

Intro to Data Science for Law Librarians Resource List

Sarah Lin, Information Architect & Digital Librarian at RStudio, PBC

9/24/2020

Datasets & R packages

- State of California COVID Dashboard
 - Interactive Shiny application
 - * <https://calcat.covid19.ca.gov/cacovidmodels/>
 - GitHub repository
 - * <https://github.com/StateOfCalifornia/CalCAT>
- Washington Post Investigation
 - “How Jared Kushner built a luxury skyscraper using loans meant for job-starved areas” (2017-05-31)
 - * https://www.washingtonpost.com/investigations/jared-kushner-and-his-partners-used-a-program-meant-for-job-starved-areas-to-build-a-luxury-skyscraper/2017/05/31/9c81b52c-4225-11e7-9869-bac8b446820a_story.html
 - GitHub repository of data and code for the investigation
 - * https://github.com/wpinvestigative/kushner_eb5_census
 - Census gerrymandering loophole was closed in July 2019
 - * <https://www.uscis.gov/news/news-releases/new-rulemaking-brings-significant-changes-to-eb-5-program>
- Racial disparity in arrests
 - Interactive Shiny application
 - * <https://redbird.shinyapps.io/police-bias-map/>
 - Research paper
 - * <https://www.ipr.northwestern.edu/our-work/working-papers/2020/wp-20-28.html>
- Election Data
 - <http://freerangestats.info/blog/2016/10/29/538-pollsters>
 - <https://www.computerworld.com/article/3139884/r-resources-for-election-night.html>
 - <https://walkerke.github.io/tidycensus/>
- Comprehensive R Archive Network: a listing of all packages available in R
 - <https://cran.r-project.org/>

Improving Data Management skills

- MarcEdit
 - <https://maredit.reeset.net>
- Library Carpentry lesson on OpenRefine
 - <https://librarycarpentry.org/lc-open-refine/>
- The Carpentries: Software, Data, & Library Carpentry
 - <https://carpentries.org>

Improving statistical thinking

- *Naked Statistics* (Wheelan 2013)
- *How to Lie with Statistics* (Huff 1993)
- *Weapons of Math Destruction* (O’Neil 2016)
- Khan Academy AP Statistics
 - <https://www.khanacademy.org/math/ap-statistics/>
- Stat545: former statistics course at the University of British Columbia
 - <https://stat545.com/>
- *Doing Data Science* (O’Neil & Schutt 2014)

Learning R

- Download the RStudio IDE (Integrated Development Environment)
 - <https://rstudio.com/products/rstudio/download/>
- *Hands on Programming with R* (Grolemund 2014)
 - <https://rstudio-education.github.io/hopr/>
- The Carpentries: Software, Data, & Library Carpentry
 - <https://carpentries.org>
- RStudio Education website
 - <https://rstudio.com/education>
- RStudio’s IDE in the cloud
 - <https://rstudio.cloud>
- *R for Data Science* (Wickham & Grolemund 2017)
 - <https://r4ds.had.co.nz/>
- *R Markdown* (Xie, Allaire & Grolemund 2019)
 - <https://bookdown.org/yihui/rmarkdown/>
- *Practical R for Mass Communication and Journalism* (Machlis 2019)
- *Text Mining with R* (Silge & Robinson 2017)
 - <https://www.tidytextmining.com/>
- Tidymodels: packages for machine learning in R
 - <https://tidymodels.org>

Articles & blog posts

- “Dispelling R and Python myths” (2020-09-10)
 - <https://blog.rstudio.com/2020/09/10/dispelling-r-and-python-myths-qanda/>
- “When an algorithm helps send you to prison” (2017-10-26)
 - <https://www.nytimes.com/2017/10/26/opinion/algorithm-compas-sentencing-bias.html>
- “Machine bias” (2016-05-23)
 - <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- “Analysis: AI tech use is common, but do lawyers understand it?” (2020-08-27)
 - <https://news.bloomberglaw.com/bloomberg-law-analysis/analysis-ai-tech-use-is-common-but-do-lawyers-understand-it>
- “Predictive policing algorithms are racist. They need to be dismantled.” (2020-07-17)
 - <https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/?fbclid=IwAR3zTH9U0OrjaPPqifYSjldzgqyIbag6m-GYKBAPQ7jo488SYYl5NbfzrjI>
- Articles on algorithmic bias by Susan Nevelow Mart
 - <http://epubs.aallnet.org/i/862080-aall-spectrum-september-october-2017-volume-22-number-1/41?>
 - <https://www.aallnet.org/wp-content/uploads/2019/10/SPECTRUM-Inside-the-Black-Box-of-Search-Algorithms.pdf>
 - <https://scholar.law.colorado.edu/articles/755/>
- “How to make a racist AI without really trying” (2017-07-13)
 - <https://blog.conceptnet.io/posts/2017/how-to-make-a-racist-ai-without-really-trying>
- “The problems AI has today go back centuries” (2020-17-31)
 - <https://www.technologyreview.com/2020/07/31/1005824/decolonial-ai-for-everyone>
- “An understanding of AI’s limitations is starting to sink in” (2016-06-11)
 - <https://www.economist.com/technology-quarterly/2020/06/11/an-understanding-of-ais-limitations-is-starting-to-sink-in>
- “Semantics derived automatically from language corpora contain human-like biases” (2017-04-17)
 - DOI: 10.1126/science.aal4230
- “Artificial intelligence policy: a primer and roadmap” (Dec 2017)
 - https://lawreview.law.ucdavis.edu/issues/51/2/Symposium/51-2_Calo.pdf
- R and Excel
 - <https://education.rstudio.com/blog/2020/08/spreadsheets-using-r/>
 - <https://rstudio-conf-2020.github.io/r-for-excel/index.html>

Organizations and other resources

- Ryan Timpe
 - Website
 - * <https://ryantimpe.com>
 - rstudio::conf talk: “Learning R wit Humhorus Side Projects”
 - * <https://rstudio.com/resources/rstudioconf-2020/learning-r-with-humorous-side-projects/>
- The Data-Sitters Club (computational text analysis for digital humanities)
 - <https://datasittersclub.github.io/site/>
- Data for Black Lives
 - <https://d4bl.org/>
- *Algorithms of Oppression* (Nobel 2018)
- *Race after Technology* (Benjamin 2019)
- Machine Learning for Everyone
 - https://vas3k.com/blog/machine_learning/

About RStudio

- RStudio’s website
 - <https://rstudio.com>
- rstudio::conf Keynote presentation by RStudio CEO JJ Allaire
 - <https://rstudio.com/resources/rstudioconf-2020/open-source-software-for-data-science/>