

# Chapter 14

## Bibliography

- [Abe13] Andrew Abela. *Advanced Presentations by Design: Creating Communication that Drives Action*. Pfeiffer, 2nd edition, 2013.
- [Ans73] Francis J Anscombe. Graphs in statistical analysis. *The American Statistician*, 27(1):17–21, 1973.
- [Bab11] Charles Babbage. *Passages from the Life of a Philosopher*. Cambridge University Press, 2011.
- [Bar10] James Barron. Apple’s new device looks like a winner. From 1988. *The New York Times*, January 28, 2010.
- [Ben12] Edward A Bender. *An Introduction to Mathematical Modeling*. Courier Corporation, 2012.
- [Bis07] Christopher Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, New York, 2007.
- [BK07] Robert M. Bell and Yehuda Koren. Lessons from the Netflix prize challenge. *ACM SIGKDD Explorations Newsletter*, 9(2):75–79, 2007.
- [BK16] Benjamin Bengfort and Jenny Kim. *Data Analytics with Hadoop: An Introduction for Data Scientists*. O’Reilly Media, Inc., 2016.
- [BP98] Serge Brin and Larry Page. The Anatomy of a Large-Scale Hypertextual Web Search Engine. In *Proc. 7th Int. Conf. on World Wide Web (WWW)*, pages 107–117, 1998.
- [Bra99] Ronald Bracewell. *The Fourier Transform and its Applications*. McGraw-Hill, 3rd edition, 1999.
- [Bri88] E. Oran Brigham. *The Fast Fourier Transform*. Prentice Hall, Englewood Cliffs NJ, facimile edition, 1988.
- [BSC<sup>+</sup>11] M. Borjesson, L. Serratos, F. Carre, D. Corrado, J. Drezner, D. Dugmore, H. Heidbuchel, K. Mellwig, N. Panhuyzen-Goedkoop, M. Papadakis, H. Rasmussen, S. Sharma, E. Solberg, F. van Buuren, and A. Pelliccia. Consensus document regarding cardiovascular safety at sports arenas. *European Heart Journal*, 32:2119–2124, 2011.
- [BT08] Dimitri Bertsekas and John Tsitsklis. *Introduction to Probability*. Athena Scientific, 2nd edition, 2008.

- [BVS08] Mikhail Bautin, Lohit Vijayarenu, and Steven Skiena. International Sentiment Analysis for News and Blogs. In *Proceedings of the International Conference on Weblogs and Social Media*, Seattle, WA, April 2008.
- [BWPS10] Mikhail Bautin, Charles B Ward, Akshay Patil, and Steven S Skiena. Access: news and blog analysis for the social sciences. In *Proceedings of the 19th International Conference on World Wide Web*, pages 1229–1232. ACM, 2010.
- [CPS<sup>+</sup>08] J Robert Coleman, Dimitris Papamichail, Steven Skiena, Bruce Fitcher, Eckard Wimmer, and Steffen Mueller. Virus attenuation by genome-scale changes in codon pair bias. *Science*, 320(5884):1784–1787, 2008.
- [CPS15] Yanqing Chen, Bryan Perozzi, and Steven Skiena. Vector-based similarity measurements for historical figures. In *International Conference on Similarity Search and Applications*, pages 179–190. Springer, 2015.
- [dBvKOS00] Mark de Berg, Mark van Kreveld, Mark Overmars, and Otfried Schwarzkopf. *Computational Geometry: Algorithms and Applications*. Springer, 2nd edition, 2000.
- [DDKN11] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning future media environments*, pages 9–15. ACM, 2011.
- [Don15] David Donoho. 50 years of data science. Tukey Centennial Workshop, Princeton NJ, 2015.
- [DP12] Thomas H Davenport and DJ Patil. Data scientist. *Harvard Business Review*, 90(5):70–76, 2012.
- [EK10] David Easley and Jon Kleinberg. *Networks, Crowds, and Markets: Reasoning about a highly connected world*. Cambridge University Press, 2010.
- [ELLS11] Brian Everitt, Sabine Landau, Mmerven Leese, and Daniel Stahl. *Cluster Analysis*. Wiley, 5th edition, 2011.
- [ELS93] Peter Eades, X. Lin, and William F. Smyth. A fast and effective heuristic for the feedback arc set problem. *Information Processing Letters*, 47:319–323, 1993.
- [FCH<sup>+</sup>14] Matthew Faulkner, Robert Clayton, Thomas Heaton, K Mani Chandy, Monica Kohler, Julian Bunn, Richard Guy, Annie Liu, Michael Olson, MingHei Cheng, et al. Community sense and response systems: Your phone as quake detector. *Communications of the ACM*, 57(7):66–75, 2014.
- [Few09] Stephen Few. *Now You See It: simple visualization techniques for quantitative analysis*. Analytics Press, Oakland CA, 2009.
- [FHT01] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The Elements of Statistical Learning*. Springer, 2001.
- [FPP07] David Freedman, Robert Pisani, and Roger Purves. *Statistics*. WW Norton & Co, New York, 2007.
- [Gay14] C. Gayomali. NYC taxi data blunder reveals which celebs don’t tip and who frequents strip clubs. <http://www.fastcompany.com/3036573/>, October 2. 2014.

