

MA Seminar

Computational Social Science: Theory & Application (Spring 2021)

Description:	The seminar introduces students to the basic concepts underlying social data science in the age of big data.
Time and place:	Online: SOWI-ZOOM-02
Instructor:	Paul C. Bauer; the University of Mannheim, Mannheim Centre of European Social Research
Email:	paul.bauer@mzes.uni-mannheim.de
Office hours:	Online; Write me to make an appointment

Course outline

In the wake of the digital revolution, societies store an ever-increasing amount of data on humans and their behavior. In parallel, advances in computational power & methods allow for meaningful interpretations of such data. This enables social scientists to approach old questions with new methods but also to study entirely new questions.

The seminar introduces students to different aspects of this “big data revolution”. It comprises theoretical sessions that discuss the implications, such as the societal and scientific opportunities and challenges of new forms of data and methods (from social media, communications platforms, Internet of Things devices, sensors/wearables, and mobile phones, digitized old data records, machine learning). Besides, it comprises lab sessions in which we learn – hands-on – how such new forms of data can be captured, curated, and analyzed using computational methods. Students apply what they have learned in their projects.

Intended Learning Objectives

The goals for this course are twofold. First, I hope you will gain a **solid understanding of how access to big data (digital traces) is changing the social sciences** in terms of **a) new substantial and theoretical insights**, and in terms of **b) new methodologies**. Second, I hope you will learn which and how big data could be used to answer further pressing questions you might encounter in the future.

- Knowledge
 - Overview of seminal studies using big (digital trace) data to study social science questions.
 - Overview of key insights generated by big data analyses that go beyond what was formerly known.
 - Overview of the most important big data sources (e.g., Wikipedia, Google Trends, Twitter, etc.).
 - Overview of the major challenges and opportunities for social science research emerging from the proliferation of big data and CSS methods.
- Skills:
 - Students will be able to read and comprehend advanced big (digital trace) data analyses without knowing all underlying theoretical and methodological specifics.

- Students will be able to assess (i.e., judge the theoretical and methodological quality) the strength and limitations of such analyses.
- Students will learn how to handle and analyze big datasets, work with APIs and how to write API calls, and use simple machine learning techniques in their research.
- Competences:
 - Students will increase their analytical, methodological, logical, and creative cognitive capacities, that is, their “imagination”.
 - Students will be enabled to identify big (digital trace) data sources that might help to answer a given research question.
 - Students will be enabled to explain how a classical sociological could be extended through the analysis of big (digital trace) data.
 - Students will be enabled to present key aspects of a study in a 5-10 minute presentation.

Participation requirements

In order to obtain credits & a grade participants are required to

1. ...attend regularly (at least 80% of the sessions) and participate actively. A maximum of two sessions can be missed without excuse. Absence in further sessions can only be excused in case of illness (i. e. with a medical certificate).
2. ...finish and hand in any exercises given during the seminar.
3. ...give 2 short 5-15-minute presentations on research papers in class.
 - a. Choose papers to present in the schedule below (pick one of the [blue papers/chapters marked with two stars “**”](#) leaving a comment with your name).
 - b. Research question? Use of CSS methods? Relevance? Innovation?
4. ...hand in a 1-2-page description of an API (data types, rate limit etc.)
5. ...hand in a research design and present it in class (**Deadline: 1.6.2021**). [[Guidelines](#)]
 - a. Sign-up for office hours to discuss plan [here](#).
6. ...write a final data science report on your project (**Deadline: MA 25.7.2021; PhD 25.7.2021/second deadline in fall**). [[Guidelines](#)]

