References

1. ACEMS: The Australian research council (arc) centre of excellence for mathematical and statistical frontiers (2014). URL www.acems.org.au/
2. Agarwal, R., Dhar, V.: Editorial-Big data, data science, and analytics: The opportunity and challenge for IS research. Information Systems Research **25**(3), 443–448 (2014)
3. Agency, X.N.: The 13th five-year plan for the national economic and social development of the peoples' republic of China (2016). URL http://news.xinhuanet.com/politics/2016lh/2016-03/17/c_1118366322.htm
4. AGIMO: AGIMO big data strategy - issues paper (2013). URL www.finance.gov.au/files/2013/03/Big-Data-Strategy-Issues-Paper1.pdf
5. Anderson, P.E., Bowring, J.F., McCauley, R., Pothering, G., Starr, C.W.: An undergraduate degree in data science: Curriculum and a decade of implementation experience. In: Computer Science Education: Proceedings of the 45th ACM Technical Symposium (SIGCSE'14), pp. 145–150 (2014)
6. Anderson, P.E., Turner, C., Dierksheide, J., McCauley, R.: An extensible online environment for teaching data science concepts through gamification. In: 2014 IEEE Frontiers in Education Conference (FIE), pp. 1–8 (2014)
7. Anya, O., Moore, B., Kieliszewski, C., Maglio, P., Anderson, L.: Understanding the practice of discovery in enterprise big data science: An agent-based approach. In: 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, vol. 3, pp. 882–889 (2015)
8. Apache: Apache opennlp (2016). URL https://opennlp.apache.org/
9. Apache: Apache spark mllib (2016). URL http://spark.apache.org/mllib/
10. Apache: The apache software foundation (2018). URL https://www.apache.org/
11. ARC: Codes and guidelines | Australian research council (2017). URL www.arc.gov.au/codes-and-guidelines
12. ASA: ASA views on data science (2015). URL http://magazine.amstat.org/?s=data+science&x=0&y=0
13. ASA: Ethical guidelines for statistical practice, American statistical association (2016). URL https://www.certifiedanalytics.org/ethics.php
14. AU: Data-matching program (1990). URL http://www.comlaw.gov.au/Series/C2004A04095
15. AU: Declaration of open government (2010). URL http://agimo.gov.au/2010/07/16/declaration-of-open-government/
16. AU: Attorney-General's department (2013). http://www.attorneygeneral.gov.au/Mediareleases/Pages/2013/Second%20quarter/22May2013-AustraliajoinsOpenGovernmentPartnership.aspx

L. Cao, *Data Science Thinking*, Data Analytics,
https://doi.org/10.1007/978-3-319-95092-1

17. AU: Australia big data (2016). URL http://www.finance.gov.au/big-data/
18. Auschitzky, E., Hammer, M., Rajagopaul, A.: How big data can improve manufacturing? (2016). URL http://www.mckinsey.com/business-functions/operations/our-insights/how-big-data-can-improve-manufacturing
19. Ayankoya, K., Calitz, A., Greyling, J.: Intrinsic relations between data science, big data, business analytics and datafication. ACM International Conference Proceeding Series **28**, 192–198 (2014)
20. Bailer, J., Hoer, R., Madigan, D., Montaquila, J., Wright, T.: Report of the asa workgroup on master's degrees (2012). URL http://magazine.amstat.org/wp-content/uploads/2013an/masterworkgroup.pdf
21. Barber, D.: Bayesian Reasoning and Machine Learning. Cambridge University Press (2012)
22. Batini, C., Scannapieco, M.: Data and Information Quality: Dimensions, Principles and Techniques. Springer (2016)
23. Baumer, B.: A data science course for undergraduates: Thinking with data. The American Statistician **69**(4), 334–342 (2015)
24. BBC: Facebook-Cambridge analytica data scandal (2018). URL http://www.bbc.com/news/topics/c81zyn0888lt/facebook-cambridge-analytica-data-scandal
25. BDL: Big data landscape (2016). URL www.bigdatalandscape.com
26. BDL: Big data landscape 2016 (version 3.0) (2016). URL http://mattturck.com/2016/02/01/big-data-landscape/
27. BDSS: Big data social science. URL http://bdss.psu.edu/
28. BESC: Behavioral, economic, socio-cultural computing. URL http://www.behaviorinformatics.org/
29. Beyer, M.A., Laney, D.: The importance of 'big data': A definition (2012). URL https://www.gartner.com/doc/2057415. Gartner
30. Bhardwaj, A., Bhattacherjee, S., Chavan, A., Deshp, A., Elmore, A.J., Madden, S., Parameswaran, A.: Datahub: Collaborative data science & dataset version management at scale. In: In CIDR (2015)
31. BI: Behavioral insights (2014). URL http://www.behaviouralinsights.co.uk/
32. BigML: Bigml (2016). URL https://bigml.com/
33. BIID: Beyond iid in information theory (biid) workshop (2013). URL https://sites.google.com/site/beyondiid4/biid-conference-series
34. Boccara, N.: Modeling Complex Systems. Springer (2003)
