Visual Data Analysis

UNIVERSITY OF SKÖVDE

Visualization for decision-making and exploration

EBRAHIM GOLRIZ

GROUP 5: LORENZ KOCH, DALIBOR MIRKOVIC, EBRAHIM GOLRIZ

Visual Data Analysis report
Spring term 2025

Table of Contents

1. Introduction

1.1 One research question

The research question that was chosen for this project is "*Which factors have the highest influence on the happiness score across different regions for the period 2005-2022?*". The project focuses on evaluating the effectiveness of different visualization techniques used to answer this question.

1.2 Motivation and aim

The goal of the project is to effectively visualize the multi-dimensional World Happiness Report dataset to determine which factors most influence happiness scores across different regions over nearly two decades, as simply presenting raw data fails to convey the complex relationships and varying regional importance of these factors. Visualization is essential to translate these numerical relationships into accessible visual patterns, but it faces challenges in representing multiple variables, showing the strength of influence, enabling regional comparisons, avoiding clutter, and ensuring interpretability for people without a technical background. This project aims to compare visualization techniques like radar chart and heatmaps to identify the most suitable approach for revealing and comparing the relative influence of various factors on happiness across different geographical regions based on clarity and user understanding.

2. Data

The dataset used for this analysis is the World Happiness Report, 2005-Present, sourced from Kaggle. This dataset aggregates annual data for numerous countries from 2005 to 2022, primarily based on the Gallup World Poll and supplemented by data from organizations like the World Bank.

The core of the analysis focuses on understanding the factors influencing the Life Ladder score (which we reference as *Happiness Score* across all the visualizations), which serves as the dependent variable representing national happiness or life evaluation (measured on a 0-10 scale via the Cantril ladder question). To answer our research question, we focused on the regions included in this dataset.

The features that were used for analysis: GDP Per Capita, Social Support, Healthy Life Expectancy at Birth, Freedom to Make Life Choices, Generosity, Perceptions of Corruption, Positive Affect, Negative Affect, Confidence in National Government.

3. Evaluation

To evaluate the effectiveness and usability of the three developed visualizations (two in Tableau, one in Power BI) in addressing the research question, a user evaluation was conducted with 10 participants using a task-based survey method. Participants were asked to interact with the dashboards to identify the factors most influential on happiness scores across Western Europe and Southeast Asia between 2005 and 2022. These two regions were selected as representative examples to provide users with a specific task. Following this task, they completed an anonymous questionnaire based on a modified System Usability Scale (SUS). This survey included eight quantitative questions rated on a scale (1-10, Strongly disagree to Strongly agree), this measures the ease of information retrieval, future use intention, navigation intuitiveness, information sufficiency, user confidence, design efficiency, terminology clarity, and visual appeal, alongside one open-ended question that they can give suggestions for usability enhancements. The primary aim was to gather empirical feedback on how well the visualizations supported the analytical task. the collected results (detailed in Section 4) provide insights into overall usability and potential areas for improvement for each visualization.

4. Visualization comparison and evaluation results

The selection process, comparison of the four visualizations, and evaluation results are presented in Table 1.

4.1 Justification of the final selected visualization technique

The chosen visualization effectively addresses the research question, it allows the users to choose the regions they want to compare, and with the use of radar chart, they can clearly see the relative influence of each factor to the happiness score of each region. The influence of each factor being a percentage helps users to easily compare different regions together and to compare how each factor impacts the happiness in each region. This approach effectively addresses the multidimensional nature of the data and presents it in an easily understandable format. To ensure fair comparison, all factor data has been normalized to eliminate the influence of varying scales on their perceived impact.

As discussed in Lecture *"Information Visualization (1)",* a primary goal of using visualization instead of providing raw data, is to present large volume of data in a compact way so users can interact with the visualization and get the desired information. In contrast to the selected visualization, fig. 2 and fig. 3 present data in a less compact manner, the users are presented with the information of all regions simultaneously and they need to find the regions and obtain the information themselves; while the selected visualization lets the users choose the regions and only shows them the information for those regions so they can easily compare those and get the insight they are looking for. This is especially the case for visualization 3 (fig. 3) which only has a heatmap with all the information already inside it, this may reduce user engagement and interactivity.