Prerequisites

- Some introductory statistics course(s) covering regression analysis, frequentist statistics, frequentist hypothesis testing etc.
- Basic knowledge of R obtained through self-study on [DataCamp](#) (access to Datacamp is organized for the course duration).
- **Required R courses in case of no R experience (March 13th 2021):**
 - [Introduction to R](#) (4h)
 - [Intermediate R](#) (<6h), Chapter 1 (Conditionals and Control Flow) & 2 (Loops)
 - [Introduction to the Tidyverse](#) (<4h), Chapter 1 (Data wrangling) & 3 (Grouping and summarizing)
- **Additional R courses (voluntary but recommended!):**
 - [Introduction to Importing Data in R](#), Chapter 1 (Importing data from flat files with utils) & 2 (readr & data.table)
 - [Cleaning Data in R](#), Chapter 1 (Common Data Problems) & 2 (Categorical and Text Data)
 - [Introduction to Data Visualization with ggplot2](#)
 - Rmarkdown (will be introduced by Frederic)
 - See also the [overview](#) and the corresponding datacamp course: [Introduction to RMarkdown](#) (Ch. 3 in Course “Communicating with Data in the Tidyverse”)

Evaluation criteria

You need to present papers, submit paper summaries, exercises and a project plan/research design to pass the course. The final report/project will constitute your entire grade for the course. You may write the project individually or in groups of max. 2 people. All members of a group will get the same grade.

- Computational social science project/term paper (90% of the grade)
 - Around ~15-20 pages text without graphs/tables; Accompanied by a replication zip folder with code/data to reproduce the analysis in R;
 - Written in the style of a scientific paper.
 - Describes the data source(s) you are using.
 - Starts from a research question (can be descriptive)
 - A descriptive analysis is sufficient for your data science report (does not have to be causal).

Material & Readings

The seminar is based on the lecture/lab script "[Computational Social Science: Theory & Application](#)". The script will be continuously updated throughout the seminar, containing all the slides and illustrations of code. There are various additional readings. These comprise fundamental texts but also studies which we will evaluate in our discussions. I will distribute a link via email through which you have access to the literature. **Please contact me in case you didn't receive the link.**

Books & script (see schedule below for actual required readings):

- Script: "[Computational Social Science: Theory & Application](#)"
- Mayer-Schönberger, Viktor, and Kenneth Cukier. 2012. "Big Data: A Revolution That Transforms How We Work, Live, and Think." Houghton Mifflin Harcourt Boston.
- Salganik, Matthew J. *Bit by bit: social research in the digital age*. Princeton University Press, 2017. (<https://www.bitbybitbook.com/en/1st-ed/preface/>)
- James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning: With Applications in R*. Springer Texts in Statistics. Springer.

Further skim readings:

- Burrows, Roger, and Mike Savage. 2014. "After the Crisis? Big Data and the Methodological Challenges of Empirical Sociology." *Big Data & Society* 1 (1): 2053951714540280.
- Mützel, Sophie. 2015. "Facing Big Data: Making Sociology Relevant." *Big Data & Society* 2 (2).
- Halford, Susan, and Mike Savage. 2017. "Speaking Sociologically with Big Data: Symphonic Social Science and the Future for Big Data Research." *Sociology* 51 (6): 1132–48.
- Varian, Hal R. 2014. "Big Data: New Tricks for Econometrics." *The Journal of Economic Perspectives: A Journal of the American Economic Association* 28 (2): 3–28.
- Tinati, Ramine, Susan Halford, Leslie Carr, and Catherine Pope. 2014. "Big Data: Methodological Challenges and Approaches for Sociological Analysis." *Sociology* 48 (4): 663–81.
- Radford, Jason, and David Lazer. 2019. "Big Data for Sociological Research." *The Wiley Blackwell Companion to Sociology*. Wiley.
- Brady, Henry E. 2019. "The Challenge of Big Data and Data Science." *Annual Review of Political Science* 22 (1): 297–323.
- Golder, Scott A., and Michael W. Macy. 2014. "Digital Footprints: Opportunities and Challenges for Online Social Research." *Annual Review of Sociology* 40 (1): 129–52.
- Keuschnigg, Marc, Niclas Lovsjö, and Peter Hedström. 2018. "Analytical Sociology and Computational Social Science." *SIAM Journal on Scientific Computing: A Publication of the Society for Industrial and Applied Mathematics* 1 (1): 3–14.
- McCay-Peet, Lori, and Anabel Quan-Haase. 2017. "What Is Social Media and What Questions Can Social Media Research Help Us Answer." *The SAGE Handbook of Social Media Research Methods*, 13–26.

