Data ANalysis and Visualization

A critical look on the Covid-19 visualizations

**Name:**

Sadri Yazidi

**Matriculation Number:**

UPS10684018

**Course:**

DLBDSEDAV01 – Explorative Data Analysis and Visualization

**Course Of Study:**

Data Analyst – Python

**Date:**

15.01.2025

**Tutor:**

Visieu Lac

**Link to GitHub Repository:**

[1992SY/Data\_Visualization](https://github.com/1992SY/Data_Visualization)

Table of Contents

[1. Introduction 2](#_Toc187834337)

[1.1 Motivation 2](#_Toc187834338)

[1.2 Objective 2](#_Toc187834339)

[1.3 Approach 2](#_Toc187834340)

[2. Baseline Visualization and Dataset 3](#_Toc187834341)

[2.1 Dataset Description 3](#_Toc187834342)

[2.2 Baseline Visualization 3](#_Toc187834343)

[2.3 Literature on Visualization Principles 3](#_Toc187834344)

[3. Alternative Visualizations 4](#_Toc187834345)

[3.1 Principles for Improvement 4](#_Toc187834346)

[3.2 Visualization 1: Normalized Bar Chart 4](#_Toc187834347)

[3.3 Visualization 2: Time-Series Line Chart 5](#_Toc187834348)

[3.4 Visualization 3: Vaccination vs. Deaths Scatterplot 5](#_Toc187834349)

[4. Discussion 5](#_Toc187834350)

[4.1 Comparison with Baseline Visualization 5](#_Toc187834351)

[4.2 Comparison of Alternative Visualizations 6](#_Toc187834352)

[4.3 Application of Literature 6](#_Toc187834353)

[5. Conclusion and Outlook 6](#_Toc187834354)

[5.1 Summary 6](#_Toc187834355)

[5.2 Future Work 7](#_Toc187834356)

[6. References 7](#_Toc187834357)

[7. Appendix 7](#_Toc187834358)

[8. Notes 7](#_Toc187834359)

1. Introduction

1.1 Motivation

In the era of information overload, effective data visualization has become a cornerstone of data communication. Simply presenting data is not enough anymore; the story behind the data must be conveyed effectively. Storytelling transforms raw numbers into compelling narratives, making data more accessible and engaging (Cairo, 2016). The COVID-19 pandemic has highlighted the critical need for clear and insightful data visualizations to communicate trends and inform decision-making effectively. Poor visualizations can mislead, confuse, or overwhelm audiences, diminishing their impact and utility (Healy, 2018).

Effective storytelling in data visualization bridges the gap between raw data and actionable insights by emphasizing relevant aspects and minimizing distractions (McCandless, 2012). This study focuses on enhancing a suboptimal COVID-19 visualization by applying principles of storytelling and effective design, thus demonstrating how thoughtful improvements can transform data presentations into powerful narratives.

1.2 Objective

The objective of this paper is to demonstrate how a poorly designed COVID-19 data visualization can be improved to better convey its intended message. By creating alternative visualizations, I aim to highlight techniques for enhancing clarity, focus, and engagement. This paper will also critically examine how principles from the literature, such as those articulated by Few (2012) and Evergreen (2016), can be applied to refine and elevate the communicative power of data visualizations.

1.3 Approach

The approach involves identifying a suboptimal COVID-19 visualization, analyzing its shortcomings based on data visualization principles, and creating three improved versions. Each alternative visualization will emphasize different aspects of storytelling, followed by a critical discussion of their effectiveness. Additionally, the visualizations are grounded in principles of cognitive load minimization, color theory and data ethics to ensure their accessibility and inclusiveness (Knaflic, 2015).

2. Baseline Visualization and Dataset

2.1 Dataset Description

The dataset for this study was obtained from “Our World in Data” and includes global COVID-19 data such as case counts, testing statistics and vaccination rates. The dataset is well-structured, with daily updates for multiple countries. Key variables include:

* Total confirmed cases
* Total deaths
* Vaccination rates
* Testing rates

This dataset provides a comprehensive view of the pandemic’s impact and progression, making it a valuable resource for creating meaningful visualizations.

2.2 Baseline Visualization

The original visualizations selected for improvement are as follows:

**Normalized Bar Chart:** showing the number of COVID-19 cases per million people.

**Time-Series Line Chart:** showing the temporal progression of cases for selected countries.

**Scatterplot:** comparing vaccination rates to deaths per million people.

The following are identified issues with the baseline visualizations:

**Lack of Focus:** The bar chart includes too many countries, making it overwhelming.

**Lack of Clarity:** The time-series chart doesn’t effectively highlight trends or key events.

**Suboptimal Color Choice:** The scatterplot uses default colors that do not help differentiate countries or highlight relationships between the data points.

These shortcomings hinder the visualizations' ability to effectively communicate insights and present a coherent narrative. They serve as the starting point for exploring how design improvements can enhance the effectiveness of the visualizations.

2.3 Literature on Visualization Principles

Effective visualizations adhere to principles such as simplicity, clarity, and focus (Tufte, 2001). Additionally, storytelling techniques, such as emphasizing key data points and using annotations, are essential for engaging the audience (Few, 2012). Research by Evergreen (2016) emphasizes the importance of aligning design elements with the intended message, while Knaflic (2015) highlights the role of color and layout in guiding the audiences attention. These techniques and principles guide the improvements in this study.

