Visualization Project Preliminary Report

Abisheka Pitumpe, Priscilla Danso

Abstract

The World Happiness Report is a landmark survey of the state of global happiness. It ranks countries based on their happiness levels, analyzing data from various sources, including GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, and perceptions of corruption. Our project aims to visualize this complex dataset to uncover patterns, trends, and insights into the factors influencing global happiness. By employing modern data visualization techniques, we aspire to make the findings accessible and engaging to a broader audience, including the younger generation, through innovative means such as interactive dashboards and social media-friendly visuals.

1 Background

The World Happiness Report [1] provides a comprehensive dataset that reflects the global state of happiness, offering an opportunity to explore how economic, social, and environmental factors converge to shape the well-being of nations. Identifying patterns within this data enhances our understanding of happiness and supports the development of policies aimed at improving the overall quality of life. Visualizing this data in an accessible and compelling way can bridge the gap between academic research and public awareness, fostering a deeper appreciation of the complex interplay of factors contributing to happiness around the world.

2 Problem

- Demonstrate available data on the world happiness index in a well-organized and understandable manner.

- Through the aid of powerful visualizations, highlight happiness scores by country, and age and view the variations of properties above through the years.

- Use modern visual means such as emojis and such social media-related visuals to reach younger, internet-age audiences.

- Provide a platform to create correlations between factors that affect the happiness index such as GDP, life expectancy, and population.

- Identify relationships(clusters) that are formed when considering similar happiness index scores.

- Explore the correlation between happiness and positive economic outcomes in a selected subset of the countries in the top of the world happiness index and in the bottom of the happiness index.

3 Approach

- Review existing visualizations created using the dataset provided by the World Happiness Index.

- Understand and extract useful attributes for our visualization.

- Prepare the dataset for initial analysis using data standardization, and normalization and handle possible missing values with the best imputation method.

- Choose the correct visualization methods to portray information.

- Build a dashboard containing all visualizations created

- Add dynamic interactivity for the user to select different countries, and ages and create user-interactive graphs and charts.

3.1 Dataset description

The dataset for our project is derived from the World Happiness Report, which compiles global happiness metrics by country, based on factors such as GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, and perceptions of corruption. The dataset spans several years, allowing for longitudinal analysis of trends and changes in happiness scores. Each entry in the dataset represents a country-year observation with associated happiness scores and factor scores. The data is comprehensive, covering over 150 countries, making it an extensive source for analysis. For this project, the dataset will be cleaned, standardized, and normalized to prepare it for analysis and visualization.

Table 1: Data preprocessing of data

Without Preprocessing		Assign Emojis		Impute Numeric Null values	
Country	0	Country	0	Country	0
Region	0	Region	0	Region	0
Happiness Rank	0	Happiness Rank	0	Happiness Rank	0
Ladder score	0	Ladder score	0	Ladder score	0
Economy	3	Economy	3	**Economy**	**0**
Social support	3	Social support	3	**Social support**	**0**
Health	4	Health	4	**Health**	**0**
Freedom	3	Freedom	3	**Freedom**	**0**
Trust	4	Trust	4	**Trust**	**0**
Generosity	3	Generosity	3	**Generosity**	**0**
Dystopia Residual	316	Dystopia Residual	316	**Dystopia Residual**	**0**
Year	0	Year	0	Year	0
Emoji	312	**Emoji**	**0**	Emoji	0

3.1.1 Preliminary Dataset Analysis

- To get started with this project, we're using data from the World Happiness Report spanning from 2015 to 2024. Each year's data is available in it's own CSV file. After downloading the datasets, the next step is to consolidate these ten datasets into one central CSV file, pooling together all the valuable information.

- Upon delving into the datasets, we've noticed a bit of a naming conundrum with attributes varying inconsistently across the years. Take GDP for instance, which goes by different monikers like "GDP per Capita" or "Economy (GDP per capita)." Thus, it's imperative to standardize these attribute names across all datasets before diving into deeper analysis.

- We've devised a fun way to depict the Happiness score using emojis based on their numerical values. Scores falling between 6 and 8 get adorned with a "big happy smile" emoji, while those ranging from 4.5 to just below 6 receive a different emoji treatment. Anything from 1 up to 4.5 gets represented with a sad or angry emoji.

- Wrangling the data involved some tidying up, particularly with numeric missing values. We opted to fill in these gaps by imputing the median for "Dystopia Residual" and the mean for all other columns. The specifics of this processing, including the number of null values addressed, are outlined in Table 1.