- Singleton, Alex, and Daniel Arribas-Bel. 2021. "Geographic Data Science." *Geographical Analysis* 53 (1): 61–75.

Non-scientific readings

- Wylie, Christopher. 2019. *Mindf*ck: Cambridge Analytica and the Plot to Break America*. Random House Publishing Group.

Documentaries and other videos

1. [The great leap forward](#) (Deutsche Welle, 2019, 42min) [The big reset 2.0 How artificial intelligence is changing our society](#) (Deutsche Welle, 2020, 42min)
2. [AlphaGo](#) (free, 90 minutes)
3. [The social dilemma](#) (94 minutes, Netflix)
4. [Artificial intelligence and its ethics](#) (Deutsche Welle, 2019, 42min)
5. [The Age of A.I.](#) (free, ~30-50 minutes, Youtube series, Youtube is owned by Google!)
6. [The dissident](#) (117 minutes, [Amazon](#))
7. [You're Being Watched Right No](#) (12 Minutes, New York times)
8. [How China Tracks Everyone](#) (Video surveillance) (13 mins, 2019, VICE)
9. [Next-Level Surveillance: China Embraces Facial Recognition](#) (5 minutes, 2017, Wall Street Journal)
10. [Slaughterbots](#) (8 mins, 2017, Stop Autonomous Weapons) and the [Meet the dazzling flying machines of the future](#) (11 minutes, TED)
11. [Machine Learning: Living in the Age of AI](#) (free, 41 minutes, WIRED)
12. [Lo And Behold: Reveries of the Connected World](#) (98 minutes)
13. [Elon Musk's Last Warning About Artificial Intelligence](#) (24 minutes)
14. [A.I. Is Making it Easier to Kill \(You\). Here's How](#) (19 minutes, New York Times)
15. [The accelerating power of technology](#) (free, 22 min, a bit of imagination by [Ray Kurzweil](#))

Schedule

Introduction: Social science in the age of Big Data and AI	5
Big data & new data sources	6
Big data & new data sources	6
Data collection: Platform APIs	7
Data collection: Platform APIs	8
Data collection: Scraping data from websites	8
Analyzing Big Data: Basics	8
Machine learning: Prediction	9
Machine learning Applications: Classification	9
Machine learning: Text classification	10
Machine learning: Text classification	10
Machine learning: Using APIs & Deep learning	10
Big Data & ML: The dark side	11
Various posts on Privacy project (NYT)	11
Project presentations	11

Schedule

*** = required readings, ** = readings for presentations

Nr.	Date	Content & Readings
1	2.03.2020	Introduction: Social science in the age of Big Data and AI - Readings & discussion REVIEW: Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-Laszlo Barabasi, Devon Brewer, Nicholas Christakis, et al. 2009. "Social Science. Computational Social Science." <i>Science</i> 323 (5915): 721–23.*** BOOK: Mayer-Schönberger, Viktor, and Kenneth Cukier. 2012. "Big Data: A Revolution That Transforms How We Work, Live, and Think." Houghton Mifflin Harcourt Boston. [Ch. 1, p.6-19]*** BOOK: Salganik, Matthew J. Bit by bit: social research in the digital age. Princeton University Press, 2017. (Chapter 1.1, Chapter 1.2, Chapter 2.1, Chapter 2.2)*** REVIEW: McFarland, Daniel A., Kevin Lewis, and Amir Goldberg. 2016. "Sociology in the Era of Big Data: The Ascent of Forensic Social Science." <i>The American Sociologist</i> 47 (1): 12–35. - Watch the following video (Gary King) and answer the following questions:*** - What is more important? The data itself or the analytics? So what is the revolution? - What does he say about survey research? - What research areas does he name in which big data plays a role? - What is the example of the Chinese word "freedom" about? What is the 50cent party?

