

Semantic content and information visualization: A proposal for a data-driven communication narrative

Mario Pérez-Montoro

Universitat de Barcelona, Spain

<https://orcid.org/0000-0003-2426-8119>

Pérez-Montoro, M. (2025). Semantic content and information visualization: A proposal for a data-driven communication narrative. In J. Guallar, M. Váñez, & A. Ventura-Cisquella (Coords). *Digital communication. Trends and good practices* (pp. 85-103). Ediciones Profesionales de la Información. <https://doi.org/10.3145/cuvicecom.07.eng>

Abstract

Charts and graphs are one of the main visual tools used to represent and convey information. A significant part of our daily work involves searching for, obtaining, and refining a dataset that allows us to tell a story. However, data does not speak for itself. It needs to be visualized to reveal the underlying analysis and bring to light the patterns it contains. In this work, we aim to characterize the main communicative contexts and content types that can be conveyed through graphs. We will also explore the most suitable visual approaches to efficiently and effectively fulfill these communicative purposes. In other words, in this chapter, we seek to present a structured methodology commonly used to properly implement effective information visualization. First, we identify the type of information to be encoded; second, we choose the type of graph that best represents that information; and finally, we design the visual elements that compose the graph.

Keywords

Information visualization; Data visualization; Visual information communication; Information design; Graphs; Charts; Visual representation of information.

1. Introduction

Charts and graphs are one of the main visual tools used to represent and communicate information. A significant part of our daily work involves searching for, obtaining, and refining datasets that allow us to tell a story. However, data does not speak for itself; it needs to be visualized to reveal derived analyses and uncover the patterns it holds.

In this work, we aim to characterize the variety of contexts and communicative purposes that can be conveyed using a graph or chart. We will also explore which visual approach is most suitable to efficiently and effectively fulfill these communicative intentions. In other words, this chapter seeks to systematically present the methodology commonly used to implement effective information visualization. First, the identification of the type of information to be encoded; second, the selection of the type of graph or chart that best represents that information; and finally, the design of the visual elements of the graph.

To achieve this goal and following this structure, we will address the following points. In the first section (Section 2), we will discuss the main basic types of information or semantic content that can be represented in a graph or chart, illustrating them with examples organized into tables for easier understanding. In Section 3, we will examine which type of visualization is best suited to communicate each of these types of informational content.

2. Semantic content

Intuitively, any communication strategy is based on three basic elements: the message or semantic content to be conveyed, the intended audience, and the context surrounding the communicative act.

Communication through graphs or charts is no different. It depends on the message or informational content we aim to deliver to our audience, who we want to reach, and the context

(the document where the graph is embedded, accompanying text, transmission channel, communicative intent, or the timing of its consumption, among many other factors) that frames the communication effort. For now, let us focus exclusively on the content we wish to convey.

Typically, the semantic content communicated through a graph is a proposition (an idea, in a non-technical sense) asserting the existence of a relationship between two or more types of information. Specifically, it asserts a relationship between two (or more) items of quantitative and/or qualitative information. In other words, it expresses some kind of relationship between values belonging to two or more quantitative variables, between the categories or attributes of two or more qualitative variables, or between categories or attributes and quantitative values.

In this sense, it is possible to classify the main types of relationships represented in the semantic content of graphs (Zelazny, 2001; Few, 2012; Shneiderman, 1996; Pérez-Montoro, 2022). These include, for example, nominal comparison, time series, ranking, part-to-whole, deviation, distribution, and correlation relationships.

2.1. Nominal comparison

Nominal comparison relationships could be considered the most basic and common semantic content represented using graphs.

From a technical perspective, this semantic content is understood as a proposition (an idea) that establishes a relationship in which a quantitative value is assigned or connected to each of the items (values or categories) of a qualitative variable to facilitate the comparison of those numeric values associated with those items.

Figure 1
Job positions and salaries of ACME company employees.

Job position	Salary
Manager	150.000
Analyst	40.000
Documentalist	40.000
Administrative Assistant	32.000
Junior Staff	25.000
Intern	8.000

(Own elaboration, simulated data)

An example of this type of semantic content can be seen in Figure 1. This figure presents a table illustrating the relationship between the various items, categories, or values (manager, analyst, documentalist, administrative assistant, junior staff, and intern) of a qualitative variable (job positions or profiles in ACME company) and a quantitative value (gross salary received by employees based on their roles).

2.2. Time series

Time series relationships are another common type of semantic content often communicated using graphs.

Conceptually, this content is identified as a proposition (an idea) that establishes a relationship where a quantitative value is assigned or connected to each of the items (values or categories) of a temporal variable, facilitating the comparison of those numeric values associated with those time units.

Figure 2
Quarterly sales during the previous fiscal year.

Quarter	Sales
First	456.876
Second	189.387
Third	63.829
Fourth	122.345

(Own elaboration, simulated data)

An example of this type of semantic content can be found in Figure 2. The figure presents a table showing the relationship between the different items, categories, or values (first, second, third, and fourth) of a temporal variable (quarters of the previous year) and a quantitative value (sales generated by the company during those time periods).

2.3. Ranking

This semantic content refers to a proposition (an idea) that establishes a relationship in which a quantitative value is assigned or connected to each item (values or categories) of a qualitative variable, and this assignment of numeric values creates an order (from highest to lowest or vice versa) among the items of the qualitative variable.

Figure 3
Sales by province in Catalonia.

Ranking	Province	Sales
1	Barcelona	456.876
2	Tarragona	189.387
3	Gerona	122.345
4	Lérida	63.829

(Own elaboration, simulated data)

Figure 3 illustrates this type of semantic content. It shows a table that represents the relationship between the various items, categories, or values (Barcelona, Tarragona, Girona, and Lleida) of a qualitative variable (the provinces of Catalonia) and a quantitative value (sales achieved in each province). This assignment of numeric values establishes an order among the provinces.

2.4. Part-to-whole

The part-to-whole relationship is another common type of semantic content often represented using graphs. Conceptually, this semantic content refers to a proposition (an idea) that establishes a relationship where a quantitative value is assigned or connected to each item (values or categories) of a qualitative variable. It highlights that the aggregation of all these

qualitative items forms a whole or entity. Thus, each of these categories is identified as a part of that whole. It is also worth noting that percentages are the most commonly used unit of measurement for expressing the contribution of parts to a whole, where the whole is represented as 100%, with each part as a portion of that total.

Figure 4

Activities carried out during a day.

Activity	Percentage
Eating	7,40%
TV/Internet	8,20%
Socializing	9,40%
Sleeping	29,20%
Working	45,80%
Total	100,00%

(Own elaboration, simulated data)

An example of this type of semantic content is shown in Figure 4, which is a table illustrating the relationship between various items, categories, or values (working, eating, socializing, watching TV/Internet, and sleeping) of a qualitative variable (different activities performed throughout the day) and a quantitative value (the percentage of daily time allocated to each activity). This numerical assignment expresses the quantitative relationship of each part (its weight, so to speak) relative to the whole and facilitates comparisons between these numerical values.

2.5. Deviation

Intuitively, the deviation relationship expresses how a set of quantitative data varies relative to a specific numeric reference value.

From a technical perspective, this semantic content is a proposition (an idea) that establishes a relationship where a quantitative value is assigned or connected to each item (values or categories) of a qualitative variable. These quantitative values are compared to a specific numeric reference value, which is typically assigned to a qualitative item (from the same or another qualitative variable). In some cases, there may be two or more reference values for comparison.

Figure 5

Comparison of last year's sales between sector competitors and our company.

Company	Sales
Competitor A	280%
Competitor B	220%
Competitor C	115%
Competitor D	60%
Competitor E	20%

(Own elaboration, simulated data)

An example of this type of semantic content is presented in Figure 5, which is a table illustrating the relationship between various items, categories, or values (Competitor A, Competitor B, Competitor C, Competitor D, and Competitor E) of a qualitative variable (different companies competing in the same sector as ours) and a quantitative value (the percentage of sales of each company relative to our company's sales). This numeric assignment expresses the quantitative relationship between each company's sales and ours, facilitating comparisons between these values. In this case, the reference value (our company's sales) is identified as 100%, with the remaining values (competitors' sales) expressed as proportions or percentages relative to this primary reference value.

