



HAL
open science

Why do Data speak for themselves? A semio-pragmatic approach of data visualization as story telling medium

Philippe Useille

► To cite this version:

Philippe Useille. Why do Data speak for themselves? A semio-pragmatic approach of data visualization as story telling medium. Data Power Conference 2015, The University of Sheffield Faculty of social Sciences, Jun 2015, Sheffield - UK, United Kingdom. hal-04527527

HAL Id: hal-04527527

<https://hal.science/hal-04527527>

Submitted on 31 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Data Power Conference 2015.
Faculty of Social Sciences at The University of Sheffield, Sheffield, U.K.

**

Why do Data speak for themselves?
A semio-pragmatic approach of data visualization as story telling medium.

Philippe USEILLE
Assistant Professor
Laboratoire DeVisu
University of Valenciennes et du Hainaut-Cambrésis (UVHC) -FRANCE

Introduction

Data journalism utilizes data visualizations as an alternative method of storytelling in response to the issue of information overload and the spread of misinformation. Online journalism faces the challenges of being as up to date as possible, which does not allow for deeper and more reflective reporting. Data journalism provides a solution to this issue. News visually without oversimplifying it. Furthermore, media users may perceive data journalism as a way of ensuring reliable news reporting. Data visualization can enhance the objectivity of journalism by allowing visual language and figures to speak for themselves, rather than relying on the potentially unreliable voice of the journalist. This approach can help to address the challenges of digital journalism and the broader issue of distrust towards news and journalists.

The question remains: how do data appear to speak for themselves? This working paper highlights the fact that data and their visualization involve various forms of mediation. Following the semiology of graphics approach, the paper emphasizes the different types of mediation that data journalism employs to create a renewed storytelling, from data production to the process of meaning-making.

1. What are the virtues of data journalism?

Discussions on data journalism often emphasize the importance of mediation, which acts as a bridge between raw data and the public. Data journalism is often likened to a refining process that transforms raw data into meaningful and relevant content. Additionally, data journalism promotes a quantitative approach to topics, utilising graphics to express correlations between facts in a clear and concise manner. The visual approach allows for the examination of complex subjects in a measurable way, combining readability and clarity to provide access to a lot of information.

In addition, data journalism allows users to be informed inductively through dynamic communication between the interface and the database (Degand et al., 2012). Similar to participative journalism, it promotes co-construction of information with the user. The use of interactive design enhances informational efficiency by facilitating data exploration and making it more engaging. According to these discourses, data journalism is an effective way to present facts and convey meaning. It is crucial to maintain objectivity and avoid subjective evaluations. However, it is important to study the properties of graphics as a medium to better understand this.

Data visualization employs images that represent objects of a certain reality using iconic signs, according to C. S. Peirce. Diagrams operate by similarity, for example, the size of the rings represents the importance of the populations they depict. This iconic dimension of the diagram enhances its transparency and, therefore, the accessibility of information.

What type of informational efficiency do they achieve? Prior to representing the world or conveying a message, an image provides a framework for the perceptual field. Therefore, it is important to study its visual properties, spatial organization, and formal structure. The theory of information visualization, as explained by the French cartographer Jacques Bertin, utilizes specific graphic sign systems such as diagrams, maps, and networks. It is one of the first systematic studies of the different ways in which graphical representations encode data.

2. Semiology approach to graphics by Jacques Bertin

In his 1967 book, *Semiology of graphics* (Bertin, 1967), Jacques Bertin explores two main questions: when do we need a graphic representation, and what kind of representation do we need? Graphics benefit from the properties of visual perception and serve as a language for the eye. As a monosemic system, it is unambiguous and forms the rational part of the world of images. The system uses visual signs to show abstract relationships, such as inclusion, similarity, and succession. Therefore, the graphic is considered to be objective because it offers a concrete and tangible starting point for discussion, reflection and understanding. The effectiveness criteria are measured by the length of time the graphic representation is observed. If one construction requires a shorter observation time than another to obtain a correct and complete answer to a given question, it can be considered more efficient for that question.

Jacques Bertin suggests that the creation of graphs involves the processing and selection of data and the identification of information components. For instance, when analyzing road accidents in a country, it is necessary to differentiate the element being studied (the number of victims) and the components of information, such as the type and context of the accidents. This process permits us to choose an appropriate representation: a map or a certain type of diagram. Information is used to answer a question, so reading takes place between meanings and not between signs and their meaning, which is clearly codified (and unambiguous) beforehand.

