Web and Network Data Science

Modeling Techniques in Predictive Analytics

THOMAS W. MILLER

Contents

12

What's Next for the Web?

[Bar scene. Princeton, New Jersey]

Nash: "Adam Smith needs revision."

Hansen: "What are you talking about?"

Nash: "If we all go for the blonde, we block each other and not a single one of us is going to get her. So then we go for her friends. But they will all give us the cold shoulder because nobody likes to be second choice. But what if no one goes for the blonde? We don't get in each other's way, and we don't insult the other girls. That's the only way we win. That's the only way we all get laid.

"Adam Smith said the best result comes from everyone in the group doing what is best for himself. Right? Incomplete. Incomplete. Because the best result will come from everyone in the group doing what's best for himself and the group."

Hansen: "Nash, if this is some way for you to get the blonde on your own, you can go to hell."

The psychologist Jean Piaget, inspired by biological thinking, would speak of assimilation and accommodation as two ways of learning. Assimilation involves taking in new information and superimposing it on already established cognitive structures. We take in more information, but we keep thinking the same way.

And then there is the learning that forces us to change, what Piaget called accommodation. When new information arrives that does not fit with what we know, we must revamp our cognitive structures.

A teacher's job is not to make learning easy. Transformative learning, the learning that matters, is never easy. A teacher's job is to help students change, to open their eyes to new worlds and new ways of thinking.

When I think of the teachers who made a difference in my life, I realize they were teachers who helped me think in new ways. They were not particularly strict, not taskmasters, although they certainly had standards. What they were most of all was supportive. They made it easier for me to do the hard learning I had to do.

There is learning to be done on the web. It is hard. It will take years to pay dividends, and it will never be finished. But it is learning that will change the way we live and work and think. This learning is machine learning, and it involves the *semantic web*.

We often describe web data as unstructured or semi-structured text. In this book we have worked numerous examples, crawling, scraping, parsing, and analyzing, all the while trying to make sense of things.

The Document Object Model may help us to traverse the text domain. XML and HTML and their tags for nodes provide a structure for storing and displaying disparate pieces of information. These standards for information interchange promote communication and collaboration among research entities, just as they facilitate interprocess communication among computers. But recognize that these are merely text formatting rules. They were never intended to be a map of the knowledge of the web.

We lament the difficulties of text analytics. We endure the idiosyncrasies of natural language, developing specialized utilities and processing schemes. But what if we could take text relating to what we know and store it in a way that could be more easily accessed by computers and people? What if

we revamped the cognitive structure of the web, organizing information in a systematic way, rendering it as knowledge?

We might think of the future research world as a collection of web services linking networks of information providers and clients. The World Wide Web itself holds promise as an information store or semantic web, a vast repository of machine-retrievable and machine-understandable data (Fenzel et al. 2003; Daconta, Obrst, and Smith 2003).

Structuring text to represent knowledge is the goal of the semantic web. More than an idea about what might be possible for organizing text, it is a set of technologies that are being used by researchers today. Allemang and Hendler (2007) and Wood, Zaidman, and Ruth (2014) provide an overview, with additional information about data representation and programming techniques available in various sources (Powers 2003; Lacy 2005; Segaran, Evans, and Taylor 2009; DuCharme 2013).

Semantic databases, natural language processing, and machine learning come together in the domain of question-answering. We have a long way to go before computers querying a web can match what we do as thinkers, but there has been great progress with the artificial intelligence for question-answering. Examples include Cyc (Lenat et al. 2010), Halo (Gunning et al. 2010), and IBM's Watson (Ferrucci et al. 2010; Lopez et al. 2013). And extracting linked information from the web is the purpose of open-source projects such as DBpedia (Bizar et al. 2009, Lehmann et al. 2014), a semantic data source for Watson. Imagine a Wikipedia bot that knows how facts fit together and can answer your every question. Making this happen is just a small matter of programming (we used to call it "SMOP").

Software is a collective endeavor. Sharing code in an open-source environment is the key to working efficiently today. We build on a foundation of tools created by others. The growth of the Python and R communities and the story of GitHub (Bourne 2013) show the promise of collaborative software environments.

Evan "Ev" Williams and Jack Dorsey talk about the origin of the microblogging that made their firm famous. Jack says he was the inventor. Ev replies, "No, you didn't invent Twitter. I didn't invent Twitter either. Neither did Biz [Christopher "Biz" Stone]. People don't invent things on the Internet. They just expand on an idea that already exists" (Bilton 2013, p. 203).

This is the way of the web—collaboration, teamwork, sharing programs and ideas. Along the way a few people get rich, and many others are listed as contributors. The data train moves forward.

Only time will tell what new community endeavors emerge. But I suspect we will see products and services that turn information into knowledge and formulaic reasoning into intelligence.

There is a child's circle game, the story game. Boys and girls form a circle. A story starts, one child whispering in the ear of the next. Around the circle the story goes, until it returns to the first child.

"Oh, no," she says, "that isn't the story I told. Not at all."

Some think the story game a lesson in information loss or distortion. But what if the opposite were true? What if the story got better as it traveled the circle—embellished, improved, the product of a learning community? What if the lesson were about collaborative creation?

Looking north from my kitchen window, I see the San Gabriel Mountains. From a distance they are very beautiful. That is how I feel about the future. No fears. Just wonder. I wonder what is next to learn.

Bibliography

Ackland, R. 2013. *Web Social Science: Concepts, Data and Tools for Social Scientists in the Digital Age*. Los Angeles: Sage.

Adler, J. 2010. *R in a Nutshell: A Desktop Quick Reference*. Sebastopol, Calif.: O'Reilly.

Agrawal, R., H. Mannila, R. Srikant, H. Toivonen, and A. I. Verkamo 1996. Fast discovery of association rules. In U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy (eds.), *Handbook of Data Mining and Knowledge Discovery*, Chapter 12, pp. 307–328. Menlo Park, Calif. and Cambridge, Mass.: American Association for Artificial Intelligence and MIT Press.

Agresti, A. 2013. *Categorical Data Analysis* (third ed.). New York: Wiley.

Airoldi, E. M., D. Blei, E. A. Erosheva, and S. E. Fienberg (eds.) 2014. *Handbook of Mixed Membership Models and Their Applications*. Boca Raton, Fla.: CRC Press.

Albert, J. 2009. *Bayesian Computation with R*. New York: Springer. 244

Albert, R. and A.-L. Barabási 2002, January. Statistical mechanics of complex networks. *Reviews of Modern Physics* 74(1):47–97. 99

Alexa Internet, I. 2014. The top 500 sites on the web. Retrieved from the World Wide Web, October 21, 2014, at `http://www.alexa.com/topsites`.

Alfons, A. 2014a. *cvTools: Cross-Validation Tools for Regression Models*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/cvTools/cvTools.pdf`.

Alfons, A. 2014b. *simFrame: Simulation Framework*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/simFrame/simFrame.pdf`. 250

Alfons, A., M. Templ, and P. Filzmoser 2014. *An Object-Oriented Framework for Statistical Simulation: The R Package simFrame*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/simFrame/vignettes/simFrame-intro.pdf`. 250

Allemang, D. and J. Hendler 2007. *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL* (second ed.). Boston: Morgan Kaufmann.

Allison, P. D. 2010. *Survival Analysis Using SAS: A Practical Guide* (second ed.). Cary, N.C.: SAS Institute Inc.

Amor, D. 2002. *The E-business (R)evolution: Living and Working in an Interconnected World* (second ed.). Upper Saddle River, N.J.: Prentice Hall.

Andersen, P. K., Ø. Borgan, R. D. Gill, and N. Keiding 1993. *Statistical Models Based on Counting Processes*. New York: Springer.

Appleton, D. R. 1995. May the best man win? *The Statistician* 44(4):529–538. 103

Asur, S. and B. A. Huberman 2010. Predicting the future with social media. In *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 01*, WI-IAT '10, pp. 492–499. Washington, DC, USA: IEEE Computer Society. http://dx.doi.org/10.1109/WI-IAT.2010.63. 133, 134

Auguie, B. 2014. *Package gridExtra: Functions in grid graphics*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/gridExtra/gridExtra.pdf`.

Baayen, R. H. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics with R*. Cambridge, UK: Cambridge University Press. 259

Bache, K. and M. Lichman 2013. UCI machine learning repository. `http://archive.ics.uci.edu/ml`. 296

Baeza-Yates, R. and B. Ribeiro-Neto 1999. *Modern Information Retrieval*. New York: ACM Press.

Barabási, A.-L. 2003. *Linked: How Everything is Connected to Everything Else*. New York: Penguin/Plume.

Barabási, A.-L. 2010. *Bursts: The Hidden Pattern Behind Everything We Do*. New York: Penguin/Dutton.

Barabási, A.-L. and R. Albert 1999. Emergence of scaling in random networks. *Science* 286 (5439):509–512. 224

Bates, D. and M. Maechler 2014. *Matrix: Sparse and Dense Matrix Classes and Methods*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/support.CEs/support.CEs.pdf`. 210

Bates, D. M. and D. G. Watts 2007. *Nonlinear Regression Analysis and Its Applications*. New York: Wiley.

Beazley, D. M. 2009. *Python Essential Reference* (fourth ed.). Upper Saddle River, N.J.: Pearson Education.

Beazley, D. M. and B. K. Jones 2013. *Python Cookbook* (third ed.). Sebastopol, Calif.: O'Reilly.

Becker, R. A. and W. S. Cleveland 1996. *S-Plus Trellis^{TM} Graphics User's Manual*. Seattle: MathSoft, Inc. 252

Behnel, S. 2014, September 10. lxml. Supported at `http://lxml.de/`. Documentation retrieved from the World Wide Web, October 26, 2014, at `http://lxml.de/3.4/lxmldoc-3.4.0.pdf`.

Belew, R. K. 2000. *Finding Out About: A Cognitive Perspective on Search Engine Technology and the WWW*. Cambridge: Cambridge University Press.

Belsley, D. A., E. Kuh, and R. E. Welsch 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley.

Berk, R. A. 2008. *Statistical Learning from a Regression Perspective*. New York: Springer.

Berners-Lee, T. 2000. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web*. New York: HarperBusiness.

Berry, M. and M. Browne 2005, October. Email surveillance using non-negative matrix factorization. *Computational and Mathematical Organization Theory* 11(3):249–264. 81

Berry, M. W. and M. Browne 1999. *Understanding Search Engines: Mathematical Modeling and Text Retrieval*. Philadelphia: Society for Industrial and Applied Mathematics.

Berthold, M. R., N. Cebron, F. Dill, T. R. Gabriel, T. Kötter, T. Meinl, P. Ohl, C. Sieb, K. Thiel, and B. Wiswedel 2007. KNIME: The Konstanz Information Miner. In *Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007)*. New York: Springer. ISSN 1431-8814. ISBN 978-3-540-78239-1.

Bilton, N. 2013. *Hatching Twitter: A True Story of Money, Power, Friendship, and Betrayal*. New York: Portfolio/Penguin.

Bird, S., E. Klein, and E. Loper 2009. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. Sebastopol, Calif.: O'Reilly. `http://www.nltk.org/book/`. 172

Bishop, Y. M. M., S. E. Fienberg, and P. W. Holland 1975. *Discrete Multivariate Analysis: Theory and Practice*. Cambridge: MIT Press.

Bizer, C., J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann 2009. DBpedia—A crystallization point for the web of data. *Web Semantics* 7:154–165.

Blei, D., A. Ng, and M. Jordan 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research* 3:993–1022. 176, 251

Bogdanov, P. and A. Singh 2013. Accurate and scalable nearest neighbors in large networks based on effective importance. *CIKM '13*:1–10. ACM CIKM '13 conference paper retrieved from the World Wide Web at `http://www.cs.ucsb.edu/~petko/papers/bscikm13.pdf`.

Bojanowski, M. 2014. *Package intergraph: Coercion Routines for Network Data Objects in R*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/intergraph/intergraph.pdf`.

Bollen, J., H. Mao, and X. Zeng 2011. Twitter mood predicts the stock market. *Journal of Computational Science* 2:1–8.

