NOVA IMS Information Management School

# Data Visualization

## MASTER'S DEGREE PROGRAM IN DATA SCIENCE AND ADVANCED ANALYTICS

# Final Project

Amanda França, number: 20220708

Nevena Cukrov, number: 20221373

Virgínia Aguiar, number: 20220707

April 2023

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

## 1. Introduction

The previous year, 2022, marked the 10th anniversary of the *World Happiness Report* (WHR) - a widely recognized and celebrated annual publication that ranks countries based on their average happiness levels - and of the proclamation of 20 March to be observed as the International Day of Happiness by the United Nations General Assembly (Helliwell et al., 2022). Since then, happiness has become an important indicator of a country's development and success (Bris, 2022). To further explore this topic we developed a data visualization project on the World Happiness Report 2023 as described in this document.

Using data from the Gallup World Poll (GWP) surveys, the publication ranks countries based on six different attributes: GDP per Capita, Social Support, Healthy Life Expectancy, Freedom to Make Life Choices, Generosity, and Perception of Corruption (Helliwell et al., 2023).

The Schneiderman's Mantra "Overview first, zoom and filter, details on demand" aligns with the Martini Glass structure of the visualization project (Fitzgerald, 2016). It provides an overview of the data followed by more detailed visualizations that offer deeper insights. With a mix of reader-driven and author-driven approaches, the project offers interactivity within the context of a structured narrative, from which readers can gain a better understanding of the factors that contribute to a country's happiness and the changes that have occurred over the years.

## 2. Data Methods

### *2.1. Dataset Description*

For the development of this visualization project, the public data was retrieved from the WHR 2023. The final cleaned and preprocessed dataset had 10 variables covering the years from 2015 to 2022 in 156 countries, as described in Table 1 in the Appendix section.

### *2.2. Technical aspects*

To develop a dynamic, web-based dashboard that allowed for interactive data exploration we employed the open-source library Dash, powered by Python using Pycharm as the integrated development environment (IDE) for its implementation. The Plotly library was leveraged to generate data visualizations. The dashboard is publicly accessible on GitHub. Link to the GitHub repository can be found in the Appendix section.

## 3. Visualizations and Interactions

The final dashboard is a comprehensive display of 10 visualizations and 5 interactive filters, segmented into two parts. The first part provides a high-level overview of the present state and previous years' trends, while the second part lays out more detailed drill-down options.

### *3.1. First Part - Overview*

At the outset of the first part, users can view cards that showcase the happiest and least happy country of 2022 along with their respective *Life Ladder* scores. Below the cards, a Choropleth Map based on the happiness score for each country in 2022 presents a quick and current worldwide assessment. By hovering over the map, the country's name and *Life Ladder* score appear. Then, a dropdown filter enables users to select specific countries to analyze through the remaining visualizations. Towards the end of the first part, a line chart with Life Ladder scores from 2015 to 2022 for the chosen countries is displayed.

### *3.2. Second Part - Details on Demand*

The second part replicates the visualizations and filters side by side, facilitating a comparison of the selected countries in two different years that can be selected from the dropdown filter at the top. Following the calendar year selections, the tables present the rankings and the 6 factors for each of the chosen countries and years. The scatter plots present an interactive y-axis that displays the selected attribute in the dropdown filter while the x-axis portrays the *Life Ladder* score. On hovering over the scatter plot, the country name and *Life Ladder* appear.

## 4. Results - Reading the Visualization

### *4.1. Encoding*

<u>Score cards</u>

The score cards encode the categorical attribute Country Name and the quantitative attribute Life Ladder using the numbers of the Life Ladder score displayed in score card format as marks.

The Color Hue channel encodes whether the categorical attribute Country Name has the lowest Life Ladder score (representing the saddest country) or the highest (representing the happiest country).

We took into consideration that the Color Hue channel is very effective for encoding categorical attributes.

Choropleth

The Choropleth world map encodes the categorical attribute Country Name and the quantitative attribute Life Ladder using area marks, represented by polygons corresponding to different countries around the world. This is achieved using the Spatial Region and Color Hue channels.

The color hue channel, through the color scale chosen, with lighter colors encodes higher values of the quantitative attribute, Life Ladder (happy country) and darker colors encodes lower values (sad country).

The Spatial Region channel allows users to quickly understand the general spatial pattern of Life Ladder values across the world.

For categorical attributes the channels Spatial region and Color Hue are considered the most effective.

Line chart

The Line chart encodes one categorical attribute, Country name, one quantitative attribute, Life Ladder, and one ordinal attribute, Year.

The line marks encode the quantitative attribute, Life Ladder, over time for each country through the range of the ordinal attribute, year.

The Line chart uses the position channel, the horizontal position (x- axis) encodes the ordinal attribute, year, while the vertical position (y-axis) encodes the quantitative attribute Life Ladder score.

In addition to the position channel, the Line chart uses the Color hue channel to differentiate between multiple line marks for up to five countries on the same chart. This allows users to compare the Life Ladder score trends for different countries.

In terms of Effectiveness principle, the position channel is considered the most effective for quantitative attributes.

Table

The Table encodes the data in a structured format using rows and columns. Each row represents an item, which is a data point for a country (max 5 countries), while each column represents seven quantitative attributes, Life Ladder, Log GDP per Capita, Social Support, Healthy Life Expectancy, Freedom to Life Choices, Generosity and Perception of Corruption.

The marks are the cells that contain the individual data values for the intersection of rows and columns representing the attribute measured.

The channels used are the position channel to encode the location of each cell within the table, the row and column headers give the semantics of items (Country name) and attributes (Life Ladder, Log GDP per Capita, Social Support, Healthy Life Expectancy, Freedom to Life Choices, Generosity and Perception of Corruption) for data interpretation.