Jacques Bertin identified the visual features of the graphic system as a mediator: the two dimensions of the plane, the aspects of the mark (size, height, surface, number of signs), the value (white, grey, black), the grain and the colors, the direction and, finally, the shape. Bertin also identified different types of "implantations" (the point, the line), which can be arranged in different ways (straight line, circle). By combining these rules, various graphic shapes can be produced: columns, bars, cubes, semi-circles... Based on these general properties, Jacques Bertin distinguishes three groups of graphic representation: diagrams, networks and maps. He presents the rules of graphic mediation of information and theorizes their readability and efficiency (objectivity, monosemic system compared to the current storytelling in the case of data journalism).

3. Data, information and news.

The theory proposed by Jacques Bertin includes a definition of data and information. According to his semiology of graphics, data is expressed as the relationship between two elements. Information, on the other hand, connects various concepts. For instance, a variation of quantity could be linked to time. Therefore, data visualization, and consequently data journalism, implies an underlying vision of information, as Jacques Bertin clearly expressed.

If information consists of data, these data are not raw, unprocessed. In the field of philosophy of information, Luciano Floridi explains that well-formed data means "that the data are rightly put together, according to the rules (syntax) that govern the chosen system, code, or language being used" (Floridi, 2010, p. 20). Floridi adds that "meaningful data must comply with the semantic of the chosen system, code or language" (Floridi, 2010, p. 21). This is particularly

important in the case of statistical data, which is used in graphs and data journalism: data are clearly well-formed and meaningful, not raw. The data should be well-formed and meaningful, rather than raw, and should integrate mediations related to their production, including the scientific, methodological, and social aspects.

Graphic information processing involves establishing relationships between primary data to construct information that emerges during user reading. Bertin's semiology can be seen as immanent (the information is in the text) and does not emphasize the role of the reader. Nowadays, it is commonly thought that information is produced through the interaction of a text with a user (the sense-making level), including interrelationships with other texts (not only graphic representation in the case of data journalism) in the specific context of a user. To fully understand the role of data visualization in relation to different kinds of storytelling formats, scholars need to study the reader's informational process.

The use of graphics, based on the properties of visual thinking and the rationality of the graphic (derived from numbers and a monosemic system), is a language which breaks with linear modes of representation, mainly verbal. Images make it easier to perceive "patterns", as suggested by Gestalt theory (laws of proximity and similarity, for example). The power of visual signs lies in their analogy to the complex system to be decoded. It is widely accepted that graphic representation mediates an epistemic point of view about what is represented. For instance, the representation of the "tree" combines unity and multiplicity in an appropriate way for a genealogy or a classification, two cases that call for order, symmetry, and regularity. But it also links centralism, finality, and essentialism to the network, as Manuel Lima argues (Lima, 2013). Therefore, visual representation is not neutral, but in many cases biased. Today, advocates of the "science of complexity", such as Manuel Lima, are using data visualization to represent all sorts of networks. So the answer to Jacques Bertin's question "What kind of graphic representation do we need?" is not just technical, but implies an epistemological vision that involves data and the information process. As Alfred Korzybski said, 'A map is not the territory' and it shows both things and the way we think about things (Korzybski, 2013).

To what extent does the use of data alter journalism's relationship with facts and the notion of truth? In fact, not much. Journalistic mediations involve many filters by which data are selected and shaped to become information (in the mind of the reader). This process is not very different in the case of data journalism, even if quantification and visual representation produce a different language. The question of credibility is not only a matter of the format or form of the presentation. As noted by Bill Kovach and Tom Rosenstiel, journalists need to ask 'What does my audience need to know in order to judge the value of this information' (Kovach and Rosenstiel, 2014)?

References

- Bertin, J., 1967. *Sémiologie graphique: les diagrammes*. Mouton.
- Degand, A., Grevisse, B., Singer, J.B., Kayser-Bril, N., 2012. *Journalisme en ligne: pratiques et recherches*. (No Title).
- Floridi, L., 2010. *Information: a very short introduction*, Very short introductions. Oxford University Press, Oxford ; New York.
- Korzybski, A., 2013. *Une carte n'est pas le territoire: Prolégomènes aux systèmes non-aristotéliens et à la sémantique générale*.
- Kovach, B., Rosenstiel, T., 2014. *The elements of journalism: what newspeople should know and the public should expect*, Revised and updated third edition. ed. Three Rivers Press, New York.
- Lima, M., 2013. *Cartographie des réseaux: l'art de représenter la complexité*. Eyrolles, Paris.