- [GBC16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.
- [GFH13] Frank R. Giordano, William P. Fox, and Steven B. Horton. *A First Course in Mathematical Modeling*. Nelson Education, 2013.
- [Gle96] James Gleick. A bug and a crash: sometimes a bug is more than a nuisance. *The New York Times Magazine*, December 1, 1996.
- [GMP<sup>+</sup>09] Jeremy Ginsberg, Matthew H Mohebbi, Rajan S Patel, Lynnette Brammer, Mark S Smolinski, and Larry Brilliant. Detecting influenza epidemics using search engine query data. *Nature*, 457(7232):1012–1014, 2009.
- [Gol16] David Goldenberg. The biggest dinosaur in history may never have existed. FiveThirtyEight, <http://fivethirtyeight.com/features/the-biggest-dinosaur-in-history-may-never-have-existed/>, January 11, 2016.
- [Gru15] Joel Grus. *Data Science from Scratch: First principles with Python*. O'Reilly Media, Inc., 2015.
- [GSS07] Namrata Godbole, Manja Srinivasaiah, and Steven Skiena. Large-scale sentiment analysis for news and blogs. *Int. Conf. Weblogs and Social Media*, 7:21, 2007.
- [HC88] Diane F Halpern and Stanley Coren. Do right-handers live longer? *Nature*, 333:213, 1988.
- [HC91] Diane F Halpern and Stanley Coren. Handedness and life span. *N Engl J Med*, 324(14):998–998, 1991.
- [HS10] Yancheng Hong and Steven Skiena. The wisdom of bookies? sentiment analysis vs. the NFL point spread. In *Int. Conf. on Weblogs and Social Media*, 2010.
- [Huf10] Darrell Huff. *How to Lie with Statistics*. WW Norton & Company, 2010.
- [Ind04] Piotr Indyk. Nearest neighbors in high-dimensional spaces. In J. Goodman and J. O'Rourke, editors, *Handbook of Discrete and Computational Geometry*, pages 877–892. CRC Press, 2004.
- [Jam10] Bill James. *The New Bill James Historical Baseball Abstract*. Simon and Schuster, 2010.
- [JLSI99] Vic Jennings, Bill Lloyd-Smith, and Duncan Ironmonger. Household size and the Poisson distribution. *J. Australian Population Association*, 16:65–84, 1999.
- [Joa02] Thorsten Joachims. Optimizing search engines using clickthrough data. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 133–142. ACM, 2002.
- [Joh07] Steven Johnson. *The Ghost Map: The story of London's most terrifying epidemic – and how it changed science, cities, and the modern world*. Riverhead Books, 2007.
- [JWHT13] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning*. Springer-Verlag, sixth edition, 2013.
- [Kap12] Karl M Kapp. *The Gamification of Learning and Instruction: Game-based methods and strategies for training and education*. Wiley, 2012.