Borg, I. and P. J. F. Groenen 2010. *Modern Multidimensional Scaling: Theory and Applications* (second ed.). New York: Springer.

Borgatti, S. P., M. G. Everett, and J. C. Johnson 2013. *Analyzing Social Networks*. Thousand Oaks, Calif.: Sage. 102

Boser, B. E., I. M. Guyon, and V. N. Vapnik 1992. A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Conference on Computational Learning Theory*, pp. 144–152. Association for Computing Machinery Press. 129

Bourne, W. 2013, February 27. 2 reasons to keep and eye on github. *Inc.com*. Retrieved from the World Wide Web, November 6, 2014, at `http://www.inc.com/magazine/201303/will-bourne/2-reasons-to-keep-an-eye-on-github.html`. 235

Box, G. E. P. and D. R. Cox 1964. An analysis of transformations. *Journal of the Royal Statistical Society, Series B (Methodological)* 26(2):211–252.

Boztug, Y. and T. Reutterer 2008. A combined approach for segment-specific market basket analysis. *European Journal of Operational Research* 187(1):294–312. 210

Bradley, R. A. 1976, June. Science, statistics, and paired comparisons. *Biometrics* 32(2):213–239.

Bradley, R. A. and M. E. Terry 1952, December. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika* 39(3/4):324–345.

Bradley, S. P., A. C. Hax, and T. L. Magnanti 1977. *Applied Mathematical Programming*. Reading, Mass.: Addison-Wesley. 225

Brandes, U. and T. Erlebach (eds.) 2005. *Network Analysis: Methodological Foundations*. New York: Springer.

Brandes, U., L. C. Freeman, and D. Wagner 2014. Social networks. In R. Tamassia (ed.), *Handbook of Graph Drawing and Visualization*, Chapter 26, pp. 805–840. Boca Raton, Fla.: CRC Press/Chapman & Hall.

Breese, J. S., D. Heckerman, and C. M. Kadie 1998, May. Empirical analysis of predictive algorithms for collaborative filtering. Technical Report MSR-TR-98-12, Microsoft Research. 18 pp. `http://research.microsoft.com/apps/pubs/default.aspx?id=69656`. 210, 296

Breiman, L. 2001a. Random forests. *Machine Learning* 45(1):5–32.

Breiman, L. 2001b. Statistical modeling: The two cultures. *Statistical Science* 16(3):199–215.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone 1984. *Classification and Regression Trees*. New York: Chapman & Hall.

Brin, S. and L. Page 1998. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems* 30:107–117. 46, 103

Brownlee, J. 2011. *Clever Algorithms: Nature-Inspired Programming Recipies*. Melbourne, Australia: Creative Commons. `http://www.CleverAlgorithms.com`. 251

Bruzzese, D. and C. Davino 2008. Visual mining of association rules. In S. Simoff, M. H. Böhlen, and A. Mazeika (eds.), *Visual Data Mining: Theory, Techniques and Tools for Visual Analytics*, pp. 103–122. New York: Springer.

Buchanan, M. 2002. *Nexus: Small Worlds and the Groundbreaking Science of Networks*. New York: W.W. Norton and Company.

Bühlmann, P. and S. van de Geer 2011. *Statistics for High-Dimensional Data: Methods, Theory and Applications*. New York: Springer. 249

Burt, J. M. and M. B. Garman 1971. Conditional Monte Carlo: A simulation technique for stochastic network analysis. *Management Science* 18(3):207–217. 225

Butts, C. T. 2008a, May 5. network: A package for managing relational data in R. *Journal of Statistical Software* 24(2):1–36. `http://www.jstatsoft.org/v24/i02`. Updated version available at `http://cran.r-project.org/web/packages/network/vignettes/networkVignette.pdf`.

Butts, C. T. 2008b, May 8. Social network analysis with sna. *Journal of Statistical Software* 24(6):1–51. `http://www.jstatsoft.org/v24/i06`.

Butts, C. T. 2014a. *Package network: Classes for Relational Data*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/network/network.pdf`.

Butts, C. T. 2014b. *Package snad: Statistical Analysis of Network Data with R*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/sand/sand.pdf`.

Butts, C. T. 2014c. *Package sna: Tools for Social Network Analysis*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/sna/sna.pdf`. Also see the statnet website at `http://www.statnetproject.org/`.

Buvač, V. and P. J. Stone 2001, April 2. The General Inquirer user's guide. Software developed with the support of Harvard University and The Gallup Organization.

Cairo, A. 2013. *The Functional Art: An Introduction to Information Graphics and Visualization*. Berkeley, Calif: New Riders.

Cami, A. and N. Deo 2008. Techniques for analyzing dynamic random graph models of web-like networks: An overview. *Networks* 51(4):211–255.

Campbell, S. and S. Swigart 2014. *Going beyond Google: Gathering Internet Intelligence* (5 ed.). Oregon City, Oreg.: Cascade Insights. `http://www.cascadeinsights.com/gbgdownload`. 59

Cantelon, M., M. Harter, T. J. Holowaychuk, and N. Rajlich 2014. *Node.js in Action*. Shelter Island, N.Y.: Manning. 3

Carley, K. M. and D. Skillicorn 2005, October. Special issue on analyzing large scale networks: The Enron corpus. *Computational and Mathematical Organization Theory* 11(3): 179–181. 291

Carlin, B. P. 1996, February. Improved NCAA basketball tournament modeling via point spread and team strength information. *The American Statistician* 50(1):39–43. 103

Carlin, B. P. and T. A. Louis 1996. *Bayes and Empirical Bayes Methods for Data Analysis*. London: Chapman & Hall. 244

Carr, D., N. Lewin-Koh, and M. Maechler 2014. *hexbin: Hexagonal Binning Routines*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/hexbin/hexbin.pdf`. 252

Carrington, P. J., J. Scott, and S. Wasserman (eds.) 2005. *Models and Methods in Social Network Analysis*. Cambridge, UK: Cambridge University Press.

Casciano, C. 2011. *The CSS Pocket Guide*. Upper Saddle River, N.J.: Pearson/Peachpit.

Cass, S. 2014, July. The top 10 programming languages: Spectrum's 2014 ranking. *IEEE Spectrum* 51(7):68. 298

Chakrabarti, S. 2003. *Mining the Web: Discovering Knowledge from Hypertext Data*. San Francisco: Morgan Kaufmann.

Chambers, J. M. and T. J. Hastie (eds.) 1992. *Statistical Models in S*. Pacific Grove, Calif.: Wadsworth & Brooks/Cole. Champions of S, S-Plus, and R call this "the white book." It introduced statistical modeling syntax using S3 classes.

Chan, E. P. 2009. *Quantitative Trading: How to Build Your Own Algorithmic Trading Business*. New York: Wiley.

Chan, E. P. 2013. *Algorithmic Trading: Winning Strategies and Their Rationale*. New York: Wiley.

Chang, W. 2013. *R Graphics Cookbook*. Sebastopol, Calif.: O'Reilly.

Chapanond, A., M. S. Krishnamoorthy, and B. Yener 2005, October. Graph theoretic and spectral analysis of Enron email data. *Computational and Mathematical Organization Theory* 11(3):265–281. 81

Charmaz, K. 2002. Qualitative interviewing and grounded theory analysis. In J. F. Gubrium and J. A. Holstein (eds.), *Handbook of Interview Research: Context and Method*, Chapter 32, pp. 675–694. Thousand Oaks, Calif.: Sage. 279

Charniak, E. 1993. *Statistical Language Learning*. Cambridge: MIT Press.

Chatterjee, S. and A. S. Hadi 2012. *Regression Analysis by Example* (fifth ed.). New York: Wiley.

Chau, M. and H. Chen 2003. Personalized and focused web spiders. In N. Zhong, J. Liu, and Y. Yao (eds.), *Web Intelligence*, Chapter 10, pp. 198–217. New York: Springer. 59

Cherny, L. 1999. *Conversation and Community: Chat in a Virtual World*. Stanford, Calif.: CSLI Publications. 265

Chodorow, K. 2013. *MongoDB: The Definitive Guide* (second ed.). Sebastopol, Calif.: O'Reilly. 3

Christakis, N. A. and J. H. Fowler 2009. *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*. New York: Little Brown and Company.

Christensen, R. 1997. *Log-Linear Models and Logistic-Regression* (second ed.). New York: Springer.

Chun, W. J. 2007. *Core Python Programming* (second ed.). Upper Saddle River, N.J.: Pearson Education.

Cleveland, W. S. 1993. *Visualizing Data*. Murray Hill, N.J.: AT&T Bell Laboratories. Initial documentation for trellis graphics in S-Plus. 252

Clifton, B. 2012. *Advanced Web Metrics with Google Analytics* (third ed.). New York: Wiley. 19

Cohen, J. 1960, April. A coefficient of agreement for nominal data. *Educational and Psychological Measurement* 20(1):37–46. 248

Conway, D. and J. M. White 2012. *Machine Learning for Hackers*. Sebastopol, Calif.: O'Reilly.

Cook, R. D. 1998. *Regression Graphics: Ideas for Studying Regressions through Graphics*. New York: Wiley.

Cook, R. D. 2007. Fisher lecture: Dimension reduction in regression. *Statistical Science* 22: 1–26.

Cook, R. D. and S. Weisberg 1999. *Applied Regression Including Computing and Graphics*. New York: Wiley.

Copeland, R. 2013. *MongoDB Applied Design Patterns* (second ed.). Sebastopol, Calif.: O'Reilly. 3

Cox, T. F. and M. A. A. Cox 1994. *Multidimensional Scaling*. London: Chapman & Hall.

Cranor, L. F. and B. A. LaMacchia 1998. Spam! *Communications of the ACM* 41(8):74–83.

Cristianini, N. and J. Shawe-Taylor 2000. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge: Cambridge University Press.

Crnovrsanin, T., C. W. Mueldera, R. Faris, D. Felmlee, and K.-L. Ma 2014. Visualization techniques for categorical analysis of social networks with multiple edge sets. *Social Networks* 37:56–64.

Crockford, D. 2008. *JavaScript: The Good Parts*. Sebastopol, Calif.: O'Reilly.

Croll, A. and S. Power 2009. *Complete Web Monitoring: Watching Your Visitors, Performance, Communities & Competitors*. Sebastopol, Calif.: O'Reilly.

Csardi, G. 2014a. *Package igraphdata: A Collection of Network Data Sets for the igraph Package*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/igraphdata/igraphdata.pdf`.

Csardi, G. 2014b. *Package igraph: Network Analysis and Visualization*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/igraph/igraph.pdf`.

Daconta, M. C., L. J. Obrst, and K. T. Smith 2003. *The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management*. New York: Wiley.

Dale, J. and S. Abbott 2014. *Qual-Online The Essential Guide: What Every Researcher Needs to Know about Conducting and Moderating Interviews via the Web*. Ithaca, N.Y.: Paramount Market Publishing.

Datta, A., S. Shulman, B. Zheng, S.-D. Lin, A. Sum, and E.-P. Lim (eds.) 2011. *Social Informatics: Proceedings of the Third International Conference SocInfo 2011*. New York: Springer.

Davenport, T. H. and J. G. Harris 2007. *Competing on Analytics: The New Science of Winning*. Boston: Harvard Business School Press. 239

Davenport, T. H., J. G. Harris, and R. Morison 2010. *Analytics at Work: Smarter Decisions, Better Results*. Boston: Harvard Business School Press. 239

David, H. A. 1963. *The Method of Paired Comparisons*. London: Charles Griffin & Company Limited.

Davidson, R. R. and P. H. Farquhar 1976, June. A bibliography on the method of paired comparisons. *Biometrics* 32(2):241–252.

Davis, B. H. and J. P. Brewer 1997. *Electronic Discourse: Linguistic Individuals in Virtual Space*. Albany, N.Y.: State University of New York Press. 265

Davison, M. L. 1992. *Multidimensional Scaling*. Melbourne, Fla.: Krieger.

de Nooy, W., A. Mrvar, and V. Batagelj 2011. *Exploratory Social Network Analysis with Pajek*. Cambridge, U.K.: Cambridge University Press.