As for pop-out effect which was mentioned in Lecture *"Human Perception and Cognition"*, visualization 2(fig. 2) fails at showcasing the information required for the user to answer the question effectively. Many of the bubbles that are representing the influence of each factor have the same or near identical size that makes it difficult to see which factors are actually more impactful, the lack of a sorting feature also makes it difficult to easily see which factors have the biggest impact. In comparison, the selected visualization makes use of radar chart which effectively conveys the relative influence of each factor and makes it easy to compare different regions.

The selected visualization got the highest overall score from users, and particularly received a very high score regarding user interactivity and clarity of the presented information, which boosted user confidence in using the system.

5. Discussion

The source dataset can best be characterized as a multidimensional table. Units of the original table are country-year observations. Key attributes are Country (Categorical), Year (Quantitative, Ordered), and Region (Categorical), along with numerous quantitative value attributes like Life Ladder score and its potential drivers (GDP, Social Support, etc.). The dataset is available, stable from 2005 to 2022. For the purpose of this visualization, this was transformed into a derived table where the items

represent Regions (Categorical Key), with attributes for Factors (Categorical Key) and their calculated Relative Influence (Quantitative Value).

To answer the research question, relative influence of all the factors on happiness score should be considered. The information obtained must be conveyed in such a way that is both easy to understand and comprehend without the need to get into detail and spend time figuring out the results. Also, that information must be presented in a way that lets us compare different regions together easily. while this makes it a necessity to use visualization techniques to clearly answer the research question, it results in a challenging project as different dimensions of information must be considered to have a visualization that is effective.

The selected visualization addresses many of the challenges and tries to present the data to users in an interactive and easy to understand manner. It utilizes linking between different charts to convey the information more intuitively as mentioned in Lecture *"Information Visualization (1)"*. One weakness of the radar chart used in the selected visualization is that it becomes completely uninterpretable if many regions are selected simultaneously; in that case users are forced to use the less intuitive bar chart to get their desired insights.

The radar chart uses colored shapes for each region. The different factors are shown around the circle. How far a shape goes out from the middle shows how much that factor influences happiness for that region. Different colors are used for different regions. The bar chart uses line marks(bars) to show how much each factor influences happiness. The length of the bar shows how strong the influence is. The bars are grouped by region, and then by factor. Vertical position is used to separate both the categorical Factors (within regions) and the categorical Regions themselves. Different colors, corresponding to the legend, are used to show the different factors. Coordination between views relies on linked highlighting based on the selected regions in the filter panel. Distinct color hues are used to differentiate between the selected regions on the radar chart. In the bar chart, distinct color hues are used to differentiate between the various happiness factors, as indicated by the legend.

According to the insights obtained from the selected visualization, Log GDP Per Capita is an extremely influential variable in most parts of the world, especially in Eastern & Central Europe (35% relative influence) and East Asia (38%). It is also highly influential in Latin America & Caribbean (28%), Sub-Saharan Africa (24%), and Southeast Asia (25%). Its influence is, however, insignificant in Western Europe (2%) and North America/ANZ (4%). In other parts of the world, other variables dominate. Social Support is also most crucial in the Commonwealth of Independent States (22%) and Southeast Asia (22%). Affective variables are also most critical, where Positive Affect weighs most heavily in Latin America & Caribbean (20%) and the Middle East & North Africa (21%). Surprisingly, Western Europe's happiness scores seem most shaped by those variables traditionally regarded negatively or linked to government: Negative Affect (25%) and Perceptions of Corruption (24%). North America and ANZ point to a grouping of key factors like Healthy Life Expectancy (20%), Positive Affect (19%), Social Support (18%), and Perceptions of Corruption (18%). These regional differences highlight the fact that economic well-being is likely to be robust, but not perhaps everywhere; other determinants associated with social cohesion, good health, positive feelings, and governance could be as effective or even more so depending on regions.

A larger dataset with more regions would make the radar chart too cluttered. Potential solutions include grouping regions or letting users filter to specific areas. For very complex data with many factors, the current chart might be too much; so, we need simpler views or ways to reduce the number of factors shown, for instance, by using clustering.

One interesting finding during the process of visual analysis was the significant difference in the insights we would gain by using different methods for getting the influence values of each factor. For the selected visualization, the data was first grouped by region, and for each region, a linear regression model was built using standardized features to calculate standardized coefficients that account for different scales. The relative importance of each factor was then computed as a percentage of the total absolute impact. This approach, which uses regression modeling based on all features, differs from relying on pairwise Pearson Coefficients between a single feature and the happiness score at a time; while Pearson Coefficient is simple and intuitive for identifying pairwise relationships, regression considers multiple factors simultaneously and accounts for interdependencies between variables. This makes it a better option for answering the research question, therefore it would be the more suitable approach for next similar projects.

