

Infographic views and interactive cartography: tools to ensure greater access to official statistics

Abstract

To achieve a good and effective visualization of a phenomenon (static or dynamic) means knowing how to choose the right data on the basis of the element to be analyzed and what you want to communicate.

This way, anyone who reads a report, listens to a presentation, or looks at a graph can immediately understand its content.

Then, among the advantages of an effective visualization, we must add the ease of communicating the content we intend to convey to our audiences.

Illustrating data means sending information to the recipient.

In other words, it means telling a story: that's why we often talk about Data Storytelling.

Introduction

"If you asked me to describe the emerging philosophy of the day, I'd say it's data-ism," David Brooks wrote in the New York Times in February 2013.

Brooks said that in a world of increasing complexity, relying on data could reduce cognitive bias and "illuminate patterns of behavior we haven't yet noticed."

The concept of dataism is echoed by Yuval Noah Harari in his 2016 book *Homo Deus: A Brief History of Tomorrow*.

Harari argues that "dataism declares that the universe consists of streams of data and the value of any phenomenon or entity is determined by its contribution to data processing, so we can interpret the entire human species as a single data processing system, with individual humans acting as chips."

All our actions generate data: the use of digital services, public and private, searches on the internet and presences in social networks. Therefore, data is the basis from which we can learn about the world today.

We might assume that having data that measure a phenomenon will allow us to know and understand it, but in reality a multitude of data might instead confuse us.

If, however, the data are represented in a visual form they really help us to understand the phenomenon to which they refer.

When did graphs first appear in history?

In 1786, the era of charts (or data visualization) began with the publication of a book by the researcher William Playfair.

The title of his book was "The Commercial and Political Atlas". No one before Playfair had ever thought of this: he translated the map of the economy and businesses into numbers and these ones into bar or pie charts.

His great insight was the one that led him to create the first line and bar graphs and it was as follows: since in geography latitude and longitude are quantities, they can be replaced by any other quantity. For example, you can use years instead of longitude (horizontal axis) and exports/imports instead of latitude (vertical axis).

What is data visualization?

As we already saw, visualization is a necessary mediation for accessing data, because a deluge of data often does not make it understandable to measure the phenomena

to which it relates and with a multitude of data we can feel lost, without a compass. But at this point it is important to understand what is meant by Data Visualization.

(SLIDE 4) Data visualization is a way to represent information graphically, highlighting patterns and trends in data and helping the reader to achieve quick insights. Also known as "interactive visual exploration," it allows visualization of data through graphic images, color, brightness, size, shape, and movement of visual objects that represent aspects of the dataset under analysis.

It includes a number of visualization options beyond those of pie, bar and line graphs, also including heat and tree maps, geographic maps, scatter plots and other special visualizations. These tools allow users to analyze data by interacting directly with a visual representation of them.

(SLIDE 5) Moreover, as knowledge increases among humans and transactions multiply, it becomes increasingly advisable to shorten and facilitate the ways in which information is conveyed from one person to another and from one individual to many.

Visual rendering is governed by a very strong proportionality, a link of need between the data and the way they are represented.

Infographics, charts, dashboard, images, photos: therefore, they all become our **best allies**.

(SLIDE 6) To visualize is to produce a representation of a network of relationships that is unlike anything existing just because it creates its own object, giving structure and body, visually, to a series of data. A diagrammatic representation is creative because it makes visible what is invisible, but it takes on the burden of representing, in a regulated way, the logical connection that relates certain elements and gives meaning to one through the other. For this reason, the more numerous and complex are the data we have to describe a phenomenon, the more visual mediation between

us and that data will be necessary to easily and quickly understand the phenomenon they represent.

Besides, nowadays time is an increasingly precious commodity and it's crucial to tap into knowledge in a matter of seconds

(SLIDE 7) Some generic types of data visualization formats are, as seen: diagrams, tables, charts, maps, infographics and dashboards.

Some more specific examples may include:

- Area chart
- Bar chart
- Bubble diagram
- Horizontal bar chart
- Cartogram
- Pie chart
- Point distribution map
- Gantt chart
- Histogram Matrix
- Network diagram
- Polar area
- Radial shaft
- Scatter plot (2D or 3D)
- Time line
- Tree diagram
- Word cloud
- and any combination of them in a dashboard or an infographic.

(From now until and including slide 10, a few examples of data

visualization, simpler and more elaborate in terms of both implementation and reading)

(SLIDE 11) Infographics

(SLIDE 12) As the name suggests, an infographic represents a hybrid, containing within it two different components that, combined, allow it to represent an entire story.

The two parts are one informative (data, concepts and their connection) and the other visual, graphic (colors, shapes and layout).

The simultaneous use of these two components provides a way to build a visual narration that is not only engineering, but also creative (**Datatelling**).

Therefore, it is imperative that the connection between the two components is well thought to make the final narrative work.

(SLIDE 13) Alberto Cairo highlights another concept: Visualization as functional art. "It doesn't matter if you consider yourself an engineer or an artist: if you create graphics and other visualization products, the balance you achieve between these two dimensions will determine the validity of your work."