Dean, J. and S. Ghemawat 2004. MapReduce: Simplifed Data Processing on Large Clusters. Retrieved from the World Wide Web at `http://static.usenix.org/event/osdi04/tech/full_papers/dean/dean.pdf`.

Dehuri, S., M. Patra, B. B. Misra, and A. K. Jagadev (eds.) 2012. *Intelligent Techniques in Recommendation Systems*. Hershey, Pa.: IGI Global.

Delen, D., R. Sharda, and P. Kumar 2007. Movie forecast guru: A Web-based DSS for Hollywood managers. *Decision Support Systems* 43(4):1151–1170. 133

Demšar, J. and B. Zupan 2013. Orange: Data mining fruitful and fun—A historical perspective. *Informatica* 37:55–60. 250, 299

Dey, I. 1999. *Grounding Grounded Theory: Guidelines for Qualitative Inquiry*. New York: Wiley. 279

Di Battista, G., P. Eades, R. Tomassia, and I. G. Tollis 1999. *Graph Drawing: Algorithms for the Visualization of Graphs*. Upper Saddle River, N.J.: Prentice Hall.

Di Battista, G. and M. Rimondini 2014. Computer networks. In R. Tamassia (ed.), *Handbook of Graph Drawing and Visualization*, Chapter 25, pp. 763–804. Boca Raton, Fla.: CRC Press/Chapman & Hall.

Diesner, J., T. L. Frantz, and K. M. Carley 2005, October. Communication networks from the Enron email corpus. *Computational and Mathematical Organization Theory* 11(3): 201–228. 81

Dippold, K. and H. Hruschka 2013. Variable selection for market basket analysis. *Computational Statistics* 28(2):519–539. `http://dx.doi.org/10.1007/s00180-012-0315-3`. ISSN 0943-4062.   210

Dover, D. 2011. *Search Engine Optimization Secrets: Do What You Never Thought Possible with SEO*. New York: Wiley.

Downing, D., M. Covington, M. Covington, C. A. Barrett, and S. Covington 2012. *Dictionary of Computer and Internet Terms* (eleventh ed.). Hauppauge, N.Y.: Barron's Educational Series.

Draper, N. R. and H. Smith 1998. *Applied Regression Analysis* (third ed.). New York: Wiley.

DuCharme, B. 2013. *Learning SPARQL*. Sebastopol, Calif.: O'Reilly.   235

Duda, R. O., P. E. Hart, and D. G. Stork 2001. *Pattern Classification* (second ed.). New York: Wiley.

Dumais, S. T. 2004. Latent semantic analysis. In B. Cronin (ed.), *Annual Review of Information Science and Technology*, Volume 38, Chapter 4, pp. 189–230. Medford, N.J.: Information Today.

Easley, D. and J. Kleinberg 2010. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge, UK: Cambridge University Press.

Efron, B. 1986. Why isn't everyone a Bayesian (with commentary). *The American Statistician* 40(1):1–11.

Eichenwald, K. 2005. *Conspiracy of Fools: A True Story*. New York: Broadway Books/Random House.

Ellis, D., R. Oldridge, and A. Vasconcelos 2004. Community and virtual community. In B. Cronin (ed.), *Annual Review of Information Science and Technology*, Volume 38, Chapter 3, pp. 145–186. Medford, N.J.: Information Today.

Engelbrecht, A. P. 2007. *Computational Intelligence: An Introduction* (second ed.). New York: Wiley.   251

Erdős, P. and A. Rényi 1959. On random graphs. *Publicationes Mathematicae* 6:290–297.

Erdős, P. and A. Rényi 1960. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 5:17–61.

Erdős, P. and A. Rényi 1961. On the strength of connectedness of a random graph. *Acta Mathematica Scientia Hungary* 12:261–267.

Erwig, M. 2000. The graph Voronoi diagram with applications. *Networks* 36(3):156–163.

Esipova, N., T. W. Miller, M. D. Zarnecki, J. Elzaurdia, and S. Ponnaiya 2004. Exploring the possibilities of online focus groups. In T. W. Miller and J. Walkowski (eds.), *Qualitative Research Online*, Chapter A, pp. 107–128. Manhattan Beach, Calif.: Research Publishers.

European Parliament 2002. Directive on privacy and electronic communications. Retrieved from the World Wide Web, October 21, 2014, at `http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32002L0058:en:HTML`.   19

Everitt, B. and G. Dunn 2001. *Applied Multivariate Data Analysis* (second ed.). New York: Wiley.   176

Everitt, B. and S. Rabe-Hesketh 1997. *The Analysis of Proximity Data*. London: Arnold.

Everitt, B. S., S. Landau, M. Leese, and D. Stahl 2011. *Cluster Analysis* (fifth ed.). New York: Wiley.

Fawcett, T. 2003, January 7. ROC graphs: Notes and practical considerations for researchers. http://www.hpl.hp.com/techreports/2003/HPL-2003-4.pdf.

Feinerer, I. 2012. *Introduction to the wordnet Package*. Comprehensive R Archive Network. 2012. `http://cran.r-project.org/web/packages/wordnet/vignettes/wordnet.pdf`. 259

Feinerer, I. 2014. *Introduction to the tm Package*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/tm/vignettes/tm.pdf`.

Feinerer, I. and K. Hornik 2014a. *tm: Text Mining Package*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/tm/tm.pdf`.

Feinerer, I. and K. Hornik 2014b. *wordnet: WordNet Interface*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/wordnet/wordnet.pdf`. 259

Feinerer, I., K. Hornik, and D. Meyer 2008, 3 31. Text mining infrastructure in R. *Journal of Statistical Software* 25(5):1–54. `http://www.jstatsoft.org/v25/i05`. ISSN 1548-7660.

Feldman, R. 2002. Text mining. In W. Klösgen and J. M. Żytkow (eds.), *Handbook of Data Mining and Knowledge Discovery*, Chapter 38, pp. 749–757. Oxford: Oxford University Press.

Fellbaum, C. 1998. *WordNet: An Electronic Lexical Database*. Cambridge, Mass.: MIT Press. 259

Fellows, I. 2014a. wordcloud makes words less cloudy. Retrieved from the World Wide Web at `http://blog.fellstat.com/`. 176, 259

Fellows, I. 2014b. *wordcloud: Word Clouds*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/wordcloud/wordcloud.pdf`. 176, 259

Fenzel, D., J. Hendler, H. Lieberman, and W. Wahlster (eds.) 2003. *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*. Cambridge: MIT Press.

Ferrucci, D., E. Brown, J. Chu-Carroll, J. Fan, D. Gondek, A. A. Kalyanpur, A. Lally, J. W. Murdock, E. Nyberg, J. Prager, N. Schlaefer, and C. Welty 2010, Fall. Building Watson: An overview of the DeepQA project. *AI Magazine* 31(3):59–79.

Few, S. 2009. *Now You See It: Simple Visualization Techniques and Quantitative Analysis*. Oakland, Calif.: Analytics Press.

Fielding, N. G. and R. M. Lee 1998. *Computer Analysis and Qualitative Research*. London: Sage. 279

Fienberg, S. E. 2007. *Analysis of Cross-Classified Categorical Data* (second ed.). New York: Springer.

Firth, D. 1991. Generalized linear models. In D. Hinkley and E. Snell (eds.), *Statistical Theory and Modeling: In Honour of Sir David Cox, FRS*, Chapter 3, pp. 55–82. London: Chapman and Hall.

Fisher, R. A. 1970. *Statistical Methods for Research Workers* (fourteenth ed.). Edinburgh: Oliver and Boyd. First edition published in 1925. 244

Fisher, R. A. 1971. *Design of Experiments* (ninth ed.). New York: Macmillan. First edition published in 1935. 244

Flam, F. 2014, September 30. The odds, continually updated. *The New York Times*:D1. Retrieved from the World Wide Web at `http://www.nytimes.com/2014/09/30/science/the-odds-continually-updated.html?_r=0`. 244

Flanagan, D. 2011. *JavaScript: The Definitive Guide* (sixth ed.). Sebastopol, Calif.: O'Reilly. 3

Fox, J. 2002, January. Robust regression: Appendix to an R and S-PLUS companion to applied regression. Retrieved from the World Wide Web at `http://cran.r-project.org/doc/contrib/Fox-Companion/appendix-robust-regression.pdf`. 249

Fox, J. 2014. *car: Companion to Applied Regression*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/car/car.pdf`.

Fox, J. and S. Weisberg 2011. *An R Companion to Applied Regression* (second ed.). Thousand Oaks, Calif.: Sage. 128

Franceschet, M. 2011, June. PageRank: Standing on the shoulders of giants. *Communications of the ACM* 54(6):92–101.

Franks, B. 2012. *Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics*. Hoboken, N.J.: Wiley. 239

Freeman, L. C. 2004. *The Development of Social Network Analysis: A Study in the Sociology of Science*. Vancouver, B.C.: Empirical Press.

Freeman, L. C. 2005. Graphic techniques for exploring social network data. In P. J. Carrington, J. Scott, and S. Wasserman (eds.), *Models and Methods in Social Network Analysis*, Chapter 12, pp. 248–269. Cambridge, UK: Cambridge University Press.

Friedl, J. E. F. 2006. *Mastering Regular Expressions* (third ed.). Sebastopol, Calif.: O'Reilly. 259

Fruchterman, T. M. J. and E. M. Reingold 1991. Graph drawing by force-directed placement. *Software—Practice and Experience* 21(11):1129–1164.

Fry, B. 2008. *Visualizing Data: Exploring and Explaining Data with the Processing Environment*. Sebastopol, Calif.: O'Reilly.

Gabriel, K. R. 1971. The biplot graphical display of matrices with application to principal component analysis. *Biometrika* 58:453–467.

Garcia-Molina, H., J. D. Ullman, and J. Widom 2009. *Database Systems: The Complete Book* (second ed.). Upper Saddle River, N.J.: Prentice-Hall.

Gasston, P. 2011. *The Book of CSS3: A Developer's Guide to the Future of Web Design*. San Francisco: No Starch Press.

Gasston, P. 2013. *The Modern Web: Multi-Device Web Development with HTML5, CSS3, and JavaScript*. San Francisco: No Starch Press.

Geisser, S. 1993. *Predictive Inference: An Introduction*. New York: Chapman & Hall. 244

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin 1995. *Bayesian Data Analysis*. London: Chapman & Hall. 244

Gelman, A., J. Hill, Y.-S. Su, M. Yajima, and M. G. Pittau 2014. *mi: Missing Data Imputation and Model Checking*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/mi/mi.pdf`.

Georgakopoulou, A. and D. Goutsos 1997. *Discourse Analysis: An Introduction*. Edinburgh, U.K.: Edinburgh University Press. 279

Gibson, H., J. Faith, and P. Vickers 2013, 07. A survey of two-dimensional graph layout techniques for information visualisation. *Information Visualization* 12(3-4):324–357.

Glaser, B. G. and A. L. Strauss 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Hawthorne, N.Y.: Aldine de Gruyter. 279

Gnanadesikan, R. 1997. *Methods for Statistical Data Analysis of Multivariate Observations* (second ed.). New York: Wiley. 176

Goldenberg, A., A. X. Zheng, S. E. Fienberg, and E. M. Airoldi 2009, January. A survey of statistical network models. *Foundations and Trends in Machine Learning* 2(2):129–233.

Golombisky, K. and R. Hagen 2013. *White Space is Not Your Enemy: A Beginner's Guide to Communicating Visually through Graphic, Web, & Multimedia Design* (second ed.). Burlington, Mass.: Focal Press/Taylor & Francis.

Goodreau, S. M., M. S. Handcock, D. R. Hunter, C. T. Butts, and M. Morris 2008, 5 8. A statnet tutorial. *Journal of Statistical Software* 24(9):1–26. `http://www.jstatsoft.org/v24/i09`. CODEN JSSOBK. ISSN 1548-7660.

Gower, J. C. and D. J. Hand 1996. *Biplots*. London: Chapman & Hall.

Graybill, F. A. 1961. *Introduction to Linear Statistical Models, Volume 1*. New York: McGraw-Hill.

Graybill, F. A. 2000. *Theory and Application of the Linear Model*. Stamford, Conn.: Cengage Learning.