2.6. Distribution

The distribution relationship is a type of semantic content corresponding to a proposition (an idea) that establishes a relationship where a quantitative value is assigned or connected to each item (values or categories) of an interval variable. This facilitates comparisons between these numeric values associated with those interval units. This content expresses how the quantitative values of a dataset are distributed or spread over a range (the interval between the minimum and maximum values), from the lowest to the highest, across the categories of that interval variable.

Figure 6

Number of patients by age group in a population of epilepsy patients.

Age group	Patients
1-10	15
11-20	30
21-30	55
31-40	75
41-50	25

(Own elaboration, simulated data)

An example of this type of semantic content is shown in Figure 6. The figure presents a table illustrating the relationship between various items, categories, or values (0–15, 16–30, 31–45, 46–60, and 61–70) of an interval variable (age ranges in a population of epilepsy patients) and a quantitative value (the number of patients). This numeric assignment expresses the distribution of the number of patients (how they are spread) across age ranges in a population of 200 epilepsy patients.

2.7. Correlation

Intuitively, the correlation relationship seeks to express whether two sets of quantitative values are related and, if so, how one changes relative to the other.

From a more conceptual perspective, and in its simplest form, this semantic content is a proposition (an idea) that establishes a relationship where two quantitative values (each belonging to a different quantitative variable) are assigned or connected to each item (values or categories) of a qualitative variable. This facilitates the identification of patterns in the variation between these two types of numeric values. It seeks to determine whether a systematic relationship exists between the values of one quantitative variable and those of another.

Figure 7

Investment and profits by manufactured product.

Product	Investment	Profits
Product 01	22.864	678.543
Product 02	45.789	778.765
Product 03	50.678	783.213
Total	119.331	2.240.521

(Own elaboration, simulated data)

An example of this type of semantic content is presented in Figure 7. The figure shows a table illustrating the relationship between various items, categories, or values (Product 1, Product 2, and Product 3) of a qualitative variable (different products manufactured by ACME company) and two numeric values, each belonging to a different quantitative variable (investment in the production and distribution of a product, and total profits from its commercial life).

3. Visualization of content

Once we have reviewed the main semantic content we aim to communicate through graphics, the next step is to address the process of visualizing these elements or propositional units using graphical representation. In other words, we need to determine the best graphical solutions to effectively encode each type of semantic content.

One critical aspect of creating an effective visualization is selecting the chart that best communicates the information efficiently and effectively. This is not a straightforward task. It depends on several factors, particularly the type of information being represented and the communicative intent. Each type of semantic content requires a specific graphical representation, depending on the communication objectives.

Most data management programs do not provide guidance on selecting the most appropriate chart for a given visualization (Hugues and van Dam, 2013; Hearn, 2011; Shirley, 2009). Only a few software tools offer limited (and often incomplete) recommendations based on the structure of the data being represented (Cherven, 2015; Jones, 2014; Khan, 2016).

In this section, we aim to address this issue. We will provide a set of resources and guidelines to help you select the most communicatively appropriate chart for representing different types of semantic content. Specifically, we will introduce the best practices for visualizing nominal comparisons, time series relationships, rankings, part-to-whole relationships, deviations, distributions, and correlations.

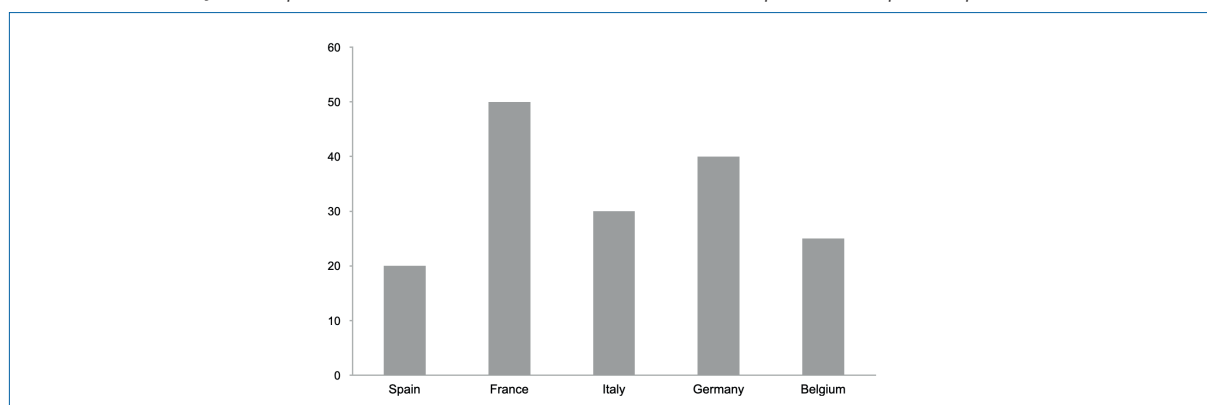
3.1. Visualizing nominal comparison

A chart that accurately represents this type of semantic content must use visual elements that simultaneously encode numeric values and their associated qualitative items. It must also ensure that each resulting pair (quantitative value and qualitative category) is independent, with no intrinsic connection between the pairs.

Given this semantic requirement, the best visual elements for this relationship are bars (or columns) (Figure 8) and points. These elements display the pairs (quantitative value and qualitative item) as visually independent, thus facilitating comparisons between the numeric values.

Figure 8

Goals conceded by the top five national football teams in the last European Championship.

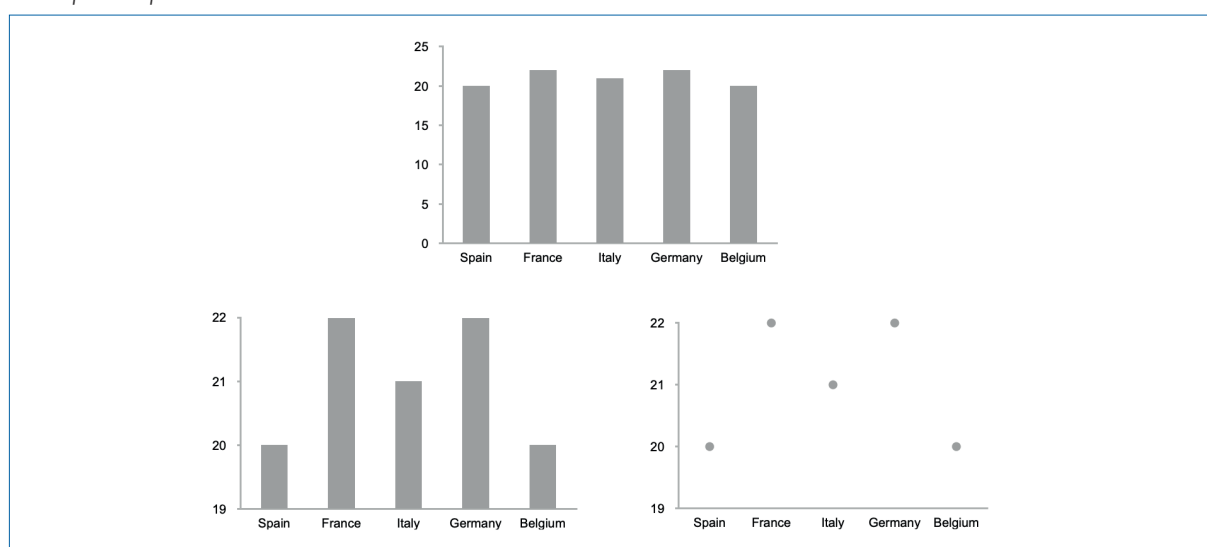


(Own elaboration, simulated data)

In certain contexts, point charts can be a good alternative to column (or bar) charts for this semantic content—specifically when the differences between quantitative values are minimal or when the range of values is narrow, indicating low variability.

Figure 9

Column and point charts showing goals conceded by the top five national football teams in the last European Championship.



