MGMT 59000 Visual Analytics

MSBAIM

Data Visualization Project Part III

Proposal and Plan

Due Date: 11:59 pm, Oct. 13th, 2023

Team Members:

Alexander Scarcelli

Avishek Dasgupta

Hummarah Shahzad

Nikolai Saporoschetz

Samarth Bansal

Siddhant Treasure

# Introduction and Background

The happiness of individuals in a nation impacts the overall health and productivity of the citizens of that country. “In the 18th century, the Enlightenment ushered in the notion that happiness was the attainment of a worthy life.” (Stearns, 2012). Following the enlightenment, the pursuit of happiness became a guiding principle for the foundation of many countries’ democratic systems of government. However, many world leaders have been measuring the success of their nation’s policies through how affluent the economy is.

In 2012 the United Nations began collecting data to collate the World Happiness Report. “This is a report that examines how citizens worldwide rank their subjective level of happiness.” (Bertolucci, 2018). This report was developed with the intention of shifting the way world leaders evaluate the success of their policies from the economy alone to a more holistic picture of how satisfied individuals are with the life they lead.

Happiness is extremely important, but it is most important to our health and well-being, so understanding what factors drive happiness enables leaders in society to better serve their constituents. “Happiness lowers your risk for cardiovascular disease, lowers your blood pressure, enables better sleep, improves your diet, allows you to maintain a normal body weight through regular exercise and reduces stress.”  (Northwestern Medicine, 2015)

Our motivation for choosing this topic originated from our interest in determining whether economic or social factors played a larger role in determining a country’s level of happiness. The United Kingdom’s Office for National Statistics shares this interest of looking beyond economic factors for measuring a country’s well-being, “Within the UK, there is a commitment to developing wider measures of well-being so that government policies can be more tailored to the things that matter.” (UK government web archive, 2016).

We believe that helping policy makers better understand the drivers of happiness will lead to a happier and thus healthier and more productive society. In doing our research we hope to accomplish the following:

* Assess the relationship between socio-economic indicators (GDP etc) and the happiness index.
* Identify other factors such as education, unemployment, human development index, political stability to determine the level to which they contribute to a country’s happiness.
* Compare happiness scores and other important indicators across nations to determine if similar patterns emerge in different countries.

# Dataset Information:

## Dataset 1: World Happiness Report (2013-2023)

|  |  |
| --- | --- |
| **Question** | **Information** |
| **Where did you find the dataset?** | We found the original data set on Kaggle. |
| **Why was the data created and for what purpose? Who collected the data?** | This data has been curated by the United Nations for the purpose of bringing more attention to happiness as a criterion for government policy. According to the World Happiness Report the report, “reviews the state of happiness in the world today and shows how the science of happiness explains personal and national variations in happiness,” (World Happiness Report). The happiness data is based on results from the Gallup World Poll which is where the data is originally collected. Additionally, the World Happiness Report uses volunteers for their publications, but their administrative and research costs are funded by their partners which include Unilever; The Blue-Chip Foundation; The Happier Way Foundation; and The Regeneration Society Foundation |
| **What is the timeline or lineage of the data?** | The timeline of the Happiness Report data spans from 2013 to 2023, or 10 years |
| **How large is the dataset (cases, variables)?** | 4 Variables: Country, Year, Index, Rank/ 167 Countries represented over ten years/ 1670 Total Countries Represented over all ten years |
| **What locations are included in the dataset? Describe how the dataset will help you achieve your goals.** | 167 unique countries are included in the dataset. We will be able to make graphs at a national/ world level. This dataset provides us with a response variable of sorts as we are trying to uncover what drives happiness |

## Dataset 2: GDP

|  |  |
| --- | --- |
| **Question** | **Information** |
| **Where did you find the dataset?** | We found the original data set on the United Nations Development Program website. |
| **Why was the data created and for what purpose? Who collected the data?** | The data was taken from United Nations systems and has been created to track economic output (GDP) over time across countries. We are comparing the sum of gross value added by all resident producers in the economy between the nations. The UNDP draws data from across the UN system and partners. |
| **What is the timeline or lineage of the data?** | The data we used is from 2018-2023. |
| **How large is the dataset (cases, variables)?** | There are around 170 cases, and we are using 2 variables i.e. Countries, Year, GDP. |
| **What locations are included in the dataset? Describe how the dataset will help you achieve your goals.** | The datasets compare the GDP per capita for all nations. We will be able to make a map or bar graph at the country and world level. This is the first economic indicator people think of when it comes to a country’s economic performance. Thus it will tell an easy-to-understand story when we develop visuals. |

