, how satisfied are you with your life as a whole these days? Using this card on which 1 means you are "completely dissatisfied" and 10 means you are "completely satisfied" where would you put your satisfaction with your life as a whole?
- **Perceptions of Corruption:** For this metric, we take a weighted average of the following five questions:
    - Q112. Now I'd like you to tell me your views on corruption – when people pay a bribe, give a gift or do a favor to other people in order to get the things they need to be done or the services they need. How would you place your views on

corruption in your country on a 10-point scale where “1” means “there is no corruption in my country” and “10” means “there is abundant corruption in my country”. If your views are somewhat mixed, choose the appropriate number in between. (*weighted at 50%)*

    - Q113 to Q117: Among the following groups of people, how many do you believe are involved in corruption? Tell me for each group if you believe it is none of them, few of them, most of them, or all of them? (state authorities, business executives, Local authorities, civil service providers, journalists, and media) (*weighted at 10% each)*
- **Positive Affect:** Q46. Taking all things together, would you say you are (readout and code one answer): 1 Very happy 2 Rather happy 3 Not very happy 4 Not at all happy

We merge these self-report measurements from the WVS with the objective variables from our baseline model to create model 2, and we are currently in the process of building this set of happiness rankings.

***Cleaning:***

During preprocessing, we took the following steps to clean the data:

- To fill in missing data that was unavailable for 2018, we rely on the most recently available data for that measurement (i.e.. if data for 2017 was available, we use the 2017 data). This is the same approach that the WHR uses to fill in missing data. If data from within the last five years are unavailable, we imputed the value using the column median.
- Certain country names were marked differently in the UN data and the World Bank data (ex: Saint Lucia vs St. Lucia), so we processed duplicate country names or countries that had slightly different names in the two datasets to make the names the same. The same has to be done with WHR data as well, as it will be used to compare ranks.
- *For WVS Data:* The WVS measures data on individual people rather than nations. Thus, to pre-process the data, we took an average of all responses from a given nation to create one indicator per variable.
- *For WVS Data:* There were two countries that had missing names and country codes (Macau and Serbia), so we manually replaced those missing entries.

***Additional Data & Next Steps:***

We will build at least one more set of rankings for European countries specifically, which seeks to understand how the use of the third source of self-report dataset influences happiness rankings. We are also in the process of finding more *negative* indicators. By this we mean, negative factors in a country that might *reduce* the happiness of its citizens. We have one negative metric in our baseline model (suicide rates) but we are currently looking into additional metrics that might negatively affect happiness.

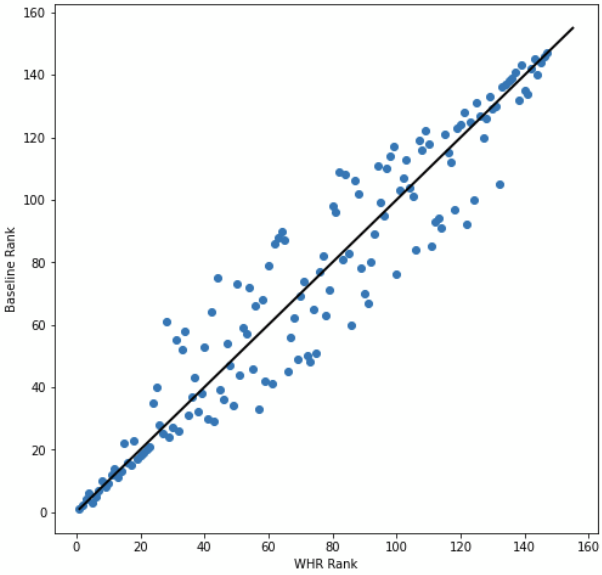
## Baseline Model Implementation

The WHR model considers an additive approach with their selected approach. They mention that positive affects have a large and highly significant impact on their score than the negative ones. Our baseline variables contribute positively as well negatively to happiness, and we treat them differently. They are normalized first, to get them all in the same range of 0 to 1. Here, we have 4 out of the 10 variables leading to positive affects while the remaining 6 have negative affects, and their breakdown is mentioned above while describing the dataset. The baseline scoring function can be defined as follows:

$$Baseline\ Score = Ladder\ Score + Positive\ affect\ Variables\ - Negative\ affect\ variable$$

| | WHR Top 10 Countries | Baseline Model Top 10 |
|---|---|---|
| 1 | Finland | Finland |
| 2 | Denmark | Denmark |
| 3 | Switzerland | Norway |
| 4 | Iceland | Switzerland |
| 5 | Norway | Netherlands |
| 6 | Netherlands | Iceland |
| 7 | Sweden | Sweden |
| 8 | New Zealand | Austria |
| 9 | Austria | Luxembourg |
| 10 | Luxembourg | New Zealand |