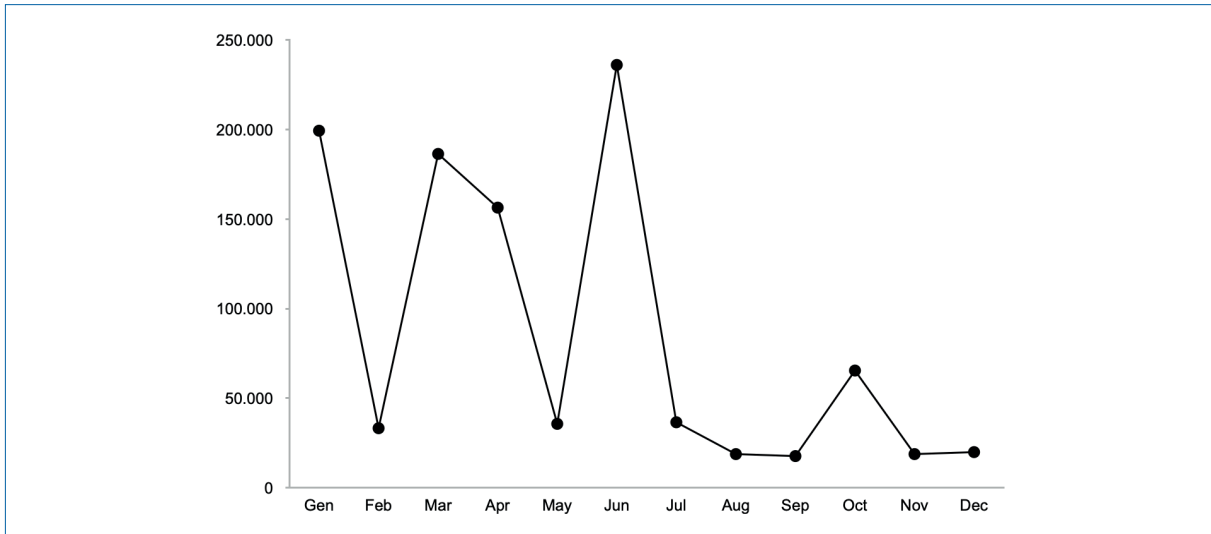
(Own elaboration, simulated data)

Using a column (or bar) chart in these cases might make it harder to perceive differences in the lengths of the bars, thus complicating visual comparisons. Attempting to address this by starting the vertical (or horizontal) axis at a value greater than zero is an incorrect practice. Bars encode quantitative values based on their length, and this requires starting the axis at zero. Point charts, by contrast, avoid this issue because quantitative and qualitative values are encoded by their position on the X and Y axes (Figure 9).

3.2. Visualizing time series

Time series relationships link a quantitative value to each item (value or category) of a temporal variable, facilitating the comparison of numeric values associated with these temporal units.

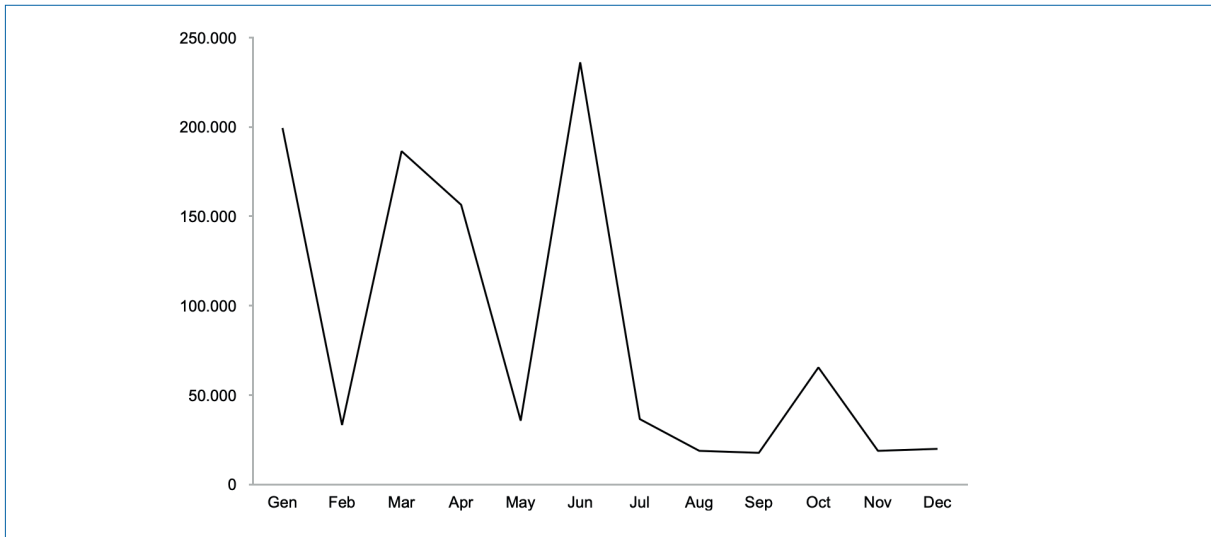
Figure 10
Monthly sales of ACME during the previous year.



(Own elaboration, simulated data)

A chart for this type of semantic content must encode numeric values alongside their associated temporal categories. One axis should represent the temporal variable, with labels for each time unit (year, month, etc.), while respecting the natural order of time. In Western cultures, this order is conventionally displayed horizontally, from left to right, along the X-axis.

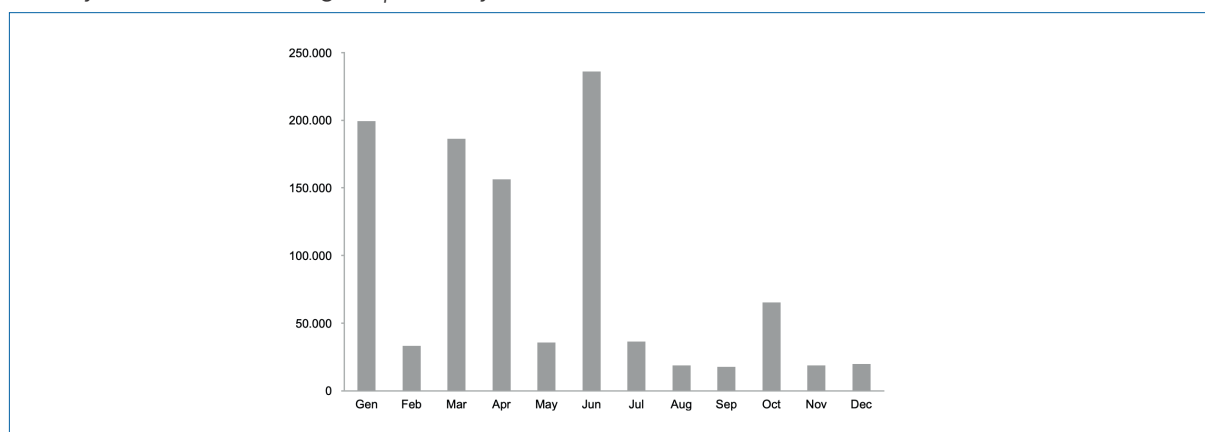
Figure 11
Monthly sales of ACME during the previous year.



(Own elaboration, simulated data)

The best visual elements for this relationship are combinations of points and lines (Figure 10), standalone lines (Figure 11), and columns. These elements connect each pair (quantitative value and temporal item) to the next, illustrating the temporal continuity.

Figure 12
Monthly sales of ACME during the previous year.



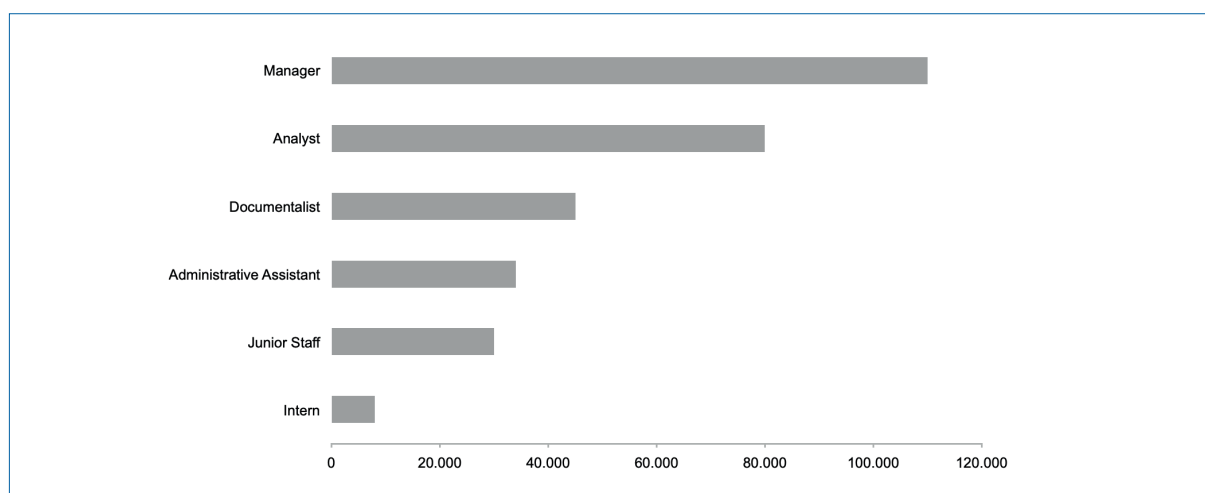
(Own elaboration, simulated data)

Column charts are better suited when the focus is on individual quantitative values associated with specific time points, rather than on the overall trend of the series (Figure 12).

