Short Course: **Data Journalism**

Theresa Gessler (→✉) and Lucas Leemann (→✉)

1   Overview

This two-day short course provides a first contact with data journalism. We start with a short section on data journalism (DJ) and what makes DJ distinct from journalism. In a nutshell, DJ can be seen as what is created when a data scientist becomes a journalists and starts writing. She will bring a large backpack full of tools and these can be applied to various parts of the production of insight and news. We decided that in this two-day course we will try and give you a sense of four relevant tools: public opinion polling, mapping the political space, web scraping, and text analysis. We chose these four areas because they allow us to highlight some relevant points about DJ and at the same time you learn how to analyze data. Each block concludes with empirical examples and code so that you have already in this course applied the tool for the first time. We will also provide to each block a list with additional resources if you wish you dig deeper.

2   Learning Objectives

Students who participate in **Short Course Data Journalism** can expect to achieve the following goals:

- To develop an understanding of what data journalism is.
- To develop a better sense of what a data scientist does.
- Learn tools for four highly relevant fields:
    - Public opinion and surveys.
    - Mapping the political space.
    - Harvesting online data.
    - Text as data.

3   Course Materials

Texts

All papers with a **(*)** sign should be read in advance of class. For some texts this is straight-forward but other texts are more challenging. Try to understand the general strategy even if the specific technical details are not always clear. Write down your questions so that we can cover them in class. Texts with no **(*)** sign are meant as additional resources and can be consulted after the class.

All texts will be shared two weeks in advance. You can find all materials here on →DropBox.

Software

It is crucial that you have access to RStudio, since we shall be using this throughout the course. We assume that everybody is familiar with RStudio.

4   Covid-related Adjustments

Due to the ongoing public health challenges we are restricting this class to Zoom exclusively.

! Respect and patience. Digital teaching and learning is challenging for everybody. Hence, we want to be patient and respectful with all people in class.

! Students are asked to turn their camera on and to mute their microphone. For teaching it is essential that we can see your faces and gauge from it how well the content is understood or not. Feel free to use an alternative background or to blur the background.

! One common problem is how easily we get distracted in online teaching. To prevent this we will engage in "cold-calling" where we randomly call upon a student to answer a question during class. While we usually would not do this, we will try this out in the hope that it helps all to stay more focused and engaged. Research shows that it is a highly effective measure leading to better learning outcomes (→ Link).

Day 1

21.04.2021: 08.30-9.00

General Short Course Introduction: What is Data Journalism? Tools matter.

- Media examples:
  - (*) Blumer, Claudia. "Frauen kommen besser weg" *Tages Anzeiger* 24.08.2019. (`Tages-Anzeiger 1.html`)
  - (*) "No Place for Foreigners" https://www.hanna-und-ismail.de/english/index.html SPIEGEL ONLINE.

21.04.2021: 9.15-10.00

Public Opinion and Surveys. What do we know about public opinion? How do measure it?

- Groves, Robert M and Lars Lyberg. 2010. "Total survey error: Past, present, and future" *Public Opinion Quarterly* 74(5): 849-879.
- Stefan Bühler. "Stimmfaule Jugend? Ganz im Gegenteil!" *Neue Zürcher Zeitung* 13.04.2014 (`NZZ 1.pdf`).

21.04.2021: 10.15-12.00 (with breaks)

Public Opinion and Surveys. What do we know about public opinion? How do measure it?

- Ansolabehere, Stephen, and Douglas Rivers. 2013. "Cooperative survey research" *Annual Review of Political Science* 16: 307-329.
- Media example:
  - (*) (how not to do it) von Burg und Tischhauser. "No Billag-Befürworter liegen deutlich vorne" *Sonntagszeitung* 03.12.2017 `SoZe 1.html`
  - (*) (watch out) Nate Cohn. "How One 19-Year-Old Illinois Man Is Distorting National Polling Averages" *New York Times* 12.10.2016 `NYT 2.pdf`
- **Lab in RStudio:** Work with public opinion data from landmark initiative votes.

Lunch Break

21.04.2021: 13.00-14.45

Mapping the political space. How do we measure where people stand in the political space?

- (*) Clinton, Joshua D, Simon Jackman and Doug Rivers. 2004a. " 'The Most Liberal Senator'? Analyzing and Interpreting Congressional Roll Calls." *Political Science and Politics* 37(04):805-811.
- Clinton, Joshua, Simon Jackman and Doug Rivers. 2004b. "The Statistical Analysis of Roll Call Voting: A Unified Approach." *American Political Science Review* 98:355-370.