35. Bono, E.D.: Lateral Thinking: Creativity Step by Step. Harper & Row (1970)
36. Borne, K.D., Jacoby, S., Carney, K., Connolly, A., Eastman, T., Raddick, M.J., Tyson, J.A., Wallin, J.: The revolution in astronomy education: Data science for the masses (2010). URL http://arxiv.org/pdf/0909.3895v1.pdf
37. Bothun, D., Vollmer, C.A.H.: 2016 entertainment & media industry trends (2016). URL https://www.strategyand.pwc.com/media/file/2016-Entertainment-and-Media-Trends.pdf
38. Boulding, K.: Notes on the information concept. Exploration (Toronto) **6**(103-112, CP IV), 21–32 (1955)
39. Boyer, S., Gelman, B.U., Schreck, B., Veeramachaneni, K.: Data science foundry for MOOCs. In: IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp. 1–10 (2015)
40. BPMM: Business process maturity model^{TM} (bpmm^{TM}) (2008). URL http://www.omg.org/spec/BPMM/
41. Brain, G.: Tensorflow (2016). URL https://www.tensorflow.org/
42. Breiman, L.: Statistical modeling: The two cultures. Statist. Sci. **16**(3), 199–231 (2001)
43. Brockman, J.: What to Think About Machines That Think: Today's Leading Thinkers on the Age of Machine Intelligence. Harper Perennial (2015)
44. Broman, K.: Data science is statistics (2013). URL https://kbroman.wordpress.com/2013/04/05/data-science-is-statistics/

45. Brown, G.: Review of education in mathematics, data science and quantitative disciplines: Report to the group of eight universities (2009). URL https://go8.edu.au/publication/go8-review-education-mathematics-data-science-and-quantitative-disciplines
46. Burnes, J.: 10 steps to improve your communication skills. URL https://www.aim.com.au/blog/10-steps-improve-your-communication-skills
47. Burtch, L.: The burtch works study: Salaries of data scientists (2014). URL http://www.burtchworks.com/files/2014/07/Burtch-Works-Study_DS_final.pdf
48. Bussaban, K., Waraporn, P.: Preparing undergraduate students majoring in computer science and mathematics with data science perspectives and awareness in the age of big data. In: 7th World Conference on Educational Sciences, vol. 197, pp. 1443–1446 (2015)
49. Bynum, T.: Computer and information ethics. In: The Stanford encyclopedia of philosophy (ed. Zalta EN) (2015). URL See http://plato.stanford.edu/archives/win2015/entries/ethics-computer/
50. CA: Canada capitalizing on big data (2016). URL http://www.sshrc-crsh.gc.ca/news_room-salle_de_presse/latest_news-nouvelles_recentes/big_data_consultation-donnees_massives_consultation-eng.aspx
51. Campbell, J.D., Molines, K., Swarth, C.: Data mining for ecological field research: Lessons learned from amphibian and reptile activity analysis. In: NSF Symposium on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation (2007)
52. Campbell, J.: *Environmental Informatics*. http://www.environmentalinformatics.com/Bibliography.php
53. Cao, L.: Data Mining and Multi-agent Integration (edited). Springer (2009)
54. Cao, L.: Domain driven data mining: Challenges and prospects. IEEE Trans. on Knowledge and Data Engineering **22**(6), 755–769 (2010)
55. Cao, L.: In-depth behavior understanding and use: The behavior informatics approach. Information Science **180**(17), 3067–3085 (2010)
56. Cao, L.: Strategic recommendations on advanced data industry and services for the yanhuang science and technology park (2011)
57. Cao, L.: Actionable knowledge discovery and delivery. Wiley Interdisc. Rew.: Data Mining and Knowledge Discovery **2**(2), 149–163 (2012)
58. Cao, L.: Social security and social welfare data mining: An overview. IEEE Trans. Systems, Man, and Cybernetics, Part C **42**(6), 837–853 (2012)
59. Cao, L.: Combined mining: Analyzing object and pattern relations for discovering and constructing complex but actionable patterns. WIREs Data Mining and Knowledge Discovery **3**(2), 140–155 (2013)
60. Cao, L.: Non-iidness learning in behavioral and social data. The Computer Journal **57**(9), 1358–1370 (2014)
61. Cao, L.: Coupling learning of complex interactions. J. Information Processing and Management **51**(2), 167–186 (2015)
62. Cao, L.: Metasynthetic Computing and Engineering of Complex Systems. Springer (2015)
63. Cao, L.: Data science: A comprehensive overview. Submitted to ACM Computing Survey pp. 1–37 (2016)
64. Cao, L.: Data science: Challenges and directions (2016). Technical Report, UTS Advanced Analytics Institute
65. Cao, L.: Data science: Nature and pitfalls (2016). Technical Report, UTS Advanced Analytics Institute
66. Cao, L.: Data science: Profession and education (2016). Technical Report, UTS Advanced Analytics Institute
67. Cao, L.: Data Science: Techniques and Applications (2018)