3.3. Visualizing ranking

Ranking relationships link a quantitative value to each item (value or category) of a qualitative variable, establishing an order (ascending or descending) among the items.

Figure 13
Salary by job position at ACME

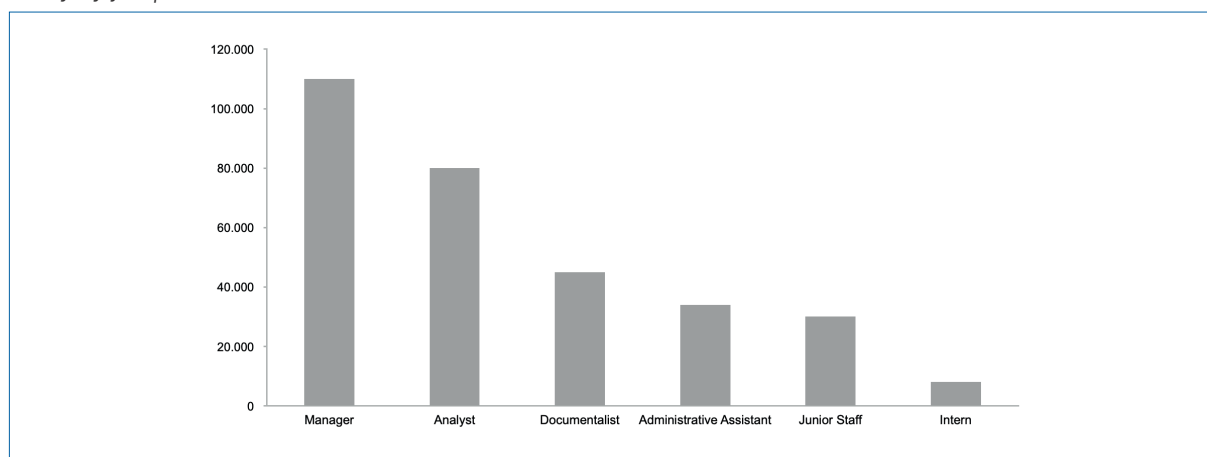


(Own elaboration, simulated data)

A chart that accurately represents this type of semantic content must encode numeric values and their associated categories while visually displaying the order of the qualitative items. This requires placing the qualitative variable along one axis in the correct order, with the quantitative scale on the other.

Figure 14

Salary by job position at ACME.

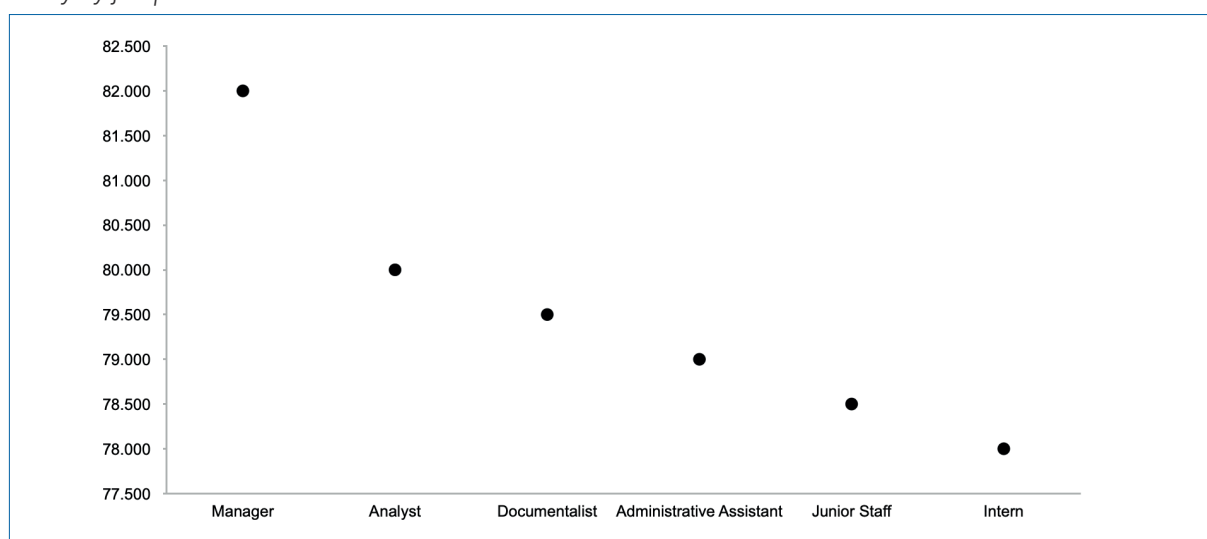


(Own elaboration, simulated data)

Bars (Figure 13), columns (Figure 14), and points are the most suitable visual elements for this type of relationship. They encode pairs (quantitative value and qualitative item) while visually displaying the order of the qualitative items.

Figure 15

Salary by job position at ACME.



(Own elaboration, simulated data)

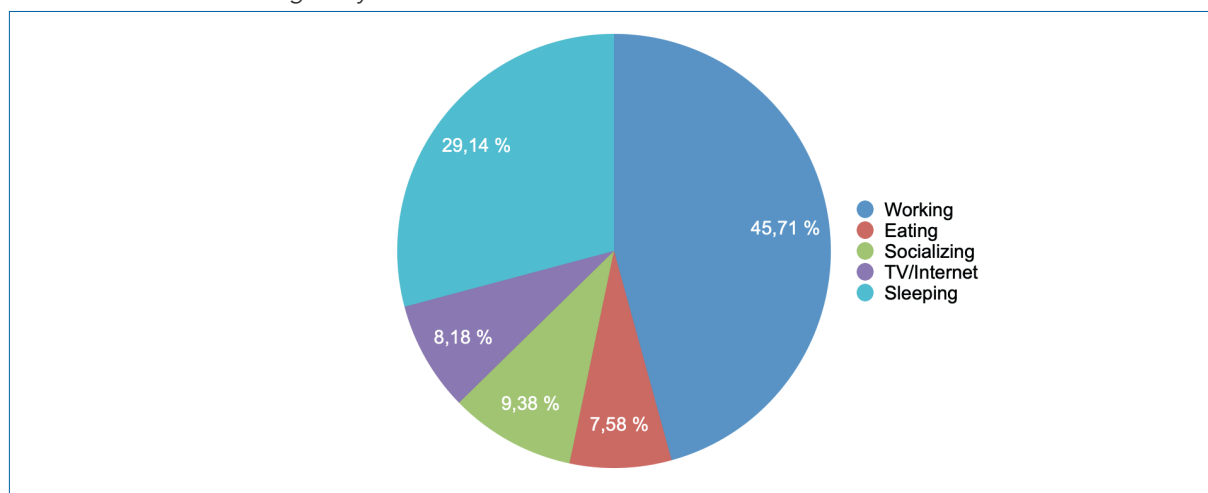
Point charts are preferable in cases where the differences between quantitative values are minimal, making it easier to visualize rankings clearly (Figure 15).

3.4. Visualizing part-to-whole

Part-to-whole relationships link quantitative values to qualitative items that collectively form a whole. A suitable chart must encode these values and categories while illustrating the quantitative relationship between the parts and the whole, facilitating comparisons.

In many communication contexts, pie charts and stacked bar (or column) charts are commonly used.

