Data visualization and communication

Dimitris Karlis, Professor, AUEB, karlis@aueb.gr

Nikos Platis, Assistant Professor, University of the Peloponnese, nplatis@uop.gr

Overview

Communicating clearly and effectively about the patterns we find in data is a key skill for a successful data scientist. Visualizations are graphical depictions that can improve the analysis and comprehension of data. In this course, the principles for creating effective visualizations will be analysed and a multitude of ways for visualizing various types of data will be showcased; different tools such as R and Tableau will be used to transform data and create visualizations. Visualizations will be paired with verbal analyses and reporting to create "data stories" that convey a message or information effectively. Assignments will allow students to gain experience with reporting on complex patterns and present results with graphics and prose.

Key Outcomes

After completing the course, the students will be able to:

- understand how data visualization works, in terms of human visual perception and cognition
- learn how data visualization is done in practice, including methods to plot various types of data, interaction techniques, etc.
- create data visualizations using popular software tools, such as Tableau, R, and/or other
- deliver effective and compelling ways of communicating / presenting the visualized data following best practice guidelines

Requirements and Prerequisites

This course combines theory and practice, including multiple individual exercises and culminating in a hands-on group project.

The course does not assume any prior experience in R, Tableau or the other software tools to be presented. However, basic knowledge of programming and computer science concepts is required.

Required Course Materials

There is no required textbook. All course materials will be provided in class and available for downloading.

Students will need to bring their laptops in class in order to try out interactively the material being presented.

Books

There are many books on the subject; the following selection provides a good foundation for those students who wish to delve deeper on the topics discussed in class:

- Bederson, B.B. & Shneiderman, B., 2003. The Craft of Information Visualization: Readings and Reflections, Morgan Kaufmann Publishers.
- Chen, C., Hardle, W. K., & Unwin, A. (2007). Handbook of data visualization. Springer Science & Business Media.
- Ward, M.O., Grinstein, G. & Keim, D., 2010. Interactive Data Visualization: Foundations, Techniques, and Applications, A K Peters Ltd.
- Ware, C., 2004. Information Visualization, Second Edition: Perception for Design 2nd ed., Morgan Kaufmann.

The following books can provide further insights into the topics of the course:

- Cleveland, W.S., 1993. Visualizing Data 1st ed., Summit, NJ, USA: Hobart Press.
- Card, S.K., Mackinlay, J. & Shneiderman, B., 1999. Readings in Information Visualization: Using Vision to Think S. K. Card, J. Mackinlay, & B. Shneiderman, eds., Morgan Kaufmann.
- Johnson, J., 2010. Designing with the Mind in Mind: Simple Guide to Understanding User Interface Design Rules, Morgan Kaufmann Publishers Inc.
- Tufte, E.R., 2001. The Visual Display of Quantitative Information 2nd ed., Graphics Press.
- Cleveland, W.S., 1994. The Elements of Graphing Data 2nd ed., Summit, NJ, USA: Hobart Press.

Software/Computing requirements

- R free software environment for statistical computing and graphics. https://www.r-project.org/. Special packages will be used such as ggplot2, lattice. Students are required to download an up-to-date version of R.
- Tableau, http://www.tableau.com/. Educational licences will be provided.

Grading

Students will be graded as follows:

- Participation - 10% (individual assessment).
- Mini assignments - 30%. These are individual assignments, such as simple exercises, small essays, work with data, search for cases, etc. that relate to each lecture and aim at establishing the student's comprehension of the lecture. They have to be submitted after each lecture.
- Group Project - 50%. There will be one group project throughout the semester. Groups will be randomly selected. The group project includes the submission of documentation in the form of a detailed report.
- Final presentation – 10%. It refers to the presentation of the group projects above. Here the interest lies on the personal oral communication of the findings/ methodology and the entire story in a scientifically correct and fascinating manner.

The course does not have exams.

Participation

In-class contribution accounts to a 10% of your grade and is an important part of our shared learning experience. Your active participation helps us to evaluate your overall performance. You can excel in this area if you come to class on time and contribute to the course by:

- Providing strong evidence of having thought through the material.
- Advancing the discussion by contributing insightful comments and questions.
- Listening attentively in class.
- Demonstrating interest in your peers' comments, questions, and presentations.
- Giving constructive feedback to your peers when appropriate.

