Computing Everywhere proposal
Data and computational journalism
Daniel Trielli

About the topic

Elements that we now recognize as hallmarks of data journalism have always been a factor in journalism. Statistical analysis of numbers, visualization through charts and maps are features have been turning up in newspapers since the 19th century. But in more recent years, as more data become publicly accessible and a wider variety of computational and analytical tools are easily available, data journalism and its most advanced branch, computational journalism, gain more and more prominence. As audiences become more aware and attuned to data journalism, news production institutions require more practitioners of it, generating a demand for better data-qualified journalists and journalism-aligned scientists. However, understanding data and computational journalism – its opportunities, concerns and limitations – it is useful not only for participants interested in working with it, but anyone drawn to data analysis, journalism, and even in a wider perspective, anyone who is keen on handling data and translating their findings to a wider audience.

Overview of approach

The approach for this session will be a combination of lectures and exercises so participants can have a full overview of what is data and computational journalism. We will discuss how data journalism is done, what are the challenges, and what to look for when practicing it. We will also do a series of hands-on exercises using tools that are used in real life. Finally, we will discuss the next frontiers in the relationship between technology and journalism, in the shape of computational journalism.

Requirements

As previously noted, although this session is ostensibly about data and computational journalism, participants can be from any area where there is interest in working with data and making sense of it to reach a wider audience. Participants will only need a laptop computer with Excel, R and RStudio installed.

Learning objectives

By the end of the session, participants shall be able to:
- Understand important concepts and mindsets of data journalism
- Be aware of and have contact with tools for data analysis and visualization
- Explore the combination of journalism and technology in the field of computational journalism

About the instructor

Daniel Trielli is a PhD student at the Media, Technology and Society program at Northwestern University. He is researching how news reaches the public in our algorithmically-defined information space, and how computational journalism can be applied to investigate complex issues. He is part of the team that created AlgorithmTips.org, an online database of government-run algorithms, with the goal of providing better access to journalists and researchers interested in covering this beat. Daniel worked as a journalist for over a decade in Brazil.

Lesson plan

Part I: Fundamentals of data journalism

Lecture and discussion 1 - What is data journalism (10 minutes)
- **Topic:** The motivation and history of data journalism

Exercise 1: Excel like a pro (20 minutes)
- **Overview:** We are going to learn how use Excel by transforming data into statistics.
- **Discuss:** Looking at the summary statistics for each crime, is there anything journalistically interesting? Is there something we can investigate further to look for a narrative?

Part II: Thinking critically about statistics

Lecture and discussion 2: How we can extract meaning from data (10 minutes)
- **Topic:** What do data journalists look for when they look at statistics; trends versus outliers (and how both can be interesting); the same number can tell many stories

Exercise 2: Brief introduction to R (30 minutes)
- **Overview:** Using code to generate data analysis
- **Discuss:** Coding is faster, better and more transparent

Lecture and discussion 3: Avoiding pitfalls (10 minutes)
- **Topic:** Becoming aware of the errors that can lead to erroneous conclusions; correlation versus causation; limitations of data; explaining the data through visualization

Exercise 3 (whole room): What's wrong with this chart? (15 minutes)
- **Overview:** Let's look at bad examples of data visualization and think how they can be improved
- **Discuss:** What are the hallmarks of a good data viz?

Part III: Computation and journalism

Lecture and discussion 4: Computational journalism and its applications (10 minutes)
- **Topic:** Exploring the combination of journalism and technology

Exercise 4 (groups of 4): Imagining Computational Journalism tools (15 minutes)
- **Overview:** Thinking about the data we have of Chicago arrests and thinking about what computational journalism can do, let's brainstorm some ideas of computational journalism projects

Further readings and references

- Gray, Jonathan, Lucy Chambers, and Liliana Bounegru. "The data journalism handbook: how journalists can use data to improve the news." O'Reilly Media, Inc., 2012.(Link)
- Cohen, Sarah, James T. Hamilton, and Fred Turner. "Computational journalism." *Communications of the ACM*, 2011. (Link)
- Berret, Charles and Phillips, Cheryl. "Teaching data and computational journalism." Columbia Journalism School, 2016. (Link)