3. Alternative Visualizations

3.1 Principles for Improvement

The following principles were applied to improve the original visualizations:

**Contextualization**: Normalizing data by population and focusing on selected countries to provide fair comparisons.

**Focus**: Highlighting critical countries and regions, making the visualization easier to interpret.

**Clarity**: Reducing clutter, using effective color schemes and emphasizing key data points to improve readability.

**Accessibility:** Ensuring the visualizations are readable for audiences with diverse abilities, such as using colorblind-friendly palettes.

3.2 Visualization 1: Normalized Bar Chart

Description: This visualization normalizes total cases by population, displaying cases per million people. It limits the number of countries to improve readability and uses a gradient color scheme to emphasize countries with higher case rates.

**Design Choices**

**Color Scheme**: A shades-of-blue palette emphasizes severity.

**Annotations:** The country with the highest number of case per million is easy to see for the viewer, because it is highlighted in red.

**Evaluation:** This visualization provides a fairer comparison of countries by normalizing the data by population, offering a cleaner and more focused presentation than the baseline chart. It highlights key insights and allows for more meaningful comparisons.

3.3 Visualization 2: Time-Series Line Chart

**Description:**This line chart displays the temporal progression of COVID-19 cases for a selection of countries. The chart emphasizes trends over time, making it ideal for telling a narrative about how different countries experienced the pandemic.

**Design Choices**

**Color Scheme:** Each country is represented by a distinct color to differentiate the lines.

**Logarithmic Scale:** Because in this graphic the numbers are growing exponentially, I decided to use a logarithmic scale for the y-axis. In this way, it is easier to get an overview of the different cases in the different countries.

**Gridlines**: Faster recognition for the viewer, how many cases at a certain time were reported.

**Evaluation:** This visualization is ideal for the trend in reported Covid-19 cases and understanding how the pandemic unfolded in different countries. It effectively conveys the temporal nature of the data and highlights differences in responses.

3.4 Visualization 3: Vaccination vs. Deaths Scatterplot

**Description:** This scatterplot compares vaccination rates with death rates per million people across various countries. It uncovers potential relationships between vaccination rates and COVID-19 mortality.

**Design Choices**

**Color Scheme:** Countries are color-coded by continent, adding context and visual appeal.

**Size of Markers**: Larger markers are used for countries with higher populations.

**Tooltips**: Hovering over the data points provides additional details about each country.

**Evaluation:** This visualization reveals insights into the relationship between vaccination rates and mortality. It could be enhanced with interactivity for deeper exploration and offers valuable analytical insights into the pandemic’s outcomes.

4. Discussion

4.1 Comparison with Baseline Visualization

The baseline visualizations failed to provide context or clarity, making it difficult to derive meaningful insights. The alternative visualizations address these issues effectively:

The normalized bar chart offers a fair comparison by accounting for population differences.

The time-series chart illustrates the progression of cases over time, adding a dynamic narrative element.

The scatterplot uncovers relationships between vaccination rates and deaths, providing a deeper understanding of the factors influencing pandemic outcomes.

These improvements demonstrate the power of thoughtful design in transforming data into actionable insights.

4.2 Comparison of Alternative Visualizations

Each alternative visualization serves a different purpose:

The normalized bar chart excels in simplicity and clarity, making it suitable for audiences seeking an overview.

The time-series chart is ideal for narrative-driven insights, providing a dynamic view of the pandemic’s progression.

The scatterplot is effective for exploring relationships, offering depth and context for analytical audiences.

4.3 Application of Literature

By applying principles from Tufte (2001), Few (2012), Evergreen (2016), and Knaflic (2015), the alternative visualizations prioritize simplicity, focus, and storytelling. The use of annotations and appropriate color schemes enhances the audience’s understanding, while interactivity and accessibility considerations further improve engagement.

5. Conclusion and Outlook

5.1 Summary

This study demonstrates that thoughtful design choices can transform suboptimal visualizations into compelling narratives. By normalizing data, focusing on key insights, and leveraging storytelling techniques, the improved visualizations provide clarity and engagement. These examples underscore the importance of aligning visualization design with communication goals.

5.2 Future Work

Future research could explore interactive visualizations and the use of advanced tools like Tableau or D3.js to further enhance storytelling. Additionally, incorporating user feedback into the design process could improve visualization effectiveness. The integration of AI-driven visualization tools may also offer new opportunities for automating and optimizing design choices.

6. References

Cairo, A. (2016). The Truthful Art: Data, Charts, and Maps for Communication. New Riders.

Evergreen, S. D. H. (2016). Effective Data Visualization: The Right Chart for the Right Data. SAGE Publications.

Few, S. (2012). Show Me the Numbers: Designing Tables and Graphs to Enlighten. Analytics Press.

Healy, K. (2018). Data Visualization: A Practical Introduction. Princeton University Press.

Knaflic, C. N. (2015). Storytelling with Data: A Data Visualization Guide for Business Professionals. Wiley.

Tufte, E. R. (2001). The Visual Display of Quantitative Information. Graphics Press.

Our World in Data. (n.d.). COVID-19 Data Repository. Retrieved from https://github.com/owid/covid-19-data

7. Appendix

Figure List

Figure 1: Normalized Bar Chart of Cases per Million.

Figure 2: Time-Series Line Chart of Case Progression.

Figure 3: Scatterplot of Vaccination Rates vs. Deaths per Million.

8. Notes

Figures are available on GitHub alongside the Python code.

Code repository: GitHub Repository