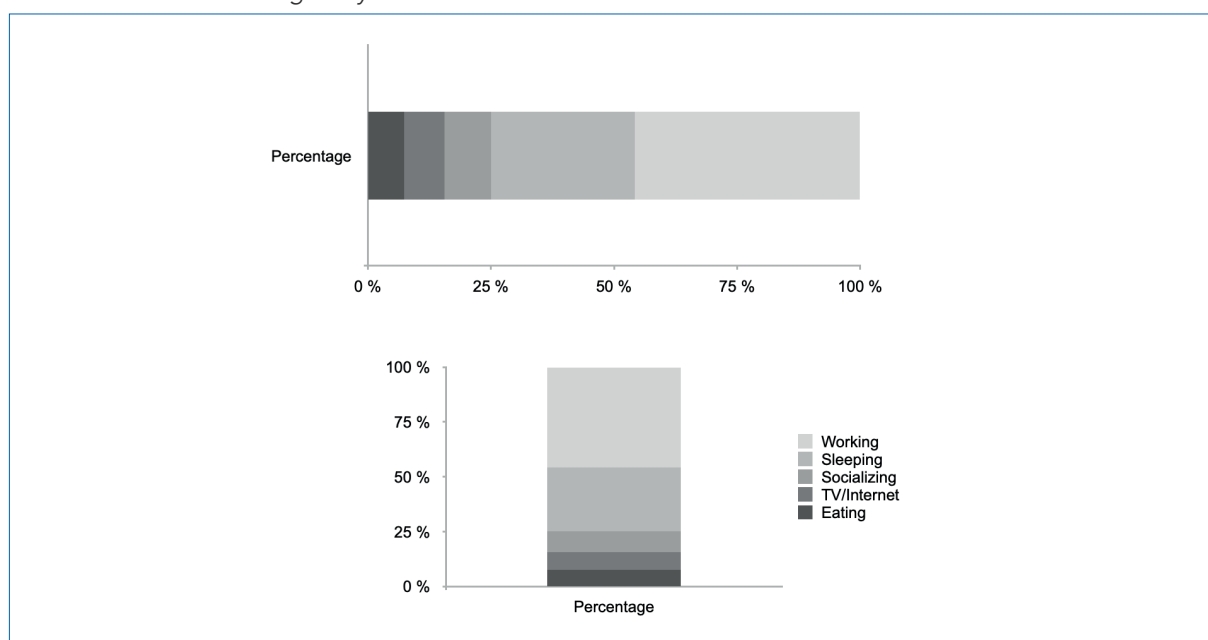
Figure 16
 Activities carried out during a day.



(Own elaboration, simulated data)

In pie charts, each segment represents a quantitative value (based on the angle between radii) associated with a qualitative item (indicated by color). These charts highlight the relationships between parts and the whole (Figure 16).

Figure 17
 Activities carried out during a day.



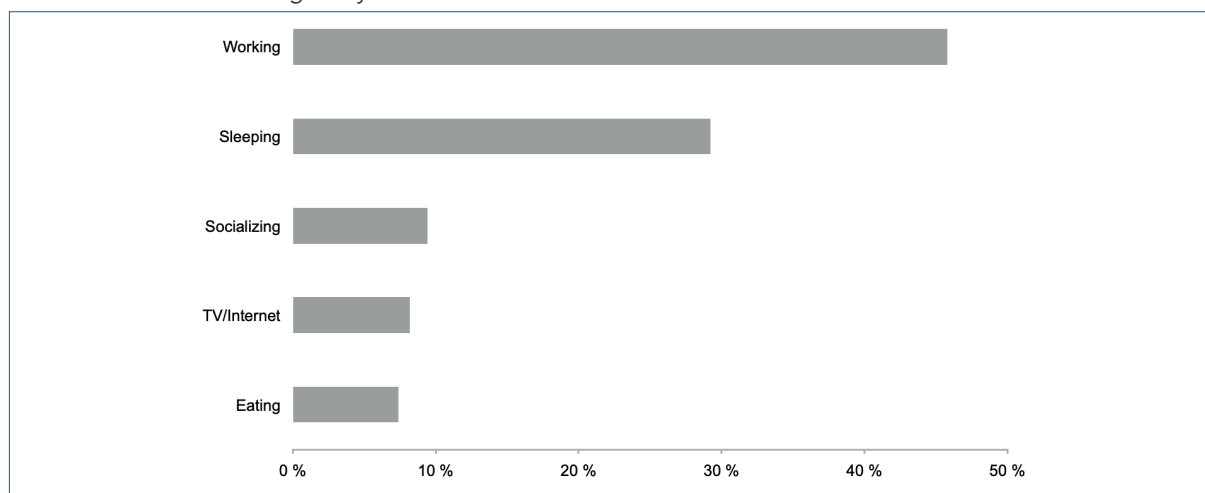
(Own elaboration, simulated data)

Stacked bar (or column) charts represent the whole as a rectangular bar, with individual segments showing the parts. These charts similarly convey the relationships between parts and the whole (Figure 17).

However, unstacked bars (Figure 18) and columns (Figure 19) are often the most effective visual elements for representing part-to-whole relationships, as they avoid the visual limitations of pie and stacked charts.

Figure 17

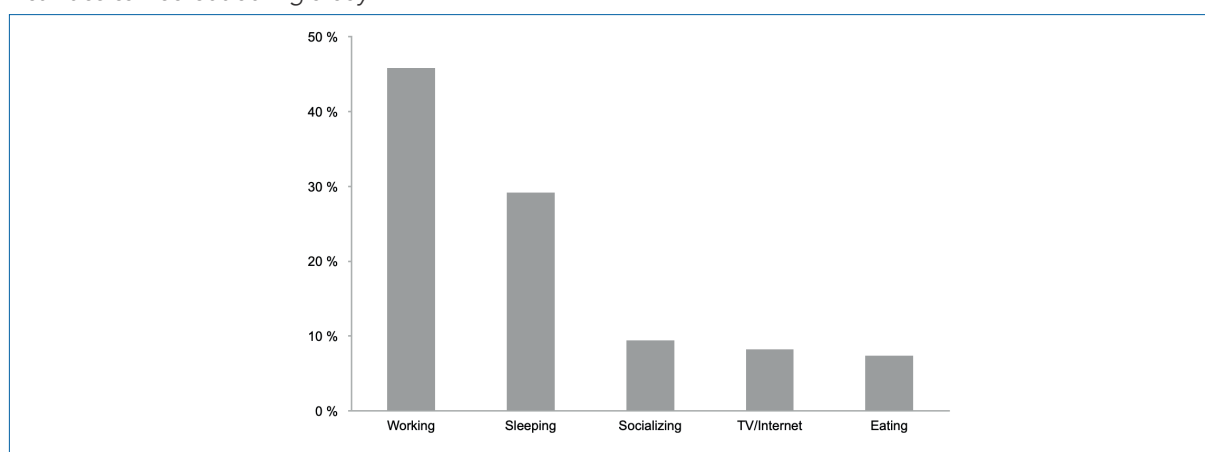
Activities carried out during a day.



(Own elaboration, simulated data)

Figure 19

Activities carried out during a day.



(Own elaboration, simulated data)

It is worth noting that lines and points (or combinations of lines and points) are not suitable for visualizing part-to-whole relationships.