Gries, S. T. 2009. *Quantitative Corpus Linguistics with R: A Practical Introduction*. New York: Routledge.

Gries, S. T. 2013. *Statistics for Linguistics with R: A Pratical Introduction* (second revised ed.). Berlin: De Gruyter Mouton. 259

Grieve, T. 2003, October 14. The decline and fall of the Enron empire. *Salon*:165–178. Retrieved from the World Wide Web at `http://www.salon.com/2003/10/14/enron_22/`.

Groeneveld, R. A. 1990, November. Ranking teams in a league with two divisions of *t* teams. *The American Statistician* 44(4):277–281. 103

Grolemund, G. and H. Wickham 2014. *lubridate: Make Dealing with Dates a Little Easier*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/lubridate/lubridate.pdf`. 62

Gross, J. L., J. Yellen, and P. Zhang (eds.) 2014. *Handbook of Graph Theory* (second ed.). Boca Raton, Fla.: CRC Press/Chapman & Hall.

Grothendieck, G. 2014a. *gsubfn: Utilities for Strings and Function Arguments*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/gsubfn/gsubfn.pdf`.

Grothendieck, G. 2014b. *gsubfn: Utilities for Strings and Function Arguments (Vignette)*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/gsubfn/vignettes/gsubfn.pdf`.

Groves, R. M., F. J. Fowler, Jr., M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau 2009. *Survey Methodology* (second ed.). New York: Wiley.

Guilford, J. P. 1936. *Psychometric Methods*. New York: McGraw-Hill.

Gunning, D., V. K. Chaudhri, P. Clark, K. Barker, S.-Y. Chaw, M. Greaves, B. Grosof, A. Leung, D. McDonald, S. Mishra, J. Pacheco, B. Porter, A. Spaulding, D. Tecuci, and J. Tien 2010, Fall. Project Halo update—Progress toward digital Aristotle. *AI Magazine* 31(3):33–58.

Hagbert, A. and D. Schult 2014. *NetworkX: Python Software for Complex Networks*. NetworkX Development Team. 2014. Retrieved from the World Wide Web at `https://github.com/networkx/`.

Hahsler, M. 2014a. *recommenderlab: A Framework for Developing and Testing Recommendation Algorithms*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/recommenderlab/vignettes/recommenderlab.pdf`.

Hahsler, M. 2014b. *recommenderlab: Lab for Developing and Testing Recommender Algorithms*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/recommenderlab/recommenderlab.pdf`.

Hahsler, M., C. Buchta, B. Grün, and K. Hornik 2014a. *arules: Mining Association Rules and Frequent Itemsets*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/arules/arules.pdf`.

Hahsler, M., C. Buchta, B. Grün, and K. Hornik 2014b. *Introduction to arules: A Computational Environment for Mining Association Rules and Frequent Itemsets*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/arules/vignettes/arules.pdf`.

Hahsler, M., C. Buchta, and K. Hornik 2008. Selective association rule generation. *Computational Statistics* 23:303–315. 210

Hahsler, M. and S. Chelluboina 2014a. *arulesViz: Visualizing Association Rules and Frequent Itemsets*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/arulesViz/arulesViz.pdf`.

Hahsler, M. and S. Chelluboina 2014b. *Visualizing Association Rules: Introduction to the R-extension Package arulesViz*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/arulesViz/vignettes/arulesViz.pdf`.

Hahsler, M., S. Chelluboina, K. Hornik, and C. Buchta 2011. The arules R-package ecosystem: Analyzing interesting patterns from large transaction data sets. *Journal of Machine Learning Research* 12:2021–2025.

Hahsler, M., B. Grün, and K. Hornik 2005, September 29. arules: A computational environment for mining association rules and frequent item sets. *Journal of Statistical Software* 14(15):1–25. `http://www.jstatsoft.org/v14/i15`.

Hale, C. and J. Scanlon 1999. *Wired Style: Principles of English Usage in the Digital Age*. New York: Broadway Books.

Hand, D. J. 1997. *Construction and Assessment of Classification Rules*. New York: Wiley.

Handcock, M. S. and K. Gile 2007, April 28. Modeling social networks with sampled or missing data. `http://www.csss.washington.edu/Papers/wp75.pdf`. 109

Handcock, M. S., D. R. Hunter, C. T. Butts, S. M. Goodreau, and M. Morris 2008, May 5. statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software* 24(1):1–11. `http://www.jstatsoft.org/v24/i01`.

Hanneman, R. A. and M. Riddle 2005. *Introduction to Social Network Methods*. Riverside, Calif.: University of California Riverside. `http://www.faculty.ucr.edu/~hanneman/nettext/`.

Harrell, Jr., F. E. 2001. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer.

Harrington, P. 2012. *Machine Learning in Action*. Shelter Island, N.Y.: Manning. 210

Hart, R. P. 2000a. *Campaign Talk: Why Elections Are Good for Us*. Princeton, N.J.: Princeton University Press.

Hart, R. P. 2000b. *DICTION 5.0: The Text Analysis Program*. Thousand Oaks, Calif.: Sage.

Hart, R. P. 2001. Redeveloping Diction: theoretical considerations. In M. D. West (ed.), *Theory, Method, and Practice in Computer Content Analysis*, Chapter 3, pp. 43–60. Westport, Conn.: Ablex.

Hastie, T., R. Tibshirani, and J. Friedman 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (second ed.). New York: Springer.

Hausser, R. 2001. *Foundations of Computational Linguistics: Human-Computer Communication in Natural Language* (second ed.). New York: Springer-Verlag.

Hearst, M. A. 1997. TextTiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics* 23(1):33–64. 259

Heer, J., M. Bostock, and V. Ogievetsky 2010, May 1. A tour through the visualization zoo: A survey of powerful visualization techniques, from the obvious to the obscure. *acmqueue: Association for Computing Machinery*:1–22. Retrieved from the World Wide Web at `http://queue.acm.org/detail.cfm?id=1805128`.

Hemenway, K. and T. Calishain 2004. *Spidering Hacks: 100 Industrial-Strength Tips & Tools*. Sabastopol, Calif.: O'Reilly. 59

Herman, A. and T. Swiss (eds.) 2000. *The World Wide Web and Contemporary Cultural Theory*. New York: Routledge. 265

Hinkley, D. V., N. Reid, and E. J. Snell (eds.) 1991. *Statistical Theory and Modeling*. London: Chapman and Hall. 244

Ho, Q. and E. P. Xing 2014. Analyzing time-evolving networks using an evolving cluster mixed membership blockmodel. In E. M. Airoldi, D. Blei, E. A. Erosheva, and S. E. Fienberg (eds.), *Handbook of Mixed Membership Models and Their Applications*, Chapter 23, pp. 489–525. Boca Raton, Fla.: CRC Press.

Hoberman, S. 2014. *Data Modeling for MongoDB: Building Well-Designed and Supportable MongoDB Databases*. Basking Ridge, N.J.: Technics Publications. 3

Hoerl, A. E. and R. W. Kennard 2000. Ridge regression: biased estimation for non-orthogonal problems. *Technometrics* 42(1):80–86. Reprinted from *Technometrics*, volume 12. 249

Hoffman, P. and Scrapy Developers 2014, June 26. Scrapy documentation release 0.24.0. Supported at `http://scrapy.org/`. Documentation retrieved from the World Wide Web, October 26, 2014, at `https://media.readthedocs.org/pdf/scrapy/0.24/scrapy.pdf`.

Holden, D. 2006. Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. *IEEE Transactions on Visualization and Computer Graphics* 12(5):741–748.

Holeton, R. (ed.) 1998. *Composing Cyberspace: Identity, Community, and Knowledge in the Electronic Age*. Boston: McGraw-Hill. 265

Holland, P. W., K. B. Laskey, and S. Leinhardt 1983. Stochastic blockmodels: First steps. *Social Networks* 5:109–138.

Honaker, J., G. King, and M. Blackwell 2014. *Amelia II: A Program for Missing Data*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/Amelia/Amelia.pdf`.

Hornik, K. 2014a. *RWeka Odds and Ends*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/RWeka/vignettes/RWeka.pdf`.

Hornik, K. 2014b. *RWeka: R/Weka Interface*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/RWeka/RWeka.pdf`.

Hosmer, D. W., S. Lemeshow, and S. May 2013. *Applied Survival Analysis: Regression Modeling of Time to Event Data* (second ed.). New York: Wiley.

Hosmer, D. W., S. Lemeshow, and R. X. Sturdivant 2013. *Applied Logistic Regression* (third ed.). New York: Wiley. 128

Houston, B., L. Bruzzese, and S. Weinberg 2002. *The Investigative Reporter's Handbook: A Guide to Documents, Databases and Techniques* (fourth ed.). Boston: Bedford/St. Martin's.

Hu, M. and B. Liu 2004, August 22–25. Mining and summarizing customer reviews. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD-2004)*. Full paper available from the World Wide Web at `http://www.cs.uic.edu/~liub/publications/kdd04-revSummary.pdf` Original source for opinion and sentiment lexicon, available from the World Wide Web at `http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon`.

Huberman, B. A. 2001. *The Laws of the Web: Patterns in the Ecology of Information*. Cambridge: MIT Press. 265

Huet, S., A. Bouvier, M.-A. Poursat, and E. Jolivet 2004. *Statistical Tools for Nonlinear Regression: A Practical Guide with S-Plus and R Examples* (second ed.). New York: Springer.

Hughes-Croucher, T. and M. Wilson 2012. *Node Up and Running*. Sebastopol, Calif.: O'Reilly. 3

Huisman, M. 2010. Imputation of missing network data: Some simple procedures. *Journal of Social Structure* 10. `http://www.cmu.edu/joss/content/articles/volume10/huisman.pdf`. 109

Hyslop, B. 2010. *The HTML Pocket Guide*. Upper Saddle River, N.J.: Pearson/Peachpit.

Ihaka, R., P. Murrell, K. Hornik, J. C. Fisher, and A. Zeileis 2014. *colorspace: Color Space Manipulation*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/colorspace/colorspace.pdf`.

Indurkhya, N. and F. J. Damerau (eds.) 2010. *Handbook of Natural Language Processing* (second ed.). Boca Raton, Fla.: Chapman and Hall/CRC.

Ingersoll, G. S., T. S. Morton, and A. L. Farris 2013. *Taming Text: How to Find, Organize, and Manipulate It*. Shelter Island, N.Y.: Manning. 176, 251

Izenman, A. J. 2008. *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*. New York: Springer. 176

Jackson, M. O. 2008. *Social and Economic Networks*. Princeton, N.J.: Princeton University Press.

James, G., D. Witten, T. Hastie, and R. Tibshirani 2013. *An Introduction to Statistical Learning with Applications in R*. New York: Springer.

Janssen, J., M. Hurshman, and N. Kalyaniwalla 2012. Model selection for social networks using graphlets. *Internet Mathematics* 8(4):338–363.

Johnson, K. 2008. *Quantitative Methods in Linguistics*. Malden, Mass.: Blackwell Publishing. 259

Johnson, R. A. and D. W. Wichern 1998. *Applied Multivariate Statistical Analysis* (fourth ed.). Upper Saddle River, N.J.: Prentice Hall. 176

Johnson, S. 1997. *Interface Culture*. New York: Basic Books. 265

Jones, S. G. (ed.) 1997. *Virtual Culture: Identity & Communication in Cybersociety*. Thousand Oaks, Calif.: Sage. 265

Jones, S. G. (ed.) 1999. *Doing Internet Research: Critical Issues and Methods for Examining the Net*. Thousand Oaks, Calif.: Sage. 265

Jonscher, C. 1999. *The Evolution of Wired Life: From the Alphabet to the Soul-Catcher Chip—How Information Technologies Change Our World*. New York: Wiley. 265

Joula, P. 2008. *Authorship Attribution*. Hanover, Mass.: Now Publishers. 259

Jung, C. G. 1968. *The Archetypes of the Collective Unconscious (Collected Works of C. G. Jung, Vol. 9, Part 1)* (second ed.). Princeton, N.J.: Princeton University Press.