3.2 Preliminary Work

The charts and graphs below are currently under development and will be integrated to construct the final dashboard.
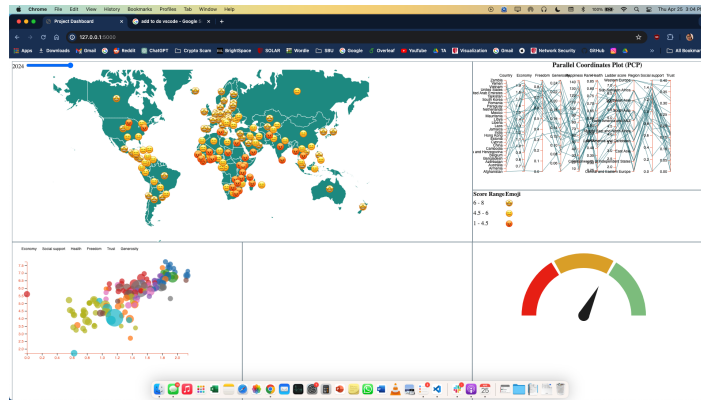

Figure 1: Current Dashboard

- **World Map:** A primary component of the dashboard, the world map summarizes the World Happiness Report data, using emojis to represent happiness scores. These emojis reflect common expressions of happiness, neutrality, and anger.
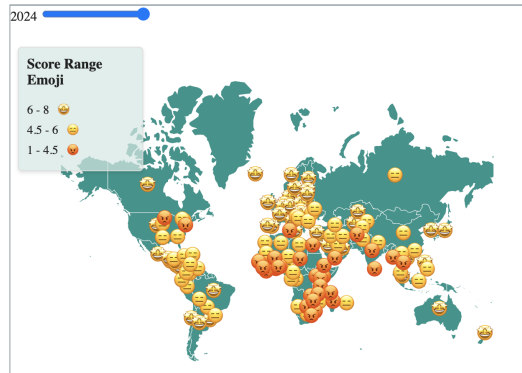

Figure 2: World Map

- **Parallel Coordinates Chart:** This chart demonstrates patterns across different dimensions of the dataset. It includes functionalities for brushing and rearranging, and we plan to add color highlighting upon user interaction.
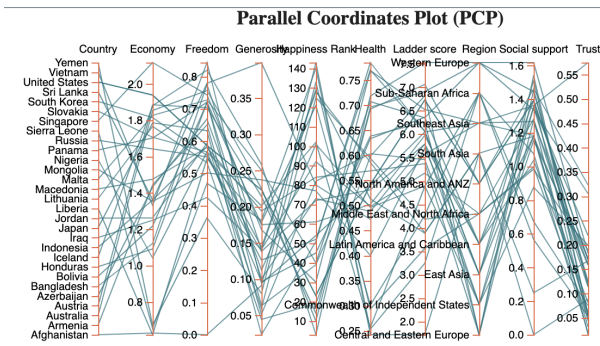
Figure 3: Parallel Coordinates Chart

- **Bubble Chart:** This chart displays various metrics such as scores for economy, freedom, and trust from the World Happiness Report dataset. The bubbles are grouped and colored by continent, illustrating the relationship between happiness and other factors. Future enhancements will allow the chart to update interactively based on selections made on the world map.
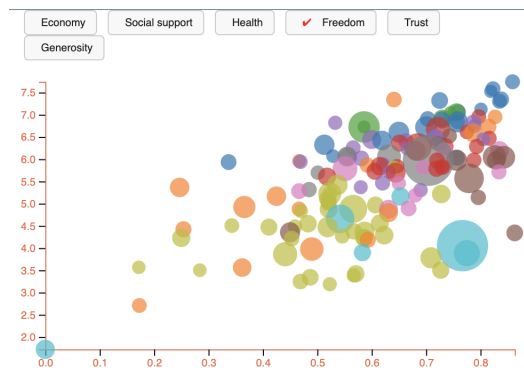

Figure 4: Bubble Chart

- **Gauge:** This visualization will indicate the happiness score associated with a selected country, represented by a color on the gauge.

Figure 5: Gauge

We are also planning to add a horizontal bar chart and a stacked line graph to further enrich the dashboard's storytelling capabilities.

4 Conclusion and Next Steps

- Refine the existing charts/graphs and implement two new visualizations.
- Design a visually appealing and organized dashboard.
- Enhance user interactions with the graphs and integrate these interactions across different visual components.
- Fully utilize all attributes of the dataset to effectively communicate insights through the dashboard.

References

[1] World Happiness Report. World Happiness Report Data Dashboard — worldhappiness.report. https://worldhappiness.report/data/, 2024. [Accessed 15-04-2024].