

More Than Words: Introduction to Quantitative Text Analysis

Friedolin Merhout, Department of Sociology

Spring 2021, Block 4

Our contemporary, increasingly digital societies generate vast amounts of textual data that provide a rich source for sociological research. The scale of these novel data however poses a challenge to the approaches sociologists traditionally use to study texts. In response, automated methods of text analysis are becoming increasingly popular and the command of these methods a valuable skill in academic environments as well as on the private industry job market.

This course introduces students to quantitative text analysis, reviews selected methods falling within this category of approaches, and illustrates their implementation in the statistical programming language R. Students will learn about the origins of quantitative approaches to studying text and how they complement traditional, qualitative methodologies. Using recent peer-reviewed publications students will gain an understanding of how these methodological approaches can be used to answer sociological questions and, in hands-on lab session, students will learn to implement selected techniques in R.

After successful participation, students will be comfortable reading current sociological research using quantitative text analysis, have an understanding of the landscape of tools used within the literature, and will have gained experience with their implementation in R.¹

Intended Learning Outcomes

- *Knowledge*
 - Successful participation in the class will prepare students to account for the need to apply novel methodologies to large-scale text data
 - Additionally, students will be able to identify methods of quantitative text analysis suited to answer sociological question of large-scale text data
- *Skills*
 - Students will gain facility with working with R in particular as it relates to quantitative text analysis
 - Students will be able to evaluate and put into perspective the benefits and complementarities of quantitative text analysis with traditional forms of text analysis
- *Competencies*
 - Students will be able to plan sociological studies that leverage the potential of modern large-scale text data
 - Students will be able to specialize in cutting-edge methodologies in quantitative text analysis

¹ *Acknowledgments:* This syllabus takes inspiration from similar course offerings by Christopher A. Bail, Bart Bonikowski, Kevin Munger, Stefan Müller, Martijn Schoonvelde, and Marshall A. Taylor.

Course Components and Expectations

This course is comprised of methodological and substantive lectures, on the one hand, and hands-on R lab session, on the other. For successful participation in this course there are three expectations: First, thoroughly prepare for each class with the assigned readings. Second, actively engage in discussion sections of lectures and provided R scripts in lab sessions. Three, prepare and submit a portfolio exam consisting of three separate assignments. Further details for each expectation are provided in the following.

Course Literature

The central aim of this course is to introduce students to the fundamentals of quantitative text analysis in the social science and provide hands-on experience with the implementation of selected techniques and methods in the R programming language. Given this hands-on focus, the required readings for each class are relatively light with a focus on central concepts, methods, techniques, and their implementation in R. Further readings with substantive applications for each method are provided separately below.

The central textbook for the course is “Text Mining: A Guidebook for the Social Sciences” (Ignatow and Mihalcea 2017). The book is available from *Academic Books* on the City Campus. Additional readings include open-source online resources, short book excerpts, and peer-reviewed journal articles which will be available through Absalon.

Required readings for each session will usually include a broad methodological and or substantive overview and an example implementation of the content in R. Make sure to read the overview in-depth, to get a general sense of the session’s content. Read the implementation with a primary focus on the involved packages, functionality, syntax, and workflow, but do not feel obligated to execute any code. We will do this collectively in the lab sessions.

Lab Sessions

Each substantive session is complemented by a hands-on lab session which will revolve around R scripts meant to create familiarity with the implementations of the covered methods and techniques. These lab sessions are intended as guided introductions to the methods and techniques in R, and meant to provide the foundation for further, independent learning.

Disclaimer

This syllabus is subject to change as required by course progression or unforeseen events. In case of changes, I will upload the updated syllabus with an Announcement to Absalon.

Exam

The exam for this course is a portfolio exam consisting of three assignments to be handed in throughout the course. The three assignments will have to be revised based on the feedback provided, combined into a coherent whole, and submitted at the end of the course. Both individual and group assignments are possible. The page limit for individual assignments is 10 and group assignments add 5 pages per additional student. The submission deadline for the exam is at noon Danish time on **June 8, 2021**. For further details on exam submission, exam policies, and more please refer to the [Department of Sociology's exam website](#).