Jünger, M. and P. Mutzel (eds.) 2004. *Graph Drawing Software*. New York: Springer. 82

Jurafsky, D. and J. H. Martin 2009. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (second ed.). Upper Saddle River, N.J.: Prentice Hall.

Kadushin, C. 2012. *Understanding Social Networks*. New York: Oxford University Press.

Kamada, T. and S. Kawai 1989. An algorithm for drawing general undirected graphs. *Information Processing Letters* 31(1):7–15.

Kaufman, L. and P. J. Rousseeuw 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: Wiley.

Kaushik, A. 2010. *Web Analytics 2.0: The Art of Online Accountability & Science of Customer Centricity*. New York: Wiley/Sybex.

Kay, M. 2008. *XSLT 2.0 and XPath 2.0 Programmer's Reference* (fourth ed.). New York: Wiley/Wrox. 29

Keener, J. P. 1993, March. The Perron-Frobenius theorem and the ranking of football teams. *SIAM Review* 35(1):80–93.

Keila, P. and D. B. Skillicorn 2005, October. Structure of the Enron email dataset. *Computational and Mathematical Organization Theory* 11(3):183–199. 81

Kelle, U. (ed.) 1995. *Computer-Aided Qualitative Data Analysis: Theory, Methods and Practice*. Thousand Oaks, Calif.: Sage. 279

Kendall, M. G. and B. B. Smith 1940, March. On the method of paired comparisons. *Biometrika* 31(3/4):324–345.

Kirk, R. E. 2013. *Experimental Design: Procedures for the Behavioral Sciences* (fourth ed.). Thousand Oaks, Calif.: Sage.

Klimmt, B. and Y. Yang 2004. Introducing the Enron corpus. `http://ceas.cc/2004/168.pdf`.

Kolaczyk, E. D. 2009. *Statistical Analysis of Network Data: Methods and Models*. New York: Springer. 224

Kolaczyk, E. D. and G. Csárdi 2014. *Statistical Analysis of Network Data with R*. New York: Springer.

Koller, M. 2014. *Simulations for Sharpening Wald-type Inference in Robust Regression for Small Samples*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/robustbase/vignettes/lmrob_simulation.pdf`. 249

Koller, M. and W. A. Stahel 2011. Sharpening Wald-type inference in robust regression for small samples. *Computational Statistics and Data Analysis* 55(8):2504–2515. 249

Kossinets, G. 2006. Effects of missing data in social networks. *Social Networks* 28(3):247–268. 109

Kratochvíl, J. (ed.) 1999. *Graph Drawing: 7th International Symposium / Proceedings GD'99*. New York: Springer.

Krippendorff, K. H. 2012. *Content Analysis: An Introduction to Its Methodology* (third ed.). Thousand Oaks, Calif.: Sage. 133

Krug, S. 2014. *Don't Make Me Think: A Common Sense Approach to Web Usability* (third ed.). Upper Saddle River, N.J.: Pearson/New Riders.

Kuhn, M. 2014. *caret: Classification and Regression Training*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/caret/caret.pdf`.

Kuhn, M. and K. Johnson 2013. *Applied Predictive Modeling*. New York: Springer.

Kutner, M. H., C. J. Nachtsheim, J. Neter, and W. Li 2004. *Applied Linear Statistical Models* (fifth ed.). Boston: McGraw-Hill.

Kvale, S. 1996. *InterViews: An Introduction to Qualitative Research Interviewing*. Thousand Oaks, Calif.: Sage. 275

Lacy, L. W. 2005. *Owl: Representing Information Using the Web Ontology Language*. Victoria, B.C.: Trafford. 235

Lamp, G. 2014. *ggplot for Python*. GitHub. 2014. `https://github.com/yhat/ggplot`. 252

Landauer, T. K., P. W. Foltz, and D. Laham 1998. An introduction to latent semantic analysis. *Discourse Processes* 25:259–284. Retrieved from the World Wide Web, October 31, 2014, at `http://lsa.colorado.edu/papers/dp1.LSAintro.pdf`. 176

Landauer, T. K., D. S. McNamara, S. Dennis, and W. Kintsch (eds.) 2014. *Handbook of Latent Semantic Analysis* (reprint ed.). New York: Psychology Press. 176

Lander, J. P. 2014. *R for Everyone: Advanced Analytics and Graphics*. Upper Saddle River, N.J.: Pearson Education.

Lang, D. T. 2014a. *Package RCurl: General Network (HTTP/FTP/...) Client Interface for R*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/RCurl/RCurl.pdf`. 62

Lang, D. T. 2014b. *Package XML: Tools for Parsing and Generating XML within R and S-Plus*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/XML/XML.pdf`. 62

Langville, A. N. and C. D. Meyer 2006. *Google's Page Rank and Beyond: The Science of Search Engine Rankings*. Princeton, N.J.: Princeton University Press.

Langville, A. N. and C. D. Meyer 2012. *Who's 1?: The Science of Rating and Ranking*. Princeton, N.J.: Princeton University Press.

Lantz, B. 2013. *Machine Learning with R*. Birmingham, U.K.: Packt Publishing.

Laursen, G. H. N. and J. Thorlund 2010. *Business Analytics for Managers: Taking Business Intelligence Beyond Reporting*. Hoboken, N.J.: Wiley. 239

Law, A. M. 2014. *Simulation Modeling and Analysis* (fifth ed.). New York: McGraw-Hill.

Le, C. T. 1997. *Applied Survival Analysis*. New York: Wiley.

Leber, J. 2013. The immortal life of the Enron e-mails. *MIT Technology Review*. `http://www.technologyreview.com/news/515801/the-immortal-life-of-the-enron-e-mails/`. 291

Ledoiter, J. 2013. *Data Mining and Business Analytics with R*. New York: Wiley.

Leetaru, K. 2011. *Data Mining Methods for Content Analysis: An Introduction to the Computational Analysis of Content*. New York: Routledge. 133

Lefkowitz, M. 2014. Twisted documentation. Contributors listed as Twisted Matrix Laboratories at `http://twistedmatrix.com/trac/wiki/TwistedMatrixLaboratories`. Documentation available at `http://twistedmatrix.com/trac/wiki/Documentation`.

Lehmann, J., R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. van Kleef, S. Auer, and C. Bizer 2014. DBpedia—A large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web Journal*:in press. Retrieved from the World Wide Web at `http://svn.aksw.org/papers/2013/SWJ_DBpedia/public.pdf`.

Leisch, F. and B. Gruen 2014. *CRAN Task View: Cluster Analysis & Finite Mixture Models*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/views/Cluster.html`.

Lenat, D., M. Witbrock, D. Baxter, E. Blackstone, C. Deaton, D. Schneider, J. Scott, and B. Shepard 2010, Fall. Harnessing Cyc to answer clinical researchers' ad hoc queries. *AI Magazine* 31(3):13–32.

Lengauer, T. and R. E. Tarjan 1979. A fast algorithm for finding dominators in a flow-graph. *ACM Transactions on Programming Languages and Systems (TOPLAS)* 1(1):121–141. 82

Leskovec, J. and C. Faloutsos 2006. Sampling from large graphs. Retrieved from the World Wide Web at `http://cs.stanford.edu/people/jure/pubs/sampling-kdd06.pdf`. 109

Leskovec, J., D. Huttenlocher, and J. Kleinbert 2010a. Predicting positive and negative links in online social networks. Retrieved from the World Wide Web at `http://cs.stanford.edu/people/jure/pubs/triads-chi10.pdf`.

Leskovec, J., D. Huttenlocher, and J. Kleinbert 2010b. Signed networks in social media. Retrieved from the World Wide Web at `http://cs.stanford.edu/people/jure/pubs/triads-chi10.pdf`.

Leskovec, J., K. J. Lang, A. Dasgupta, and M. W. Mahoney 2009. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics* 6(1):29–123. `http://www.technologyreview.com/news/515801/the-immortal-life-of-the-enron-e-mails/`.

Lévy, P. 1997. *Collective Intelligence: Mankind's Emerging World in Cyberspace*. Cambridge, Mass.: Perseus Books. 265

Lévy, P. 2001. *Cyberculture*. Minneapolis: University of Minnesota Press. 265

Lewin-Koh, N. 2014. *Hexagon Binning: an Overview*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/hexbin/vignettes/hexagon_binning.pdf`. 252

Lewis, T. G. 2009. *Network Science: Theory and Applications*. New York: Wiley. 224

Liaw, A. and M. Wiener 2014. *randomForest: Breiman and Cutler's Random Forests for Classification and Regression*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/randomForest/randomForest.pdf`.

Lindahl, C. and E. Blount 2003. Weblogs: simplifying web publishing. *Computer* 36(11): 114–116.

Little, R. J. A. and D. B. Rubin 1987. *Statistical Analysis with Missing Data*. New York: Wiley.

Liu, B. 2010. Sentiment analysis and subjectivity. In N. Indurkhya and F. J. Damerau (eds.), *Handbook of Natural Language Processing* (second ed.)., pp. 627–665. Boca Raton, Fla.: Chapman and Hall/CRC.

Liu, B. 2011. *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*. New York: Springer. 251

Liu, B. 2012. *Sentiment Analysis and Opinion Mining*. San Rafael, Calif.: Morgan & Claypool.

Lopez, V., C. Unger, P. Cimiano, and E. Motta 2013. Evaluating question answering over linked data. *Web Semantics* 21:3–13.

Lord, F. M. and M. R. Novick 1968. *Statistical Theories of Mental Test Scores*. Reading, Mass.: Addison-Wesley.

Lovink, G. (ed.) 2011. *Networks Withoug a Cause: A Critique of Social Media*. Cambridge, UK: Polity Press. 134

Luangkesorn, L. 2014. *Simulation Programming with Python*. University of Pittsburgh. 2014. Retrieved from the World Wide Web at `http://users.iems.northwestern.edu/~nelsonb/IEMS435/PythonSim.pdf`. Translation of chapter 4 of Nelson (2014).

Luger, G. F. 2008. *Artificial Intelligence: Structures and Strategies for Complex Problem Solving* (sixth ed.). Boston: Addison-Wesley. 251

Lumley, T. 2010. *Complex Surveys: A Guide to Analysis Using R*. New York: Wiley.

Lumley, T. 2014. *mitools: Tools for Multiple Imputation of Missing Data*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/mitools/mitools.pdf`.

Maas, A. L., R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts 2011, June. Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 142–150. Portland, Ore.: Association for Computational Linguistics. Retrieved from the World Wide Web at `http://ai.stanford.edu/~amaas/papers/wvSent_acl2011.pdf`.

Macal, C. M. and M. J. North 2006. *Introduction to Agent-Based Modeling and Simulation*. `http://www.docstoc.com/docs/36015647/Introduction-to-Agent-based-Modeling-and-Simulation`.

Maechler, M. 2014a. *Package cluster*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/cluster/cluster.pdf`.

Maechler, M. 2014b. *robustbase: Basic Robust Statistics*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/robustbase/robustbase.pdf`. 249

Maisel, L. S. and G. Cokins 2014. *Predictive Business Analytics: Forward-Looking Capabilities to Improve Business Performance*. New York: Wiley. 239

Manly, B. F. J. 1994. *Multivariate Statistical Methods: A Primer* (second ed.). London: Chapman & Hall. 176

Mann, C. and F. Stewart 2000. *Internet Communication and Qualitative Research: A Handbook for Researching Online*. London: Sage. 265

Mann, C. and F. Stewart 2002. Internet interviewing. In J. F. Gubrium and J. A. Holstein (eds.), *Handbook of Interview Research: Context and Method*, Chapter 29, pp. 603–627. Thousand Oaks, Calif.: Sage.

Manning, C. D. and H. Schütze 1999. *Foundations of Statistical Natural Language Processing*. Cambridge: MIT Press.

Markham, A. N. 1998. *Life Online: Researching Real Experience in Virtual Space*. Walnut Creek, Calif.: AltaMira Press. 265

Maronna, R. A., D. R. Martin, and V. J. Yohai 2006. *Robust Statistics Theory and Methods*. New York: Wiley. 249

Matloff, N. 2011. *The Art of R Programming*. San Francisco: no starch press.