- What is the end of the quantitative-qualitative divide?
- Watch documentaries under 1. above (Deutsche Welle)
 - What actors appear? What data types are discussed? What technology is described? Which points struck you as the most interesting in the documentary?
- Cross-read wikipedia on [Big Data](#), [Machine Learning](#), on [Deep Learning](#)
- [Barberá, Pablo. "Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data." Political Analysis 23.1 \(2015\): 76-91. \(TWITTER\)**](#)
- Homework (until 9.3.2020)
 - [Datacamp: Introduction to SQL](#)

2 9.03.2020 **Big data & new data sources (1)**

- Readings & presentations
 - **BOOK:** Salganik, Matthew J. Bit by bit: social research in the digital age. Princeton University Press, 2017. ([Chapter 2.3](#): Ten common characteristics of big data)***
 - **REVIEW:** Edelmann, Achim, Tom Wolff, Danielle Montagne, and Christopher A. Bail. 2020. "Computational Social Science and Sociology." *Annual Review of Sociology* 46 (1): 61–81.***
 - **REVIEW:** Grimmer, Justin. 2015. "We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together." *PS, Political Science & Politics* 48 (1): 80–83.***
 - **REVIEW:** Swan, Melanie. 2013. "The Quantified Self: Fundamental Disruption in Big Data Science and Biological Discovery." *Big Data* 1 (2): 85–99.**
 - **M-PHONE DATA:** Blumenstock, Joshua, Gabriel Cadamuro, and Robert On. 2015. "Predicting Poverty and Wealth from Mobile Phone Metadata." *Science* 350 (6264): 1073–76.**
 - **NY SERVICE REQUESTS:** Legewie, Joscha, and Merlin Schaeffer. 2016. "Contested Boundaries: Explaining Where Ethnoracial Diversity Provokes Neighborhood Conflict." *AJS; American Journal of Sociology* 122 (1): 125–61.
- Skim readings
 - **REVIEW:** Lazer, David, and Jason Radford. 2017. "Data Ex Machina: Introduction to Big Data." *Annual Review of Sociology* 43 (1): 19–39.
- ~~Lab~~
 - ~~Data management: Handling big datasets in R (SQL & local databases)~~
- Homework (until 16.3.2020)
 - [Setup Google Cloud Research Credits](#)

3 16.03.2020 **Big data & new data sources (2)**

- Readings & presentations
 - **BOOK:** Salganik, Matthew J. Bit by bit: social research in the digital age. Princeton University Press, 2017. ([Chapter 2.4](#): Research strategies)***
 - **GOOGLE SEARCH:** Mellon, Jonathan. 2013. "Where and When Can We Use Google Trends to Measure Issue Salience?" *PS, Political Science & Politics* 46 (2): 280–90.**
 - **GOOGLE SEARCH:** Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. "Big Data. The Parable of Google Flu: Traps in Big Data Analysis." *Science* 343 (6176): 1203–5.**

- Kandula, Sasikiran, and Jeffrey Shaman. 2019. "Reappraising the Utility of Google Flu Trends." *PLoS Computational Biology* 15 (8): e1007258.
- Skim readings
 - **GOOGLE SEARCH:** Stephens-Davidowitz, Seth. 2014. "The Cost of Racial Animus on a Black Candidate: Evidence Using Google Search Data." *Journal of Public Economics* 118 (October): 26–40.
 - **GOOGLE SEARCH:** Bail, Christopher A., Taylor W. Brown, and Andreas Wimmer. 2019. "Prestige, Proximity, and Prejudice: How Google Search Terms Diffuse across the World." *The American Journal of Sociology* 124 (5): 1496–1548.
 - **GOOGLE SEARCH:** Bail, Christopher A., Friedolin Merhout, and Peng Ding. 2018. "Using Internet Search Data to Examine the Relationship between Anti-Muslim and pro-ISIS Sentiment in U.S. Counties." *Science Advances* 4 (6): eaao5948.
- Lab
 - Data management: Handling big datasets in R (SQL & local databases) & data warehouses (**Google BigQuery**)
- Homework
 - [Exercise: SQL database](#)