3.5. Visualizing deviation

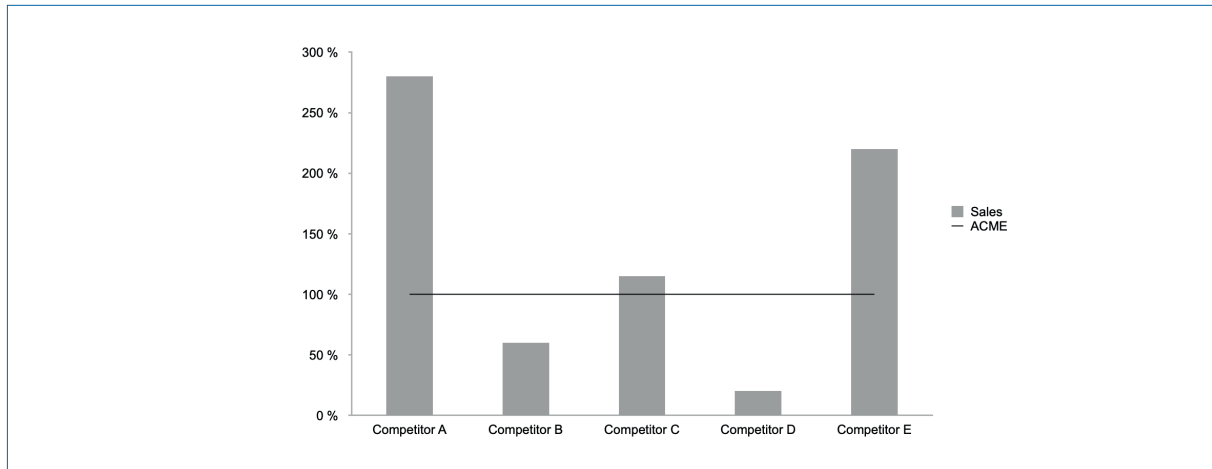
As previously stated, a deviation relationship assigns or connects a quantitative value to each item (value or category) of a qualitative variable. This relationship compares the associated quantitative values to a specific numeric reference value.

A chart that effectively represents this type of semantic content must use visual objects that simultaneously encode the numeric value and the associated category, while enabling quantitative comparison of those numeric data points against the reference value.

The visual objects that best encode this relationship are bars and columns as they adequately represent these pairs (quantitative value and qualitative item) and visually illustrate the quantitative relationship between the values and the reference value.

Figure 20

Comparison of last year's sales between sector competitors and our company.

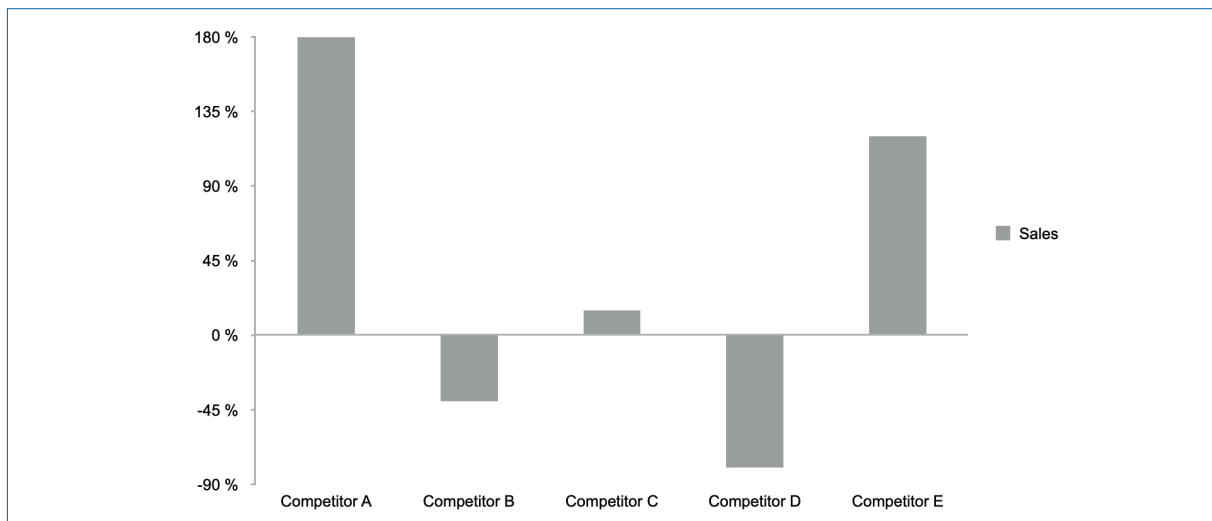


(Own elaboration, simulated data)

In a bar chart (or column chart), each rectangular figure encodes a quantitative value (along the X-axis for bars and the Y-axis for columns) associated with a qualitative item (along the Y-axis for bars and the X-axis for columns). If the reference value is represented with a line or aligned with one of the axes (the Y-axis for bars and the X-axis for columns), it also conveys the quantitative relationship between the numeric values and the reference value (Figure 20).

Figure 21

Comparison of last year's sales between sector competitors and our company.



(Own elaboration, simulated data)

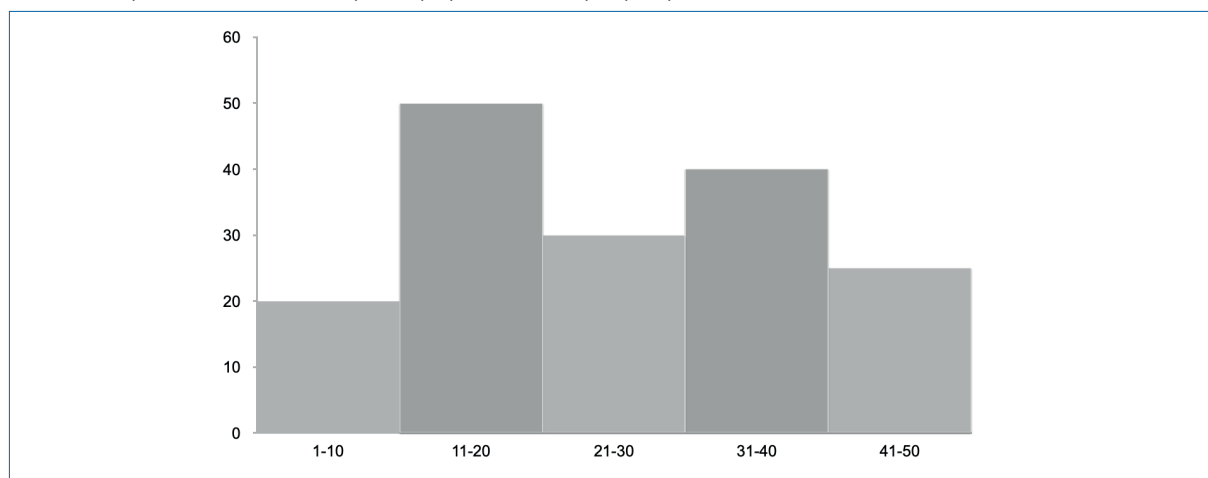
As an alternative to the mixed graph, we can provide a version without a line, where the columns are the only visual objects. In this case, to avoid using a line, the reference value (ACME's sales) is aligned with the X-axis (Figure 21).

3.6. Visualizing distribution

A distribution relationship, as previously defined, assigns or connects a quantitative value to each item (value or category) of an interval variable, facilitating the comparison of numeric values across the interval units.

Figure 22

Number of patients by age group in a population of epilepsy patients .

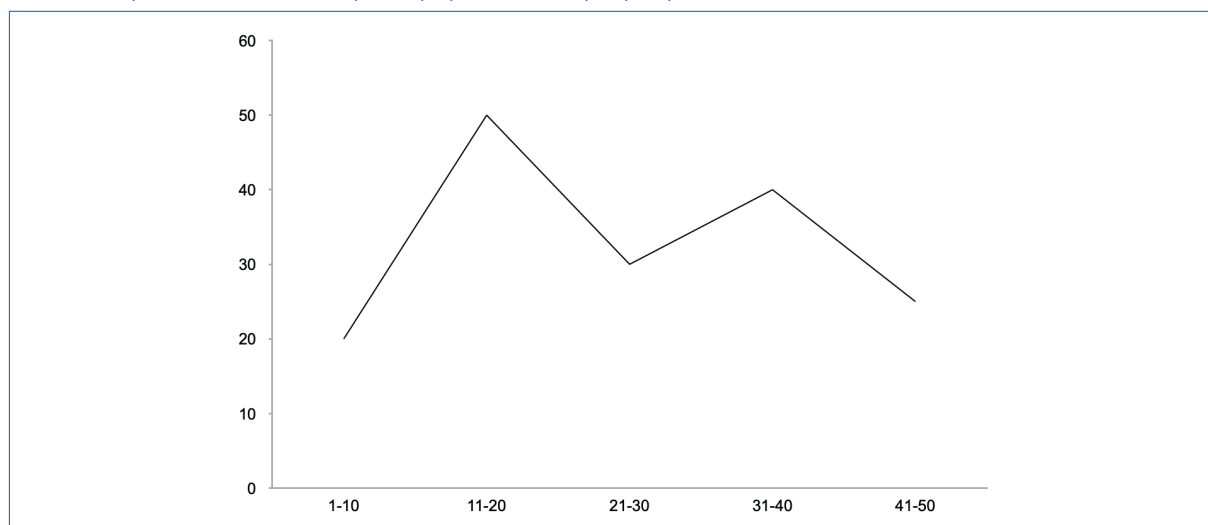


(Own elaboration, simulated data)

A chart that effectively represents this type of semantic content must use visual objects that simultaneously encode the numeric value and the associated category while illustrating how the numeric values are distributed across the range of the set they belong to, through the categories that comprise the interval variable.

