

Fundamentals of Data Visualisation (gdv)

Noah Leuenberger

October 15, 2023

Contents

1	Introduction	3
1.1	Introduction to the Gapminder Dataset	3
2	LO1: Visualization basics, chart types	4
2.1	Line Charts	4
2.2	Pie Charts	4
2.3	Bar Charts	4
2.4	Boxplots	4
2.5	Bubble Charts	5

1 Introduction

This report provides an in-depth overview of data visualization, covering fundamental concepts and the optimal use-cases for various chart types, while also considering the needs of different user demographics. The science of visual perception is examined, emphasizing its role in effective design. Additionally, we explore the relationship between data and design principles, stressing the importance of consistency between data preprocessing decisions and design choices. The report presents the 'Grammar of Graphics,' which is a framework that simplifies the design process. It then explores various techniques for assessing the efficacy of data visualizations. The overall objective of this report is to offer a comprehensive comprehension of data visualization by merging its theoretical, practical, and evaluative components.

1.1 Introduction to the Gapminder Dataset

The Gapminder dataset provides a comprehensive examination of numerous socio-economic and health-related aspects across countries and selected timeframes. The variables consist of country name, year, population, continent, life expectancy, and GDP per capita [1]. The following is a concise explanation of the dataset variables:

- **Country:** The name of the country.
- **Year:** The year the data was recorded.
- **Population:** The total population of the country for the corresponding year.
- **Continent:** The continent to which the country belongs.
- **Life Expectancy:** The average life expectancy in years for people born in that year.
- **GDP per Capita:** The Gross Domestic Product per capita, which serves as an indicator of the economic performance of a country.

2 LO1: Visualization basics, chart types

2.1 Line Charts

Line charts are often the first choice for illustrating temporal data (Figure 1). This is because they excel at showing trends over time, making them ideal for financial, scientific, and statistical data. The linear representation helps the audience quickly identify increases or decreases in values over a specified period. They are also useful for showing the relationship between two variables, such as the correlation between GDP and life expectancy.

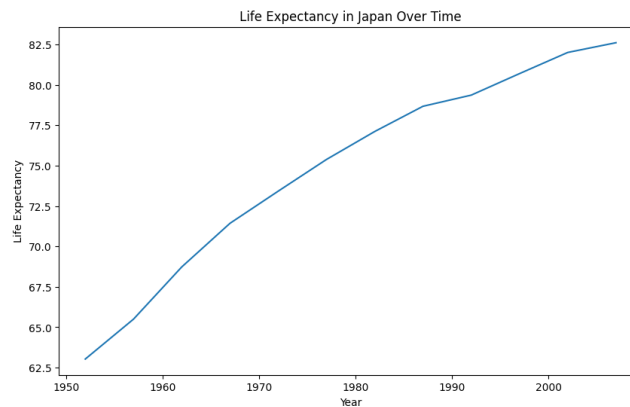


Figure 1: Line Chart of the Life Expectancy in Japan Over Time

However, line charts are less appropriate for comparing individual categories side-by-side. This is because lines imply a connection or progression between points, which may not be meaningful for unrelated categories [2].

2.2 Pie Charts

Pie charts are one of the most commonly misused chart types (Figure 2). They are best used to represent proportional or percentage data, often for a total that amounts to 100%. However, they can be misleading as the human eye struggles to accurately compare the area or angle of pie slices, especially as the number of categories increases [3].

2.3 Bar Charts

Bar charts are highly versatile and can be used in various applications (Figure 3). The reason they are so effective for comparisons is that the length of bars allows for easy, accurate judgments of value differences. This makes them suitable for comparing data across a large number of different categories.

Bar charts are less suitable for showing trends over time, as they do not imply a temporal progression between categories [4].

2.4 Boxplots

Boxplots are a powerful tool for understanding the distribution of data across different categories (Figure 4). They display a five-number summary of the data: the minimum, first quartile, median,