68. Cao, L.: Data Science Thinking: The Next Scientific, Technological and Economic Revolution. Springer (2018)
69. Cao, L., Dai, R.: Open Complex Intelligent Systems. Post & Telecom Press (2008)

70. Cao, L., Dai, R., Zhou, M.: Metasynthesis: M-Space, M-Interaction and M-Computing for open complex giant systems. IEEE Trans. On Systems, Man, and Cybernetics–Part A **39**(5), 1007–1021 (2009)
71. Cao, L., Dong, X., Zheng, Z.: e-NSP: Efficient negative sequential pattern mining. Artificial Intelligence **235**, 156–182 (2016)
72. Cao, L., (Eds), P.S.Y.: Behavior Computing: Modeling, Analysis, Mining and Decision. Springer (2012)
73. Cao, L., Gorodetsky, V., Mitkas, P.A.: Agent mining: The synergy of agents and data mining. IEEE Intelligent Systems **24**(3), 64–72 (2009)
74. Cao, L., Ou, Y., Yu, P.S.: Coupled behavior analysis with applications. IEEE Trans. on Knowledge and Data Engineering **24**(8), 1378–1392 (2012)
75. Cao, L., Weiss, G., Yu, P.S.: A brief introduction to agent mining. Autonomous Agents and Multi-Agent Systems **25**(3), 419–424 (2012)
76. Cao, L., Yu, P.S., Kumar, V.: Nonoccurring behavior analytics: A new area. IEEE Intelligent Systems **30**(6), 4–11 (2015)
77. Cao, L., Yu, P.S., Zhang, C., Zhao, Y.: Domain Driven Data Mining. Springer (2010)
78. Cao, L., Zhang, C.: Knowledge actionability: Satisfying technical and business interestingness. International Journal of Business Intelligence and Data Mining **2**(4), 496–514 (2007)
79. Cao, L., Zhao, Y., Zhang, C.: Mining impact-targeted activity patterns in imbalanced data. IEEE Trans. on Knowledge and Data Engineering **20**(8), 1053–1066 (2008)
80. Cao, W., Cao, L.: Financial crisis forecasting via coupled market state analysis. IEEE Intelligent Systems **30**(2), 18–25 (2015)
81. Capterra: Top project management tools (2016). URL http://www.capterra.com/project-management-software/
82. Capterra: Top reporting software products (2016). URL http://www.capterra.com/reporting-software/
83. Casey, E.: The growing importance of data science in digital investigations. Digital Investigation **14**, A1–A2 (2015)
84. CBDIO: China big data industrial observation (2016). URL www.cbdio.com
85. CCF-BDTF: China computer federation task force on big data (2013). URL http://www.bigdataforum.org.cn/
86. Ceglar, A., Roddick, J.: Association mining. ACM Computing Surveys **38**(2), 5 (2006)
87. Chambers, J.M.: Greater or lesser statistics: A choice for future research. Statistics and Computing **3**(4), 182–184 (1993)
88. Chandrasekaran, S.: Becoming a data scientist (2013). URL http://nirvacana.com/thoughts/becoming-a-data-scientist/
89. Chawla, S., Hartline, J., Nekipelov, D.: Mechanism design for data science. In: Economics and computation: Proceedings of the Fifteenth ACM Conference, pp. 711–712 (2014)
90. Chemuturi, M.: Mastering Software Quality Assurance: Best Practices, Tools and Techniques for Software Developers. J. Ross Publishing (2010)
91. Chen, H., Chiang, R.H.L., Storey, V.C.: Business intelligence and analytics: From big data to big impact. MIS Quarterly **36**(4), 1165–1188 (2012)
92. Chen, Z., Liu, B.: Lifelong Machine Learning. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool (2016)
93. Clancy, T.R., Bowles, K.H., Gelinas, L., Androwich, I., Delaney, C., Matney, S., Sensmeier, J., Warren, J., Welton, J., Westra, B.: A call to action: Engage in big data science. Nursing Outlook **62**(1), 64–65 (2014)
94. Classcentral: Data science and big data | free online courses (2016). URL https://www.class-central.com/subject/data-science
95. Clauset, A., Larremore, D.B., Sinatra, R.: Data-driven predictions in the science of science. Science **355**, 477–480 (2017)
96. Clay, K.: Ces 2013: The year of the quantified self? (2013). URL http://www.forbes.com/sites/kellyclay/2013/01/06/ces-2013-the-year-of-the-quantified-self/#4cf4d2b55e74

97. Cleveland, W.S.: Data science: An action plan for expanding the technical areas of the field of statistics. International Statistical Review **69**(1), 21–26 (2001). doi:10.1111/j.1751-5823.2001.tb00477.x. URL http://dx.doi.org/10.1111/j.1751-5823.2001.tb00477.x
98. CMIST: China will establish a series of national labs (2016). URL http://news.sciencenet.cn/htmlnews/2016/4/344404.shtm
99. CNSF: National natural science foundation of China (2015). URL http://www.nsfc.gov.cn/
100. Cohen, R., Havlin, S.: Complex Networks: Structure, Robustness and Function, 1st Edition. Cambridge University Press (2010)
101. Commission, E.: Commission urges governments to embrace potential of big data (2014). URL www.europa.eu/rapid/press-release_IP-14-769_en.htm