Portfolio Items

The overall aim for the exam is for you to (1) identify a sociological phenomenon that could benefit from quantitative text analysis, (2) identify a data source and method that is suitable to study the phenomenon you identified, and (3) prepare the data for analysis, conduct the analysis, to the extent possible, and discuss challenges and potential findings in relation to your chosen data, method, and phenomenon.

The first portfolio item should introduce your chosen sociological phenomenon relating it briefly to the existing literature, describe the data you intend to use to study it, and make the case for the data's suitability. Portfolio item number two should present your data with appropriate summary statistics and descriptive visualization, discuss how it was collected or you collected it, and how you prepared it for analysis. The last portfolio item should introduce an analytical method suited for your phenomenon and data, present analysis results if you have them *or* how you would go about implementing the analysis, and conclude with a discussion reflecting on the overall process, including any results where applicable.

Portfolio Schedule

Separate portfolio items will have to be submitted through Absalon by the 8 PM (Danish time) indicated in the table below.

Portfolio Item #	Due Date	Content
1	May 2, 2021	Introduction of phenomenon and data
2	May 16, 2021	Data pre-processing and description
3	May 30, 2021	Method description, (analysis result), and discussion

Course Outline

Session	Title	Content	Readings
1	Words of Welcome	Introduction, history, and terminology	Ignatow and Mihalcea (2017, Ch. 1); Underwood (2012)
2	Starting with Strings	Basic string methods, regular expressions, tidy text format	Silge and Robinson (2017:1-7); Wickham and Grolemund (2016, Ch. 14)
3	Trying out Texts	Text pre-processing and corpora construction	Benoit et al. (2018); Ignatow and Mihalcea (2017:52-9); Larson (2016)
4	Data Collection I	(Online) archives and application programming interfaces	Schumacher et al. (2020); Veltri (2020:48-51)
5	Descriptive Inference	Term frequency, keywords in context, and similarity	Ignatow and Mihalcea (2017, Ch. 10); Silge and Robinson (2017, Ch. 3)
6	Data Collection II	Web scraping	Ignatow and Mihalcea (2017, Ch. 3); McNulty (2019)
7	Dictionary Methods	Sentiment analysis, LIWC, topic dictionaries	Ignatow and Mihalcea (2017, Ch. 4 & 14); Silge and Robinson (2017, Ch. 2)
8	Beyond Semantics	Part of speech tagging, named-entity recognition	Ignatow and Mihalcea (2017:59-61 & Ch. 12); Raja (2018)
9	Teaching Machines I	Basics of supervised learning	Hvitfeldt and Silge (2020, Ch. 6.1-4 & 6.11); Ignatow and Mihalcea (2017, Ch. 6)
10	Teaching Machines II	Classification and model evaluation	Hvitfeldt and Silge (2020, Ch. 7.1-4, 7.9, & 7.11); Ignatow and Mihalcea (2017, Ch. 11)
11	Learning from Machines I	Topic modeling	Bail (2018b); Ignatow and Mihalcea (2017, Ch. 15)
12	Learning from Machines II	Word embeddings	Huang (2017); Hvitfeldt and Silge (2020, Ch. 5)
13	Emerging Techniques	Text networks, image and audio as text	Bail (2018a); Steinert-Threlkeld (2019)
14	Closing Chapter	Conclusion, potential catch up, and review	Ignatow and Mihalcea (2017, Ch. 16)

Further Readings

The following readings include substantive applications of the different methods covered in this course and provide an insight into the breadth of sociological and social scientific scholarship more broadly drawing on quantitative text analysis. Many of the readings relate to more than the specific content they are grouped under.

