Visual Analytics V24 Group 8

Group Member: Magnus Tomter Ouren (Matricola: 2181751)
Github link: https://github.com/magnusouren/world-happiness-report

Introduction

Happiness is a complex and multi-dimensional concept that reflects the well-being of individuals and societies. The World Happiness Report serves as a comprehensive dataset that captures various factors influencing happiness levels across countries over a significant time span (2006–2023). The diverse nature of the dataset, coupled with its global reach, makes it an invaluable resource for understanding how different regions of the world have developed and how key global events have shaped happiness over the years.

This project leverages the World Happiness Report to explore and visualize global happiness trends. The dataset provides a rich structure, with columns representing a wide array of influential factors such as GDP per capita, social support, life expectancy, and perceptions of corruption. These variables offer a multi-faceted view of happiness, enabling an in-depth analysis of patterns, outliers, and clusters across countries and continents.

The primary motivation for choosing this dataset lies in its potential to reveal significant differences between countries and regions, providing an opportunity to examine how diverse factors influence happiness. By analyzing a dataset with a large time span, the project aims to uncover how situations have impacted different parts of the world.

The main goals of the project include:

- Comparing happiness scores across continents to identify regional patterns.
- Understanding the underlying reasons behind the scores of high-performing and low-performing countries.
- Visualizing happiness scores on an interactive map to highlight global trends and outliers.
- Using Principal Component Analysis (PCA) to reduce dimensionality, allowing for simplified comparisons between countries and identifying clusters or anomalies.

Through interactive visualizations and analytical techniques, this project offers insights into the factors driving happiness globally and provides a platform for further exploration and decision-making by stakeholders such as policymakers and researchers.

The intended users of the application are researchers, policymakers, and data enthusiasts interested in exploring global happiness trends and understanding the factors influencing them. The tool is designed to provide an intuitive and interactive platform for analyzing regional patterns, correlations, and anomalies in happiness data over time.

Related Work

The official World Happiness Report website (www.worldhappiness.report/data) provides a visualization tool for exploring happiness scores across countries. While the tool is visually appealing and informative, it offers limited functionality for understanding the underlying reasons behind the scores. This limitation inspired me to create a visualization system that not only highlights global trends and clusters but also provides deeper insights into how various factors influence happiness.

A simple Google search revealed several projects that visualize similar datasets. Many of these projects utilize maps with color-coded values to represent happiness scores across regions, a technique that I found particularly effective for understanding how continents and areas perform relative to each other. Other visualizations focused on ranking the top- and bottom-performing countries and others is using scatterplots to explore relationships between specific parameters.

The primary motivation was to combine these approaches into a more comprehensive tool. By integrating maps and scatterplots, the aim was to provide a platform where users could explore both the geographic distribution of happiness scores and the relationships between factors affecting those scores. Furthermore, the visualization should be customizable, allowing users to select specific parameters of interest for deeper analysis. For example, users can compare countries on a color-coded map while simultaneously analyzing correlations between parameters such as GDP per capita, life expectancy, and social support through scatterplots.

This combination of techniques allows for a more intuitive and detailed exploration of the data, bridging the gap between existing visualization tools and the need for a better understanding of global happiness trends and their driving factors.

Data and Methods

The dataset used in this project was sourced from Kaggle: World Happiness Report (2006–2024). It contains yearly happiness data for countries worldwide, offering 2358 rows and 9 key columns. These columns include:

- **Country name**: The name of the country.
- **Year**: The year of data collection.
- **Life Ladder**: A subjective measure of happiness.
- **Log GDP per capita**: A measure of economic performance.
- **Social support**: A measure of societal support.
- **Healthy life expectancy at birth**: A proxy for public health.
- **Freedom to make life choices**: A measure of individual freedom.
- **Generosity**: A measure of societal generosity.
- **Perceptions of corruption**: A proxy for trust in institutions.
- **Positive affect** and **Negative affect**: Indicators of emotional states.

The dataset spans multiple years and countries, making it a resource for analyzing global happiness trends. However, some missing values and rows posed challenges. For instance, when entire rows for certain years were missing, no imputation was applied. For individual missing values in numeric columns, the mean of the respective column was used to fill gaps. This approach ensured minimal data loss while maintaining consistency. To enhance usability, a new column indicating the continent for each country was added. This labeling makes colorization by continent possible.

Preprocessing

The numerical values in the dataset were normalized using z-score normalization (Gaussian scaling). This step standardized the data, making variables with different scales comparable, which is particularly important for clustering and dimensionality reduction techniques. The normalized values were appended as additional columns rather than replacing the original data, preserving the integrity of the dataset.