102. Commission, E.: Towards a thriving data-driven economy (2014). URL https://ec.europa.eu/digital-single-market/en/towards-thriving-data-driven-economy
103. Coursera: Coursera (2016). URL www.coursera.org/data-science
104. CRISP-DM: CRISP-DM (2016). URL www.sv-europe.com/crisp-dm-methodology
105. Crowston, K., Qin, J.: A capability maturity model for scientific data management: Evidence from the literature. In: Proceedings of the American Society for Information Science and Technology, vol. 48, pp. 1–9 (2011)
106. CSC: Big data universe beginning to explode (2012). URL http://www.csc.com/insights/flxwd/78931-big_data_growth_just_beginning_to_explode
107. CSNSTC: Harnessing the power of digital data for science and society (2009). URL https://www.nitrd.gov/About/Harnessing_Power_Web.pdf. Report of the Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council
108. CSS: Computational social science. URL http://www.gesis.org/en/institute/gesis-scientific-departments/computational-social-science/
109. Csurka, G.: Domain adaptation for visual applications: A comprehensive survey. CoRR **abs/1702.05374** (2017). URL http://arxiv.org/abs/1702.05374
110. CTM: Defining critical thinking (1987). URL https://www.criticalthinking.org/pages/defining-critical-thinking/766
111. Cummings, T., Jones, Y.: Creating actionable knowledge. URL http://meetings.aomonline.org/2004/theme.htm
112. Cuzzocrea, A., Gaber, M.M.: Data science and distributed intelligence: Recent developments and future insights. Studies in Computational Intelligence **446**, 139–147 (2013)
113. DABS: Data analytics book series (2016). URL http://www.springer.com/series/15063
114. Dai, R., Wang, J., Tian, J.: Metasynthesis of Intelligent Systems. Zhejiang Sci. Technol. Press, Hangzhou, China (1995)
115. DARPA: DARPA xdata program (2016). URL www.darpa.mil/program/xdata
116. Data61: Data61 (2016). URL https://www.data61.csiro.au/
117. DataRobot: Datarobot (2016). URL https://www.datarobot.com/
118. Datasciences.org: Datasciences.org (2005). URL www.datasciences.org
119. Dataversity: How data scientists can improve communications skills. URL http://www.dataversity.net/how-data-scientists-can-improve-communications-skills/
120. Daumé III, H., Marcu, D.: Domain adaptation for statistical classifiers. J. Artif. Int. Res. **26**(1), 101–126 (2006)
121. Davenport, T.H., Patil, D.: Data scientist: The sexiest job of the 21st century. Harvard Business Review pp. 70–76 (2012)
122. Davis, J.: 10 programming languages and tools data scientists used (2016). URL http://www.informationweek.com/devops/programming-languages/10-programming-languages-and-tools-data-scientists-use-now/d/d-id/1326034
123. Deeplearning: Deeplearning (2016). URL www.deeplearning.net/
124. Deo, N.: Graph Theory with Applications to Engineering and Computer Science (Prentice Hall Series in Automatic Computation). Prentice-Hall, Upper Saddle River, NJ, USA (1974)

125. Desale, D.: Top 30 social network analysis and visualization tools (2015). URL http://www.kdnuggets.com/2015/06/top-30-social-network-analysis-visualization-tools.html
126. Dhar, V.: Data science and prediction. Communications of the ACM **56**(12), 64–73 (2013)
127. Dierick, H.A., Gabbiani, F.: Drosophila neurobiology: No escape from 'big data' science. Current Biology **25**(14), 606–608 (2015)
128. Diggle, P.J.: Statistics: A data science for the 21st century. Journal of the Royal Statistical Society: Series A (Statistics in Society) **178**(4), 793–813 (2015)
129. Donoho, D.: 50 years of data science (2015). URL http://courses.csail.mit.edu/18.337/2015/docs/50YearsDataScience.pdf
130. Dorr, B.J., Greenberg, C.S., Fontana, P., Przybocki, M.A., Bras, M.L., Ploehn, C.A., Aulov, O., Michel, M., Golden, E.J., Chang, W.: The NIST data science initiative. In: 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp. 1–10 (2015)
131. Dowden, B.H.: Logical Reasoning. Philosophy Department, California State University Sacramento (2017). URL http://www.csus.edu/indiv/d/dowdenb/4/logical-reasoning.pdf
132. Drew, C.: Data science ethics in government. Phil. Trans. R. Soc. A **374** (2016)
133. DSA: Data science association (2016). URL http://www.datascienceassn.org/
134. DSA: Data science code of professional conduct, data science association (2016). URL http://www.datascienceassn.org/code-of-conduct.html
135. DSAA: IEEE/ACM/ASA international conference on data science and advanced analytics (2014). URL www.dsaa.co
136. DSC: College & university data science degrees (2016). URL http://datascience.community/colleges (accessed on 16 April 2016.)