Content	Readings
Approaches to text-as-data in social science	DiMaggio (2015); Evans and Aceves (2016); Gentzkow, Kelly, and Taddy (2019); Grimmer and Stewart (2013); Ignatow (2015); Lucas et al. (2015); Mützel (2015); O'Connor, Bamman, and Smith (2011); Petchler and Gonzalez-Bailon (2015); Roberts (2000)
Text pre-processing and corpora construction	Schoonvelde, Schumacher, and Bakker (2019); Blinder and Allen (2016); Denny and Spirling (2018)
(Online) archives and APIs	Bernau (2018); Freelon et al. (2020); Müller and Schwarz (2020)
Term frequency, keywords in context, and similarity	Gentzkow and Shapiro (2010); Harwood (2019); Lizardo et al. (2018); Schwemmer and Wieczorek (2020)
Web scraping	Long and Eveland (2018); Schwemmer and Jungkunz (2019)
Sentiment analysis, LWIC, topic dictionaries	Bonikowski and Gidron (2016); Danowski, Yan, and Riopelle (2021); Flores (2017); Gorman and Seguin (2015); Spörlein and Schlueter (2020)
Part of speech tagging, named entity recognition	Martin, Rafail, and McCarthy (2017); van de Rijt et al. (2013)
Supervised learning	Carlsen, Toubøl, and Ralund (2020); Hanna (2013); Hjorth et al. (2015); Hopkins and King (2010); Kim (2021)
Topic modeling	Chakrabarti and Frye (2017); DiMaggio, Nag, and Blei (2013); Farrell (2016); Edelmann, Moody, and Light (2017); Light and Odden (2017); Lindstedt (2019); Moeller, Munksgaard, and Demant (2017); Mohr and Bogdanov (2013); Mohr et al. (2013)
Word embeddings	Joseph and Morgan (2020); Kozlowski, Taddy, and Evans (2019); Stoltz and Taylor (2019a)
Text networks, image and audio as text	Adams and Roscigno (2005); Bail (2016); Bearman and Stovel (2000); Fuhse et al. (2020); Hoffman (2019); Hoffman et al. (2018); Joo and Steinert-Threlkeld (2018); Light (2014); Stoltz and Taylor (2019b); Rule, Cointet, and Bearman (2015)
Automated text-generation	Schwemmer et al. (2020); Vries, Schoonvelde, and Schumacher (2018)

References

- Adams, Josh, and Vincent J. Roscigno. 2005. "White Supremacists, Oppositional Culture and the World Wide Web." *Social Forces* 84(2):759–78. doi: [10.1353/sof.2006.0001](https://doi.org/10.1353/sof.2006.0001).
- Bail, Christopher A. 2016. "Combining Natural Language Processing and Network Analysis to Examine How Advocacy Organizations Stimulate Conversation on Social Media." *Proceedings of the National Academy of Sciences* 113(42):11823–28. doi: [10.1073/pnas.1607151113](https://doi.org/10.1073/pnas.1607151113).
- Bail, Christopher A. 2018a. "Text Networks." Retrieved March 1, 2021 (https://cbail.github.io/SICSS_Text_Networks.html).
- Bail, Christopher A. 2018b. "Topic Modeling." Retrieved March 1, 2021 (https://cbail.github.io/SICSS_Topic_Modeling.html).
- Bearman, Peter S., and Katherine Stovel. 2000. "Becoming a Nazi: A Model for Narrative Networks." *Poetics* 27(2–3):69–90. doi: [10.1016/S0304-422X\(99\)00022-4](https://doi.org/10.1016/S0304-422X(99)00022-4).
- Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. 2018. "Quanteda: An R Package for the Quantitative Analysis of Textual Data." *Journal of Open Source Software* 3(30):774. doi: [10.21105/joss.00774](https://doi.org/10.21105/joss.00774).