The insights that can be gained from this visualization (given we use a richer dataset), would allow us to understand what significantly impacts happiness in different regions. This understanding is crucial for governments and organizations to make informed policy decisions. By identifying key happiness drivers in specific areas, they can better allocate resources and design effective interventions to improve well-being. For example, if strong social support consistently correlates with high happiness, policies can focus on strengthening communities, while a strong negative impact of corruption can inform anti-corruption efforts.

As Eberhard (2023) points out, those responsible for creating visualizations must be well-trained and aware of their potential impact, as poorly designed visuals can introduce bias. This responsibility is amplified when dealing with sensitive topics like national happiness, where insights derived from visualizations, especially when based on data that must be ensured to be fair and without bias, can influence policies impacting millions of people. Therefore, careful design and critical evaluation of both visualization techniques and underlying data are crucial to avoid misrepresentation and ensure positive outcomes.

The visualization comparison highlighted key learnings: different techniques have trade-offs (for example, radar charts for detailed comparison vs. clutter), user-centric design and interactivity are crucial for effectiveness (as seen in the high score for the interactive radar chart and low score for static heatmap), representing multidimensional data requires careful consideration of visual encoding, clear labeling is essential to avoid misinterpretation, and user feedback is invaluable for iterative design and justifying visualization choices.

6. References

Eberhard, K. (2023). The effects of visualization on judgment and decision-making: a systematic literature review. Management Review Quarterly, 73(1), 167–214. https://doi.org/10.1007/s11301-021-00235-8

Matzen, L. E., Howell, B. C., Trumbo, M. C. S., & Divis, K. M. (2023). Numerical and visual representations of uncertainty lead to different patterns of decision making. IEEE Computer Graphics and Applications, 43(5), 72–82. https://doi.org/10.1109/MCG.2023.3299875

Padilla, L. M., Creem-Regehr, S. H., Hegarty, M., & Stefanucci, J. K. (2018). Decision making with visualizations: a cognitive framework across disciplines. Cognitive Research: Principles and Implications, 3(1), 1–25. https://doi.org/10.1186/S41235-018-0120-9/FIGURES/18

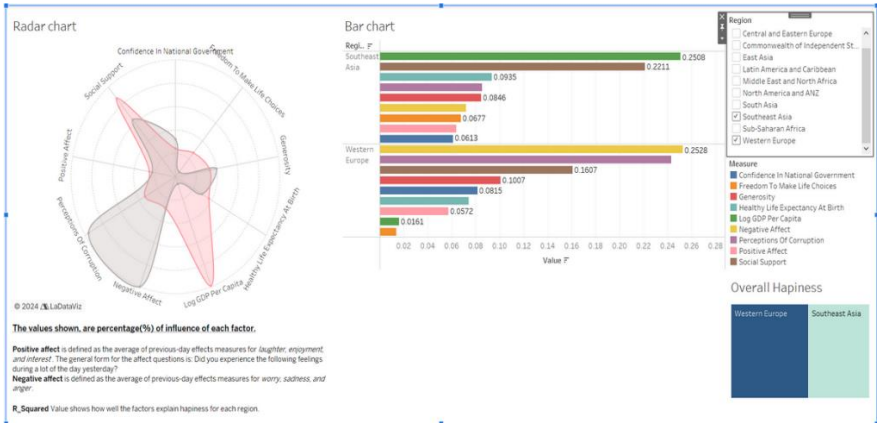
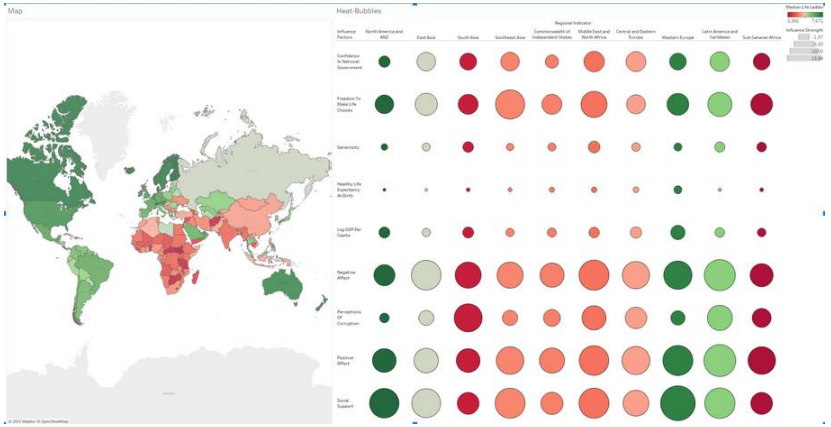
Wang, J., Shu, X., Bach, B., & Hinrichs, U. (2025). Visualization atlases: Explaining and exploring complex topics through data, visualization, and narration. IEEE Transactions on Visualization and Computer Graphics, 31(1), 437–447. https://doi.org/10.1109/TVCG.2024.3456311