Maybury, M. T. (ed.) 1997. *Intelligent Multimedia Information Retrieval*. Menlo Park, Calif./ Cambridge: AAAI Press / MIT Press.

McArthur, R. and P. Bruza 2001. The ABCs of online community. In N. Zhong, Y. Yao, J. Liu, and S. Ohsuga (eds.), *Web Intelligence: Research and Development*, pp. 141–147. New York: Springer.

McCallum, Q. E. (ed.) 2013. *Bad Data Handbook*. Sebastopol, Calif.: O'Reilly.

McCullagh, P. and J. A. Nelder 1989. *Generalized Linear Models* (second ed.). New York: Chapman and Hall.

McGrayne, S. B. 2011. *The Theory that Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines and Emerged Triumphant from Two Centuries of Controversy*. New Haven, Conn.: Yale University Press. 244

McKellar, J. and A. Fettig 2013. *Twisted Network Programming Essentials: Event-driven Network Programming with Python* (second ed.). Sebastopol, Calif.: O'Reilly.

McLean, B. and P. Elkind 2003. *The Smartest Guys in the Room: The Amazing Rise and Scandalous Fall of Enron*. New York: Penguin.

McTaggart, R. and G. Daroczi 2014. *Package Quandl: Quandl Data Connection*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/Quandl/Quandl.pdf` with additional online documentation available at `https://www.quandl.com/`. 62

Meadow, C. T., B. R. Boyce, and D. H. Kraft 2000. *Text Information Retrieval Systems* (second ed.). San Diego: Academic Press.

Merkl, D. 2002. Text mining with self-organizing maps. In W. Klösgen and J. M. Żytkow (eds.), *Handbook of Data Mining and Knowledge Discovery*, Chapter 46.9, pp. 903–910. Oxford: Oxford University Press.

Mertz, D. 2002. *Charming Python: SimPy Simplifies Complex Models*. IBM developerWorks. 2002. Retrieved from the World Wide Web at `http://www.ibm.com/developerworks/linux/library/l-simpy/l-simpy-pdf.pdf`.

Meyer, D. 2014a. *Proximity Measures in the proxy Package for R*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/proxy/vignettes/overview.pdf`.

Meyer, D. 2014b. *proxy: Distance and Similarity Measures*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/proxy/proxy.pdf`.

Meyer, D., E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch 2014. *e1071: Misc Functions of the Department of Statistics (e1071), TU Wien*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/e1071/e1071.pdf`.

Michalawicz, Z. and D. B. Fogel 2004. *How to Solve It: Modern Heuristics*. New York: Springer. 251

Mikowski, M. S. and J. C. Powell 2014. *Single Page Web Applications JavaScript End-to-End*. Shelter Island, N.Y.: Manning. 3

Milborrow, S. 2014. *rpart.plot: Plot rpart models. An Enhanced Version of plot.rpart*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/rpart.plot/rpart.plot.pdf`.

Milgram, S. 1967. The small world problem. *Psychology Today* 1(1):60–67. 99

Miller, D. and D. Slater 2000. *The Internet: An Ethnographic Approach*. Oxford: Berg. 265

Miller, G. A. 1995. Wordnet: A lexical database for English. *Communications of the ACM* 38 (11):39–41. 259

Miller, J. H. and S. E. Page 2007. *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton, N. J.: Princeton University Press.

Miller, J. P. 2000a. *Millennium Intelligence: Understanding and Conducting Competitive Intelligence in the Digital Age*. Medford, N.J.: Information Today, Inc. 59

Miller, T. W. 2000b. Marketing research and the information industry. *CASRO Journal 2000*:21–26.

Miller, T. W. 2001, Summer. Can we trust the data of online research. *Marketing Research*: 26–32. Reprinted as "Make the call: Online results are a mixed bag," *Marketing News*, September 24, 2001:20–25.

Miller, T. W. 2002. Propensity scoring for multimethod research. *Canadian Journal of Marketing Research* 20(2):46–61.

Miller, T. W. 2005. *Data and Text Mining: A Business Applications Approach*. Upper Saddle River, N.J.: Pearson Prentice Hall.

Miller, T. W. 2015a. *Modeling Techniques in Predictive Analytics with Python and R: A Guide to Data Science*. Upper Saddle River, N.J.: Pearson Education. `http://www.ftpress.com/miller`.

Miller, T. W. 2015b. *Modeling Techniques in Predictive Analytics: Business Problems and Solutions with R* (revised and expanded ed.). Upper Saddle River, N.J.: Pearson Education. `http://www.ftpress.com/miller`.

Miller, T. W. and P. R. Dickson 2001. On-line market research. *International Journal of Electronic Commerce* 5(3):139–167.

Miller, T. W., D. Rake, T. Sumimoto, and P. S. Hollman 2001. Reliability and comparability of choice-based measures: Online and paper-and-pencil methods of administration. *2001 Sawtooth Software Conference Proceedings*:123–130. Published by Sawtooth Software, Sequim, WA.

Miller, T. W. and J. Walkowski (eds.) 2004. *Qualitative Research Online*. Manhattan Beach, Calif.: Research Publishers LLC. 261, 313

Miniwatts Marketing Group 2014. Internet users of the world. Retrieved from the World Wide Web, October 21, 2014, at `http://www.internetworldstats.com/stats.htm`. 45

Mitchell, M. 1996. *An Introduction to Genetic Algorithms*. Cambridge: MIT Press. 251

Mitchell, M. 2009. *Complexity: A Guided Tour*. Oxford, U.K.: Oxford University Press. 227

Moran, M. and B. Hunt 2009. *Search Engine Marketing, Inc.: Driving Search Traffic to Your Company's Web Site* (second ed.). Boston: IBM Press/Pearson Education.

Moreno, J. L. 1934. Who shall survive?: Foundations of sociometry, group psychotherapy, and sociodrama. Reprinted in 1953 (second edition) and in 1978 (third edition) by Beacon House, Inc., Beacon, N.Y. 96

Mosteller, F. and D. L. Wallace 1984. *Applied Bayesian and Classical Inference: The Case of "The Federalist" Papers* (second ed.). New York: Springer. Earlier edition published in 1964 by Addison-Wesley, Reading, Mass. The previous title was *Inference and Disputed Authorship: The Federalist*.

Müller, K., T. Vignaux, and C. Chui 2014. *SimPy: Discrete Event Simulation for Python*. SimPy Development Team. 2014. Retrieved from the World Wide Web at `https://simpy.readthedocs.org/en/latest/` with code available at `https://bitbucket.org/simpy/simpy/`.

Murphy, K. P. 2012. *Machine Learning: A Probabilistic Perspective*. Cambridge, Mass.: MIT Press. 176, 244, 251

Murrell, P. 2011. *R Graphics* (second ed.). Boca Raton, Fla.: CRC Press.

Nair, V. G. 2014. *Getting Started with Beautiful Soup: Build Your Own Web Scraper and Learn All About Web Scraping with Beautiful Soup*. Birmingham, UK: PACKT Publishing.

Nelson, B. L. 2013. *Foundations and Methods of Stochastic Simulation: A First Course*. New York: Springer. Supporting materials available from the World Wide Web at `http://users.iems.northwestern.edu/~nelsonb/IEMS435/`.

Nelson, W. B. 2003. *Recurrent Events Data Analysis for Product Repairs, Disease Recurrences, and Other Applications*. Series on Statistics and Applied Probability. Philadelphia and Alexandria, Va.: ASA-SIAM.

Neuendorf, K. A. 2002. *The Content Analysis Guidebook*. Thousand Oaks, Calif.: Sage.

Neuwirth, E. 2014. *Package RColorBrewer: ColorBrewer palettes*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/RColorBrewer/RColorBrewer.pdf`.

Newman, M. E. J. 2010. *Networks: An Introduction*. Oxford, UK: Oxford University Press. 224

Nolan, D. and D. T. Lang 2014. *XML and Web Technologies for Data Sciences with R*. New York: Springer. 29

North, M. J. and C. M. Macal 2007. *Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and Simulation*. Oxford, U.K.: Oxford University Press. 226

Nunnally, J. C. 1967. *Psychometric Theory*. New York: McGraw-Hill.

O'Hagan, A. 2010. *Kendall's Advanced Theory of Statistics: Bayesian Inference*, Volume 2B. New York: Wiley. 244

Osborne, J. W. 2013. *Best Practices in Data Cleaning: A Complete Guide to Everything You Need to Do Before and After Collecting Your Data*. Los Angeles: Sage.

Osgood, C. 1962. Studies in the generality of affective meaning systems. *American Psychologist* 17:10–28. 133

Osgood, C., G. Suci, and P. Tannenbaum (eds.) 1957. *The Measurement of Meaning*. Urbana, Ill.: University of Illinois Press. 133

Pang, B. and L. Lee 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* 2(1–2):1–135.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12:2825–2830.

Penenberg, A. L. and M. Barry 2000. *Spooked: Espionage in Corporate America*. Cambridge, Mass.: Perseus Publishing. 56

Pentland, A. 2014. *Social Physics: How Good Ideas Spread—The Lessons from a New Science*. New York: Penguin.

Peters, G. and J. T. Woolley 2014. State of the Union Addresses and Messages. Online papers available from The American Presidency Project. Retrieved from the World Wide Web, November 1–14, 2014 at `http://www.presidency.ucsb.edu`. Documentation provided at `http://www.presidency.ucsb.edu/sou.php`. 295

Pinker, S. 1994. *The Language Instinct*. New York: W. Morrow and Co.

Pinker, S. 1997. *How the Mind Works*. New York: W.W. Norton & Company.

Pinker, S. 1999. *Words and Rules: The Ingredients of Language*. New York: HarperCollins.

Popping, R. 2000. *Computer-Assisted Text Analysis*. Thousand Oaks, Calif.: Sage. 133, 279

Porter, D. (ed.) 1997. *Internet Culture*. New York: Routledge. 265

Potts, C. 2011. On the negativity of negation. In *Proceedings of Semantics and Linguistic Theory 20*, pp. 636–659. CLC Publications. Retrieved from the World Wide Web at `http://elanguage.net/journals/salt/article/view/20.636/1414`.

Powazak, D. M. 2001. *Design for Community: The Art of Connecting Real People in Virtual Places*. Indianapolis: New Riders.

Powers, S. 2003. *Practical RDF*. Sebastopol, Calif.: O'Reilly. 235

Preece, J. 2000. *Online Communities: Designing Usability, Supporting Sociability*. New York: Wiley.

Priebe, C. E., J. M. Conroy, D. J. Marchette, and Y. Park 2005, October. Scan statistics on Enron graphs. *Computational and Mathematical Organization Theory* 11(3):229–247.

Priebe, J. 2009, April 29. A study of Internet users' cookie and JavaScript settings. Retrieved from the World Wide Web, October 21, 2014, at `http://smorgasbork.com/component/content/article/84-a-study-of-internet-users-cookie-and-javascript-settings`. 16

Provost, F. and T. Fawcett 2014. *Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking*. Sebastopol, Calif.: O'Reilly. 239

Pustejovsky, J. and A. Stubbs 2013. *Natural Language Annotation for Machine Learning*. Sebastopol, Calif.: O'Reilly.

Putler, D. S. and R. E. Krider 2012. *Customer and Business Analytics: Applied Data Mining for Business Decision Making Using R*. Boca Raton, Fla: Chapman & Hall/CRC.

Radcliffe-Brown, A. R. 1940. On social structure. *Journal of the Royal Anthropological Society of Great Britain and Ireland* 70:1–12. 96

Rajaraman, A. and J. D. Ullman 2012. *Mining of Massive Datasets*. Cambridge, UK: Cambridge University Press. 210

Reingold, E. M. and J. S. Tilford 1981. A fast algorithm for finding dominators in a flow-graph. *IEEE Transactions on Software Engineering* 7:223–228. 82

Reitz, K. 2014a, October 15. Python guide documentation (release 0.0.1). Retrieved from the World Wide Web, October 23, 2014, at `https://media.readthedocs.org/pdf/python-guide/latest/python-guide.pdf`.