Dimensionality Reduction

Principal Component Analysis (PCA) was applied to reduce the dataset's dimensionality and enable effective comparison between countries. By reducing the 6 numerical normalized variables to 2 principal components, PCA highlighted clusters and outliers in the data while retaining most of the variance. This approach is effective for high-dimensional datasets like this, as it simplifies analysis and visualization without sacrificing critical information.

Visualizations

The visual analytics application incorporates:

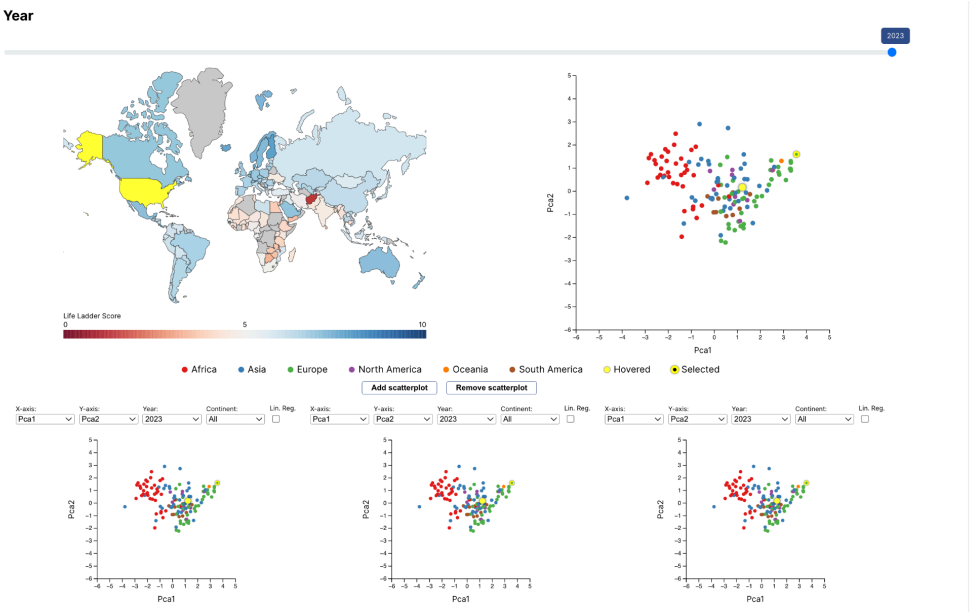
1. **Map-based Visualization**: Countries are color-coded based on their happiness scores. This enables a quick understanding of global trends and regional differences. The scaling used in the map is the diverging scale *interpolateRdBu* from d3.js (https://d3js.org/d3-scale-chromatic/diverging). The scale domain is set to [1,8], based on the rounded global minimum and maximum values. This adjustment enhances color differentiation on the map, making variations in happiness scores more visually distinct.
2. **Scatterplots**: Users can select any two parameters (e.g., GDP per capita and life expectancy) to explore correlations and trends. User can add as many scatterplots as it wants be clicking on a button. At the bottom a big scatterplot is added for more detailed analysis on a singular set of columns. All countries are colorized by the continent they belongs by using the categorical colorscheme *schemeSet1* by d3.js. (https://d3js.org/d3-scale-chromatic/categorical)
3. **PCA Visualization**: A scatterplot of PCA results shows clusters and anomalies among countries for the given year.
4. **Table:** A table listing all columns for a selected country for deeper understanding of all data belonging to a country for a given year.

All interaction with a visualization is shown in the other visualizations simultaneously. When holding the mouse over a country in the map or a scattersplot, it is highlighted in yellow in all

visualizations. When clicking on a country it's highlighted in yellow in the scatterplot, and added to the table at the bottom. Filtering by year provides a temporal perspective, making it easier to study trends and analyze the effects of specific historical events.

System Design and Implementation

The application is an interactive visual analytics platform designed to explore global happiness trends using the World Happiness Report dataset. It integrates map-based visualizations, scatterplots, and dimensionality reduction to allow users to compare happiness scores and analyze relationships between factors.

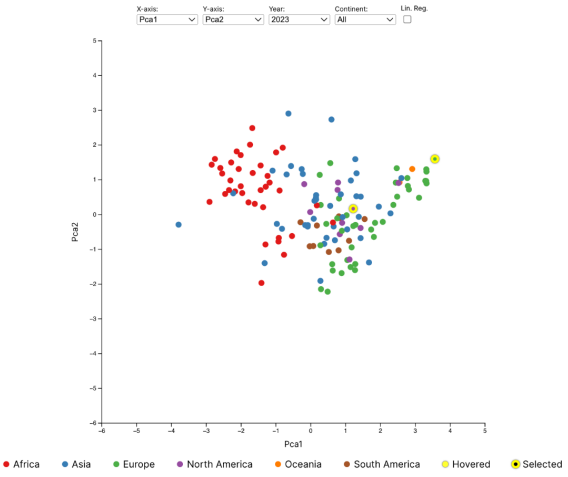


Selected Countries

Table showing the data for the selected countries for the selected year.