Scatter plot

The scatter plot encodes the two quantitative attributes using point marks. Each point on the scatter plot represents a pair of values for two quantitative attributes, with one attribute selected from six possible options (Log GDP per Capita, Social Support, Healthy Life Expectancy, Freedom to Life Choices, Generosity, Perception of Corruption) and the other attribute being Life Ladder.

The data for each point is previously filtered by the categorical attribute Country name and the ordinal attribute Year using the dropdown above the scatter plot.

The position channel is used to encode the data, with one vertical spatial position for one quantitative attribute from the six options (Log GDP per Capita, Social Support, Healthy Life Expectancy, Freedom to Life Choices, Generosity, Perception of Corruption) and one horizontal spatial position for the quantitative attribute Life Ladder.

In addition to the position channel, the Color hue channel is used to encode the quantitative attribute Log GDP per Capita, allowing for differentiation between selected countries. The color scale used represents higher values of Log GDP per Capita with lighter colors and lower values with darker colors.

### *4.2. Filtering*

In the first part of the dashboard, located above the Line Chart, there is a dropdown that allows the user to filter the categorical attribute "Country name". We recommend selecting up to 5 countries for better chart readability. This selection determines the countries displayed in the Line Chart, as well as the left and right side tables and scatter plots.

The second part enables users to explore and compare two years. There are two dropdowns located on the left and right sides, which allow users to select a year for comparison. This year dropdown filters the data by year for the left and right side tables and scatter plots.

In the bottom of the second part of the dashboard features a scatter plot, which includes a dropdown for selecting one factor out of the six available factors (Log GDP per Capita, Social Support, Healthy Life Expectancy, Freedom to Life Choices, Generosity, Perception of Corruption) that explain the Life Ladder Score.

## 5. Discussion and Conclusion

In conclusion, our data visualization project aimed to provide insights into the world's happiness by analyzing data from the World Happiness Report 2023. We created a dashboard that allows users to explore how happiness has changed over the years, the factors that contribute to happiness, and how different countries compare in terms of happiness.

Overall, we believe that we have achieved our objective of creating a dashboard that communicates information to users and tells stories through data. By following a Martini Glass structure along with Schneiderman's Mantra, the dashboard provides an overview of the data followed by more detailed visualizations that offer deeper insights. The use of interactive filters and visualizations allows users to explore the data on their own and gain a better understanding of the factors that contribute to happiness.

However, there are limitations to the project. For example, the data is limited to the years and countries included in the World Happiness Report, which allowed us to make a dashboard with a limited scope. Additionally, the data is self-reported and may be subject to biases, and errors.

Future work could include incorporating external datasets into the dashboard to explore the impact of events on happiness, such as natural disasters and conflicts.

## 6. References

Bris, A. (2022, January 10). Happy people, happy economy? moving beyond GDP in measuring economic success in 2022. I by IMD. Retrieved April 7, 2023, from https://www.imd.org/ibyimd/finance/happy-people-happy-economy-moving-beyond-gdp-in-measuring-economic-success-in-2022

Fitzgerald, B. (2016, February 26). Schneiderman's Mantra. Hamp Data Visualization. Retrieved April 7, 2023, from https://hampdatavisualization.wordpress.com/2016/02/26/schneidermans-mantra/

Helliwell, J. F., Layard, R., Sachs, J. D., Neve, J.-E. D., Aknin, L. B., & Wang, S. (2022, March 18). World happiness report 2022. The World Happiness Report. Retrieved April 7, 2023, from https://worldhappiness.report/ed/2022/

Helliwell, J. F., Layard, R., Sachs, J. D., Neve, J.-E. D., Aknin, L. B., & Wang, S. (2023, March 20). World happiness report 2023. The World Happiness Report. Retrieved April 7, 2023, from https://worldhappiness.report/ed/2023/

World Happiness Report. (n.d.). FAQ. World Happiness Report FAQ. Retrieved April 7, 2023, from https://worldhappiness.report/faq/

## 7. Appendix

### 7. 1 Table 1 - Dataset description

| Name | Category | Description |
|---|---|---|
| Country Name | Categorical | Country's name following the GeoJson format |
| Year | Ordinal | Year of country's data |
| Life Ladder | Quantitative | Final score accounting for the six factors evaluated in the WHR |
| Log GDP per Capita | Quantitative | How much each country produces is divided by the number of people in the country. It is adjusted for Purchasing Power Parity and presented in constant 2017 international dollars. The natural log of it is used as it fits the data better |
| Social Support | Quantitative | The national average of the binary responses to the GWP question "If you were in trouble, do you have relatives or friends you can count on to help you whenever you need them, or not?" |
| Healthy Life Expectancy | Quantitative | Time series for healthy life expectancy at birth that is built based on data from the World Health Organization Global Health Observatory data repository and takes into account physical and mental health |
| Freedom to Life Choices | Quantitative | The country's average of binary responses to the GWP question "Are you satisfied or dissatisfied with your freedom to choose what you do with your life?" |
| Generosity | Quantitative | The residual of regressing the national mean of GWP responses to the donation question "Have you donated money to a charity in the past month?" |

| | | |
|---|---|---|
| | | on log GDP per capita |
| Perception of Corruption | Quantitative | The mean of binary answers to two GWP questions: “Is corruption widespread throughout the government in this country or not?” and “Is corruption widespread within businesses in this country or not?” |
| Ranking | Quantitative | Country’s position regarding the final Life Ladder score |

### 7. 2 Github Link

https://github.com/CukrovNevena/DataVisualization