Please arrive to class on time and stay to the end of the class period. Chronically arriving late or leaving class early is unprofessional and disruptive to the entire class. Repeated tardiness will have an impact on your grade. *Turn off all electronic devices prior to the start of class. Cell phones, tablets and other electronic devices are a distraction to everyone.*

You will be informed in advance for lectures that will require the use of your laptop.

Assignments

Late assignments will either not be accepted or will incur a grade penalty unless due to documented serious illness or family emergency. Exceptions to this policy for reasons of civic obligations will only be made available when the assignment cannot reasonably be completed prior to the due date, you make suitable arrangements, and give notice for late submission in advance.

Attendance Requirements

Class attendance is essential to succeed in this course and is part of your grade. An excused absence can only be granted in cases of serious illness or grave family emergencies and must be documented. Job interviews and incompatible travel plans are considered unexcused absences. Where possible, please notify the instructor in advance of an excused absence.

Students are responsible for keeping up with the course material, including lectures, from the first day of this class, forward. It is the student's obligation to bring oneself up to date on any missed coursework.

Code of Ethics

Students may not work together on individual graded assignments unless the instructor gives explicit permission.

Exercise integrity in all aspects of one's academic work including, but not limited to, the preparation and completion of all other course requirements by not engaging in any method or means that provides an unfair advantage. In any case of doubt, students must be able to prove that they are the sole authors of their work by demonstrating their knowledge to the instructor.

Clearly acknowledge the work and efforts of others when submitting written work as one's own. Ideas, data, direct quotations (which should be designated with quotation marks), paraphrasing, creative expression, or any other incorporation of the work of others should be fully referenced. No plagiarism of any sort will be tolerated. This includes any material found on the internet. Reuse of material found in question and answer forums, code repositories, other lecture sites, etc., is unacceptable. You may use online material to deepen your understanding of a concept, not for finding answers.

Please report observed violations of this policy. Any violations will incur a fail grade at the course and reporting to the senate for further disciplinary action.

Course Syllabus

The course comprises **eleven** units of three hours each.

Unit 1: Introduction – visual perception

Introduction to the course (purpose, goals, structure).

The human brain's perceptual and cognitive characteristics. Understand the basic principles of spatial, temporal, and color processing by the human visual system and how these relate to creating effective information visualizations.

Presentation of real-world examples & critique.

Unit 2: The basics of data visualization

Basic design principles for creating plots with data.

Do's and don'ts, examples of good use and misuse of plots. Best practices and different methods to plot different categories of data (e.g., symbolic, tabular, networked, hierarchical, textual datasets).

Units 3-4: Data visualization elements

More advanced galleries: new plots for specific types of data such as dependent data, time series, spatial data, categorical data, social networks, etc. Plotting multidimensional / multivariate data: how to summarize them and how this relates to dimension reduction (not the methods but the output). These lectures may include issues about spatial/geographical information representation. R illustrations.

Units 5-6: Interactive data visualization

Enhancing the known techniques with interactivity: Direct Manipulation, Immediate Feedback, Linked Displays, Animate Shift of Focus, Dynamic Queries, Semantic Zoom, Focus+Context, Details-on-Demand, Output -> Input, etc. Dynamic Data Manipulation and Presentation. Interactive plots.

Interactive Infographics on the WWW and Animation. Tableau / R implementations and illustrations.

Unit 7: Interactive data visualization - implementations

Tableau / R implementation workshop. More practical examples with real data.

Unit 8: Communicating data

Dashboards and story points. Infographics. User-centered design (UCD) - focus on communicating results to specific audiences. Understanding your stakeholders. Storytelling, how to produce effective data analysis reports and compelling presentations, or persuasive “data stories” that use visualizations, capitalizing on well-tested methods and design principles.

Unit 9: Layered grammar of graphics

Tools that enable the user to concisely describe the components of a graphic. The theory of layered grammar of graphics: basic ideas, transformations, geometric objects, facets, aesthetics. Examples. Implementations in R through ggplot2.

Unit 10: Hands-on Project

The students will be given a dataset few days before and come to the class to work on teams on questions that will be given in the class. The students will have 2 hours to work with the already known data to produce a small presentation and 10 minutes for each team to show the findings. (This is one of the mini assignments)

Unit 11: Final presentation of projects

Each team presents their course project to peers and instructors.

The final project will involve:

- performing data analyses and creating graphs from a real-world dataset, using tools
- designing and organizing a presentation of the date (developing a data story) following best practice guidelines
- choosing the graphs and charts that best support the proposed presentation
- communicating the data in a way that would be clear to the targeted users/stakeholders.

Written documentation due.