4 23.03.2020 **Data collection: Platform APIs**

- Readings & presentations
 - **REVIEW:** Puschmann, Cornelius, and Julian Ausserhofer. 2017. "Social Data APIs: Origin, Types, Issues." ***
 - **BOOK:** Nyhuis, Dominic. 2021. "Application Programming Interfaces and Web Data for Social Research."***
 - **WIKIPEDIA:** Göbel, Sascha, and Simon Munzert. 2018. "Political Advertising on the Wikipedia Marketplace of Information." *Social Science Computer Review* 36 (2): 157–75.**
 - **WIKIPEDIA:** Shi, Feng, Misha Teplitskiy, Eamon Duede, and James A. Evans. 2019. "The Wisdom of Polarized Crowds." *Nature Human Behaviour* 3 (4): 329–36.**
 - Skim readings
 - Lomborg, Stine, and Anja Bechmann. 2014. "Using APIs for Data Collection on Social Media." *The Information Society* 30 (4): 256–65.
 - Lab
 - Data collection: Understanding APIs and making API calls with R
 - Homework
 - Go through the [BigQuery Lab](#) (parts on querying public datasets).
 - Datacamp
 - Downloading Files and Using API Clients (see [here](#))
 - Using httr to interact with APIs directly (see [here](#))
 - Handling JSON and XML (see [here](#))
 - Pick an API below ([sign-up table](#)).
 - [Google Trends API](#)
 - [Google Static Maps API](#)
 - [Google Places API](#)
 - [Youtube Data API](#)
 - [Instagram Graph API](#)
-

- [Instagram Basic Display API](#)
- [Facebook Graph API](#)
- [Wikipedia API](#)
- [CrowdTangle API](#)
- etc. (an [ProgrammableWeb](#) for an overview)

30.03.2020
entfällt

6.04.2021
entfällt

- | | | |
|---|------------|---|
| 5 | 13.04.2021 | <p>Data collection: Platform APIs</p> <ul style="list-style-type: none"> ● Readings & presentations <ul style="list-style-type: none"> ○ REVIEW: Jungherr, Andreas. 2016. "Twitter Use in Election Campaigns: A Systematic Literature Review." <i>Journal of Information Technology & Politics</i> 13 (1): 72–91. ○ TWITTER: Tumasjan, Andranik, Timm Sprenger, Philipp Sandner, and Isabell Welp. 2010. "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment." <i>Proceedings of the International AAAI Conference on Web and Social Media</i> 4 (1). https://ojs.aaai.org/index.php/ICWSM/article/view/14009.** (10 minutes) <ul style="list-style-type: none"> ■ TWITTER: Jungherr, Andreas, Pascal Jürgens, and Harald Schoen. 2012. "Why the Pirate Party Won the German Election of 2009 or the Trouble with Predictions: A Response to Tumasjan, A., Sprenger, To, Sander, Pg, & Welp, Im 'predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment.'" <i>Social Science Computer Review</i> 30 (2): 229–34.** (2 minutes) ● Lab <ul style="list-style-type: none"> ○ Data collection: Understanding APIs and making API calls with R ● Homework <ul style="list-style-type: none"> ○ APIs: In teams of max. 2 people write a 2-pager about the API you picked in the Google Doc. <ul style="list-style-type: none"> ■ Example: Nyhuis 2021: 4.A MORE COMPLEX EXAMPLE: COLLECTING DATA FROM TWITTER ○ Try out uploading/downloading table to bigquery (see here) |
|---|------------|---|