137. DSC: The data science community (2016). URL http://datasciencebe.com/
138. DSCentral: Data science central (2016). URL http://www.datasciencecentral.com/
139. DSE: Data science and engineering (2015). URL http://link.springer.com/journal/41019
140. DSJ: Data science journal (2014). URL www.datascience.codata.org
141. DSKD: Data science and knowledge discovery lab, university of technology Sydney (2007). URL http://www.uts.edu.au/research-and-teaching/our-research/quantum-computation-and-intelligent-systems/data-sciences-and
142. DSSG: Data science for social good fellowship. URL https://dssg.uchicago.edu/
143. Duncan, D.E.: Experimental Man: What One Man's Body Reveals about His Future, Your Health, and Our Toxic World. New York: Wiley & Sons (2009)
144. Dunleavy, P.: The social science of human-dominated & human-influenced systems: Annual lecture report (2014). URL http://www.acss.org.uk/news/annual-lecture-report/
145. van Dyk, D., Fuentes, M., Jordan, M.I., Newton, M., Ray, B.K., Lang, D.T., Wickham, H.: ASA statement on the role of statistics in data science (2015). URL http://magazine.amstat.org/blog/2015/10/01/asa-statement-on-the-role-of-statistics-in-data-science/
146. (Ed.), M.P.: Similarity-based pattern analysis and recognition. Springer (2013)
147. Edx: EDX courses (2016). URL https://www.edx.org/course?search_query=data+science
148. Ehling, M., Korner, T.: Handbook on Data Quality Assessment Methods and Tools (eds.). EUROSTAT, Wiesbaden (2007)
149. Elder, L., Paul, R.: The Thinker's Guide to Analytic Thinking: How to Take Thinking Apart and What to Look for When You Do (2nd ed.). Foundation for Critical Thinking (2007)
150. EMC: Data science revealed: A Data-Driven glimpse into the burgeoning new field (2011). URL www.emc.com/collateral/about/news/emc-data-science-study-wp.pdf
151. EPJDS: EPJ data science (2012). URL http://epjdatascience.springeropen.com/
152. ESF: Research integrity : European science foundation (2016). URL www.archives.esf.org/coordinating-research/mo-fora/research-integrity.html
153. ESRC: Big data and the social sciences: A perspective from the esrc. URL https://www2.warwick.ac.uk/fac/soc/economics/research/centres/cage/events/conferences/peuk/peter_elias_big_data_and_the_social_sciences_pe_final.pdf
154. EU-DSA: The European data science academy (2016). URL www.edsa-project.eu
155. EU-OD: The European union open data portal (2016). URL https://open-data.europa.eu/
156. Facebook: Facebook data (2016). URL https://www.facebook.com/careers/teams/data/

157. Faghmous, J.H., Kumar, V.: A big data guide to understanding climate change: The case for theory-guided data science. Big Data **2**(3), 155–163 (2014)
158. Fairfielda, J., Shteina, H.: Big data, big problems: Emerging issues in the ethics of data science and journalism. Journal of Mass Media Ethics **29**(1), 38–51 (2014)
159. Faris, J., Kolker, E., Szalay, A., Bradlow, L., Deelman, E., Feng, W., Qiu, J., Russell, D., Stewart, E., Kolker, E.: Communication and data-intensive science in the beginning of the 21st century. A Journal of Integrative Biology **15**(4), 213–215 (2011)
160. Fawcett, T.: Mining the quantified self: Personal knowledge discovery as a challenge for data science. Big Data **3**(4), 249–266 (2016)
161. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery in databases. AI Magazine **17**(3), 37–54 (1996)
162. Feng, J.X., Kusiak, A.: Data mining applications in engineering design, manufacturing and logistics. International Journal of Production Research **44**(14), 2689–2694 (2007)
163. Fichman, P., Rosenbaum, H.: Social Informatics: Past, Present and Future. Cambridge Scholars Publishing (2014)
164. Finzer, W.: The data science education dilemma. Technology Innovations in Statistics Education **7**(2) (2013). URL http://escholarship.org/uc/item/7gv0q9dc#page-1
165. Floreano, D., Mattiussi, C.: Bio-Inspired Artificial Intelligence: Theories, Methods, and Technologies. The MIT Press (2008)
166. Floridi, L.: The ethics of information. Oxford University Press (2013)
167. Floridi, L., Taddeo, M.: What is data ethics. Phil. Trans. R. Soc. A **374**(2083) (2016)
168. Forbes: The world's biggest public companies (2016). URL https://www.forbes.com/global2000/list
169. Fox, G., Maini, S., Rosenbaum, H., Wild, D.J.: Data science and online education. In: 2015 IEEE 7th International Conference on Cloud Computing Technology and Science (CloudCom), pp. 582–587 (2015)
170. Freedman, D., Pisani, R., Purves, R.: Statistics (4th edn.). W. W. Norton (2007)
171. G. Szkely, e.a.: Measuring and testing independence by correlation of distances. Annals of Statistics **35**(6), 2769–2794 (2007)