- [KARPS15] Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. Statistically significant detection of linguistic change. In *Proceedings of the 24th International Conference on World Wide Web*, pages 625–635. ACM, 2015.
- [KCS08] Aniket Kittur, Ed H Chi, and Bongwon Suh. Crowdsourcing user studies with mechanical turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 453–456. ACM, 2008.
- [KKK<sup>+</sup>10] Slava Kisilevich, Milos Krstajic, Daniel Keim, Natalia Andrienko, and Gennady Andrienko. Event-based analysis of people’s activities and behavior using flickr and panoramio geotagged photo collections. In *2010 14th International Conference Information Visualisation*, pages 289–296. IEEE, 2010.
- [Kle13] Phillip Klein. *Coding the Matrix: Linear Algebra through Computer Science Applications*. Newtonian Press, 2013.
- [KSG13] Michal Kosinski, David Stillwell, and Thore Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proc. National Academy of Sciences*, 110(15):5802–5805, 2013.
- [KTDS17] Vivek Kulkarni, Yingtao Tian, Parth Dandiwal, and Steven Skiena. Dating documents: A domain independent approach to predict year of authorship. Submitted for publication, 2017.
- [Lei07] David J. Leinweber. Stupid data miner tricks: overfitting the S&P 500. *The Journal of Investing*, 16(1):15–22, 2007.
- [Lew04] Michael Lewis. *Moneyball: The art of winning an unfair game*. WW Norton & Company, 2004.
- [LG14] Omer Levy and Yoav Goldberg. Neural word embedding as implicit matrix factorization. In *Advances in Neural Information Processing Systems*, pages 2177–2185, 2014.
- [LKKV14] David Lazer, Ryan Kennedy, Gary King, and Alessandro Vespignani. The parable of Google flu: traps in big data analysis. *Science*, 343(6176):1203–1205, 2014.
- [LKS05] Levon Lloyd, Dimitrios Kechagias, and Steven Skiena. Lydia: A system for large-scale news analysis. In *SPIRE*, pages 161–166, 2005.
- [LLM15] David Lay, Steven Lay, and Judi McDonald. *Linear Algebra and its Applications*. Pearson, 5th edition, 2015.
- [LM12] Amy Langville and Carl Meyer. *Who’s #1? The Science of Rating and Ranking*. Princeton Univ. Press, 2012.
- [LRU14] Jure Leskovec, Anand Rajaraman, and Jeffrey David Ullman. *Mining of Massive Datasets*. Cambridge University Press, 2014.
- [Mal99] Burton Gordon Malkiel. *A Random Walk Down Wall Street: Including a life-cycle guide to personal investing*. WW Norton & Company, 1999.
- [MAV<sup>+</sup>11] J. Michel, Y. Shen A. Aiden, A. Veres, M. Gray, Google Books Team, J. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, M. Nowak, and E. Aiden. Quantitative analysis of culture using millions of digitized books. *Science*, 331:176–182, 2011.

- [MBL<sup>+</sup>06] Andrew Mehler, Yunfan Bao, Xin Li, Yue Wang, and Steven Skiena. Spatial Analysis of News Sources. In *IEEE Trans. Vis. Comput. Graph.*, volume 12, pages 765–772, 2006.
- [MCCD13] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
- [McK12] Wes McKinney. *Python for Data Analysis: Data wrangling with Pandas, NumPy, and IPython*. O’Reilly Media, Inc., 2012.
- [McM04] Chris McManus. *Right Hand, Left Hand: The origins of asymmetry in brains, bodies, atoms and cultures*. Harvard University Press, 2004.
- [MCP<sup>+</sup>10] Steffen Mueller, J Robert Coleman, Dimitris Papamichail, Charles B Ward, Anjaruwee Nimnual, Bruce Fletcher, Steven Skiena, and Eckard Wimmer. Live attenuated influenza virus vaccines by computer-aided rational design. *Nature Biotechnology*, 28(7):723–726, 2010.
- [MOR<sup>+</sup>88] Bartlett W Mel, Stephen M Omohundro, Arch D Robison, Steven S Skiena, Kurt H. Thearling, Luke T. Young, and Stephen Wolfram. Tablet: personal computer of the year 2000. *Communications of the ACM*, 31(6):638–648, 1988.
- [NYC15] Anh Nguyen, Jason Yosinski, and Jeff Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*, pages 427–436. IEEE, 2015.
- [O’N16] Cathy O’Neil. *Weapons of Math Destruction: How big data increases inequality and threatens democracy*. Crown Publishing Group, 2016.
- [O’R01] Joseph O’Rourke. *Computational Geometry in C*. Cambridge University Press, New York, 2nd edition, 2001.
- [Pad15] Sydney Padua. *The Thrilling Adventures of Lovelace and Babbage: The (mostly) true story of the first computer*. Penguin, 2015.
- [PaRS14] Bryan Perozzi, Rami al Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 701–710. ACM, 2014.
- [PFTV07] William Press, Brian Flannery, Saul Teukolsky, and William T. Vetterling. *Numerical Recipes: The art of scientific computing*. Cambridge University Press, 3rd edition, 2007.
- [PSM14] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, 2014.
- [RD01] Ed Reingold and Nachum Dershowitz. *Calendrical Calculations: The Millennium Edition*. Cambridge University Press, New York, 2001.
- [RLOW15] Sandy Ryza, Uri Laserson, Sean Owen, and Josh Wills. *Advanced Analytics with Spark: Patterns for Learning from Data at Scale*. O’Reilly Media, Inc., 2015.
- [Sam05] H. Samet. Multidimensional spatial data structures. In D. Mehta and S. Sahni, editors, *Handbook of Data Structures and Applications*, pages 16:1–16:29. Chapman and Hall/CRC, 2005.