- | | | |
|---|------------|---|
| 6 | 20.04.2021 | <p>Data collection: Platform APIs</p> <ul style="list-style-type: none"> ● Readings & presentations <ul style="list-style-type: none"> ○ BOOK: Munzert, Simon, Christian Rubba, Peter Meißner, and Dominic Nyhuis. 2014. <i>Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining</i>. John Wiley & Sons. (Chapter 9/9.1: Scraping the Web/Retrieval Scenarios, Also consider reading first introductory pages of Chapter 2-7 on different protocols)*** ○ DARKNET MARKETS: Przepiorka, Wojtek, Lukas Norbutas, and Rense Corten. 2017. "Order without Law: Reputation Promotes Cooperation in a Cryptomarket for Illegal Drugs." <i>European Sociological Review</i> 33 (6): 752–64.** ○ AIRBNB: Edelman, Benjamin, Michael Luca, and Dan Svirsky. 2017. "Racial Discrimination in the Sharing Economy: Evidence from a Field |
|---|------------|---|

Experiment.” American Economic Journal. Applied Economics 9 (2): 1–22.** (10 minutes)

- Ahuja, Rishi, Ronan C. Lyons, and Others. 2017. “The Silent Treatment: LGBT Discrimination in the Sharing Economy.” No. tep1917, Trinity College Dublin, Department of Economics, Dublin. <https://www.tcd.ie/Economics/TEP/2017/tep1917.pdf>.

(2 minutes)

- [AirBnb in the pandemic](#)

- Api discussion: “APIs for Social Scientists”.
- Lab
 - APIs
- Homework
 - Finish [API review](#), Correct & provide feedback on [API review](#) (see assignments [here](#))

7 27.04.2021 **Data collection: Scraping data from websites**

- Readings & presentations
 - **BOOK:** Munzert, Simon, Christian Rubba, Peter Meißner, and Dominic Nyhuis. 2014. Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining. John Wiley & Sons. (Chapter 9/9.1: Scraping the Web/Retrieval Scenarios, Also consider reading first introductory pages of Chapter 2-7 on different protocols)***
 - Gold, Alex. (2019) “[Three Strategies for Working with Big Data in R](#).”***
 - **GEOTRACKING:** Palmer, John R. B., Thomas J. Espenshade, Frederic Bartumeus, Chang Y. Chung, Necati Ercan Ozgencil, and Kathleen Li. 2013. “New Approaches to Human Mobility: Using Mobile Phones for Demographic Research.” Demography 50 (3): 1105–28.***
 - **SP-GEO-TRACKING:** Kreuter, Frauke, Georg-Christoph Haas, Florian Keusch, Sebastian Bähr, and Mark Trappmann. 2020. “Collecting Survey and Smartphone Sensor Data With an App: Opportunities and Challenges Around Privacy and Informed Consent.” Social Science Computer Review 38 (5): 533–49.** (10 minutes)
- Lab
 - Simple scraping & Three strategies to work with big data in R
- Homework
 - Finish [API review](#), Correct & provide feedback on [API review](#) (see assignments [here](#) - SECOND round)
 - [Homework Exercise: Scraping tables](#)
 - [Halftime Feedback Survey](#)

8 4.05.2021 **Data collection: Scraping data from websites**

- Readings & presentations
 - **BOOK:** Munzert, Simon, Christian Rubba, Peter Meißner, and Dominic Nyhuis. 2014. Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining. John Wiley & Sons. (Chapter 9/9.1: Scraping the Web/Retrieval Scenarios, Also consider reading first introductory pages of Chapter 2-7 on different protocols)***

- **COMPAS/EQUIVANT DATA:** Dressel, Julia, and Hany Farid. "The accuracy, fairness, and limits of predicting recidivism." *Science advances* 4.1 (2018).** (10 minutes)¹
- Lab
 - Simple scraping
- Homework
 - Finish and correct [API review](#) (take into account new feedback)
 - [Homework Exercise: Scraping unstructured data](#)
 - [Datacamp: Introduction to RMarkdown](#)
 - [ML demos exercise: Try out the different ML demos/examples and think about the exercise questions](#)