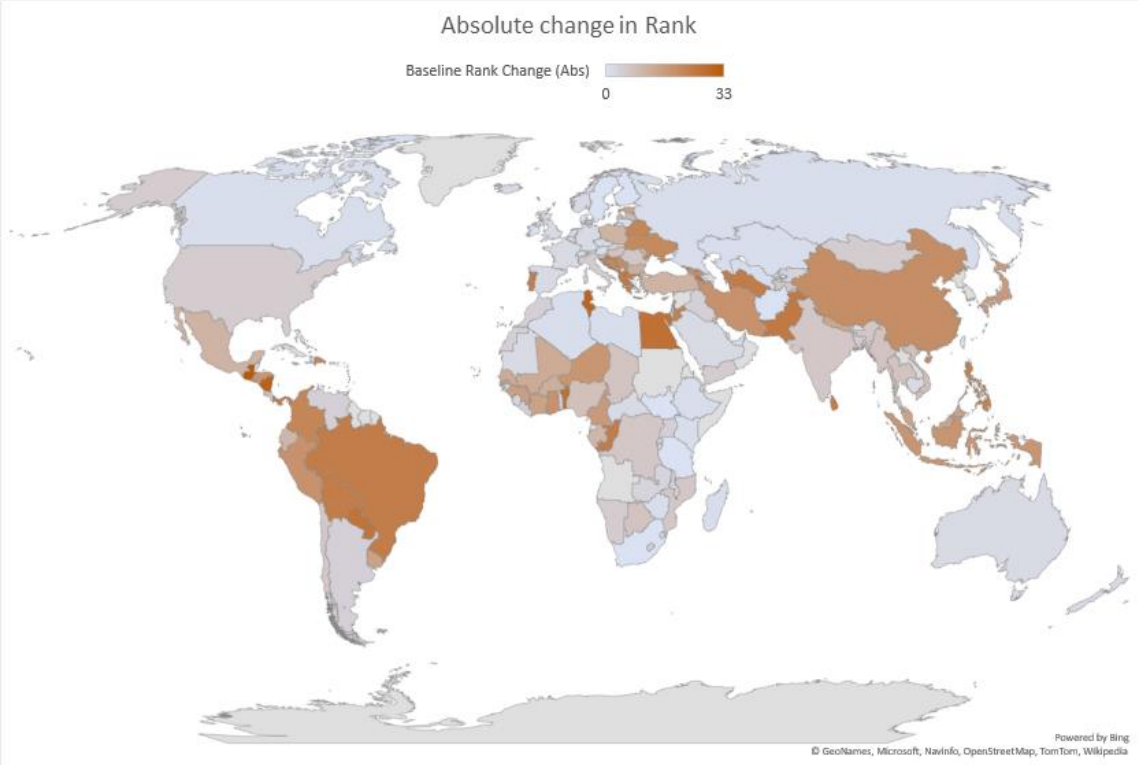
Our initial model yielded some interesting patterns. Some countries had very little change in their happiness ranking with the inclusion of additional variables, while other countries had a notable shift in their rank. Since most of our variables are negative, the baseline score normally dropped for most countries, except the ones whose positive affects are much higher than negative ones, this can be seen in the Top 10 countries. The Top 10 countries did not have much substantial change from the WHR top 10 in our baseline model (see Figure 1), though many countries shifted their position within the top 10 slightly. Though there are many countries that saw substantial shifts in their rankings and their position shifted by at least 10 ranks or more in either direction. Figure 2 displays an overview of the net change in ranking for all the countries in our model. Points above the line indicate that the country's rank increased, from the WHR ranking, and points below the line indicate countries whose ranking dropped.




There seem to be more countries that decreased in happiness rather than increased, which could present an interesting pattern. This may suggest that the WHR is inflating the happiness of certain countries by the indicators they choose to include. We are currently in the process of determining whether a notable pattern exists in the countries which get notably less happy with the inclusion of new variables, and the reverse: countries that get significantly happier under our scoring function. One of the

next steps would be to find which of our variables make a higher impact on the happiness and weight the variables to help get a better score dependent more on significant variables, be it with positive or negative affect. One way we plan to do so is by checking out variable's correlation with the WHR scores and weight them accordingly. This will help to build a better scoring function to predict a better happiness function.

Also, worth noting is the ways in which countries' happiness rankings vary by region. Figure 3 depicts the net change in ranking plotted on a map of the world. Interestingly, the only countries largely affected are in Asia, Africa, and South America, with North America and Europe largely unaffected.




## References:

[1] Kabene, S., Baadel, S., Jiwani, Z., & Lobo, V. (2017). *Women in Political Positions and Countries' Level of Happiness*. *18*(4), 10. https://vc.bridgew.edu/jiws/vol18/iss4/15/

[2] Helliwell, John F., Richard Layard, Jeffrey Sachs, and Jan-Emmanuel De Neve, eds. 2020. World Happiness Report 2020. New York: Sustainable Development Solutions Network https://worldhappiness.report/ed/2020/