Figure 23

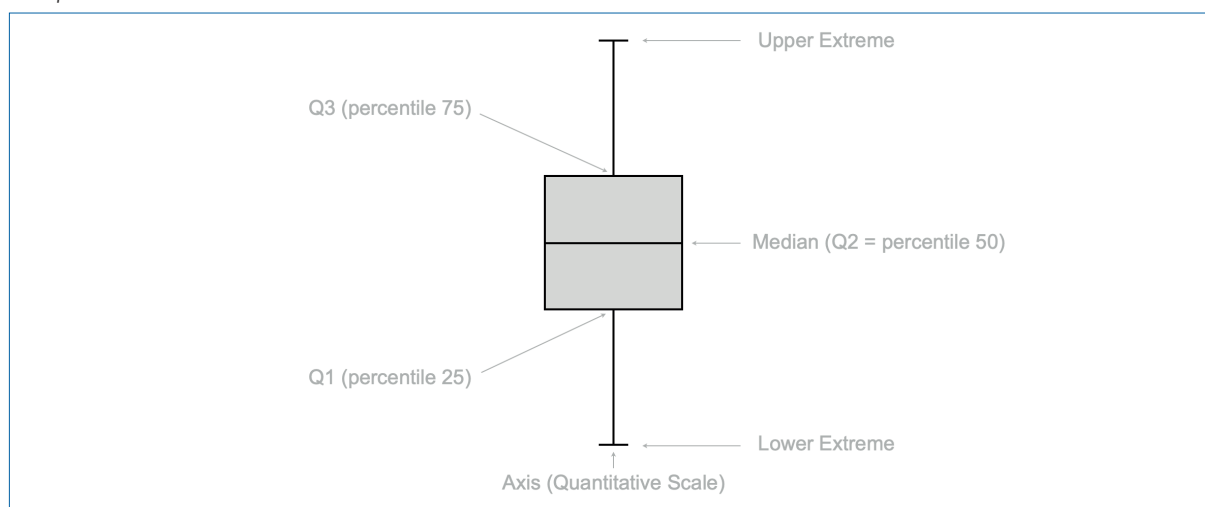
Number of patients by age group in a population of epilepsy patients.



(Own elaboration, simulated data)

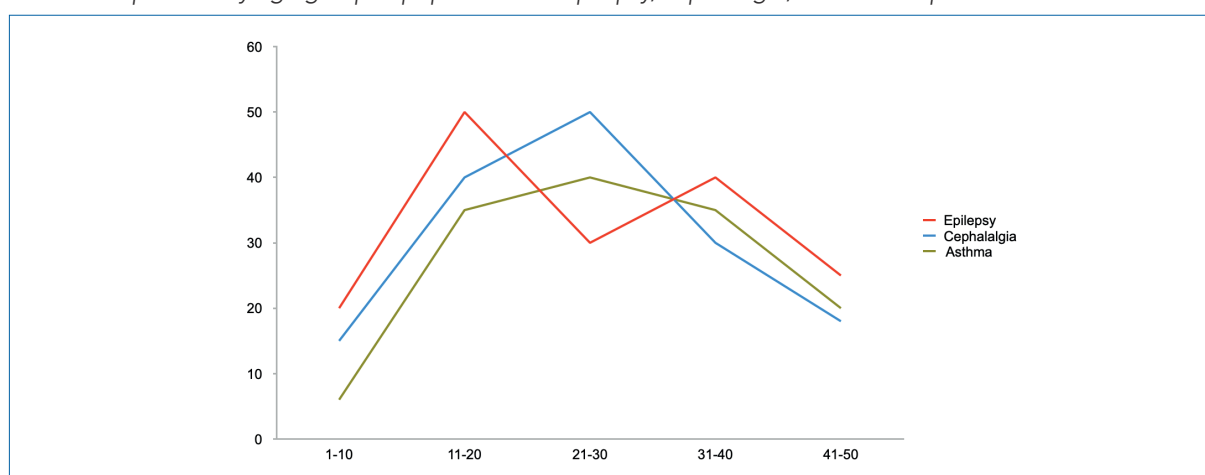
An additional consideration is whether we want to represent the distribution of a single data-set or visualize the distributions of two or more datasets within the same chart.

Figure 24
Box plot.



(Own elaboration, simulated data)

Figure 25
Number of patients by age group in populations of epilepsy, cephalalgia, and asthma patients.

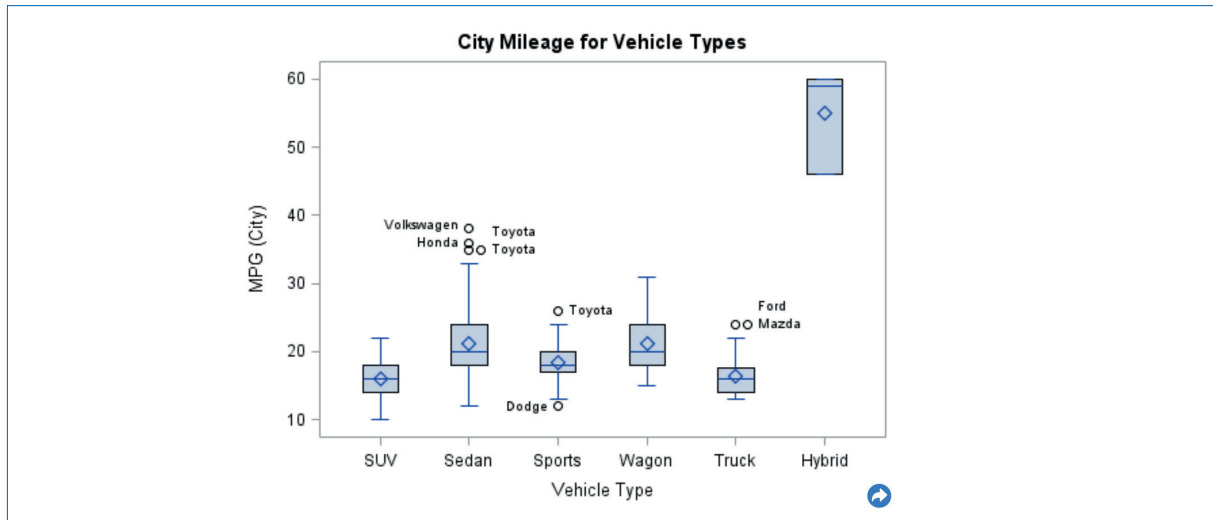


(Own elaboration, simulated data)

In the first case, when the goal is to represent the distribution of a single dataset, the charts that best encode this relationship are simple distribution charts such as the histogram (Figure 22), the frequency polygon (Figure 23), and the box plot (Figure 24).

Figure 26

Vehicle distribution by model and miles per gallon (https://support.sas.com/rnd/datavisualization/gtl/boxplot_sect2.htm).