Reitz, K. 2014b. Requests: HTTP for humans. Documentation available from the World Wide Web at `http://docs.python-requests.org/en/latest/`.

Rencher, A. C. and G. B. Schaalje 2008. *Linear Models in Statistics* (second ed.). New York: Wiley.

Resig, J. and BearBibeault 2013. *Secrets of the JavaScript Ninja*. Shelter Island, N.Y.: Manning. 3

Resnick, M. 1998. *Turtles, Termites, and Traffic Jams: Explorations in Massively Parallel Microworlds*. Cambridge, Mass.: MIT Press. 227

Rheingold, H. 2000. *The Virtual Community: Homesteading on the Electronic Frontier* (revised ed.). Cambridge: The MIT Press. 265

Ricci, F., L. Rokach, B. Shapira, and P. B. Kantor (eds.) 2011. *Recommender Systems Handbook*. New York: Springer.

Richardson, L. 2014. Beautiful soup documentation. Available from the World Wide Web at `http://www.crummy.com/software/BeautifulSoup/bs4/doc/`.

Robbins, J. N. 2003. *Learning Web Design: A Beginners Guide to HTML, CSS, JavaScript, and Web Graphics* (fourth ed.). Sebastopol, Calif.: O'Reilly.

Robert, C. P. 2007. *The Bayesian Choice: From Decision Theoretic Foundations to Computational Implementation* (second ed.). New York: Springer. 244

Robert, C. P. and G. Casella 2009. *Introducing Monte Carlo Methods with R*. New York: Springer. 244

Roberts, C. W. (ed.) 1997. *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts*. Mahwah, N.J.: Lawrence Erlbaum Associates. 133, 279

Robinson, I., J. Webber, and E. Eifrem 2013. *Graph Databases*. Sebastopol, Calif.: O'Reilly.

Rosen, J. 2001. *The Unwanted Gaze: The Destruction of Privacy in America*. New York: Vintage. 134

Rossant, C. 2014. *IPython Interactive Computing and Visualization Cookbook*. Birmingham, UK: Packt Publishing.

Rousseeuw, P. J. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20:53–65.

Rubin, D. B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

Russell, M. A. 2014. *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More*. Sebastopol, Calif.: O'Reilly. 61

Russell, S. and P. Norvig 2009. *Artificial Intelligence: A Modern Approach* (third ed.). Upper Saddle River, N.J.: Prentice Hall. 251

Rusu, A. 2014. Tree drawing algorithms. In R. Tamassia (ed.), *Handbook of Graph Drawing and Visualization*, Chapter 5, pp. 155–192. Boca Raton, Fla.: CRC Press/Chapman & Hall. 82

Ryan, G. W. and H. R. Bernard 2000. Data management and analysis methods. In N. K. Denzin and Y. S. Lincoln (eds.), *Handbook of Qualitative Research: Context and Method* (second ed.), Chapter 29, pp. 769–802. Thousand Oaks, Calif.: Sage. 279

Ryan, J. A. 2014. *Package quantmod: Quantitative Financial Modelling Framework*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/quantmod/quantmod.pdf`. 62

Ryan, J. A. and J. M. Ulrich 2014. *Package xts: eXtensible Time Series*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/xts/xts.pdf`. 62

Ryan, T. P. 2008. *Modern Regression Methods* (second ed.). New York: Wiley. 128

Šalamon, T. 2011. *Design of Agent-Based Models: Developing Computer Simulations for a Better Understanding of Social Processes*. Czech Republic: Tomáš Bruckner, Řepín-Živonín.

Samara, T. 2007. *Design Elements: A Graphic Style Manual*. Gloucester, Mass.: Rockport Publishers.

Sarkar, D. 2008. *Lattice: Multivariate Data Visualization with R*. New York: Springer.

Sarkar, D. 2014. *lattice: Lattice Graphics*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/lattice/lattice.pdf`.

Sarkar, D. and F. Andrews 2014. *latticeExtra: Extra Graphical Utilities Based on Lattice*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/latticeExtra/latticeExtra.pdf`.

Savage, D., X. Zhang, X. Yu, P. Chou, and Q. Wang 2014. Anomaly detection in online social networks. *Social Networks* 39:62–70.

Schafer, J. L. 1997. *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall.

Schauerhuber, M., A. Zeileis, D. Meyer, and K. Hornik 2008. Benchmarking open-source tree learners in R/RWeka. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme, and R. Decker (eds.), *Data Analysis, Machine Learning, and Applications*, pp. 389–396. New York: Springer.

Schaul, J. 2014. *ComplexNetworkSim Package Documentation*. pypi.python.org. 2014. Retrieved from the World Wide Web at `https://pythonhosted.org/ComplexNetworkSim/` with code available at `https://github.com/jschaul/ComplexNetworkSim`.

Schnettler, S. 2009. A structured overview of 50 years of small-world research. *Social Networks* 31:165–178. 99

Schrott, P. R. and D. J. Lanoue 1994. Trends and perspectives in content analysis. In I. Borg and P. Mohler (eds.), *Trends and Perspectives in Empirical Social Research*, pp. 327–345. Berlin: Walter de Gruyter.

Scott:, J. 2013. *Social Network Analysis* (third ed.). Thousand Oaks, Calif.: Sage.

Seale, C. 2002. Computer-assisted analysis of qualitative interview data. In J. F. Gubrium and J. A. Holstein (eds.), *Handbook of Interview Research: Context and Method*, Chapter 31, pp. 651–670. Thousand Oaks, Calif.: Sage. 279

Sebastiani, F. 2002. Machine learning in automated text categorization. *ACM Computing Surveys* 34(1):1–47.

Seber, G. A. F. 2000. *Multivariate Observations*. New York: Wiley. Originally published in 1984. 176

Segaran, T., C. Evans, and J. Taylor 2009. *Programming the Semantic Web*. Sebastopol, Calif.: O'Reilly. 235

Sen, R. and M. H. Hansen 2003, March. Predicting web user's next access based on log data. *Journal of Computational and Graphical Statistics* 12(1):143–155.

Senkul, P. and S. Salin 2012, March. Improving pattern quality in web usage mining by using semantic information. *Knowledge and Information Systems* 30(3):527–541.

SEOMoz, Inc. 2013. Search engine ranking factors: survey and correlation data. Retrieved from the World Wide Web, November 1, 2014, at `http://moz.com/search-ranking-factors`.

Shakhnarovich, G., T. Darrell, and P. Indyk (eds.) 2006. *Nearest-Neighbor Methods in Learning and Vision: Theory and Practice*. Cambridge, Mass.: MIT Press.

Sharda, R. and D. Delen 2006. Predicting box office success of motion pictures with neural networks. *Expert Systems with Applications* 30:243–254. 133

Sharma, S. 1996. *Applied Multivariate Techniques*. New York: Wiley. 176

Sherry, Jr., J. F. and R. V. Kozinets 2001. Qualitative inquiry in marketing and consumer research. In D. Iacobucci (ed.), *Kellogg on Marketing*, Chapter 8, pp. 165–194. New York: Wiley.

Shmueli, G. 2010. To explain or predict? *Statistical Science* 25(3):289–310.

Shroff, G. 2013. *The Intelligent Web: Search, Smart Algorithms, and Big Data*. Oxford, U.K.: Oxford University Press.

Siegel, E. 2013. *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die*. Hoboken, N.J.: Wiley. 239

Silverman, D. 1993. *Interpreting Qualitative Data: Methods for Analysing Talk, Text, and Interaction*. London: Sage. 279

Silverman, D. 2000a. Analyzing talk and text. In N. K. Denzin and Y. S. Lincoln (eds.), *Handbook of Qualitative Research: Context and Method* (second ed.), Chapter 31, pp. 821–834. Thousand Oaks, Calif.: Sage. 279

Silverman, D. 2000b. *Doing Qualitative Research: A Practical Approach*. Thousand Oaks, Calif.: Sage.

Silverman, D. 2001. *Interpreting Qualitative Data: Methods for Analysing Talk, Text, and Interaction* (second ed.). Thousand Oaks, Calif.: Sage.

Simpson, J. 2002. *XPath and XPointer: Locating Content in XML Documents*. Sebastopol, Calif.: O'Reilly. 29

Sing, T., O. Sander, N. Beerenwinkel, and T. Lengauer 2005. ROCR: Visualizing classifier performance in R. *Bioinformatics* 21(20):3940–3941.

Slater, J. B., P. V. M. Broekman, M. Corris, A. Iles, B. Seymour, and S. Worthington (eds.) 2013. *Proud to Be Flesh—A Mute Magazine Anthology of Cultural Politics After the Net*. London: Mute Publishing.

Smith, J. A. and J. Moody 2013. Structural effects of network sampling coverage I: Nodes missing at random. *Social Networks* 35:652–668. 109

Smith, M. A. and P. Kollock (eds.) 1999. *Communities in Cyberspace*. New York: Routledge. 265

Snedecor, G. W. and W. G. Cochran 1989. *Statistical Methods* (eighth ed.). Ames, Iowa: Iowa State University Press. First edition published by Snedecor in 1937. 244

Socher, R., J. Pennington, E. H. Huang, A. Y. Ng, and C. D. Manning 2011. Semi-Supervised Recursive Autoencoders for Predicting Sentiment Distributions. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Srivastava, A. N. and M. Sahami (eds.) 2009. *Text Mining: Classification, Clustering, and Applications*. Boca Raton, Fla.: CRC Press.

StatCounter 2014. StatCounter global stats. Retrieved from the World Wide Web, October 21, 2014, at `http://gs.statcounter.com/#browser-ww-monthly-200807-201410`.

Stefanov, S. 2010. *JavaScript Patterns: Building Better Applications with Coding and Design Patterns*. Sebastopol, Calif.: O'Reilly. 3

Stone, P. J. 1997. Thematic text analysis: New agendas for analyzing text content. In C. W. Roberts (ed.), *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts*, Chapter 2, pp. 35–54. Mahwah, N.J.: Lawrence Erlbaum Associates. 279

Stone, P. J., D. C. Dunphy, M. S. Smith, and D. M. Ogilvie 1966. *The General Inquirer: A Computer Approach to Content Analysis*. Cambridge: MIT Press.

Strauss, A. and J. Corbin 1998. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory* (second ed.). Thousand Oaks, Calif.: Sage. 279

Stuart, A., K. Ord, and S. Arnold 2010. *Kendall's Advanced Theory of Statistics: Classical Inference and the Linear Model*, Volume 2A. New York: Wiley. 244

Suess, E. A. and B. E. Trumbo 2010. *Introduction to Probability Simulation and Gibbs Sampling with R*. New York: Springer. 244

Sullivan, D. 2001. *Document Warehousing and Text Mining: Techniques for Improving Business Operations, Marketing, and Sales*. New York: Wiley.

Supowit, K. J. and E. M. Reingold 1983. The complexity of drawing trees nicely. *Acta Informatica* 18:177–392. 82

Taddy, M. 2013a. Measuring political sentiment on Twitter: factor-optimal design for multinomial inverse regression. Retrieved from the World Wide Web at `http://arxiv.org/pdf/1206.3776v5.pdf`.

Taddy, M. 2013b. Multinomial inverse regression for text analysis. Retrieved from the World Wide Web at `http://arxiv.org/pdf/1012.2098v6.pdf`.

Taddy, M. 2014. *textir: Inverse Regression for Text Analysis*. 2014. `http://cran.r-project.org/web/packages/textir/textir.pdf`.

Tamassia, R. (ed.) 2014. *Handbook of Graph Drawing and Visualization*. Boca Raton, Fla.: CRC Press/Chapman & Hall.

Tan, P.-N., M. Steinbach, and V. Kumar 2006. *Introduction to Data Mining*. Boston: Addison-Wesley. 210

Tang, W., H. He, and X. M. Tu 2012. *Applied Categorical and Count Data Analysis*. Boca Raton, Fla.: Chapman & Hall/CRC.

Tannenbaum, A. S. and D. J. Wetherall 2010. *Computer Networks* (fifth ed.). Upper Saddle River, N.J.: Pearson/Prentice Hall.