9 11.05.2021 **Machine learning: Intro & prediction & classification**

- Readings & presentations
 - **BOOK:** James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning: With Applications in R*. Springer Texts in Statistics. Springer.[Ch. 1, 2.1, 2.2, 3.1-3.4 (skim read)].***
 - **REVIEW:** Molina, Mario, and Filiz Garip. 2019. "Machine Learning for Sociology." *Annual Review of Sociology*.***
 - **TWITTER:** Bastos, Marco T., and Dan Mercea. 2019. "The Brexit Botnet and User-Generated Hyperpartisan News." *Social Science Computer Review* 37 (1): 38–54.** (10 minutes)
 - [How Pew Research Center identified bots on Twitter](#) (2 minutes)
- Skim readings
 - **REVIEW:** Athey, Susan, and Guido W. Imbens. 2019. "Machine Learning Methods That Economists Should Know About." *Annual Review of Economics* 11 (1): 685–725.
- Lab
 - Machine learning & prediction: Simple example (linear regression)
- Homework
 - Readings etc.

10 18.05.2021 **Machine learning: Intro & prediction & classification**

- Readings & presentations
 - **BOOK:** James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning: With Applications in R*. Springer Texts in Statistics. Springer.[Ch. 4.1-4.3].***
 - **FACEBOOK:** Guess, Andrew, Jonathan Nagler, and Joshua Tucker. 2019. "Less than You Think: Prevalence and Predictors of Fake News Dissemination on Facebook." *Science Advances* 5 (1).** (10 minutes)
 - **FACEBOOK:** Kramer, Adam D. I., Jamie E. Guillory, and Jeffrey T. Hancock. 2014. "Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks." *Proceedings of the National Academy of Sciences of the United States of America* 111 (24): 8788–90.** (2 minutes)
 - Lab
 - Machine learning & classification: Simple example (logistic regression)
-

¹ Follow-up study: Lin, Zhiyuan Jerry, Jongbin Jung, Sharad Goel, and Jennifer Skeem. 2020. "The Limits of Human Predictions of Recidivism." *Science Advances* 6 (7): eaaz0652.

11	25.05.2021	Machine learning: Intro & prediction & classification • Readings & presentations ○ BOOK: Silge, Julia, and David Robinson. 2017. Text Mining with R: A Tidy Approach. "O'Reilly Media, Inc." [Ch. 1-3]*** ○ REVIEW: Wilkerson, John, and Andreu Casas. 2017. "Large-Scale Computerized Text Analysis in Political Science: Opportunities and Challenges." <i>Annual Review of Political Science</i> 20 (1): 529–44.. ○ Machine Learning for Social Science: An Agnostic Approach ○ DATING PLATFORM: Curington, Celeste Vaughan, Ken-Hou Lin, and Jennifer Hicke Lundquist. 2015. "Positioning Multiraciality in Cyberspace: Treatment of Multiracial Daters in an Online Dating Website." <i>American Sociological Review</i> 80 (4): 764–88.** (10 minutes) ■ DATING PLATFORM: Bruch, Elizabeth E., and M. E. J. Newman. 2019. "Structure of Online Dating Markets in U.S. Cities." <i>Sociological Science</i> 6 (April): 219–34. (2 minutes) • Lab ○ Machine learning & classification: Simple example (logistic regression)
12	1.06.2021	Machine learning: Text classification (Camille Landesvatter) • Readings & presentations ○ BOOK: Silge, Julia, and David Robinson. 2017. Text Mining with R: A Tidy Approach. "O'Reilly Media, Inc." [Ch. 4-6]*** ○ Carsten Schwemmer & Sebastian Jungkunz (2019) Whose ideas are worth spreading? The representation of women and ethnic groups in TED talks, Political Research Exchange, 1:1, 1-23, DOI: 10.1080/2474736X.2019.1646102 ○ 50c PARTY POSTS: King, Gary, Jennifer Pan, and Margaret E. Roberts. 2017. "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument." <i>The American Political Science Review</i> 111 (3): 484–501.** (10 minutes)Lab ○ Simple examples of text classification, sentiment analysis, topic modelling in R
13	8.6.2021	Machine learning: Using APIs & Deep learning • Readings & presentations ○ BOOK: Chollet, Francois, and J. J. Allaire. 2018. <i>Deep Learning with R</i>. Manning Publications. [Ch. 1, (Ch. 2), Ch. 3]*** ○ Barari, Soubhik, Christopher Lucas, and Kevin Munger. n.d. "Political Deepfake Videos Misinform the Public, But No More than Other Fake Media." Accessed March 22, 2021. https://osf.io/cdfh3/.** (10 minutes) • APIs ○ Google AutoML Natural Language ○ Google Natural Language API ○ Google Speech-to-Text API ○ Clarifai Image recognition ○ others see above... • Lab ○ Using different platform APIs for machine learning problems in R • <u>9.6.2021/6PM: Talk & Discussion: Pablo Barberá - A career in data science</u>