- Bernau, John A. 2018. "Text Analysis with JSTOR Archives." *Socius: Sociological Research for a Dynamic World* 4:237802311880926. doi: [10.1177/2378023118809264](https://doi.org/10.1177/2378023118809264).
- Blinder, Scott, and William L. Allen. 2016. "Constructing Immigrants: Portrayals of Migrant Groups in British National Newspapers, 2010–2012." *International Migration Review* 50(1):3–40. doi: [10.1111/imre.12206](https://doi.org/10.1111/imre.12206).
- Bonikowski, Bart, and Noam Gidron. 2016. "The Populist Style in American Politics: Presidential Campaign Discourse, 1952–1996." *Social Forces* 94(4):1593–1621. doi: [10.1093/sf/sov120](https://doi.org/10.1093/sf/sov120).
- Carlsen, Hjalmar Bang, Jonas Toubøl, and Snorre Ralund. 2020. "Consequences of Group Style for Differential Participation." *Social Forces*. doi: [10.1093/sf/soaa063](https://doi.org/10.1093/sf/soaa063).
- Chakrabarti, Parijat, and Margaret Frye. 2017. "A Mixed-Methods Framework for Analyzing Text Data: Integrating Computational Techniques with Qualitative Methods in Demography." *Demographic Research* 37(1):1351–82. doi: [10.4054/DemRes.2017.37.42](https://doi.org/10.4054/DemRes.2017.37.42).
- Danowski, James A., Bei Yan, and Ken Riopelle. 2021. "A Semantic Network Approach to Measuring Sentiment." *Quality & Quantity* 55(1):221–55. doi: [10.1007/s11135-020-01000-x](https://doi.org/10.1007/s11135-020-01000-x).
- Denny, Matthew J., and Arthur Spirling. 2018. "Text Preprocessing for Unsupervised Learning: Why It Matters, When It Misleads, and What to Do about It." *Political Analysis* 26(2):168–89. doi: [10.1017/pan.2017.44](https://doi.org/10.1017/pan.2017.44).
- DiMaggio, Paul. 2015. "Adapting Computational Text Analysis to Social Science (and Vice Versa)." *Big Data & Society* 2(2):205395171560290. doi: [10.1177/2053951715602908](https://doi.org/10.1177/2053951715602908).
- DiMaggio, Paul, Manish Nag, and David Blei. 2013. "Exploiting Affinities between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U.S. Government Arts Funding." *Poetics* 41(6):570–606. doi: [10.1016/j.poetic.2013.08.004](https://doi.org/10.1016/j.poetic.2013.08.004).
- Edelmann, Achim, James Moody, and Ryan Light. 2017. "Disparate Foundations of Scientists' Policy Positions on Contentious Biomedical Research." *Proceedings of the National Academy of Sciences* 114(24):6262–67. doi: [10.1073/pnas.1613580114](https://doi.org/10.1073/pnas.1613580114).
- Evans, James A., and Pedro Aceves. 2016. "Machine Translation: Mining Text for Social Theory." *Annual Review of Sociology* 42:21–50. doi: [10.1146/annurev-soc-081715-074206](https://doi.org/10.1146/annurev-soc-081715-074206).
- Farrell, Justin. 2016. "Corporate Funding and Ideological Polarization about Climate Change." *Proceedings of the National Academy of Sciences* 113(1):92–97. doi: [10.1073/pnas.1509433112](https://doi.org/10.1073/pnas.1509433112).
- Flores, René D. 2017. "Do Anti-Immigrant Laws Shape Public Sentiment? A Study of Arizona's SB