172. Galetto, M.: Top 50 data science resources (2016). URL http://www.ngdata.com/top-data-science-resources/?
173. Galin, D.: Software Quality Assurance: From Theory to Implementation. Pearson (2003)
174. Ganis, M., Kohirkar, A.: Social Media Analytics: Techniques and Insights for Extracting Business Value Out of Social Media. IBM Press (2015)
175. Ganiz, M., George, C., Pottenger, W.: Higher order naive bayes: A novel non-iid approach to text classification. IEEE Transactions on Knowledge and Data Engineering **23**(7), 1022–1034 (2011)
176. Gavrilovski, A., Jimenez, H., Mavris, D.N., Rao, A.H., Shin, S., Hwang, I., Marais, K.: Challenges and opportunities in flight data mining: A review of the state of the art. In: AIAA Infotech Aerospace, AIAA SciTech Forum, (AIAA 2016-0923), pp. 1–66 (2016)
177. Geczy, P.: Big data characteristics. The Macrotheme Review **3**(6), 94–104 (2014)
178. Gentle, J.E.: Computational Statistics. Springer Publishing Company (2009)
179. GEO: Gene expression omnibus (2016). URL http://www.ncbi.nlm.nih.gov/geo/
180. George, G., Osinga, E., Lavie, D., Scott, B.: Big data and data science methods for management research. Academy of Management Journal **59**(5), 1493–1507 (2016). URL https://aom.org/uploadedFiles/Publications/AMJ/Oct_2016_FTE.pdf
181. Ghodke, D.: Bye bye 2015: What lies ahead for bi? (2015). URL www.ciol.com/bye-bye-2015-what-lies-ahead-for-bi/
182. Github: Data science colleges (2016). URL https://github.com/ryanswanstrom/awesome-datascience-colleges. (retrieved on 4 April 2016)
183. Github: List of recommender systems (2016). URL https://github.com/grahamjenson/list_of_recommender_systems
184. Gitub, D.: Data science for social good. URL https://github.com/dssg

185. Globalsecurity: Worldwide military command and control system (1996). URL http://www.globalsecurity.org/wmd/systems/wwmccs.htm
186. Gold, M., McClarren, R., Gaughan, C.: The lessons oscar taught us: Data science and media & entertainment. Big Data **1**(2), 105–109 (2013)
187. Golge, E.: Brief history of machine learning, a blog from a human-engineer-being. retrieved 21 march 2017 (2017). URL http://www.erogol.com/brief-history-machine-learning/
188. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press (2016). URL http://www.deeplearningbook.org
189. Google: Deepmind (2016). URL https://deepmind.com/
190. Google: Google bigquery and cloud platform (2016). URL https://cloud.google.com/bigquery/
191. Google: Google cloud prediction api (2016). URL https://cloud.google.com/prediction/docs/
192. Google: Google online open education (2016). URL https://www.google.com/edu/openonline/
193. Google: Google trends (2016). URL https://www.google.com.au/trends/explore#q=data%20\science%2C%20data%20analytics%2C%20big%20data%2C%20data%20analysis%2C%20\advanced%20analytics&cmpt=q&tz=Etc%2FGMT-11. Retrieved on 14 November 2016
194. Google: Open mobile data (2016). URL https://console.developers.google.com/storage/\browser/openmobiledata_public/
195. Government, B.M.: Beijing big data and cloud computing development action plan (2016). URL http://zhengwu.beijing.gov.cn/gh/dt/t1445533.htm
196. Government, C.: China big data (2015). URL http://www.gov.cn/zhengce/content/2015-09/05/content_10137.htm
197. Graham, M.J.: The art of data science. In: Astrostatistics and Data Mining, Volume 2 of the series Springer Series in Astrostatistics, pp. 47–59 (2012)
198. Gray, J.: escience – a transformed scientific method (2007). URL http://research.microsoft.com/en-us/um/people/gray/talks/NRC-CSTB_eScience.ppt
199. GTD: Global terrorism database (2016). URL https://www.start.umd.edu/gtd/
200. Gupta, A., Cecen, A., Goyal, S., Singh, A.K., Kalidindi, S.R.: Structure-property linkages using a data science approach: Application to a non-metallic inclusion/steel composite system. Acta Mater **91**, 239–254 (2015)
201. H. Lu, e.a.: Beyond intratransaction association analysis. ACM Transactions on Information Systems **18**(4), 423–454 (2000)
202. Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques, 3rd edn. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2011)
203. Hand, D.J.: Statistics and computing: The genesis of data science. Statistics and Computing **25**(4), 705–711 (2015)
204. Hardin: Github (2016). URL www.hardin47.github.io/DataSciStatsMaterials/
205. Hardin, J., Hoerl, R., Horton, N.J., Nolan, D.: Data science in statistics curricula: Preparing students to "think with data". The American Statistician **69**(4), 343–353 (2015)
206. Harris, H., Murphy, S., Vaisman, M.: Analyzing the Analyzers: An Introspective Survey of Data Scientists and Their Work. O'Reilly Media (2013)
207. Hastie, T., Tibshirani, R., Friedman, J.H.: The elements of statistical learning: data mining, inference, and prediction, 2nd Edition. Springer (2009)
208. Hazena, B.T., Booneb, C.A., Ezellc, J.D., Jones-Farmer, L.A.: Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. International Journal of Production Economics **154**, 72–80 (2014)