- [Sam06] Hanan Samet. *Foundations of Multidimensional and Metric Data Structures*. Morgan Kaufmann, 2006.
- [SAMS97] George N Sazaklis, Esther M Arkin, Joseph SB Mitchell, and Steven S Skiena. Geometric decision trees for optical character recognition. In *Proceedings of the 13th Annual Symposium on Computational Geometry*, pages 394–396. ACM, 1997.
- [SF12] Gail M. Sullivan and Richard Feinn. Using effect size: or why the  $p$  value is not enough. *J. Graduate Medical Education*, 4:279282, 2012.
- [Sil12] Nate Silver. *The Signal and the Noise: Why so many predictions fail-but some don't*. Penguin, 2012.
- [Ski01] S. Skiena. *Calculated Bets: Computers, Gambling, and Mathematical Modeling to Win*. Cambridge University Press, New York, 2001.
- [Ski08] S. Skiena. *The Algorithm Design Manual*. Springer-Verlag, London, second edition, 2008.
- [Ski12] Steven Skiena. Redesigning viral genomes. *Computer*, 45(3):0047–53, 2012.
- [SMB<sup>+</sup>99] Arthur G Stephenson, Daniel R Mulville, Frank H Bauer, Greg A Dukeman, Peter Norvig, Lia S LaPiana, Peter J Rutledge, David Folta, and Robert Sackheim. Mars climate orbiter mishap investigation board phase i report. *NASA, Washington, DC*, page 44, 1999.
- [SRS<sup>+</sup>14] Paolo Santi, Giovanni Resta, Michael Szell, Stanislav Sobolevsky, Steven H Strogatz, and Carlo Ratti. Quantifying the benefits of vehicle pooling with shareability networks. *Proceedings of the National Academy of Sciences*, 111(37):13290–13294, 2014.
- [SS15] Oleksii Starov and Steven Skiena. GIS technology supports taxi tip prediction. Esri Map Book, 2014 User Conference, July 14-17, San Diego, 2015.
- [Str11] Gilbert Strang. *Introduction to Linear Algebra*. Wellesley-Cambridge Press, 2011.
- [Sur05] James Surowiecki. *The wisdom of crowds*. Anchor, 2005.
- [SW13] Steven S. Skiena and Charles B. Ward. *Who's Bigger?: Where Historical Figures Really Rank*. Cambridge University Press, 2013.
- [Tij12] Henk Tijms. *Understanding Probability*. Cambridge University Press, 2012.
- [Tuc88] Alan Tucker. *A Unified Introduction to Linear Algebra: Models, methods, and theory*. Macmillan, 1988.
- [Tuf83] Edward R Tufte. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT, 1983.
- [Tuf90] Edward R Tufte. *Envisioning Information*. Graphics Press, Cheshire, CT, 1990.
- [Tuf97] Edward R Tufte. *Visual Explanations*. Graphics Press, Cheshire, CT, 1997.
- [VAMM<sup>+</sup>08] Luis Von Ahn, Benjamin Maurer, Colin McMillen, David Abraham, and Manuel Blum. recaptcha: Human-based character recognition via web security measures. *Science*, 321(5895):1465–1468, 2008.

- [Vig15] Tyler Vigen. *Spurious Correlations*. Hatchette Books, 2015.
- [Wat16] Thayer Watkins. Arrow’s impossibility theorem for aggregating individual preferences into social preferences. <http://www.sjsu.edu/faculty/watkins/arrow.htm>, 2016.
- [Wea82] Warren Weaver. *Lady Luck*. Dover Publications, 1982.
- [Wei05] Sanford Weisberg. *Applied linear regression*, volume 528. Wiley, 2005.
- [Wes00] Doug West. *Introduction to Graph Theory*. Prentice-Hall, Englewood Cliffs NJ, second edition, 2000.
- [Whe13] Charles Wheelan. *Naked Statistics: Stripping the dread from the data*. WW Norton & Company, 2013.
- [ZS09] Wenbin Zhang and Steven Skiena. Improving movie gross prediction through news analysis. In *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology-Volume 01*, pages 301–304. IEEE Computer Society, 2009.
- [ZS10] Wenbin Zhang and Steven Skiena. Trading strategies to exploit blog and news sentiment. In *Proc. Int. Conf. Weblogs and Social Media (ICWSM)*, 2010.