## Dataset 3: Human Development Index

|  |  |
| --- | --- |
| **Question** | **Information** |
| **Where did you find the dataset?** | We found the original data set on the UNDP website. |
| **Why was the data created and for what purpose? Who collected the data?** | The data was created to establish a universally recognized index of average achievement by country across key dimensions including life expectancy, education, and per capita income, collected by the UNDP data shows the Human Development Index by country which combines life expectancy, education and per capita income into a single measurement ranging from 0 to 1 by country. |
| **What is the timeline or lineage of the data?** | The data is from 1989 to 2018. |
| **How large is the dataset (cases, variables)?** | There are 4 variables: Country, ISO-3 Code, Year, Human Development Index. There are around 7,080 cases. |
| **What locations are included in the dataset? Describe how the dataset will help you achieve your goals.** | The datasets compare the HDI for all the nations. This will help us to measure HDI against an economic indicator that has a presumed relationship with happiness. This will also help us to measure happiness and the HDI across countries by year. |

## Dataset 4: Unemployment Data

|  |  |
| --- | --- |
| **Question** | **Information** |
| **Where did you find the dataset?** | International Monetary Fund (IMF). |
| **Why was the data created and for what purpose? Who collected the data?** | The data was created, and unemployment is tracked by the IMF for the purpose of studying the inverse relationship that exists between unemployment and economic growth. Increases in unemployment often occur after recessions begin and recover later in a growth cycle. The IMF believes tracking unemployment across countries provides important data points for understanding various relationships to help inform policy makers and businesses. |
| **What is the timeline or lineage of the data?** | The data is from 2018-2022. |
| **How large is the dataset (cases, variables)?** | We are using three variables i.e. Countries, Year, Unemployment Rate. There are over 500 rows of data. |
| **What locations are included in the dataset? Describe how the dataset will help you achieve your goals.** | The dataset compares the yearly unemployment rate by country across just below 170 countries. This will help us to measure happiness against an economic indicator that has a presumed relationship with happiness. We will be able to compare happiness and unemployment rates across countries. |

## Dataset 5: Political Stability

|  |  |
| --- | --- |
| **Question** | **Information** |
| **Where did you find the dataset?** | We found the original data set on the UNDP website. |
| **Why was the data created and for what purpose? Who collected the data?** | The dataset was developed by the United Nations to rank political stability and the absence of terrorism by country. The UN utilizes this data in peacemaking efforts to advance peace in political conflicts across the globe. |
| **What is the timeline or lineage of the data?** | The data is from 2018-2022. |
| **How large is the dataset (cases, variables)?** | The dataset contains 212 rows along with individual columns for the political stability rating for the years 2018-2022. Other variables include country name, country code, indicator code. |
| **What locations are included in the dataset? Describe how the dataset will help you achieve your goals.** | The dataset compares the political stability rating (0-100) by country (167). This data point provides us with another point of comparison. This indicator provides us with a glimpse into the political climate of countries and the threats of terrorism which we perceive as having a likely relationship with happiness. |

## Dataset 6: Population Data

|  |  |
| --- | --- |
| **Question** | **Information** |
| **Where did you find the dataset?** | International Monetary Fund (IMF) |
| **Why was the data created and for what purpose? Who collected the data?** | The data was created to measure the total population growth/ decline by country over time. IMF collected the data. |
| **What is the timeline or lineage of the data?** | The data is from 2018-2022. |
| **How large is the dataset (cases, variables)?** | There are around 755 cases, and we are using 3 variables i.e., Countries, Year, Population (In Millions). |
| **What locations are included in the dataset? Describe how the dataset will help you achieve your goals.** | The dataset compares the yearly total population by country. This will help us to measure population density against an economic indicator that has a presumed relationship with happiness. We will be able to compare happiness, total population, and population density across countries by year. |

## Dataset 7: Inflation Data

|  |  |
| --- | --- |
| **Question** | **Information** |
| **Where did you find the dataset?** | International Monetary Fund (IMF) |
| **Why was the data created and for what purpose? Who collected the data?** | The data was created to measure Created to track year-over-year inflation across countries over time. IMF collected the data |
| **What is the timeline or lineage of the data?** | The data is from 1980 to 2028 (Forecasted values after 2022). |
| **How large is the dataset (cases, variables)?** | There are around 227 cases, and we are using 2 variables i.e., Country Name and YoY inflation (Individual by year). |
| **What locations are included in the dataset? Describe how the dataset will help you achieve your goals.** | The dataset compares the year over year inflation by country. This will help us to measure inflation against a social indicator that has a presumed relationship with happiness. We will be able to compare happiness and inflation rate across countries by year, seeking to find a correlation with happiness |

