Bibliography

Bibliography

Introduction

Ahnert, Ruth, Emma Griffin, Mia Ridge, and Giorgia Tolfo. *Collaborative Historical Research in the Age of Big Data: Lessons from an Interdisciplinary Project*. Elements in Historical Theory and Practice. Cambridge: Cambridge University Press, 2023. https://doi.org/10.1017/9781009175548.

Algee-Hewitt, Mark. "Computing Criticism: Humanities Concepts and Digital Methods." In *Computational Humanities*, edited by Lauren Tilton, David Mimno, and Jessica Marie Johnson. Minneapolis: University of Minnesota Press, 2024.

Arnold, Taylor, and Lauren Tilton. *Distant Viewing: Computational Exploration of Digital Images*. Cambridge, MA: MIT Press, 2023. https://doi.org/10.7551/mitpress/14046.001.0001.

Alvarado, Rafael. "The Differences between Digital Humanities and Digital History." In *Debates in the Digital Humanities 2016*, edited by Stephen Robertson. Minneapolis: University of Minnesota Press, 2016. https://dhdebates.gc.cuny.edu/projects/debates-in-the-digital-humanities-2016.

Berry, David M., ed. *Understanding Digital Humanities*. Basingstoke, UK: Palgrave Macmillan, 2012.

Blaxill, Luke. "A Feminized Language of Democracy? The Representation of Women at Westminster since 1945." *Twentieth Century British History* 27, no. 3 (September 2016): 412–49. https://doi.org/10.1093/tcbh/hww028.

Blaxill, Luke. "Why Do Historians Ignore Digital Analysis? Bring on the Luddites." *The Political Quarterly* 94 (2023): 279–89. https://doi.org/10.1111/1467-923X.13267.

Blaxill, Luke. *The War of Words: The Language of British Elections, 1880–1914*. Royal Historical Society Studies in History New Series, vol. 104. Woodbridge, UK: Boydell Press, 2020.

Buongiorno, Steph. *hansardr: Access a Cleaned Version of the C19 Hansard Corpus with Improved Speaker Names in the R Environment*. R package. GitHub, 2021. https://github.com/stephbuon/hansardr.

Buongiorno, Steph, Robert Kalescky, Omar Alexander Cerpa, and Jo Guldi. "The Hansard 19th-Century British Parliamentary Debates with Improved Speaker Names: Parsed Debates, N-Gram Counts, Special Vocabulary, Collocates, and Topics." Harvard Dataverse, 2022. https://doi.org/10.7910/DVN/ZCYJH8.

Buongiorno, Steph, Ryan Schaefer, Alisha Morejon, Serena Di Martino, Shawn Chen, Wes Anderson, Matt Swigart, Chris Miller, Hailey DeMark, Grace Schultz, Francesca La Marca, Jo Guldi, and Tommy Ottolin. *Democracy Viewer*. 2024. https://democracyviewer.com/.

Colley, Linda. *Britons: Forging the Nation, 1707–1837*. New Haven, CT: Yale University Press, 1992.

Davidson, Cathy N. "Humanities 2.0: Promise, Perils, Predictions." In *Debates in the Digital Humanities*, edited by Matthew K. Gold. Minneapolis: University of Minnesota Press, 2012. https://dhdebates.gc.cuny.edu/read/untitled-88c11800-9446-469b-a3be-3fdb36bfbd1e/section/c6fc3cc1-5751-4c6b-aac5-b9890206930f#ch28.

Fischer, David Hackett. *Historians' Fallacies: Toward a Logic of Historical Thought*. New York: Harper & Row, 1970.

Grimmer, Justin, Margaret E. Roberts, and Brandon M. Stewart. *Text as Data: A New Framework for Machine Learning and the Social Sciences.* Princeton, NJ: Princeton University Press, 2022.

Guldi, Jo. *The Dangerous Art of Text Mining: A Methodology for Digital History.* Cambridge: Cambridge University Press, 2023.

Guldi, Jo. "Critical Search: A Procedure for Guided Reading in Large-Scale Textual Corpora." *Journal of Cultural Analytics* 3, no. 1 (2018). https://doi.org/10.22148/16.030.

Jockers, Matthew L., and Ted Underwood. "Text-Mining the Humanities." In *A New Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, and John Unsworth. Hoboken, NJ: Wiley, 2015. https://doi.org/10.1002/9781118680605.ch20.

