



# Introduction to Data Visualization

Simone Diniz Junqueira Barbosa<sup>(✉)</sup>  and Gabriel Diniz Junqueira Barbosa 

Pontifical Catholic University of Rio de Janeiro (PUC-Rio),  
Rua Marques de Sao Vicente, 225, Gavea, Rio de Janeiro, RJ, Brazil  
[simone@inf.puc-rio.br](mailto:simone@inf.puc-rio.br), [gabrieldjb@gmail.com](mailto:gabrieldjb@gmail.com)

**Abstract.** The objective of this course is to provide students, young researchers, and practitioners without a formal education in data visualization with an introduction and overview of the field. In addition to introducing basic concepts, the course will present diverse visualizations for different kinds of data (*e.g.*, categorical, numeric, hierarchic, network, temporal, and spatial data); and for different kinds of visualization tasks and goals (*e.g.*, retrieve value, filter, compute derived value, find extremum, sort, determine range, characterize distribution, find anomalies, cluster, and correlate). We will also discuss the role of data stories to convey data-driven insights.

**Keywords:** Data visualization · Data types · Visualization tasks · Storytelling with data

## 1 Learning Objectives

In this course, students will learn how to produce data-driven visualizations for both exploration and communication purposes, given the types of data and visualization goals and tasks. By exploring real-world examples, they will learn to identify and avoid misleading visualizations. They will also learn how to convey their insights through data stories.

## 2 Content

The course material is divided into seven sections:

1. **Why visualize:** this section argues for the need to visualize data to improve understanding and communication. By integrating statistical data and visualizations, we show how we are able to improve the reader's understanding of the underlying information.
2. **Human perceptual system:** this section presents basic concepts of the human perceptual system and how they apply to visualization, such as visual variables – *e.g.*, color (hue, luminance, and saturation), size (length, height, width, area), shape, orientation, opacity, position – and Gestalt principles.

3. **Types of data:** this section introduces different types of data, *e.g.*, categorical, numeric, hierarchic, network, temporal, and spatial data, illustrated through real-world examples.
4. **Basic charts:** this section describes widely used charts and some of their variations, *e.g.*, bar chart, line chart, histogram, dot plot, box plot, scatter plot, bubble plot, maps, trees, networks, and heat maps. We show good and bad examples, and discuss common mistakes and misleading visualizations.
5. **Visualization tasks:** this section describes how the set of suitable visualizations is determined by the combination of data types and visualization tasks – *e.g.*, retrieve value, filter, compute derived value, find extremum, sort, determine range, characterize distribution, find anomalies, cluster, and correlate [1]. As in some situations a single visualization is insufficient, we explore how visualizations may be combined to achieve an exploration or a communication goal.
6. **Interactivity in data visualization:** this section presents some interaction mechanisms for manipulating data visualizations, *e.g.*, filtering, zooming, brushing and linking. For high-volume data, complex data, or complex visualization tasks, interactivity is essential for proper understanding.
7. **Storytelling with data:** this section discusses how to tell stories by using visualizations within a narrative so as to communicate data-driven insights. As humans learn better through stories, data stories may create empathy and be more memorable than isolated data facts and visualizations.

For pedagogical reasons, these sections will not be presented in a strict sequence, *i.e.*, content from different sections may be intermingled.

### 3 Course Format

The course is formatted as a 3-h lecture split in two sessions, with several examples given throughout.

The examples will be explored in both top-down and bottom-up approaches: (i) by presenting students with visualization goals, we will explore how to produce data-driven visualizations to achieve them; and (ii) by presenting students with (well and poorly designed) visualizations, they will be provoked to reflect on their quality and consequences, as well as discuss alternative visualizations to achieve the same goal.

All slides, public data sets, and code (in R and Python) used in the course will be provided beforehand, as well as instructions on how to install the (open source) software necessary to run the examples.

### 4 Intended Audience

The intended audience is made up of students, young researchers, and professionals in information and communication technology-related fields who have not yet had a systematic exposure to data visualization. As an introductory course,

it is not suitable for people with advanced knowledge of visualization concepts and techniques.

Programming knowledge of R and Python is not required, but students are encouraged to bring their own laptops to run the course examples throughout the sessions.

## References

1. Amar, R., Eagan, J., Stasko, J.: Low-level components of analytic activity in information visualization. In: 2005 IEEE Symposium on Information Visualization, INFOVIS 2005, pp. 111–117, October 2005. <https://doi.org/10.1109/INFVIS.2005.1532136>
2. Berinato, S.: *Good Charts: The HBR Guide to Making Smarter, More Persuasive Data Visualizations*. Harvard Business Review Press, Boston (2016)
3. Bertin, J.: *Semiology of Graphics: Diagrams, Networks, Maps*, 1st edn. Esri Press, Redlands (2010)
4. Cairo, A.: *The Functional Art: An Introduction to Information Graphics and Visualization*, 1st edn. New Riders, Berkeley (2012)
5. Cairo, A.: *The Truthful Art: Data, Charts, and Maps for Communication*, 1 edn. New Riders (2016). Place of publication not identified
6. Card, S.K., Mackinlay, J.D., Shneiderman, B.: *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, Burlington (1999)
7. Few, S.: *Now You See It: Simple Visualization Techniques for Quantitative Analysis*, 1st edn. Analytics Press, Oakland (2009)
8. Few, S.: *Information Dashboard Design: Displaying Data for At-a-Glance Monitoring*, 2nd edn. Analytics Press, Burlingame (2013)
9. Kirk, A.: *Data Visualisation: A Handbook for Data Driven Design*, 1st edn. SAGE Publications Ltd., Los Angeles (2016)
10. Knaflic, C.N.: *Storytelling with Data: A Data Visualization Guide for Business Professionals*, 1st edn. Wiley, Hoboken (2015)
11. Munzner, T.: *Visualization Analysis and Design*, har/psc edn. CRC Press, Boca Raton (2014)
12. Steele, J., Iliinsky, N. (eds.): *Beautiful Visualization: Looking at Data through the Eyes of Experts*, 1st edn. O'Reilly Media, Sebastopol (2010)
13. Tufte, E.R.: *Envisioning Information*. Graphics Press, Cheshire (1990)
14. Tufte, E.R.: *Visual Explanations: Images and Quantities, Evidence and Narrative*. Graphics Press, Cheshire (1997)
15. Tufte, E.R.: *The Visual Display of Quantitative Information*, 2nd edn. Graphics Press, Cheshire (2001)
16. Tufte, E.R.: *Beautiful Evidence*. Graphics Press, Cheshire (2006)
17. Ware, C.: *Information Visualization: Perception for Design*, 3rd edn. Morgan Kaufmann, Waltham (2012)
18. Yau, N.: *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics*, 1st edn. Wiley, Indianapolis (2011)
19. Yau, N.: *Data Points: Visualization That Means Something*, 1st edn. Wiley, Indianapolis (2013)