## Dataset 8: Education Inequality

|  |  |
| --- | --- |
| **Question** | **Information** |
| **Where did you find the dataset?** | We found the original data set on the UNDP website |
| **Why was the data created and for what purpose? Who collected the data?** | The data was created to evaluate inequality in educational opportunities across countries over time. UNDP collected the data. |
| **What is the timeline or lineage of the data?** | The data is from 2010-2021 |
| **How large is the dataset (cases, variables)?** | There are around 195 cases, and we are using 6 variables i.e., ISO-3, Country, Human Development Group, UNDP Developing Regions, HDI Rank, Inequality in Education (Columns for 2010 to 2021). |
| **What locations are included in the dataset? Describe how the dataset will help you achieve your goals.** | The dataset shows a measure of inequality in education by country from 2010 to 2021 ranging from 1 to 50 by country. 50 means high levels of inequality. This will help us to measure education inequality against an economic indicator that has a presumed relationship with happiness. We will be able to compare happiness and education inequality across countries by year. |

# Data Story

## Visualization 1

A map of the world

Description automatically generated

**Dataset Breakdown:**

* **Datasets:** World Happiness Report (Variables: Country, Happiness Index), Population (Variable: Population (In Millions)
* **Variables:**
  + Country (Name of Country), Average of Happiness Index, Sum of Population (In Millions)
* **Limitations:** The map only shows us the patterns of the average happiness index by country. It is hard to determine the influence of population on happiness when the heat map corresponds to only average index.

**Analysis:**

* **Goal Achievement**: This helps us to visualize the average index of each country, allowing us to determine which countries are happiest, and which are the unhappiest to focus our analysis on. Furthermore, it helps us to identify happiness by region, allowing us to focus our analysis of other variables on these countries/ regions to determine any correlation with their happiness/ unhappiness.
* **Chart Analysis:** From the map visualization, we can determine that Africa, the middle east, and India/ South Asia are the unhappiest regions by average happiness index. Most European and Western countries (including Australia and New Zealand) appear to be happier on average compared to the rest of the world.

## Visualization 2

A table with numbers and text

Description automatically generated

**Dataset Breakdown:**

* **Datasets:** World Happiness Report, GDP, Human Development Index, Inequality in Education, Inflation and Unemployment.
* **Variables:**
  + Country, Happiness Index Ranking (1=most happy), GDP Ranking (1=highest GDP), Human Development Index Ranking (1= Most developed education, life expectancy, etc), Education Inequality Ranking (1=Most unequal in educational attainment), Inflation Ranking (1 = Lowest inflation), Unemployment Ranking (1 = Highest unemployment).
  + The Unemployment Rate measures the number of eligible workers without employment. The Human Development Index assigns scores to countries based on educational attainment and life expectancy. Happiness Index Score is a subjective rating from 0-100 of citizen happiness levels. Educational Inequality measures variances in educational opportunities by various social groups in countries. Inflation measures the increase in prices across a country. GDP measures the total economic output of a country.
* **Limitations:** For some countries we have happiness index scores, but they may not report their data to the UN or IMF for indicators such as inflation, unemployment or their human development index.

**Analysis:**

* **Goal Achievement:** This crosstab report helps us to easily identify our countries of interest and their rankings across our various indicators.
* **Chart Analysis:** This table allows us to go country by country to draw conclusions about our various indicators and how they may influence happiness. For example, across Finland, Denmark, Iceland, Switzerland and Norway we can see that they have the highest happiness index rankings, and they are all also highly ranked under the human development index. This helps us with initial analysis to identify a possible relationship between life expectancy and educational attainment, and happiness. However, Iceland which is the third happiest country is ranked very low in GDP, inflation & unemployment ranking.
* **Note:** The inequality in education ranking scale is reversed, meaning the country with Rank 1 has the highest inequality in education among different groups of population.

## Visualization 3





**Dataset Breakdown:**

* **Datasets:** World Happiness Report
* **Variables:**
  + Country, Human Development Index Score
  + The Happiness Index Score is a subjective rating from 0-100 of citizen happiness levels.

**Analysis:**

* **Goal Achievement:** These graphs allow us to visualize the trends in happiness scores across time for the top and bottom 20 countries by Happiness Index Score. This helps us narrow in on countries with notable changes to their score so we can analyze which factors may impact these changes.
* **Chart Analysis:** Viewing the top and bottom 20 countries we take note of which countries see large increases or decreases in their Happiness Index Score.
* **Outliers:** 
  + The steady increase of Luxembourg’s rank may be attributed to steady economic growth, investment in infrastructure improvements, policies to improve work-life balance and promote gender equality etc
  + The sudden dip in UK’s rank in 2020 may be attributed to Brexit & resignation of Prime Minister Theresa May
  + The constant decline in Lebanon’s rank may be attributed to severe economic crisis from 2019
  + The sudden stability in Afghanistan’s rank post 2020 may be attributed to the agreement between United States and the Taliban for the withdrawal of U.S. and NATO forces.