- Readings & presentations
 - **BOOK:** Chollet, Francois, and J. J. Allaire. 2018. Deep Learning with R. Manning Publications. [[Ch. 1](#), (Ch. 2), [Ch. 3](#)]*******
 - **REVIEW:** Tucker, J. A., Theocharis, Y., Roberts, M. E., & Barberá, P. (2017). [From Liberation to Turmoil: Social Media And Democracy](#). *Journal of Democracy*, 28(4), 46-59.*******
 - **REVIEW:** Helbing, D., Frey, B. S., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., ... & Zwitter, A. (2017). Will democracy survive big data and artificial intelligence. *Scientific American*, 25.*******
- Skim readings
 - Narayanan, Arvind, and Vitaly Shmatikov. "Robust de-anonymization of large sparse datasets." Security and Privacy, 2008. SP 2008. IEEE Symposium on. IEEE, 2008.
 - Zimmer, Michael. 2010. "'But the Data Is Already Public': On the Ethics of Research in Facebook." *Ethics and Information Technology* 12 (4): 313–25.
 - Various posts on [Privacy project \(NYT\)](#)
- Lab
 - Using different platform APIs for machine learning problems in R

Other topics

- **DATING PLATFORM:** Bruch, Elizabeth E., and M. E. J. Newman. 2019. "Structure of Online Dating Markets in U.S. Cities." *Sociological Science* 6 (April): 219–34.
- **DATING SURVEY:** Huber, Gregory A., and Neil Malhotra. 2017. "Political Homophily in Social Relationships: Evidence from Online Dating Behavior." *The Journal of Politics* 79 (1): 269–83.
- **TWITTER:** Munger, K., R. Bonneau, and J. Nagler. 2019. "Elites Tweet to Get Feet off the Streets: Measuring Regime Social Media Strategies during Protest."******
- **TEXT DATA:** Proksch, Sven-Oliver, Christopher Wratil, and Jens Wäckerle. 2019. "Testing the Validity of Automatic Speech Recognition for Political Text Analysis." *Political Analysis*. 27 (3): 339–59.******
- **FACEBOOK:** Phan, Tuan Q., and Edoardo M. Airolidi. 2015. "A Natural Experiment of Social Network Formation and Dynamics." *Proceedings of the National Academy of Sciences of the United States of America* 112 (21): 6595–6600.******
- **EMPLOYMENT DATA:** Bansak, Kirk, Jeremy Ferwerda, Jens Hainmueller, Andrea Dillon, Dominik Hangartner, Duncan Lawrence, and Jeremy Weinstein. 2018. "Improving Refugee Integration through Data-Driven Algorithmic Assignment." *Science* 359 (6373): 325–29.******
- **GEOTRACKING:** Athey, Susan, Billy A. Ferguson, Matthew Gentzkow, and Tobias Schmidt. 2020. "Experienced Segregation." w27572. National Bureau of Economic Research. <https://doi.org/10.3386/w27572>.******
- **TWITTER:** Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. 2018. "Exposure to Opposing Views on Social Media Can Increase Political Polarization." *Proceedings of the National Academy of Sciences of the United States of America* 115 (37): 9216–21.******
- **TWITTER:** Barceló, Joan, and Elena Labzina. 2020. "Do Islamic State's Deadly Attacks Disengage, Deter, or Mobilize Supporters?" *British Journal of Political Science* 50 (4): 1539–59.******