- Media example:
  - (*) Silver, Nate. "In Singling Out Murkowski, Tea Party Chose Wisely" *The New York Times* 01.09.2010. (`NYT 1.html`)
  - (*) Weber, Hannes. "So positionieren sich die Mitteparteien wirklich" *Tages Anzeiger* 27.08.2018. (`Tages-Anzeiger 2.html`)

21.04.2021: 15.00-17.00 (with breaks)

Mapping the political space II. Measures in the two-dimensional space.

- Hinich, Melvin J. and Michael C. Munger. 1997. *Analytical Politics*. Cambridge: Cambridge University Press, chapters 1&2.
- **Lab in RStudio:** Two-Dimensional Models (...the statistical house of cards...)

End of Day 1

Day 2

22.04.2021: 08.30-9.00
Introduction to Web Data.

- (*) Salganik, Matthew. 2018. *Bit by Bit: Social Research in the Digital Age*, Introduction & Observing behavior
- Media examples:
    - (*) Carolyn Said, Window into Airbnb's hidden impact on S.F., `https://www.sfchronicle.com/business/item/Window-into-Airbnb-s-hidden-impact-on-S-F-30110.php` San Francisco Chronicle
    - (*) Seth Stephens-Davidowitz, The Return of the D.I.Y. Abortion, `https://www.nytimes.com/2016/03/06/opinion/sunday/the-return-of-the-diy-abortion.html` New York Times
    - Michael D. Shear, Maggie Haberman, Nicholas Confessore, Karen Yourish, Larry Buchanan and Keith Collins, How Trump Reshaped the PResidency in Over 11,000 Tweets, Nov. 2 2019 `https://www.nytimes.com/interactive/2019/11/02/us/politics/trump-twitter-presidency.html`, New York Times

22.04.2021: 9.15-10.00
Webpage Architecture and Data Harvesting

- (*) Introduction to HTML `https://www.w3schools.com/html/html_intro.asp`, read through the following sections: Introduction, Basic, Elements, Attributes, Headings, Paragraphs, Links, Classes

22.04.2021: 10.15-12.00 (with breaks)
Data Harvesting Lab

- **Lab in RStudio:** Scraping Presidential Speeches
- Media examples:
    - (*) Barbara Maseda, "Counting words in SOTU speeches", 2018, `https://medium.com/text-data-stories/text-data-stories-counting-words-in-sotu-speeches-nlp-journalism-text-analysis-4dc487c00869` - have a closer look at at least one of the linked articles

---

Lunch Break

22.04.2021: 13.00-14.45

Text and Data Journalism

- (*) Grimmer, J., and B. M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21 (3): 267‚Äì97. https://doi.org/10.1093/pan/mps028.

- Media examples:

  - (*) Rob O'Dell and Nick Penzenstadler "Copy, paste, Legislate. You elected them to write new laws. They're letting corporations do it instead." *USA Today*, https://eu.usatoday.com/in-depth/news/investigations/2019/04/03/takata-airbag-gm-ignition-switch-recalls-used-car-dealers-sue-deaths-crashes-honda/3162202002 and Rob O'Dell " How we uncovered 10,000 times lawmakers introduced copycat model bills and why it matters", https://eu.usatoday.com/story/news/investigations/2019/04/03/how-laws-made-why-we-revealed-politicians-didnt-write-them/3162256002

  - (*) Martina Schories. "Das gehetzte Parlament" *Süddeutsche Zeitung*, https://projekte.sueddeutsche.de/artikel/politik/bundestag-das-gehetzte-parlament-e953507 and Sabrina Ebitsch, Martina Schories. "So haben wir den Bundestag ausgerechnet" *Süddeutsche Zeitung*, https://projekte.sueddeutsche.de/artikel/politik/so-haben-wir-den-bundestag-ausgerechnet-e893391

  - Trevor Martin "Dissecting Trump's Most Rabid Online Following" *FiveThirtyEight*, https://fivethirtyeight.com/features/dissecting-trumps-most-rabid-online-following

22.04.2021: 15.00-17.00 (with breaks)

Simple Methods for Text Analysis.

- **Lab in RStudio:** Analysis of Presidential Speeches

End of Day 2