## Visualization 4



**Dataset Breakdown:**

* **Datasets**: World Happiness Report, GDP
* **Variables:**
  + Country, Human Development Index Score, GDP
  + The Happiness Index Score is a subjective rating from 0-100 of citizen happiness levels. GDP measures the total economic output of a given country.

**Analysis:**

* **Goal Achievement:** This chart allows us to visualize the relationship between GDP and Happiness Index Score. Since most people equate money with happiness, this chart helps us speak to the layman in terms they will understand.
* **Chart Analysis:** The size of the bubbles represents a country’s GDP, and the color represents their Happiness Index Score. The larger the bubble the larger the GDP, and the greener the bubble the happier the country is based on their score. Here we can see a more affluent society leads to higher happiness when we view the US, Germany and the United Kingdom. However, other large economies such as China and India rate very low in terms of happiness. This indicates that although most of the large economies are happy, there are other factors that influence happiness which our analysis has shown in the form of Human Development, Educational Inequality and Unemployment as key indicators beyond GDP for driving up happiness levels.

## Visualization 5



**Dataset Breakdown:**

* **Datasets:** Inflation (Variables: Country, Inflation Rate)
* **Variables:**
  + Country (Country Name), Average Inflation Rate
  + The Average Inflation Rate is a measure of the overall average inflation by each country from 2018-2022

**Analysis:**

* **Goal Achievement:** This graph allows us to visualize the overall average inflation rate of the top and bottom 20 countries. This allows us to identify and focus in on certain countries when analyzing the overall impact of inflation on happiness.
* **Chart Analysis:** Overall, it appears African have the highest overall inflation by continent on average. On the other hand, it appears European countries have the lowest overall inflation by continent on average. The Middle East appears on both sides of the spectrum, with Lebanon and Yemen having two of the highest over average inflation rates and Jordan and Israel having two of the lowest.
* **Outliers:** 
  + USA has an average inflation rate of 3.64 and is a top 20 country in happiness index, while Tanzania which has an average inflation rate of 3.66 is a bottom 20 country in happiness index.
  + Zimbabwe has a history of hyperinflation, which has eroded the value of its currency. In the past, the country experienced hyperinflation rates so extreme that it led to the issuance of trillion-dollar banknotes.

## Visualization 6

A graph with colored dots

Description automatically generated

**Dataset Breakdown:**

* **Datasets:** World Happiness Index (Variables: Country, Happiness Index, Year), Political Stability (Variable: Political Stability)
* **Variables:** 
  + Country (Country Name), Year, Political Stability, Happiness Index
  + The chart allows you to filter by year (unshown). The chart also represents a different country by each color (unshown filter). The Happiness Index Score is a subjective rating from 0-100 of citizen happiness levels. The political stability score (0-100) by country (167).
* **Limitations:** While the graph does a good job showing an overall positive trend between happiness index and political stability score, it fails to explain the reasoning of the outliers. For example, Israel, the blue dot in the top left of the chart, has a high happiness index and low political stability, going against the trend.

**Analysis:**

* **Goal Achievement:** This graph allows us to visualize the happiness index score by each country versus political stability score. This allows us to see the overall seemingly positive trend of happiness compared to political stability. This chart also allows us to visualize each outlier individually, giving reason for further analysis into each country’s economic and social factors.
* **Chart Analysis:** The graph shows a seemingly overall positive relationship/ trend between happiness index and political stability. It seems the more stable a country is politically, the happier a country is on average. However, the outliers also allow us to determine that not all countries are a part of this trend and require further analysis into other variables.
* **Outliers:** 
  + Botswana is high on political stability because of its long history of democratic governance and a well-established and respected legal system. However, its low rank in happiness score may be attributed to problems like Income Inequality, HIV/AIDS Epidemic, Unemployment etc
  + Israel, on the other hand, ranks low on Political Stability due to conflicts with Palestine and the fact that five Knesset snap elections were held in a span of less than four years: from April 2019 to November 2022. However, Israel’s high rank on Happiness Index may be attributed to its strong tradition of upholding the rule of law, a vibrant civil society, including numerous advocacy groups, political parties, a free press and the fact that the country is self-sufficient on almost every front.