209. Hey, T., Tansley, S., (Eds.), K.T.: The Fourth Paradigm: Data-Intensive Scientific Discovery. Microsoft Research (2009). URL http://research.microsoft.com/en-us/collaboration/fourthparadigm/
210. Hey, T., Trefethen, A.: The Data Deluge: An e-Science Perspective, pp. 809–824. John Wiley & Sons (2003)

211. HLSG: Final report of the high level expert group on scientific data. In: Riding the wave: How Europe can gain from the rising tide of scientific data (2010). URL http://ec.europa.eu/information_society/newsroom/cf/document.cfm?action=display&doc_id=707
212. HLSG: An rda europe report. In: The Data Harvest: How sharing research data can yield knowledge, jobs and growth (2014). URL http://www.e-nformation.ro/wp-content/uploads/2014/12/TheDataHarvestReport_-Final.pdf
213. Hofmann, T., Schölkopf, B., Smola, A.J.: Kernel methods in machine learning (2008)
214. Horizon: European commission horizon 2020 big data private public partnership (2014). URL http://ec.europa.eu/programmes/horizon2020/en/h2020-section/information-and-communication-technologies
215. Horton, N.J., Baumer, B.S., Wickham, H.: Setting the stage for data science: Integration of data management skills in introductory and second courses in statistics. arXiv preprint arXiv:1502.00318 (2015)
216. Huber, P.J.: Data Analysis: What Can Be Learned From the Past 50 Years. John Wiley & Sons (2011)
217. IASC: International association for statistical computing (1977). URL http://www.iasc-isi.org/
218. IBM: The value of analytics in healthcare. URL https://www.ibm.com/services/us/gbs/thoughtleadership/ibv-healthcare-analytics.html
219. IBM: Capitalizing on complexity (2010). URL http://www-935.ibm.com/services/us/ceo/ceostudy2010/multimedia.html
220. IBM: Ibm analytics and big data (2016). URL http://www.ibm.com/analytics/us/en/ or http://www-01.ibm.com/software/data/bigdata/
221. IBM: Ibm watson (2016). URL https://www.ibm.com/watson/
222. IBM: What is a data scientist? (2016). URL http://www-01.ibm.com/software/data/infosphere/data-scientist/
223. IEEEBD: IEEE big data initiative (2014). URL http://bigdata.ieee.org/
224. IEMSS: The international environmental modelling & software society. URL http://www.iemss.org/society/
225. IFSC-96: Data science, classification, and related methods. In: IFSC-96 (1996). URL http://d-nb.info/955715512/04
226. IJDS: International journal of data science (2016). URL http://www.inderscience.com/jhome.php?jcode=ijds
227. IJRDS: International journal of research on data science (2017). URL http://www.sciencepublishinggroup.com/journal/index?journalid=310
228. INFORMS: Informs code of ethics for certified analytics professionals. URL https://www.certifiedanalytics.org/ethics.php
229. INFORMS: Candidate handbook (2014). URL https://www.informs.org/Certification-Continuing-Ed/Analytics-Certification/Candidate-Handbook
230. INFORMS: Institute for operations research and the management sciences (2016). URL https://www.informs.org/
231. Iwata, S.: Scientific "agenda" of data science. Data Science Journal **7**(5), 54–56 (2008)
232. J. Hair, e.a.: Multivariate data analysis (7th Edition). Prentice Hall (2009)
233. Jagadish, H., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J.M., Ramakrishnan, R., Shahabi, C.: Big data and its technical challenges. Communications of the ACM **57**(7), 86–94 (2014)
234. Jagadish, H.V.: Big data and science: Myths and reality. Big Data Research **2**(2), 49–52 (2015)
235. JDS: Journal of data science (2002). URL http://www.jds-online.com/
236. JDSA: International journal of data science and analytics (JDSA) (2015). URL http://www.springer.com/41060
237. JFDS: The journal of finance and data science (2016). URL http://www.keaipublishing.com/en/journals/the-journal-of-finance-and-data-science/

238. Johnstone, I., Roberts, F.: Data science at nsf (2014). URL http://www.nsf.gov/attachments/130849/public/Stodden-StatsNSF.pdf
239. Jones, R.P.: Foundations of Critical Thinking. Cengage Learning (2000)
240. Josephson, R., J. & G. Josephson, S.: Abductive Inference: Computation, Philosophy, Technology. Cambridge University Press, New York & Cambridge (1994)
241. Kaggle: Kaggle competition data (2016). URL https://www.kaggle.com/competitions
242. Kalidindi, S.R.: Data science and cyberinfrastructure: critical enablers for accelerated development of hierarchical materials. International Materials Reviews **60**(3), 150–168 (2015)
243. Kan, S.H.: Metrics and Models in Software Quality Engineering, 2nd Edition. Addison-Wesley Professional (2002)
244. Kanter, J.M., Veeramachaneni, K.: Deep feature synthesis: Towards automating data science endeavors. In: 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp. 1–10 (2015)
245. KDD89: IJCAI-89 workshop on knowledge discovery in databases (1989). URL http://www.kdnuggets.com/meetings/kdd89/index.html
246. KDnuggets: Visualization software (2015). URL http://www.kdnuggets.com/software/visualization.html