Appendix

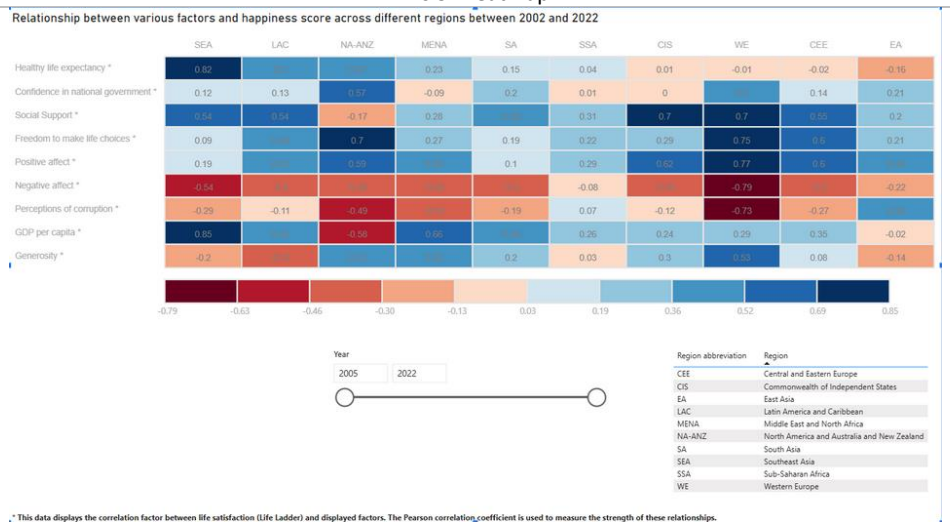
Questionnaire that was used for user evaluation:

- Using the dashboards, answer the question: Which factors have the highest influence on the happiness score across the regions of Western Europe and South East Asia between 2005-2022?
- It was easy to find the factors that have the highest influence on happiness scores in different regions.
- I would like to use this system to answer similar questions in the future.
- Navigating through the dashboard was intuitive.
- The dashboard provided sufficient information to answer the research question.
- I felt confident using this dashboard without requiring additional assistance.
- The design of the system helped me efficiently interpret the data.
- The terminology and labels used in the system were clear and understandable.
- I found the dashboard visually appealing.
- If anything, what would you improve in this system to enhance usability?

Table 1- Comparison of selected visualization techniques.

	Vis 1: Radar chart + Bar chart (Evaluators' Decision) (My selection)	Vis 2: Map + Heat Bubbles
	 Fig 1. Showing the relative influence of each factor using radar chart and bar chart	 Fig 2. Showing the influence of each factor by having them in different bubble sizes
What	This visualization is based on a multidimensional table where each row represents a region, and each column contains a quantitative attribute indicating the relative influence (in percentage) of various factors on regional happiness scores. The primary data attributes include factors like Social Support, Log GDP Per Capita, and others, all treated as quantitative variables. The categorical variable "Region" serves as a grouping dimension for comparison. Each region has a set of influence values that reflect how much each factor contributes to its happiness score, as estimated via **regression models**. The dataset structure allows for region-wise aggregation and comparison across multiple dimensions of happiness indicators.	The underlying data is a multidimensional table where each row corresponds to a region and each column to a happiness factor such as GDP, Social Support, or Positive Affect. The regions are categorical items, while the attributes are quantitative. The visualization also includes a map to show the distribution of countries. The relationship between factors and their influence on the overall happiness of a region is obtained using **Pearson Coefficient** values.
Why	The goal of this dashboard is to allow users to analyze and compare how various factors influence happiness across different world regions. Users are expected to discover region-specific patterns—such as which factors dominate in Southeast Asia versus Western Europe—and evaluate the relative weight of each driver of happiness. This provides a comparative purpose: identifying what matters most for happiness in one region versus another. The intended audience includes policymakers and researchers who want to explore regional variations in well-being in order to inform targeted policy designs.	The main goal of this visualization is to help users discover which factors have the highest influence on happiness across different regions, and to communicate those insights clearly. Users are expected to compare the strength of multiple factors within and between regions, identify which factors are most impactful in certain areas, and summarize patterns or outliers in the data. It can be useful for audiences like policymakers or researchers interested in understanding regional drivers of well-being.
How	The radar chart uses area marks (closed polygonal shapes) with color hue to represent selected regions, and radial distance from the center to indicate the strength of each factor's influence. It supports multi-dimensional comparison of factor importance per region. The bar chart on the right uses horizontal bar marks to visualize the same data with a more precise, easily readable format. Bars are grouped by region and color-coded by factor, using both length and position along a quantitative axis to encode influence.	The visualization uses a matrix of circles (bubble chart) where each row represents a specific happiness factor and each column corresponds to a region. Circle size encodes the magnitude of influence—larger circles indicate stronger influence—while circle color shows the overall happiness of each region. The map on the left provides geographic context, linking regions to their spatial locations. This encoding strategy leverages