- 1070 Using Twitter Data." *American Journal of Sociology* 123(2):333–84. doi: [10.1086/692983](https://doi.org/10.1086/692983).
- Freelon, Deen, Michael Bossetta, Chris Wells, Josephine Lukito, Yiping Xia, and Kirsten Adams. 2020. "Black Trolls Matter: Racial and Ideological Asymmetries in Social Media Disinformation." *Social Science Computer Review* 089443932091485. doi: [10.1177/0894439320914853](https://doi.org/10.1177/0894439320914853).
- Fuhse, Jan, Oscar Stuhler, Jan Riebling, and John Levi Martin. 2020. "Relating Social and Symbolic Relations in Quantitative Text Analysis. A Study of Parliamentary Discourse in the Weimar Republic." *Poetics* 78(April):101363. doi: [10.1016/j.poetic.2019.04.004](https://doi.org/10.1016/j.poetic.2019.04.004).
- Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. "Text as Data." *Journal of Economic Literature* 57(3):535–74. doi: [10.1257/jel.20181020](https://doi.org/10.1257/jel.20181020).
- Gentzkow, Matthew, and Jesse M. Shapiro. 2010. "What Drives Media Slant? Evidence From U.S. Daily Newspapers." *Econometrica* 78(1):35–71. doi: [10.3982/ECTA7195](https://doi.org/10.3982/ECTA7195).
- Gorman, Brandon, and Charles Seguin. 2015. "Reporting the International System: Attention to Foreign Leaders in the US News Media, 1950-2008." *Social Forces* 94(2):775–99. doi: [10.1093/sf/sov061](https://doi.org/10.1093/sf/sov061).
- Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21(3):267–97. doi: [10.1093/pan/mps028](https://doi.org/10.1093/pan/mps028).
- Hanna, Alexander. 2013. "Computer-Aided Content Analysis of Digitally Enabled Movements." *Mobilization: An International Quarterly* 18(4):367–88. doi: [10.17813/maiq.18.4.m1g180620x7n1542](https://doi.org/10.17813/maiq.18.4.m1g180620x7n1542).
- Harwood, Elizabeth T. 2019. "Terrorism and the Digital Right-Wing." *Contexts* 18(3):60–62. doi: [10.1177/1536504219864961](https://doi.org/10.1177/1536504219864961).
- Hjorth, Frederik, Robert Klemmensen, Sara Hobolt, Martin Ejnar Hansen, and Peter Kurrild-Klitgaard. 2015. "Computers, Coders, and Voters: Comparing Automated Methods for Estimating Party Positions." *Research and Politics* 2(2). doi: [10.1177/2053168015580476](https://doi.org/10.1177/2053168015580476).
- Hoffman, Mark Anthony. 2019. "The Materiality of Ideology: Cultural Consumption and Political Thought after the American Revolution." *American Journal of Sociology* 125(1):1–62. doi: [10.1086/704370](https://doi.org/10.1086/704370).
- Hoffman, Mark Anthony, Jean-Philippe Cointet, Philipp Brandt, Newton Key, and Peter Bearman. 2018. "The (Protestant) Bible, the (Printed) Sermon, and the Word(s): The Semantic Structure of the Conformist and Dissenting Bible, 1660–1780." *Poetics* 68(July 2017):89–103. doi: [10.1016/j.poetic.2017.11.002](https://doi.org/10.1016/j.poetic.2017.11.002).
- Hopkins, Daniel J., and Gary King. 2010. "A Method of Automated Nonparametric Content Analysis for Social Science." *American Journal of Political Science* 54(1):229–47. doi: [10.1111/j.1540-5907.2009.00428.x](https://doi.org/10.1111/j.1540-5907.2009.00428.x).
- Huang, Roger. 2017. "An Introduction to Word Embeddings." *Springboard*. Retrieved March 1, 2021 (<https://www.springboard.com/blog/introduction-word-embeddings/>).
- Hvitfeldt, Emil, and Julia Silge. 2020. *Supervised Machine Learning for Text Analysis in R*. bookdown.
- Ignatow, Gabe, and Rada Mihalcea. 2017. *Text Mining: A Guidebook for the Social Sciences*. Thousand Oaks, CA: SAGE Publications, Inc.
- Ignatow, Gabriel. 2015. "Theoretical Foundations for Digital Text Analysis." *Journal for the Theory of Social Behaviour* 17. doi: [10.1111/jtsb.12086](https://doi.org/10.1111/jtsb.12086).
- Joo, Jungseock, and Zachary C. Steinert-Threlkeld. 2018. "Image as Data: Automated Visual Content Analysis for Political Science." *ArXiv 1–38*.
- Joseph, Kenneth, and Jonathan H. Morgan. 2020. "When Do Word Embeddings Accurately Reflect Surveys on Our Beliefs about People?" *ArXiv*. doi: [10.18653/v1/2020.acl-main.405](https://doi.org/10.18653/v1/2020.acl-main.405).