Kirschenbaum, Matthew G. Mechanisms: New Media and the Forensic Imagination. MIT Press, 2012.

Piper, Andrew. *Enumerations: Data and Literary Study.* University of Chicago Press, 2018.

Said, Edward W. Orientalism. Pantheon Books, 1978.

Silge, Julia, and Robinson, David. (2016). "tidytext: Text mining and analysis using tidy data principles in R." *Journal of Open Source Software*, 1(3), 37. https://doi.org/10.21105/joss.00037

Hayles, N. Katherine. How We Think: Digital Media and Contemporary Technogenesis. University of Chicago Press, 2012.

Hicks, Michael Townsen, James Humphries, and Joe Slater. "ChatGPT Is Bullshit." *Ethics and Information Technology*, vol. 26, no. 38, 2024, https://doi.org/10.1007/s10676-024-09753-z.

Lubanovic, Bill. *Introducing Python.* 2nd ed., O'Reilly Media, 2019.

Luhmann, Jan, and Manuel Burghardt. "Digital Humanities—A Discipline in Its Own Right? An Analysis of the Role and Position of Digital Humanities in the Academic Landscape." Journal of the Association for Information Science and Technology, vol. 73, no. 1, 2022, pp. 148–162. Wiley, https://doi.org/10.1002/asi.24533.

Matthes, Eric. *Python Crash Course: A Hands-on, Project-based Introduction to Programming.* 3rd ed., No Starch Press, 2023.

Metcalf, Thomas R. Ideologies of the Raj. Cambridge University Press, 1995.

Mullen, Lincoln A. *Computational Historical Thinking: With Applications in R* (2018–): https://dh-r.lincolnmullen.com.

Ramsay, Stephen. Reading Machines: Toward an Algorithmic Criticism. Champaign: University of Illinois Press, 2011.

Robertson, Stephen. "The Differences between Digital Humanities and Digital History." Debates in the Digital Humanities 2016, edited by Lauren F. Klein and Matthew K. Gold, University of Minnesota Press, 2016, pp. 289–307. Project MUSE, https://muse.jhu.edu/book/46383.

Robertson, Stephen, and Lincoln Mullen. "Arguing with Digital History: Patterns of Historical Interpretation." Journal of Social History, vol. 54, no. 4, July 2021, pp. 1005–1022. Oxford University Press.

Ros, Ruben J. G. "Conceptualizing an Outside World." *The Case of Foreign in Dutch Newspapers 1815–1914.* Contributions to the History of Concepts. 2021.

Sinclair, Stéfan, and Geoffrey Rockwell. Voyant Tools. 2016, http://voyant-tools.org/

Spinu, Vitalie, Garrett Grolemund, and Hadley Wickham. lubridate: Make Dealing with Dates a Little Easier. Version 1.9.2, 2023. R package, https://CRAN.R-project.org/package=lubridate.

Tagliaferri, Lisa. "A Technology of the Vernacular: Re-centering Innovation within the Humanities." *Computational Humanities*, edited by Lauren Tilton, David Mimno, and Jessica Marie Johnson, University of Minnesota Press, 2024.

Toney-Wails, Autumn, and Lisa Singh. "Are You Sure About That: Eliciting Natural Language Confidence Indicators from Chatbots." HAI '24: Proceedings of the 12th International Conference on Human-Agent Interaction, 24 Nov. 2024, pp. 278–86. ACM, https://doi.org/10.1145/3687272.3688318.

Tukey, John W. Exploratory Data Analysis. Reading, MA: Addison-Wesley, 1977

Wexler, Laura, Lauren Tilton, and Taylor Arnold, editors. Photogrammar. Yale University, 2014–2016, https://photogrammar.org/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. "Welcome to the Tidyverse." *Journal of Open Source Software*, vol. 4, no. 43, 2019, p. 1686. https://doi.org/10.21105/joss.01686.

Wineburg, Sam. *Historical Thinking and Other Unnatural Acts: Charting the Future of Teaching the Past.* Temple University Press, 2001.

Zumel, Nina, and John Mount. "Practical Data Science with R." 2nd ed., Manning Publications, 2019.

Chapter 1: Counting Words in the 19th-Century British Parliamentary Debates

Buongiorno, Steph. hansardr: Access a cleaned version of the c19 Hansard corpus with improved speaker names in the R environment. 2021. GitHub, R package. https://github.com/stephbuon/hansardr.