(SLIDE 14) So, what makes an infographic a good infographic?

The truth is that there is no one valid infographic model that applies to all the phenomena that we intend to represent through it.

The communication strategy to be used depends on the specific context in which it must be conveyed and therefore in the audience we intend to address.

So, when creating infographics or any other data visualization, it's important to ask yourself some questions about your target audience:

- How do users behave?
- What are users looking for?
- How do users search?
- What do users like?

(SLIDE 15) The look of your infographic must take into account all of these variables, i.e. the goals it sets out to achieve. Only once you have established exactly what these targets are, you will be able to understand what kind of infographic is suitable for your purpose at that particular time and context.

The difficulty lies precisely in choosing the right one given the many different types of templates that exist.

(SLIDE 16) Of course, a statistical infographic relies and focuses on the data we have available to illustrate a given phenomenon and thus tell the story about it.

Don't underestimate the use of appealing graphic elements to make your infographic attractive.

(SLIDE 17) In order to choose the most appropriate infographic to tell the story you want to narrate, keep a few little tips in mind:

- Research the story your data is based on (the phenomenon that your infographic wants to tell) and make sure it is reflected in your design
- Vary the type of data visualization you use - such as charts, icons and text
- Write a descriptive title to put your data in context
- Emphasize key data by using contrasting colors or matching the number to an icon

(SLIDE 18) For example, here we have an infographic of medium complexity, aimed at a non-expert audience (but not only) in which a first, more descriptive part is followed by a second part reporting the results of the Bee Colony Mortality Study (EPILOBEE) survey, also using cartograms.

(SLIDE 19) In this second example we have instead a very simple infographic, representing 4 of the 16 principles of the European Statistical Code of Practice; each principle is narrated through a single evocative image that renders immediate perception to the viewer of what is the area to which that principle applies, even before reading the text that explains what it is.

Principle 4 uses the lens to represent quality, Principle 5 uses a lock for data protection, Principle 6 uses the scale for impartiality and Principle 7 uses the checkmark to metaphorically render the idea that European Statistics adhere to a sound methodology.

(SLIDE 20) Another example is the one used by Eurostat to represent those who apply for Asylum in the EU: here the choice of the authors has fallen on an interactive infographic.

By clicking on the name of a Member State in the column of blue squares you can read the 3 main citizenships of asylum seekers who have applied in that specific EU Member State, while clicking on one of the Countries listed in the gray squares, you can see which are the 3 main EU Member States where asylum seekers with this citizenship have applied.

<https://ec.europa.eu/eurostat/news/themes-in-the-spotlight/asylum2016>

(SLIDE 21) So far we have seen that contemporary communication lives on data, graphs and statistics and we tend to associate them with rationality, scientificity and objectivity. Precisely for this reason, when we see an infographic - which puts together numbers, graphs and visual communication - the game is up. We tend to lower our defenses made of critical sense and trust these images much more than words.

In short: numbers and graphs persuade us easily.

But what if this power of infographics and data visualizations is used to deceive us?

In other words, how do we know that an infographic is actually a “**disinfographic**” and therefore misinforms rather than inform?

(SLIDE 22) Some examples of disinfographics and why they are classifiable as such:

1) Visualization is the result of bad design.

This category includes charts that do not follow the most basic conventions of how to display data in a consistent and readable manner. For example:

- graphs in 3D, which produce distortions in the perception of quantities;
- graphs that use two y-axes with different scales, perhaps without specifying it;
- bar charts that cut off the y-axis;
- line graphs with patterns that would be more obvious if the y-axis did not start at 0;
- bar graphs that do not use consistent intervals on the x-axis;
- graphs that would be clearer with a logarithmic scale on the axes.

2) The visualization shows us data of dubious or disputable origin.

An infographic is only as good as the data it wants to communicate. The most accurate design is not enough to make reliable an infographic that shows us wrong data or coming from dubious sources. This can include:

- no source and/or methodological notes are included in the data presented;
- comparisons are shown between entities that are not comparable

3) The visualization contains insufficient data to support the thesis presented

The disinfographics that fall into this category are those that, to support the message they intend to convey:

- show us data from a specific time interval and ignore earlier or later data that would instead tell a different story;
- show us data from a specific selection of geographic areas and ignore data from other areas that, if included, would tell a different story;
- show us data that has not been normalized, for example absolute values are shown when percentages or per capita would be needed, or monetary data is not adjusted for inflation and the value of money over time.

4) The visualization hides or downplays the uncertainty present in the data

Although in most of the infographics we see the data are stated as certain, without nuance, in reality any data or statistic always contains a margin of error and is a calculation of probability. Not talking about uncertainty, especially when it is crucial to accurately understanding the phenomenon, produces misleading infographics.

Nuances, bands and annotations help to communicate to the reader the presence of uncertain data. Pay attention to data from physical measurements, surveys, and forecasting models: these are perhaps the ones where showing uncertainty is critical to a proper communication of the message.