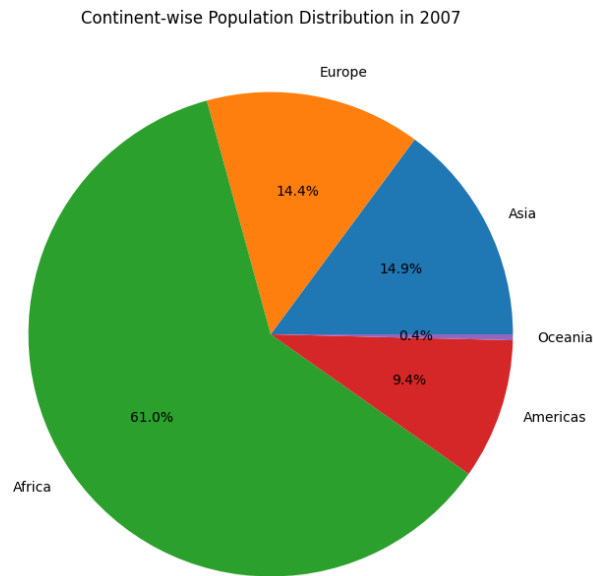


Figure 2: Pie chart of Continent-wise Population Distribution in 2007

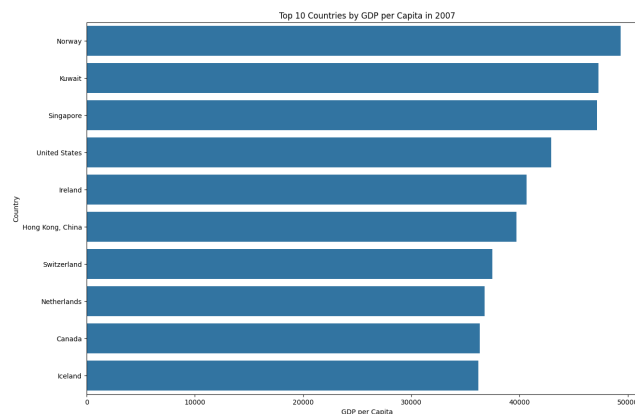


Figure 3: Bar chart of the top 10 Countries by GDP per Capita in 2007

third quartile, and maximum. This provides a robust summary of the data's central tendency, spread, and presence of outliers [5].

2.5 Bubble Charts

Bubble charts are an extension of the scatter plot (Figure 5). They allow comparison and portrayal of data points in three dimensions: x -coordinate, y -coordinate, and size of the bubble. While they are useful for evaluating complex relationships, the size variable can sometimes be misinterpreted. This is because humans are generally less adept at accurately comparing the size of 2D areas than they are at comparing lengths or heights [6].

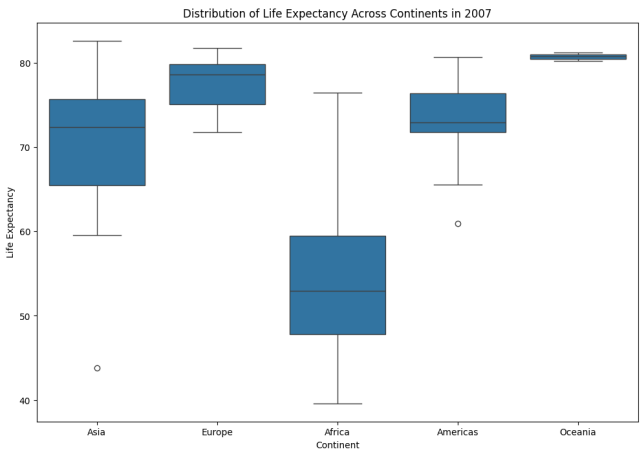


Figure 4: Box plots of the Distribution of Life Expectancy Across Continents in 2007

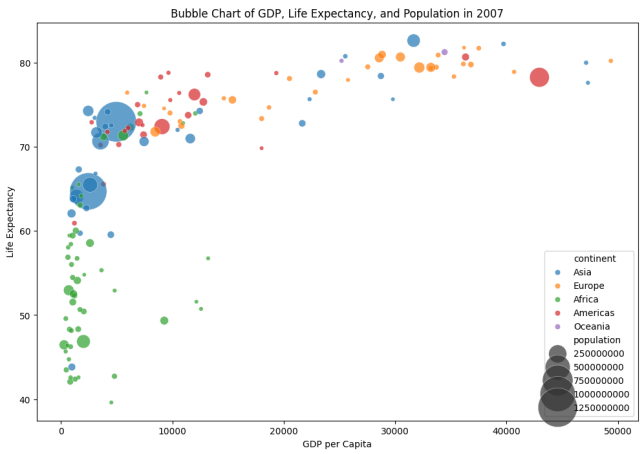


Figure 5: Bubble Chart of GDP, Life Expectancy, and Population in 2007

List of Figures

1	Line Chart of the Life Expectancy in Japan Over Time	4
2	Pie chart of Continent-wise Population Distribution in 2007	5
3	Bar chart of the top 10 Countries by GDP per Capita in 2007	5
4	Box plots of the Distribution of Life Expectancy Across Continents in 2007	6
5	Bubble Chart of GDP, Life Expectancy, and Population in 2007	6

References

- [1] “Gapminder World.” <https://www.kaggle.com/datasets/tklimonova/gapminder-datacamp-2007>.
- [2] S. Few, *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. Analytics Press, 2009.
- [3] E. R. Tufte and E. R. Tufte, *The Visual Display of Quantitative Information*. Cheshire, Connecticut: Graphics Press, second edition, ninth printing ed., 2015.
- [4] J. Heer, M. Bostock, and V. Ogievetsky, “A tour through the Visualization zoo,” *Commun. ACM*, vol. 53, pp. 59–67, June 2010.
- [5] J. W. Tukey, *Exploratory Data Analysis*. Reading: Addison-Wesley, 1977.
- [6] W. S. Cleveland and R. McGill, “Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods,” *Journal of the American Statistical Association*, vol. 79, no. 387, pp. 531–554, 1984.