Firth, J. R. *Studies in Linguistic Analysis.* Oxford University Press, 1974.

Guldi, Jo. *The Dangerous Art of Text Mining: A Methodology for Digital History.* Cambridge University Press, 2023. Cambridge Core, https://doi.org/10.1017/9781009263016.

Jockers, Matthew L. *Macroanalysis: Digital Methods and Literary History.* University of Illinois Press, 2013.

Jones, Steven E. (2016). *Roberto Busa, S.J., and the Emergence of Humanities Computing: The Priest and the Punched Cards.* Routledge.

Moretti, Franco. Distant Reading. Verso, 2013.

Silge, Julia, and Robinson, David. (2016). "tidytext: Text mining and analysis using tidy data principles in R." *Journal of Open Source Software*, 1(3), 37. https://doi.org/10.21105/joss.00037

Spinu, Vitalie, Garrett Grolemund, and Hadley Wickham. "lubridate: Make Dealing with Dates a Little Easier." Version 1.9.2, 2023. R package, https://CRAN.R-project.org/package=lubridate.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. "Welcome to the Tidyverse." *Journal of Open Source Software*, vol. 4, no. 43, 2019, p. 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Jim Hester, and Winston Chang. "devtools: Tools to Make Developing R Packages Easier." Version 2.4.5, 2022. R package, https://CRAN.R-project.org/package=devtools.

Williams, Raymond. *Keywords: A Vocabulary of Culture and Society.* Revised ed., Oxford University Press, 1985.

Chapter 2: Investigating the Memory of Events Using N-Grams, Controlled Vocabulary, and Joins

Buongiorno, Steph. hansardr: Access a cleaned version of the c19 Hansard corpus with improved speaker names in the R environment. 2021. GitHub, R package. https://github.com/stephbuon/hansardr.

Buongiorno, Steph, Serena Di Martino, Ryan Schaefer, Shawn Chen, Alisha Morejon, Hailey DeMark, Francesca La Marca, Grace Schultz, Wes Anderson, Chris Miller, and Matt Swigart. Democracy Viewer, 2023, accessed Nov. 9, 2024.

Fyfe, Paul. (2016). "An archaeology of Victorian newspapers." Victorian Periodicals Review, 49(4), 546–577.

Garnier, Simon, Noam Ross, and Robert Rudis. viridis: Colorblind-Friendly Color Maps for R. Version 0.6.4, 2024. R package, https://CRAN.R-project.org/package=viridis.

Guldi, Jo. *The Dangerous Art of Text Mining: A Methodology for Digital History.* Cambridge University Press, 2023.

Palermo, Gregory. "Transforming Text: Four Valences of a Digital Humanities Informed Writing Analytics," in The Journal of Writing Analytics, Vol 1 (2017)

Silge, Julia, and Robinson, David. (2016). "tidytext: Text mining and analysis using tidy data principles in R." *Journal of Open Source Software*, 1(3), 37. https://doi.org/10.21105/joss.00037

Spinu, Vitalie, Garrett Grolemund, and Hadley Wickham. lubridate: Make Dealing with Dates a Little Easier. Version 1.9.2, 2023. R package, https://CRAN.R-project.org/package=lubridate.

Spirling, Arthur and Any Eggers. "Electoral Security as a Determinant of Legislator Activity, 1832–1918: New Data and Methods for Analyzing British Political Development".

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. "Welcome to the Tidyverse." *Journal of Open Source Software*, vol. 4, no. 43, 2019, p. 1686. https://doi.org/10.21105/joss.01686.

Chapter 3: Investigating Speakers and Change Over Time Using Grouped Data

Borges, J. L. 1998. Collected Fictions. New York: Penguin

Buongiorno, Steph. hansardr: Access a cleaned version of the c19 Hansard corpus with improved speaker names in the R environment. 2021. GitHub, R package. https://github.com/stephbuon/hansardr.

Eggers, Andrew C., and Arthur Spirling. "Electoral Security as a Determinant of Legislator Activity, 1832-1918: New Data and Methods for Analyzing British Political Development." Legislative Studies Quarterly 39, no. 4 (2014): 593–620. http://www.jstor.org/stable/43862507.