- Kim, Jae Yeon. 2021. "Integrating Human and Machine Coding to Measure Political Issues in Ethnic Newspaper Articles." *Journal of Computational Social Science* (0123456789). doi: [10.1007/s42001-020-00097-2](https://doi.org/10.1007/s42001-020-00097-2).
- Kozlowski, Austin C., Matt Taddy, and James A. Evans. 2019. "The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings." *American Sociological Review* 84(5):905–49. doi: [10.1177/0003122419877135](https://doi.org/10.1177/0003122419877135).
- Larson, Ben. 2016. "R: Text Mining (Pre-Processing)." *Analytics4All*. Retrieved March 1, 2021 (<https://analytics4all.org/2016/12/22/r-text-mining-pre-processing/>).
- Light, Ryan. 2014. "From Words to Networks and Back." *Social Currents* 1(2):111–29. doi: [10.1177/2329496514524543](https://doi.org/10.1177/2329496514524543).
- Light, Ryan, and Colin Odden. 2017. "Managing the Boundaries of Taste: Culture, Valuation, and Computational Social Science." *Social Forces* 96(2):877–908. doi: [10.1093/sf/sox055](https://doi.org/10.1093/sf/sox055).
- Lindstedt, Nathan C. 2019. "Structural Topic Modeling For Social Scientists: A Brief Case Study with Social Movement Studies Literature, 2005–2017." *Social Currents* 6(4):307–18. doi: [10.1177/2329496519846505](https://doi.org/10.1177/2329496519846505).
- Lizardo, Omar, Dustin S. Stoltz, Marshall A. Taylor, and Michael Lee Wood. 2018. "Visualizing Bring-Backs." *Socius: Sociological Research for a Dynamic World* 4:237802311880536. doi: [10.1177/2378023118805362](https://doi.org/10.1177/2378023118805362).
- Long, Jacob A., and William P. Eveland. 2018. "Entertainment Use and Political Ideology: Linking Worldviews to Media Content." *Communication Research* 009365021879101. doi: [10.1177/0093650218791011](https://doi.org/10.1177/0093650218791011).
- Lucas, Christopher, Richard A. Nielsen, Margaret E. Roberts, Brandon M. Stewart, and Alex Storer. 2015. "Computer-Assisted Text Analysis for Comparative Politics." *Political Analysis* 23(2):254–77. doi: [10.1093/pan/mpu019](https://doi.org/10.1093/pan/mpu019).
- Martin, Andrew W., Patrick Rafail, and John D. McCarthy. 2017. "What a Story?" *Social Forces* 96(2):779–802. doi: [10.1093/sf/sox057](https://doi.org/10.1093/sf/sox057).
- McNulty, Keith. 2019. "Tidy Web Scraping in R — Tutorial and Resources." *Towards Data Science*. Retrieved March 1, 2021 (<https://towardsdatascience.com/tidy-web-scraping-in-r-tutorial-and-resources-ac9f72b4fe47>).
- Moeller, Kim, Rasmus Munksgaard, and Jakob Demant. 2017. "Flow My FE the Vendor Said: Exploring Violent and Fraudulent Resource Exchanges on Cryptomarkets for Illicit Drugs." *American Behavioral Scientist* 61(11):1427–50. doi: [10.1177/0002764217734269](https://doi.org/10.1177/0002764217734269).
- Mohr, John W., and Petko Bogdanov. 2013. "Introduction-Topic Models: What They Are and Why They Matter." *Poetics* 41(6):545–69. doi: [10.1016/j.poetic.2013.10.001](https://doi.org/10.1016/j.poetic.2013.10.001).
- Mohr, John W., Robin Wagner-Pacifi, Ronald L. Breiger, and Petko Bogdanov. 2013. "Graphing the Grammar of Motives in National Security Strategies: Cultural Interpretation, Automated Text Analysis and the Drama of Global Politics." *Poetics* 41(6):670–700. doi: [10.1016/j.poetic.2013.08.003](https://doi.org/10.1016/j.poetic.2013.08.003).
- Müller, Karsten, and Carlo Schwarz. 2020. "Fanning the Flames of Hate: Social Media and Hate Crime." *Journal of the European Economic Association* 00(0):1–37. doi: [10.1093/jeea/jvaa045](https://doi.org/10.1093/jeea/jvaa045).
- Mützel, Sophie. 2015. "Facing Big Data: Making Sociology Relevant." *Big Data & Society* 2(2):205395171559917. doi: [10.1177/2053951715599179](https://doi.org/10.1177/2053951715599179).
- O'Connor, Brendan, David Bamman, and Noah A. Smith. 2011. "[Computational Text Analysis for Social Science: Model Assumptions and Complexity](#)." *Proceedings of the NIPS Workshop on Computational Social Science and the Wisdom of Crowds* 1–8.
- Petchler, Ross, and Sandra Gonzalez-Bailon. 2015. "[Automated Content Analysis of Online Political Communication](#)." Pp. 433–50 in *Handbook of Digital Politics*, edited by S. Coleman and D.