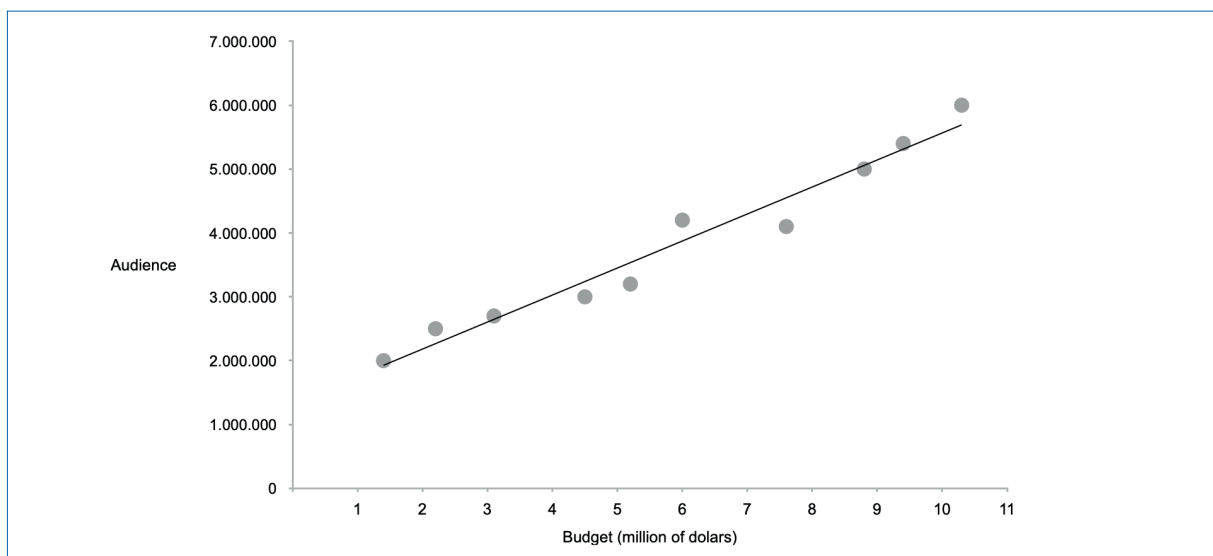
In the second case, when the goal is to simultaneously represent the distributions of two or more datasets, the charts that best encode this relationship are multiple distribution charts, such as combining different frequency polygons in one graph (Figure 25) or overlaying multiple box plots in the same visual layout (Figure 26).

3.7. Visualizing correlation

Next, we address the visualization of the final type of semantic content. A correlation relationship, as previously discussed, assigns or connects two quantitative values (each belonging to a distinct quantitative variable) to each item (value or category) of a qualitative variable. This facilitates identifying patterns (strong or weak) in the variation between these two numeric values related to those items.

Figure 27

Budget and audience for a film released last week.

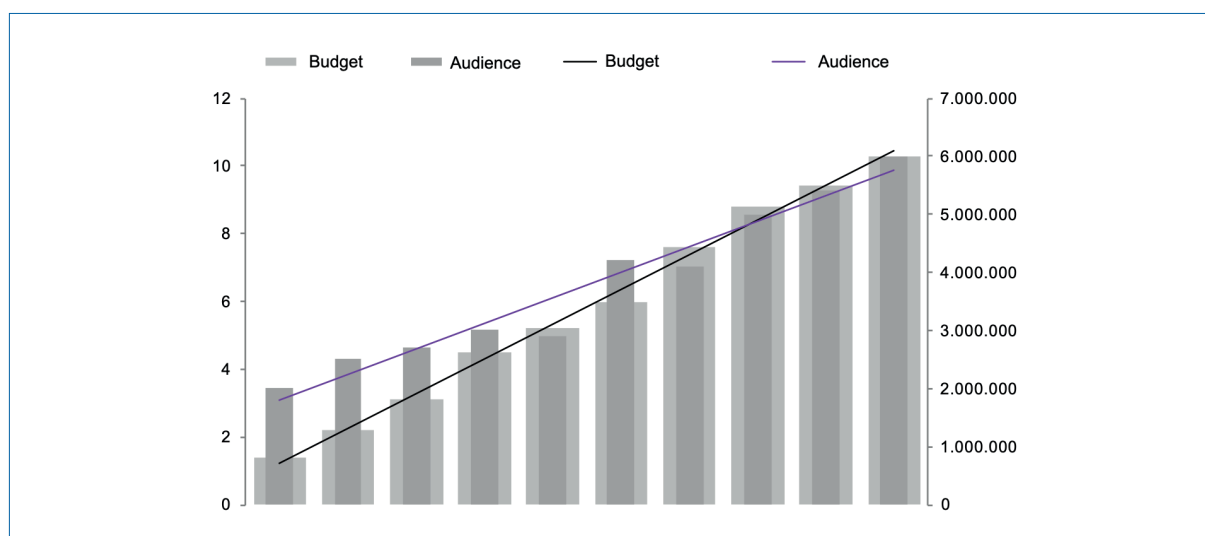


(Own elaboration, simulated data)

A chart that effectively represents this type of semantic content must use visual objects that simultaneously encode the two numeric values and the associated category while expressing whether there is a systematic relationship in which the values of one quantitative variable change relative to the values of the other.

Figure 28

Budget and audience for films released last week.

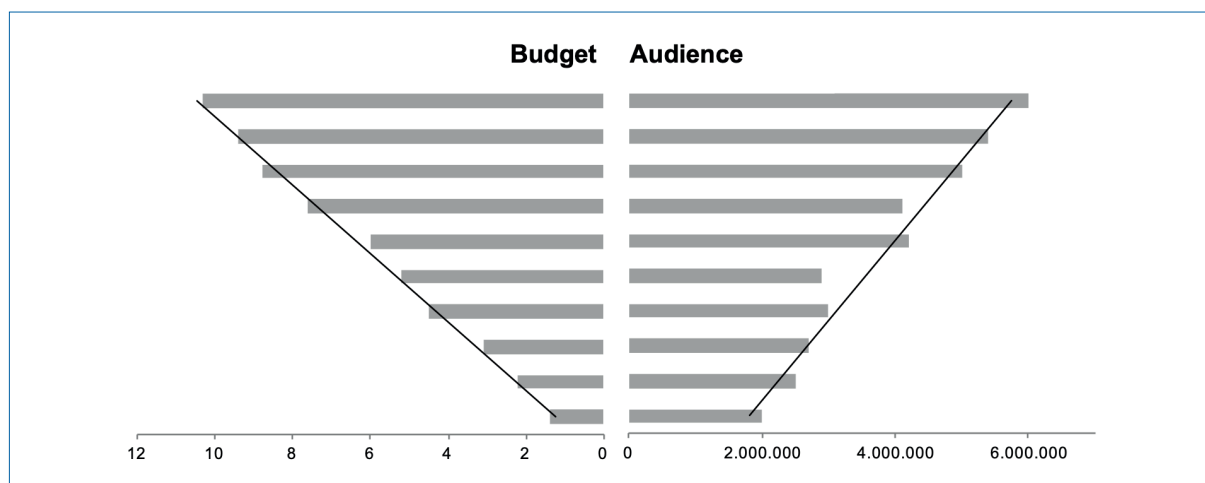


(Own elaboration, simulated data)

The most commonly used visual representation for this type of semantic content is the scatter plot, which can be understood as a specific variant of a dot chart (Figure 27).

Figure 29

Budget and audience for films released last week.



(Own elaboration, simulated data)

Some authors argue that, particularly in business contexts, scatter plots can be difficult to interpret for users unfamiliar with this type of visualization (Few, 2012). As more intuitive alternatives, when working with small datasets, it is possible to use a correlation column chart (Figure 28) or a paired bar chart (Figure 29).

4. Funding

This work is part of the Project “Parameters and strategies to increase the relevance of media and digital communication in society: curation, visualisation and visibility (CUVICOM)”. Grant PID2021-123579OB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF, EU.

5. References

- Cherven, K. (2015). *Mastering Gephi Network Visualization*. Birmingham: Ed. Packt Publishing Ltd.
- Few, S. (2012). *Show me the numbers*. Oakland: Analytics Press.
- Hearn, D. (2011). *Computer graphics with OpenGL*. Boston: Pearson.
- Hugues, J. F., & Van Dam, A. (2013). *Computer Graphics: Principles and Practice*. Boston: Addison-Wesley.
- Jones, B. (2014). *Communicating data with Tableau: Designing, developing, and delivering data visualizations*. London: Ed. Safari Books Online.
- Khan, A. (2016). *Jumpstart Tableau: A step-by-step guide to better data visualization*. London: Ed. Apress.
- Pérez-Montoro, M. (2022). *Comunicación visual de la información. Qué y cómo podemos narrar con datos*. Río de Janeiro: IBICT-UNESCO.
- Shirley, P. (2009). *Fundamentals of computer graphics*. Natick, MA: AK Peters.
- Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*, IEEE Computer Society Press, pp. 336-343.
- Zelazny, G. (2001). *Say it with charts*. New York: MacGraw-Hill.