Silge, Julia, and Robinson, David. (2016). "tidytext: Text mining and analysis using tidy data principles in R." *Journal of Open Source Software*, 1(3), 37. https://doi.org/10.21105/joss.00037

Spinu, Vitalie, Garrett Grolemund, and Hadley Wickham. lubridate: Make Dealing with Dates a Little Easier. Version 1.9.2, 2023. R package, https://CRAN.R-project.org/package=lubridate.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. "Welcome to the Tidyverse." *Journal of Open Source Software*, vol. 4, no. 43, 2019, p. 1686. https://doi.org/10.21105/joss.01686.

Winters, Jane and Martin Steer. Digging into Linked Parliamentary Data (DiLiPaD). 2013.

Chapter 4: Text Mining British Concepts in the 1833 Debates About the Abolition of Slavery

Chapter 5: Using Distinctiveness and Distance for Comparative Analyses of Historical Data

Belkin, Nicholas J., and W. Bruce Croft. "Information Filtering and Information Retrieval: Two Sides of the Same Coin?" *Communications of the ACM*, vol. 35, no. 12, Dec. 1992, pp. 29–38. https://doi.org/10.1145/138859.138861.

Brown, Margaret Wise. *Goodnight Moon.* Harper & Brothers, 1947.

Buongiorno, Steph. hansardr: Access a cleaned version of the c19 Hansard corpus with improved speaker names in the R environment. 2021. GitHub, R package. https://github.com/stephbuon/hansardr.

Domeniconi, Giacomo, Giacomo Moro, Roberto Pasolini, and Cesare Sartori. "A Comparison of Term Weighting Schemes for Text Classification and Sentiment Analysis with a Supervised Variant of tf.idf." Data Management Technologies and Applications. DATA 2015, edited by Markus Helfert, Andreas Holzinger, Oscar Belo, and Chiara Francalanci, vol. 584, Communications in Computer and Information Science, Springer, 2016, https://doi.org/10.1007/978-3-319-30162-4_4.

Fuglede, Bent, and Flemming Topsøe. "Jensen-Shannon Divergence and Hilbert Space Embedding." Proceedings. *International Symposium on Information Theory*, 2004. Chicago, IL, 2004, pp. 31–. IEEE, https://doi.org/10.1109/ISIT.2004.1365067.

Karen Spärck-Jones (1972). "A statistical interpretation of term specificity and its application in retrieval." *Journal of Documentation*.

Klingenstein, Sara, Tim Hitchcock, and Simon DeDeo. "The Civilizing Process in London's Old Bailey." *Proceedings of the National Academy of Sciences*, vol. 111, no. 26, 16 June 2014, pp. 9419–9424. https://doi.org/10.1073/pnas.1405984111.

Lin, J. (1991). "Divergence measures based on the Shannon entropy." *IEEE Transactions on Information Theory*, 37(1), 145–151. https://doi.org/10.1109/18.61115

Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze. "Scoring, Term Weighting, and the Vector Space Model." *Introduction to Information Retrieval*, Cambridge University Press, 2008, p. 100. doi:10.1017/CBO9780511809071.007.

Schnober, Carsten, and Iryna Gurevych. "Combining Topic Models for Corpus Exploration: Applying LDA for Complex Corpus Research Tasks in a Digital Humanities Project." TM '15: Proceedings of the 2015 Workshop on Topic Models: Post-Processing and Applications, 18 Oct. 2015, pp. 11–20. https://doi.org/10.1145/2809936.2809939.

Schofield, Alexandra, Måns Magnusson, and David Mimno. "Pulling Out the Stops: Rethinking Stopword Removal for Topic Models." *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*. Volume 2, Short Papers, edited by Mirella Lapata, Phil Blunsom, and Alexander Koller, Association for Computational Linguistics, Apr. 2017, Valencia, Spain, pp. 432–436. https://aclanthology.org/E17-2069/.

Seuss, Dr. *The Cat in the Hat*. Random House, 1957.

Seuss, Dr. *Green Eggs and Ham*. Random House, 1960.

Silge, Julia, and Robinson, David. (2016). "tidytext: Text mining and analysis using tidy data principles in R." *Journal of Open Source Software*, 1(3), 37. https://doi.org/10.21105/joss.00037

Spinu, Vitalie, Garrett Grolemund, and Hadley Wickham. lubridate: Make Dealing with Dates a Little Easier. Version 1.9.2, 2023. R package, https://CRAN.R-project.org/package=lubridate.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. "Welcome to the Tidyverse." *Journal of Open Source Software*, vol. 4, no. 43, 2019, p. 1686. https://doi.org/10.21105/joss.01686.

Wilke, Claus O. *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*. O'Reilly Media, 2019.