Freelon. Edward Elgar Publishing.

- Raja, Abdul Majed. 2018. "Text Analysis in R Made Easy with Udpipes." *Towards Data Science*. Retrieved March 1, 2021 (<https://towardsdatascience.com/easy-text-analysis-on-abc-news-headlines-b434e6e3b5b8>).
- van de Rijt, Arnout, Eran Shor, Charles Ward, and Steven Skiena. 2013. "Only 15 Minutes? The Social Stratification of Fame in Printed Media." *American Sociological Review* 78(2):266–89. doi: [10.1177/0003122413480362](https://doi.org/10.1177/0003122413480362).
- Roberts, Carl W. 2000. "A Conceptual Framework for Quantitative Text Analysis." *Quality and Quantity* 34(3):259–74. doi: [10.1023/A:1004780007748](https://doi.org/10.1023/A:1004780007748).
- Rule, Alix, Jean-philippe Cointet, and Peter S. Bearman. 2015. "Lexical Shifts, Substantive Changes, and Continuity in State of the Union Discourse, 1790–2014." *Proceedings of the National Academy of Sciences* 112(35):10837–44. doi: [10.1073/pnas.1512221112](https://doi.org/10.1073/pnas.1512221112).
- Schoonvelde, Martijn, Gijs Schumacher, and Bert N. Bakker. 2019. "Friends with Text as Data Benefits: Assessing and Extending the Use of Automated Text Analysis in Political Science and Political Psychology." *Journal of Social and Political Psychology* 7(1):124–43. doi: [10.5964/jspp.v7i1.964](https://doi.org/10.5964/jspp.v7i1.964).
- Schumacher, Gijs, Nicolai Berk, Christian Pipal, Martijn Schoonvelde, and Denise Traber. 2020. [A Dataset of 11,466 Speeches by European Leaders](#).
- Schwemmer, Carsten, and 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](https://doi.org/10.1080/2474736X.2019.1646102).
- Schwemmer, Carsten, Carly Knight, Emily D. Bello-Pardo, Stan Oklobdzija, Martijn Schoonvelde, and Jeffrey W. Lockhart. 2020. "Diagnosing Gender Bias in Image Recognition Systems." *Socius: Sociological Research for a Dynamic World* 6:237802312096717. doi: [10.1177/2378023120967171](https://doi.org/10.1177/2378023120967171).
- Schwemmer, Carsten, and Oliver Wieczorek. 2020. "The Methodological Divide of Sociology: Evidence from Two Decades of Journal Publications." *Sociology* 54(1):3–21. doi: [10.1177/0038038519853146](https://doi.org/10.1177/0038038519853146).
- Silge, Julia, and David Robinson. 2017. *Text Mining with R*. Sebastopol, CA: O'Reilly Media, Inc.
- Spörlein, Christoph, and Elmar Schlueter. 2020. "Ethnic Insults in YouTube Comments: Social Contagion and Selection Effects During the German 'Refugee Crisis.'" *European Sociological Review* 1–16. doi: [10.1093/esr/jcaa053](https://doi.org/10.1093/esr/jcaa053).
- Steinert-Threlkeld, Zachary C. 2019. "The Future of Event Data Is Images." *Sociological Methodology* 49(1):68–75. doi: [10.1177/0081175019860238](https://doi.org/10.1177/0081175019860238).
- Stoltz, Dustin S., and Marshall A. Taylor. 2019a. "Concept Mover's Distance: Measuring Concept Engagement via Word Embeddings in Texts." *Journal of Computational Social Science* 2(2):293–313. doi: [10.1007/s42001-019-00048-6](https://doi.org/10.1007/s42001-019-00048-6).
- Stoltz, Dustin S., and Marshall A. Taylor. 2019b. "Textual Spanning: Finding Discursive Holes in Text Networks." *Socius: Sociological Research for a Dynamic World* 5:237802311982767. doi: [10.1177/2378023119827674](https://doi.org/10.1177/2378023119827674).
- Underwood, Ted. 2012. "Where to Start with Text Mining." *The Stone and the Shell*. Retrieved March 1, 2021 (<https://tedunderwood.com/2012/08/14/where-to-start-with-text-mining/>).
- Veltri, Giuseppe A. 2020. *Digital Social Research*. Cambridge: Polity Press.
- Vries, Erik De, Martijn Schoonvelde, and Gijs Schumacher. 2018. "No Longer Lost in Translation: Evidence That Google Translate Works for Comparative Bag-of-Words Text Applications." *Political Analysis* 26(4):417–30. doi: [10.1017/pan.2018.26](https://doi.org/10.1017/pan.2018.26).
- Wickham, Hadley, and Garrett Golemund. 2016. *R for Data Science*. Sebastopol, CA: O'Reilly

Media, Inc.