	The interaction panel allows users to filter and compare specific regions, with dynamic linked highlighting across both views. The Overall Happiness section uses color-coded tiles to display comparative happiness scores between selected regions, reinforcing the relationship between influential factors and overall outcomes.	position, area, and color channels to enable effective comparison and pattern recognition across multiple dimensions of the data.
Evaluation result	**Average score: 8.84** Users' notes: While radar chart works very well if a few regions are selected, it can be cluttered and not easily interpretable if many regions are selected at the same time. The radar chart can benefit from having a legend that shows the colors of each region. So, for improvement, legend should be implemented and also there should be an identifier on the radar chart to know which country is being displayed easier.	**Average score: 7.54** Users' notes: The similar sizes of the bubbles can be confusing and difficult to know which factors are the most influential; therefore, for future improvements, numbers can be added for precise display of values and sorting based on each column can be very helpful in this visualization.
Advantage (briefly)	Users found the dashboard to be visually appealing and felt confident using it to answer the research question. The labels and terminology made it easy for users to understand the context and made it clear for them.	Although the map doesn't directly contribute to answering the research question, it makes it easy to see the allocation of regions. The coloring gives quick overview of the overall happiness of each region.
Disadvantage (briefly)	Radar chart being cluttered if many regions are chosen all at the same time and the lack of legend for radar chart.	It is difficult to determine which factors are most influential as many bubbles are very close in size to each other and there is no sorting option.

	Vis 3: Heatmap
	 Fig 3. Showing the influence of each factor using a heatmap
What	The underlying data is a multidimensional table where each row corresponds to a region and each column to a happiness factor such as GDP, Social Support, or Positive Affect. The regions are categorical items, while the attributes are quantitative. The relationship between factors and their influence on the overall happiness of a region is obtained using **Pearson Coefficient** values.	

Why	The purpose of this dashboard is to allow users—such as policymakers, researchers, and data analysts—to explore which factors most influence happiness in specific world regions and to compare these influences across selected regions. It enables users to identify patterns, such as whether economic variables dominate in one region while emotional or social variables dominate in another. The visualization aims to support informed decision-making by highlighting which factors are most impactful in driving happiness outcomes. Using heatmap is efficient for comparing the Pearson correlation coefficient among different factors and is easily understandable.
How	The visualization uses a heatmap with rectangular cell marks. The horizontal axis (X-axis) encodes categorical variables representing happiness-related factors, while the vertical axis (Y-axis) represents regions. Each cell is colored using a diverging color scale to indicate the magnitude of a factor's influence. The visual channel used for encoding quantitative values is color saturation and also which color is being used—more saturated shades represent higher influence (dark blue for positive relation and dark red for negative relation). A legend is included to help interpret the names of regions. The use of grid layout and fixed cell size supports rapid scanning.
Evaluation result	**Average score: 7.86** Users' notes: Using two different colors in this case can cause confusion, as the main goal is to see the influence of each factor regardless of its direction, so for future project, using a single color might be more efficient. Also, the year slider is not contributing to answering the research question as only the whole period is relevant. While the diverging color scale helps distinguish positive vs. negative correlations, future iterations could use a single-color scale to focus solely on strength of influence.
Advantage (briefly)	As it is using a heatmap, it is easy to answer the research question at a glance and users felt confident interacting with the visualization.
Disadvantage (briefly)	Lacks effective and clear labels and the visualization is not engaging.