Xiang, Lin. "Application of an Improved TF-IDF Method in Literary Text Classification", *Advances in Multimedia*, 2022. https://doi.org/10.1155/2022/9285324

Chapter 7: Decoding Grammar to Uncover the Language of Agency

ADD THESE:

Rackley and Auchmuty 2019 Kent 1999

Griffin 2012

(Tangherlini 2016,

Franzosi 2024).

(Allison 2018)

(Arnold 2019)

Kraicer and Piper 2018,

Underwood et. al 2018)

Allison, Sarah. *Reductive Reading: A Syntax of Victorian Moralizing.* Johns Hopkins University Press, 2018.

Arnold, Taylor, Nicolas Ballier, Paula Lissón, and Lauren Tilton. "Beyond Lexical Frequencies: Using R for Text Analysis in the Digital Humanities." Language Resources and Evaluation, vol. 53, no. 4, 2019, pp. 707–733. Springer, https://doi.org/10.1007/s10579-019-09456-6.

Bode, Katherine. *A World of Fiction: Digital Collections and the Future of Literary History.* University of Michigan Press, 2018.

Buongiorno, Steph. hansardr: Access a cleaned version of the c19 Hansard corpus with improved speaker names in the R environment. 2021. GitHub, R package. https://github.com/stephbuon/hansardr.

Cheng, Jonathan. "Fleshing out Models of Gender in English-Language Novels (1850–2000)." *Journal of Cultural Analytics*, 2020, https://doi.org/10.22148/001c.11831.

Cixous, Hélène. "The Laugh of the Medusa." Translated by Keith Cohen and Paula Cohen. Signs 1, no. 4 (1976): 875–893.

Eve Kraicer and Andrew Piper, "Social Characters: The Hierarchy of Gender in Contemporary English-Langauge Fiction," *Journal of Cultural Analytics* (2018).

Franzosi, Roberto. *From Words to Numbers: Narrative, Data, and Social Science.* Cambridge University Press, 2004.

George Eliot, *Daniel Deronda.* New York: Penguin, 1995, pp. 118.

Guldi, Jo. "Critical Search: A Procedure for Guided Reading in Large-Scale Textual Corpora." Journal of Cultural Analytics, vol. 3, no. 1, 19 Dec. 2018, https://doi.org/10.22148/16.030.

Honnibal, Matthew, and Ines Montani. 2017. "spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing." https://explosion.ai.

Irigaray, Luce. This Sex Which Is Not One. Translated by Catherine Porter and Carolyn Burke. Ithaca, NY: Cornell University Press, 1985.

Johnson, Barbara. *A World of Difference.* (Baltimore: The Johns Hopkins University, 1987).

Manning, Christopher D., Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. "The Stanford CoreNLP Natural Language Processing Toolkit." In Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, 55–60. https://aclanthology.org/P14-5010

Marcus, Mitchell P., Beatrice Santorini, and Mary Ann Marcinkiewicz. (1993). Building a large annotated corpus of English: The Penn Treebank. Computational Linguistics, 19(2), 313–330.

Mulvey, Laura. Visual Pleasure and Narrative Cinema, Screen, Volume 16, Issue 3, Autumn 1975, Pages 6–18, https://doi.org/10.1093/screen/16.3.6

Silge, Julia, and Robinson, David. (2016). "tidytext: Text mining and analysis using tidy data principles in R." *Journal of Open Source Software*, 1(3), 37. https://doi.org/10.21105/joss.00037

Spinu, Vitalie, Garrett Grolemund, and Hadley Wickham. lubridate: Make Dealing with Dates a Little Easier. Version 1.9.2, 2023. R package, https://CRAN.R-project.org/package=lubridate.

Tangherlini, Timothy R., et al. " 'Mommy Blogs' and the Vaccination Exemption Narrative: Results From a Machine-Learning Approach for Story Aggregation on Parenting Social Media Sites." JMIR Public Health and Surveillance, vol. 2, no. 2, 2016, doi:10.2196/publichealth.5054.

Underwood, Ted, David Bamman, and Sabrina Lee. "The Transformation of Gender in English-Language Fiction." Journal of Cultural Analytics, 2018, https://doi.org/10.22148/16.036.

Walkowitz, Judith R. *Prostitution and Victorian Society: Women, Class, and the State.* Cambridge University Press, 1980, pp.4.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. "Welcome to the Tidyverse." *Journal of Open Source Software*, vol. 4, no. 43, 2019, p. 1686. https://doi.org/10.21105/joss.01686.