Country	Year	Continent	Life Ladder Score	GDP	Social Support	Healthy Life Expectancy	Freedom	Generosity	Corruption	Positive Affect	Negative Affect
Norway	2023	Europe	7.25	11.13	0.95	72	0.94	0.22	0.24	0.76	0.23
United States of America	2023	North America	6.52	11.09	0.86	66	0.72	0.18	0.72	0.71	0.28

Show Descriptions



User Interface Design

1. **Map Visualization**:
   A color-coded world map displays the "Life Ladder" score, highlighting global trends and regional differences. Users can hover over countries for detailed values and trends.
2. **Scatterplots**:
   One scatterplot with two-dimensional PCA for comparison of countries and clustering of continents. Interactive scatterplots allow users to select variables for the X and Y axes, filter by year or continent, and enable linear regression to explore relationships. An additional big scatterplot was added to the bottom of the page, making more detailed lookups available.
3. **Controls**:
   A year slider and dropdown menus for continent, year and axis selection provide filtering and customization.

Implementation

The application is built using Vite.js, a fast development environment for modern web applications, and React for creating a dynamic and responsive user interface. Visual components are implemented with D3.js, enabling efficient and interactive data visualizations. Data preprocessing is performed in Python, where the initial .csv dataset is cleaned, normalized, and converted into a JSON-formatted file to streamline integration with JavaScript and ensure smooth data handling in the application.

Insights and Discussion

The application provides several key insights into global and regional happiness trends, as well as the relationships between various factors influencing happiness. These insights demonstrate the tool's effectiveness in identifying patterns, correlations, and anomalies within the dataset.

Major Insights

The clustering of countries by continent reveals significant similarities among nations within the same region, underscoring shared socio-economic and cultural influences on happiness. This clustering helps identify both *global outliers*, such as Afghanistan in recent years, and *regional anomalies*, such as Greece and Ukraine in Europe. Afghanistan's low scores from the recent years highlight its significant challenges, with notable disparities across nearly all parameters. Haiti also stands out as an outlier in earlier years, with an extremely low "Healthy Life Expectancy" of only 7 years, contrasting with other low-performing countries where the average ranges between 50 and 60 years.

Linear regression analysis highlights strong positive correlations between the "Life Ladder" score and parameters such as GDP per capita, social support, and life expectancy. This clearly shows that improved socio-economic and health factors are associated with greater happiness. Interestingly, "Generosity" displayed no significant correlation, as evidenced by a

flat regression line, raising questions about its impact on perceived happiness. Conversely, "Negative Affect" and "Perceptions of Corruption" exhibit a negative correlation, indicating that higher levels of negativity and perceived corruption detract from happiness.

Local Anomalies

In Europe, notable outliers include *Greece*, which has lower scores throughout the dataset, potentially linked to economic instability, and *Ukraine*, whose scores in recent years have been severely impacted by the war against Russia. These anomalies provide valuable insights into how regional crises affect happiness levels.

Discussion and Applications

The interactive features of the application, such as filtering by year and continent and customizing scatterplot axes, were instrumental in uncovering these insights. The ability to visualize clusters and outliers at both global and regional levels enables users to identify patterns and correlations with ease.

Users can use these findings to address specific challenges in low-performing countries. For instance, prioritizing improvements in health, economic stability, and institutional trust could significantly enhance happiness scores in these regions. Additionally, the flat regression for "Generosity" suggests the need for further exploration into why this parameter does not correlate with happiness as strongly as others.

The ease of spotting outliers is a strength of this tool. This makes it easy to spot significantly strong and bad-performing countries, and then the possibility to analyze how these outliers perform within other areas.

Conclusion

This project successfully demonstrates the power of visual analytics in exploring and understanding complex global datasets such as the World Happiness Report. By combining interactive visualizations with dimensionality reduction techniques, the tool provides valuable insights into happiness trends and their underlying factors. These findings highlight opportunities for further research and practical applications to address disparities and improve global well-being.

Future Work

For further improvements to the application, it would be interesting to implement graphical views making it easier to easier compare different countries and their data. Here a line chart with all parameters and their development throughout the years could be an interesting approach. Furthermore, more work on the map and the possibility to use zoom should improve the usability of the tool, making it easier to find the smaller countries. As it is now, it’s only the ladder score that sets the colors on the map. An interesting feature here should be the possibility to select parameter(s) to set the color value, more like the functionality in the scatter plots.

References

- World Happiness Report. (n.d.). *World Happiness Report Data*. Retrieved January 11, 2025, from https://worldhappiness.report/data/