5) The visualization suggests misleading conclusions

Included in this category are graphics that, more or less transparently, encourage the reader to be convinced of conclusions that are actually without foundation. For example:

- they propose a simple correlation as a cause-and-effect relationship;
- in establishing a correlation, they ignore a third variable that might explain the correlation (i.e.: it can be seen that those who drink more green tea have a longer life expectancy. But the benefit of green tea has not been demonstrated: the link is only there because green tea drinkers are generally more lifestyle conscious people overall, eat healthy, and exercise).

(SLIDE 23) A disinfographic example. A pie chart was used to represent the pizza tastes of Brits in a multiple choice survey.

The choice was obviously influenced by the fact that the aesthetic choice overcame the engineering choice.

The result is that we have a pie that represents much more than 100%, obviously wrong.

The pie chart can be used to represent a percentage composition (where the sum is 100), but if you have to represent the result of a multiple-choice survey, it is better to use a histogram.

(SLIDE 24) (SLIDE 25) The result of this error has been a decline in the reputation of those who produced it, in fact the infographic was published on a social network and these are the comments of users. The creator of the infographic had to apologize officially.

(SLIDE 26) Here is an example where YouGov used a histogram, but could have used a pie chart.

(SLIDE 27) An interesting point of view to conclude what has been said so far on the topic of infographics and disinfographics is that of Edward Tufte, a pioneer of data visualization as well as a statistician.

Edward R. Tufte warned against inFOGraphics, using this play on words to highlight when the visual part takes over thereby obscuring the informative part with its haze.

Tufte intended visualization as a very precise relationship between the data and the visual media employed. What is outside this connection is for him a superfluous decoration (and as such confusing).

(SLIDE 28) Interactive cartography

(SLIDE 29) An interactive map is a system of graphical representation of information that can interact with user choices or the occurrence of other external events. An interactive map is generally used in geolocation and geographic mapping systems.

It allows zooming in and out, panning around, identifying specific features, querying underlying data such as by topic or a specific indicator (e.g., socioeconomic status), generating reports and other means of using or visualising select information in the map.

(SLIDE 30) The development of statistical mapping has also been encouraged by a growing demand for spatial information. Consequently, the development of methods that allow the acquisition and publication of data at any spatial level has become one of the goals of official statistics. In addition, broader use of administrative sources by official statistics and their combination with geospatial information have enabled important changes in this field. Therefore, the use of such data makes it possible to conduct research and analysis in a much broader scope than in the past.

(SLIDE 31) In addition to the main thematic content, each thematic map also has a basic content, namely topographic elements (in the GIS context called spatial reference data). They are necessary to correctly interpret the geographical relationships of the quantitative data presented. The core content also allows the elements of the thematic content to be placed in the appropriate geographic (spatial) context. The basic content is presented on a base map that is the basis for any thematic map. The base map contains the boundaries of the presented area and its division units (e.g. administrative boundaries), as well as the complementary elements of the topographic content.

The base map has to be prepared at the appropriate scale and using a correct cartographic projection.

In addition, the base map should also be adapted to the nature of the thematic content through the appropriate selection and generalization of the basic (topographic) content elements.

(SLIDE 32) An interactive map can therefore offer the added value of placing the data in a real context, but be careful not to be turned away from the main goal, which in data visualization is always to tell the best of the phenomenon that we want to narrate through the data we have. In this sense, cartography may not always be the best choice.

(SLIDE 33) (SLIDE 34) (SLIDE 35) Eurostat example of cartography use

(SLIDE 36) There are many different tools for making interactive maps, but the macro categories within which they reside are two: those that do not require any programming and others for which programming is necessary.

(SLIDE 37) (SLIDE 38) List of cartography tools that do not require programming and those that need it.

(SLIDE 39) Among the various tools mentioned, each of them obviously has different peculiarities (see in-depth analysis in the relevant slide).

(SLIDE 40) With specific reference to Eurostat, for the realization of choropleth maps they use the web tool IMAGE and make it available to their accredited users (see in-depth analysis in the relevant slide).

(SLIDE 41) Conclusions and tips

(SLIDE 42) Visual language is of crucial importance in the creation of any type of content, both online and offline. Of the five sense organs, sight is the main one and about 90% of the inputs that come to us derive from the eyes and have a processing speed of less than a tenth of a second. Images spark interest and more trust than just text; it is more immediate to explain a concept, to build a story, even on numbers, if one accompanies it with images and charts: **the world is facing a deluge of data, that's why a well-done visualization can tell a story at a glance, that's why mastering it is imperative!**

(SLIDE 43) What to do (list the tips)

- **IMPORTANT:** if one data visualization is not the most appropriate to convey the message (even if it looks good), **choose another one** that effectively delivers it.

(SLIDE 44) (SLIDE 45) What to avoid (list the tips)

- **IMPORTANT:** doing an effective data visualization correctly is not easy, and good will and correct data are not always enough. You need to study: learning how **to avoid the most common mistakes**, of course, but also how to use some strategies correctly.