Wittig, Monique. The Straight Mind and Other Essays. Translated by Monique Wittig and others, Beacon Press, 1992.

Zeldes, Amir, and Caroline T. Schroeder. "Computational Methods for Coptic: Developing and Using Part-of-Speech Tagging for Digital Scholarship in the Humanities." Digital Scholarship in the Humanities, vol. 30, suppl. 1, Dec. 2015, pp. i164–i176. Oxford University Press, https://doi.org/10.1093/llc/fqv043.

Chapter 8: Analyzing the Language of Climate Change in the United States Congressional Record

Silge, Julia, and Robinson, David. (2016). "tidytext: Text mining and analysis using tidy data principles in R." *Journal of Open Source Software*, 1(3), 37. https://doi.org/10.21105/joss.00037

Spinu, Vitalie, Garrett Grolemund, and Hadley Wickham. lubridate: Make Dealing with Dates a Little Easier. Version 1.9.2, 2023. R package, https://CRAN.R-project.org/package=lubridate.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. "Welcome to the Tidyverse." *Journal of Open Source Software*, vol. 4, no. 43, 2019, p. 1686. https://doi.org/10.21105/joss.01686.

Chapter 9: Data

BBC News. BBC News RSS Feed. https://feeds.bbci.co.uk/news/rss.xml. Accessed 22 Mar. 2025.

Benoit, Kenneth, et al. "Quanteda: An R Package for the Quantitative Analysis of Textual Data." Journal of Open Source Software, vol. 3, no. 30, 2018, p. 774, https://doi.org/10.21105/joss.00774.

Buongiorno, Steph. Code for "Foundations and Applications of Humanities Analytics" (2023). GitHub, 2023, https://github.com/stephbuon/faha-2023. Accessed 22 Mar. 2025.

Buongiorno, Steph. hansardr: Access a cleaned version of the c19 Hansard corpus with improved speaker names in the R environment. 2021. GitHub, R package. https://github.com/stephbuon/hansardr.

City of Dallas. "Dallas Animal Shelter Data Fiscal Year 2019–2020." Dallas Open Data, https://www.dallasopendata.com/resource/7h2m-3um5.

Cordell, Ryan. " 'Q i-jtb the Raven': Taking Dirty OCR Seriously." *Book History*, vol. 20, 2017, pp. 188–225. Johns Hopkins University Press, https://doi.org/10.1353/bh.2017.0006.

Dobson, James E. Critical Digital Humanities: The Search for a Methodology. First ed., University of Illinois Press, 2019. Topics in the Digital Humanities.

Greg, W. W. "Bibliography—an Apologia," in *Collected Papers*, ed. J. C. Maxwell (Oxford: Clarendon Press, 1966), 257, 259

The Guardian. "Open Platform." https://open-platform.theguardian.com/.

Hai-Jew, Shalin, editor. Data Analytics in Digital Humanities. Springer, 2018. Multimedia Systems and Applications.

HathiTrust Research Center. HathiTrust Digital Library. https://www.hathitrust.org.

Hayler, Matt, and Gabriele Griffin, editors. *Research Methods for Creating and Curating Data in the Digital Humanities.* Illustrated ed., Edinburgh University Press, 2016.

Hitchcock, Tim, Robert Shoemaker, Clive Emsley, Sharon Howard, and Jamie McLaughlin, et al. *The Proceedings of the Old Bailey, 1674–1913: A Fully Searchable Edition of the Proceedings of London's Central Criminal Court.* Version 8.0, 2013, https://www.oldbaileyonline.org.

Kaggle. Datasets. https://www.kaggle.com/datasets.

King, Gary, et al. Harvard Dataverse Repository. https://dataverse.harvard.edu.

Kirschenbaum, Matthew. "Text Messaging." *Mechanisms: New Media and the Forensic Imagination*, MIT Press, 2008, p. 17.

Magritte, René. The Treachery of Images (This is Not a Pipe). 1929, Los Angeles County Museum of Art.

Milligan, Ian. *History in the Age of Abundance?: How the Web is Transforming Historical Research.* McGill-Queen's University Press, 2019.

Ooms, Jeroen. pdftools: Text Extraction, Rendering and Converting of PDF Documents. R package version 3.3.1, 2023, https://CRAN.R-project.org/package=pdftools.