Tanner, M. A. 1996. *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions* (third ed.). New York: Springer. 244

Tennison, J. 2001. *XSLT and XPath On The Edge*. New York: Wiley. 29

Therneau, T. 2014. *survival: Survival Analysis*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/survival/survival.pdf`.

Therneau, T., B. Atkinson, and B. Ripley 2014. *rpart: Recursive Partitioning*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/rpart/rpart.pdf`.

Therneau, T. and C. Crowson 2014. *Using Time Dependent Covariates and Time Dependent Coefficients in the Cox Model*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/survival/vignettes/timedep.pdf`.

Therneau, T. M. and P. M. Grambsch 2000. *Modeling Survival Data: Extending the Cox Model*. New York: Springer.

Thiele, J. C. 2014. *Package RNetLogo: Provides an Interface to the Agent-Based Modelling Platform NetLogo*. Comprehensive R Archive Network. 2014. User manual: `http://cran.r-project.org/web/packages/RNetLogo/RNetLogo.pdf`. Additional documentation available at `http://rnetlogo.r-forge.r-project.org/`.

Thompson, M. 1975. On any given Sunday: fair competitor orderings with maximum likelihood methods. *Journal of the American Statistical Association* 70(351):536–541. 103

Thompson, T. (ed.) 2000. *Writing about Business* (second ed.). New York: Columbia University Press.

T.Hothorn, F. Leisch, A. Zeileis, and K. Hornik 2005, September. The design and analysis of benchmark experiments. *Journal of Computational and Graphical Statistics* 14(3):675–699.

Thurstone, L. L. 1927. A law of comparative judgment. *Psychological Review* 34:273–286.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B* 58:267–288. 249

Tidwell, J. 2011. *Designing Interfaces: Patterns for Effective Interactive Design* (second ed.). Sebastopol, Calif.: O'Reilly.

Toivonen, R., L. Kovanen, J.-P. O. Mikko Kivelä, J. SaramŁki, and K. Kaski 2009. A comparative study of social network models: Network evolution models and nodal attribute models. *Social Networks* 31:240–254.

Tong, S. and D. Koller 2001. Support vector machine active learning with applications to text classification. *Journal of Machine Learning Research* 2:45–66.

Torgerson, W. S. 1958. *Theory and Methods of Scaling*. New York: Wiley.

Travers, J. and S. Milgram 1969, December. Experimental study of small world problem. *Sociometryy* 32(4):425–443. 99

Trybula, W. J. 1999. Text mining. In M. E. Williams (ed.), *Annual Review of Information Science and Technology*, Volume 34, Chapter 7, pp. 385–420. Medford, N.J.: Information Today, Inc.

Tsvetovat, M. and A. Kouznetsov 2011. *Social Network Analysis for Startups: Finding Connections on the Social Web*. Sebastopol, Calif.: O'Reilly.

Tufte, E. R. 1990. *Envisioning Information*. Cheshire, Conn.: Graphics Press.

Tufte, E. R. 1997. *Visual Explanations: Images and Quantities, Evidence and Narrative*. Cheshire, Conn.: Graphics Press.

Tufte, E. R. 2004. *The Visual Display of Quantitative Information* (second ed.). Cheshire, Conn.: Graphics Press.

Tufte, E. R. 2006. *Beautiful Evidence*. Cheshire, Conn.: Graphics Press.

Tukey, J. W. 1977. *Exploratory Data Analysis*. Reading, Mass.: Addison-Wesley.

Tukey, J. W. and F. Mosteller 1977. *Data Analysis and Regression: A Second Course in Statistics*. Reading, Mass.: Addison-Wesley.

Turney, P. D. 2002, July 8–10. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL '02)*:417–424. Available from the National Research Council Canada publications archive. 134

Turow, J. 2013. *The Daily You: How the New Advertising Industry Is Defining Your Identity and Your Worth*. New Haven, Conn.: Yale University Press. 134

Uddin, S., J. Hamra, and L. Hossain 2013. Exploring communication networks to understand organizational crisis using exponential random graph models. *Computational and Mathematical Organization Theory* 19:25–41.

Uddin, S., A. Khan, L. Hossain, M. Piraveenan, and S. Carlsson 2014. A topological framework to explore longitudinal social networks. *Computational and Mathematical Organization Theory*:21 pages. Manuscript published online by Springer Science+Business Media, New York.

Unwin, A., M. Theus, and H. Hofmann (eds.) 2006. *Graphics of Large Datasets: Visualizing a Million*. New York: Springer. 252

Vanderbei, R. J. 2014. *Linear Programming: Foundations and Extensions* (fourth ed.). New York: Springer. 225

Vapnik, V. N. 1998. *Statistical Learning Theory*. New York: Wiley. 129

Vapnik, V. N. 2000. *The Nature of Statistical Learning Theory* (second ed.). New York: Springer. 129

Venables, W. N. and B. D. Ripley 2002. *Modern Applied Statistics with S* (fourth ed.). New York: Springer-Verlag. Champions of S, S-Plus, and R call this *the mustard book*.

Vine, D. 2000. *Internet Business Intelligence: How to Build a Big Company System on a Small Company Budget*. Medford, N.J.: Information Today, Inc. 59

W3Techs 2014. Usage of traffic analysis tools for websites. Retrieved from the World Wide Web, October 21, 2014, at `http://w3techs.com/technologies/overview/traffic_analysis/all`. 3, 14

Walther, J. B. 1996. Computer-mediated communication: impersonal, interpersonal, and hyperpersonal interaction. *Communication Research* 23(1):3–41. 265

Wanderschneider, M. 2013. *Learning Node.js: A Hands-On Guide to Building Web Applications in JavaScript*. Upper Saddle River, N.J.: Pearson Education/Addison-Wesley. 3

Wang, X., H. Tao, Z. Xie, and D. Yi 2013. Mining social networks using wave propogation. *Computational and Mathematical Organization Theory* 19:569–579.

Wasserman, L. 2010. *All of Statistics: A Concise Course in Statistical Inference*. New York: Springer. 244

Wasserman, S. and C. Anderson 1987. Stochastic *a posteriori* blockmodels: Construction and assessment. *Social Networks* 9:1–36.

Wasserman, S. and K. Faust 1994. *Social Network Analysis: Methods and Applications*. Cambridge, UK: Cambridge University Press. Chapter 4: Graphs and Matrices contributed by D. Iacobucci.

Wasserman, S. and D. Iacobucci 1986. Statistical analysis of discrete relational data. *British Journal of Mathematical and Statistical Psychology* 39:41–64.

Wasserman, S. and P. Pattison 1996, September. Logit models and logistic regression for social networks: I An introduction to markov graphs and p^*. *Psychometrika* 61(3):401–425.

Watts, D. J. 1999. *Small Worlds: The Dynamics of Networks between Order and Randomness*. Princeton, N.J.: Princeton University Press. 99

Watts, D. J. 2003. *Six Degrees: The Science of a Connected Age*. New York: W.W. Norton.

Watts, D. J. and S. H. Strogatz 1998. Collective dynamics of 'small-world' networks. *Nature* 393(6684):440–442. 99, 224

Wei, T. H. 1952. The algebraic foundations of ranking theory. Ph.D. thesis, Cambridge University, Cambridge, UK.

Weiner, J. 2014. *Package riverplot: Sankey or ribbon plots*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/riverplot/riverplot.pdf`.

Weiss, S. M., N. Indurkhay, and T. Zhang 2010. *Fundamentals of Predictive Text Mining*. New York: Springer.

Weiss, S. M., N. Indurkhya, and T. Zhang 2010. *Fundamentals of Predictive Text Mining*. New York: Springer.

Weitzman, E. A. 2000. Software and qualitative research. In N. K. Denzin and Y. S. Lincoln (eds.), *Handbook of Qualitative Research: Context and Method* (second ed.)., Chapter 30, pp. 803–820. Thousand Oaks, Calif.: Sage. 279

Weitzman, E. A. and M. B. Miles 1995. *Computer Programs for Qualitative Data Analysis*. Thousand Oaks, Calif.: Sage. 279

Werry, C. and M. Mowbray (eds.) 2001. *Online Communities: Commerce, Community Action, and the Virtual University*. Upper Saddle River, N.J.: Prentice-Hall.

West, B. T. 2006. A simple and flexible rating method for predicting success in the NCAA basketball tournament. *Journal of Quantitative Analysis in Sports* 2(3):1–14. 103

West, M. D. (ed.) 2001. *Theory, Method, and Practice in Computer Content Analysis*. Westport, Conn.: Ablex. 133, 279

White, T. 2011. *Hadoop: The Definitive Guide* (second ed.). Sebastopol, Calif.: O'Reilly.

Wickham, H. 2010. stringr: Modern, consistent string processing. *The R Journal* 2(2):38–40.

Wickham, H. 2014a. *Advanced R*. Boca Raton, Fla.: Chapman & Hall/CRC.

Wickham, H. 2014b. *stringr: Make It Easier to Work with Strings*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/stringr/stringr.pdf`. 259

Wickham, H. and W. Chang 2014. *ggplot2: An Implementation of the Grammar of Graphics*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf`. 62, 252

Wikipedia 2014a. 2011 Virginia earthquake—Wikipedia, the free encyclopedia. Retrieved from the World Wide Web, September 18, 2014, at url = `http://en.wikipedia.org/w/index.php?title=2011_Virginia_earthquake&oldid=625563344`.

Wikipedia 2014b. Directory on privacy and electronic communications. Retrieved from the World Wide Web, October 21, 2014, at `http://en.wikipedia.org/wiki/Directive_on_Privacy_and_Electronic_Communication`. 19

Wikipedia 2014c. History of Wikipedia—Wikipedia, the free encyclopedia. Retrieved from the World Wide Web, September 23, 2014, at `http://en.wikipedia.org/wiki/History_of_Wikipedia`.

Wilensky, U. 1999. Netlogo. NetLogo 5.1.0 User Manual and documentation available from the Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL at `http://ccl.northwestern.edu/netlogo/`.

Wilkinson, L. 2005. *The Grammar of Graphics* (second ed.). New York: Springer.

Williams, H. P. 2013. *Model Building in Mathematical Programming* (fifth ed.). New York: Wiley. 225

Witten, I. H., E. Frank, and M. A. Hall 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. Burlington, Mass.: Morgan Kaufmann. 210, 250

Witten, I. H., A. Moffat, and T. C. Bell 1999. *Managing Gigabytes: Compressing and Indexing Documents and Images* (second ed.). San Francisco: Morgan Kaufmann.

Wolcott, H. F. 1994. *Transforming Qualitative Data: Description, Analysis, and Interpretation*. Thousand Oaks, Calif.: Sage.

Wolcott, H. F. 1999. *Ethnography: A Way of Seeing*. Walnut Creek, Calif.: AltaMira.

Wolcott, H. F. 2001. *The Art of Fieldwork*. Walnut Creek, Calif.: AltaMira.

Wood, D., M. Zaidman, and L. Ruth 2014. *Linked Data: Structured Data on the Web*. Shelter Island, N.Y.: Manning.

Yau, N. 2011. *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics*. New York: Wiley.

Yau, N. 2013. *Data Points: Visualization That Means Something*. New York: Wiley.

Youmans, G. 1990. Measuring lexical style and competence: The type-token vocabulary curve. *Style* 24(4):584–599. 259

Youmans, G. 1991. A new tool for discourse analysis: The vocabulary management profile. *Language* 67(4):763–789. 259

Zeileis, A., G. Grothendieck, and J. A. Ryan 2014. *Package zoo: S3 Infrastructure for Regular and Irregular Time Series (Z's ordered observations)*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/zoo/zoo.pdf`. 62

Zeileis, A., K. Hornik, and P. Murrell 2009, July. Escaping RGBland: Selecting colors for statistical graphics. *Computational Statistics and Data Analysis* 53(9):3259–3270.

Zeileis, A., K. Hornik, and P. Murrell 2014. *HCL-Based Color Palettes in R*. Comprehensive R Archive Network. 2014. `http://cran.r-project.org/web/packages/colorspace/vignettes/hcl-colors.pdf`.

Zubcsek, P. P., I. Chowdhury, and Z. Katona 2014. Information communities: The network structure of communication. *Social Networks* 38:50–62.