## Visualization 7





**Dataset Breakdown:**

* **Datasets:** World Happiness Report, Human Development Index, Inequality in Education, and Unemployment.
* **Variables:**
  + Country, Average Educational Inequality Score, Average Unemployment Rate, Average Human Development Index, Happiness Index Score.
  + The Unemployment Rate measures the number of eligible workers without employment. The Human Development Index assigns scores to countries based on educational attainment and life expectancy. Happiness Index Score is a subjective rating from 0-100 of citizen happiness levels. Educational Inequality measures variances in educational opportunities by various social groups in countries.
* **Limitations:** For some countries we have happiness index scores, but they may not report their data to the UN or IMF for indicators such as unemployment or educational attainment equality.

**Analysis:**

* **Goal Achievement:** These graphs allow us to visualize the top and bottom 20 countries by Happiness Index score. It also brings in Educational Inequality and Unemployment data for analysis.
* **Chart Analysis:** Viewing the top and bottom 20 countries we can see one clear relationship which is that educational inequality is much more prevalent in less happy countries than more happy countries. This indicates that educational inequality and access to equal opportunities across social groups could be a driving factor for increasing happiness scores.

# Conclusions

Before we started our analysis, we set out to understand the relationship between the happiness index and other social and economic factors. We also set out to compare these indicators across various countries. We did this with the hope of deriving valuable insights that could be used by policymakers to increase the happiness of their constituents.

Our findings indicate that overall, having a thriving economy with a high GDP does drive happiness but only to an extent. India and China are examples where this relationship is not perfect. Beyond GDP, we were able to identify educational inequality and unemployment as driving forces for happiness scores. The top 20 countries in happiness had significantly lower levels of educational inequality and boasted lower unemployment rates when compared to the 20 least happy countries. This indicates that opening education to more social groups in a country leads to a happier population. And the same goes for unemployment, where finding ways to provide jobs to citizens seems to increase happiness as indicated by our top 20 happiest countries.

Another important driver of happiness would be overall human development which considers educational attainment and life expectancy. The top 5 happiest countries also have the highest levels of human development as measured by the factors listed. This tells us that a citizenship that has high levels of educational attainment and who expect to live long lives are set up for higher levels of happiness.

Overall, we can conclude that while economic factors such as GDP and inflation might have an influential effect on happiness, socio-political factors also have a general effect, if not greater, than economic factors. This suggests that both economic factors, like GDP, and social factors, like educational inequality likely have an overall combined effect on happiness. This makes logical sense, as countries with better economies could provide better social services and more educational opportunities due to more resources and higher overall efficiency. However, this still does not explain the various outlying countries with high or low happiness scores compared to others with similar economic and social conditions. To better understand what truly influences happiness and to better explain these outlying countries, further research and analysis is required, such as regression analysis and the inclusion of other social, political and economic variables not included. Furthermore, further research into recent social and economic events is needed for each country, such as the impact of Brexit or the effect of certain conflicts within different regions of the world.

# References:

* *UK government web archive*. The National Archives. (2016). <https://webarchive.nationalarchives.gov.uk/ukgwa/20160105170225/http://www.ons.gov.uk/ons/guide-method/user-guidance/well-being/why-measure-well-being-/index.html>
* Stearns, P. N. (2012, January). *The history of happiness*. Harvard business review. <https://pubmed.ncbi.nlm.nih.gov/22299510/>
* Bertolucci, S. (2018, October 31). *Beyond GDP: Economics and Happiness – Berkeley Economic Review*. Berkley Economic Review. <https://econreview.berkeley.edu/beyond-gdp-economics-and-happiness/>
* Northwestern Medicine. (2015). *How happiness impacts health*. <https://www.nm.org/healthbeat/healthy-tips/how-happiness-impacts-health>
* World Happiness Report. (n.d.). *About*. World Happiness Report. <https://worldhappiness.report/about/>
* The Economist Newspaper. (2021). *The world in 2021*. The Economist. <https://www.economist.com/the-world-in-2021>

1. Contributions**:**

* Alexander Scarcelli: Introduction and Background, Data Story, Dataset Descriptions, Conclusions, References, Presentation assistance
* Avishek Dasgupta: Visualizations, Tableau Story, Dataset cleaning/preparation using Python scripts
* Hummarah Shahzad: Visualizations, Presentation Content and Design
* Nikolai Saporoschetz: Introduction and Background, Data Story, Dataset Descriptions, Conclusions, References, Presentation assistance
* Samarth Bansal: Dataset Information, Visualizations, Tableau Story, Dataset cleaning/preparation
* Siddhant Treasure: Visualizations, Majority of dataset cleaning/preparation, Tableau Story