Ooms, Jeroen, Duncan Temple Lang, and Lloyd Hilaiel. jsonlite: A Simple and Robust JSON Parser and Generator for R. R package version 1.8.8, 2023, https://CRAN.R-project.org/package=jsonlite.

Programming Historian. "Lesson Index: APIs." *The Programming Historian.* Accessed October 20, 2025. https://programminghistorian.org/en/lessons/?topic=api

Programming Historian. "Lesson Index: Web Scraping." *The Programming Historian.* Accessed October 20, 2025. https://programminghistorian.org/en/lessons/?topic=web-scraping

Reddit. "API Documentation." https://www.reddit.com/dev/api/.

Roued-Cunliffe, Henriette. "Open Heritage Data and APIs." Chapter. In Participatory Heritage, edited by Henriette Roued-Cunliffe and Andrea Copeland, 195–204. Facet, 2017.

Ries, Thorsten. "Digital History and Born-Digital Archives: The Importance of Forensic Methods." *Journal of the British Academy* 10 (2022): 157–185.

Thordarson, Sveinbjorn, ed. *Egil's Saga.* Reykjavík: Icelandic Saga Database. Accessed November 24, 2025. https://sagadb.org/files/pdf/egils_saga.en.pdf

Silge, Julia, and Robinson, David. (2016). "tidytext: Text mining and analysis using tidy data principles in R." *Journal of Open Source Software*, 1(3), 37. https://doi.org/10.21105/joss.00037

Silge, Julia, and David Robinson. tidytext: Text Mining Using 'dplyr', 'ggplot2', and Other Tidy Tools. R package version 0.4.1, 2022, https://CRAN.R-project.org/package=tidytext.

Spinu, Vitalie, Garrett Grolemund, and Hadley Wickham. lubridate: Make Dealing with Dates a Little Easier. Version 1.9.2, 2023. R package, https://CRAN.R-project.org/package=lubridate.

University of Virginia Library. "APIs for Scholarly Research." LibGuides at UVA Library, University of Virginia, https://guides.lib.virginia.edu/c.php?g=515084&p=10046079&utm_source

Wachowiak, Mark P. "Archives, Museums, and Other GLAM Institutions." Contemporary Digital Humanities, 2022, https://ecampusontario.pressbooks.pub/nudh1/chapter/archives-museums-and-other-glam-institutions/.

Wickham, Hadley. rvest: Easily Harvest (Scrape) Web Pages. Version 1.0.3, 2022, https://CRAN.R-project.org/package=rvest.

Wickham, Hadley. httr: Tools for Working with URLs and HTTP. Version 1.4.7, 2023, https://CRAN.R-project.org/package=httr.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. "Welcome to the Tidyverse." *Journal of Open Source Software*, vol. 4, no. 43, 2019, p. 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, and Jim Hester. fs: Cross-Platform File System Operations Based on 'libuv'. R package version 1.6.3, 2023, https://CRAN.R-project.org/package=fs.

Wickham, Hadley. stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.5.1, 2023, https://CRAN.R-project.org/package=stringr.

Glossary

Raymond Williams Keywords

Alexander, Marc. (2023) Speech in the British Hansard. In: Korhonen, M., Kotze, H. and Tyrkkö, J. (eds.) Exploring Language and Society with Big Data: Parliamentary Discourse across Time and Space. Series: Studies in corpus linguistics (111). John Benjamins: Amsterdam, pp. 17-53. ISBN 9789027214065 (doi: 10.1075/scl.111.01ale).

Demmen, Jane, Lesley Jeffries, and Brian Walker. "Charting the Semantics of Labour Relations in House of Commons Debates Spanning Two Hundred Years: A Study of Parliamentary Language Using Corpus Linguistic Methods and Automated Semantic Tagging." In Doing Politics: Discursivity, Performativity and Mediation in Political Discourse, edited by Michael Kranert and Geraldine Horan, 81–104. Discourse Approaches to Politics, Society and Culture 80. Amsterdam: John Benjamins, 2018.

In New Digital Worlds (2018), Roopika Risam argues that data in the digital humanities must be understood as a product of cultural and historical processes.

Cite this: https://www.nature.com/articles/sdata201618

Underwood novel TM dataset

For example, they might show that in the United States Congressional Record, the word man became increasingly associated with the term gay in later decades—indicating a change in how certain concepts were linked in political discourse (Buongiorno). congress viewer