247. Kdnuggets: Kdnuggets (2016). URL http://www.kdnuggets.com/
248. KDnuggets: Software suites/platforms for analytics, data mining, & data science (2017). URL http://www.kdnuggets.com/software/suites.html
249. Keller, J.M., Liu, D., Fogel, D.B.: Fundamentals of Computational Intelligence: Neural Networks, Fuzzy Systems, and Evolutionary Computation. Wiley-IEEE Press (2016)
250. Kelly, K.: The quantified century. In: Quantified Self Conference (2012). URL http://quantifiedself.com/conference/Palo-Alto-2012
251. Kenett, R.S., Shmueli, G.: Information Quality: The Potential of Data and Analytics to Generate Knowledge. Wiley (2016)
252. Khan, N., Yaqoob, I., Hashem, I.A.T., et al: Big data: Survey, technologies, opportunities, and challenges. The Scientific World Journal **2014**, 18 (2014)
253. King, J., Magoulas, R.: 2015 data science salary survey (2015). URL http://duu86o6n09pv.cloudfront.net/reports/2015-data-science-salary-survey.pdf
254. Kirk, A.: Data Visualisation: A Handbook for Data Driven Design. SAGE Publications (2016)
255. Kirk, R.E.: Experimental Design: Procedures for the Behavioral Sciences (4th Edition. SAGE Publications (2012)
256. Kirkpatrick, K.: Putting the data science into journalism. Communications of the ACM **58**(5), 15–17 (2015)
257. Kohavi, R.: Mining e-commerce data: the good, the bad, and the ugly. In: SIGKDD, pp. 8–13 (2001)
258. Kohavi, R., Mason, L., Parekh, R., Zheng, Z.: Lessons and challenges from mining retail e-commerce data. Mach. Learn. **57**(1-2), 83–113 (2004)
259. Kohavi, R., Rothleder, N.J., Simoudis, E.: Emerging trends in business analytics. Communications of the ACM **45**(8), 45–48 (2002)
260. Koller, D., Friedman, N.: Probabilistic Graphical Models: Principles and Techniques. The MIT Press (2009)
261. Kramer, A., Guillory, J., Hancock, J.: Experimental evidence of massive-scale emotional contagion through social networks. Proc. Natl. Acad. Sci. **111**(24), 8788–8790 (2014)
262. Kung, S.Y.: Kernel Methods and Machine Learning. Cambridge University Press (2014)
263. Kurzweil, R.: How to Create a Mind: The Secret of Human Thought Revealed. Penguin Books (2013)
264. Lab, A.: Mlbase (2016). URL http://mlbase.org/
265. Labrinidis, A., Jagadish, H.V.: Challenges and opportunities with big data. Proceedings of the VLDB Endowment **5**(12), 2032–2033 (2012)
266. Laney, D.: 3D data management: Controlling data volume, velocity and variety (2001). Technical Report, META Group

267. Larder, B., Summerhayes, N.: Application of smiths aerospace data mining algorithms to british airways 777 and 747 fdm data (2004). URL https://flightsafety.org/files/FDM_data_mining_report.pdf
268. Lassonde, M., Candel, S., Hacker, J., Quadrio-Curzio, A., Onishi, T., Ramakrishnan, V., McNutt, M.: G7 academies' joint statements 2017, new economic growth: The role of science, technology, innovation and infrastructure (2017)
269. Lazer, D., Kennedy, R., King, G., Vespignani, A.: The parable of google flu: Traps in big data analysis. Science **343**, 1203–1205 (2014)
270. LDC: Linguistic data consortium (2016). URL https://www.ldc.upenn.edu/about
271. Lehmann, E.L., Lehmann, J.P.: Testing Statistical Hypotheses. Springer (2010)
272. Lencioni, P.: The Five Dysfunctions of a Team: A Leadership Fable. Jossey-Bass (2002)
273. Leonelli, S.: Locating ethics in data science: responsibility and accountability in global and distributed knowledge production systems. Phil. Trans. R. Soc. A **374** (2016)
274. Leuphana: Master's programme in management & data science (2017). URL http://www.leuphana.de/en/graduate-school/master/course-offerings/management-data-science.html
275. Li, R., Shengjie Wang, e.a.: Towards social user profiling: unified and discriminative influence model for inferring home locations. In: Proceedings of KDD2012, pp. 1023–1031 (2012)
276. LinkedIn: Linkedin jobs (2016). URL https://www.linkedin.com/jobs/data-scientist-jobs
277. Loshin, D.: Enterprise Knowledge Management. Morgan Kaufmann (2001)
278. Loukides, M.: The Evolution of Data Products. O'Reilly, Cambridge (2011)
279. Loukides, M.: What is data science? O'Reilly Media, Sebastopol, CA (2012). URL http://radar.oreilly.com/2010/06/what-is-data-science.html#data-scientists
280. MacKay, D.J.C.: Information Theory, Inference & Learning Algorithms. Cambridge University Press, New York (2002)
281. Manieri, A., Brewer, S., Riestra, R., Demchenko, Y., Hemmje, M., Wiktorski, T., Ferrari, T., Frey, J.: Data science professional uncovered: How the EDISON project will contribute to a widely accepted profile for data scientists. In: 2015 IEEE 7th International Conference on Cloud Computing Technology and Science (CloudCom), pp. 588–593 (2015)
282. Manieri, A., Nucci, F.S., Femminella, M., Reali, G.: Teaching Domain-Driven data science: Public-Private co-creation of Market-Driven certificate. In: 2015 IEEE 7th International Conference on Cloud Computing Technology and Science (CloudCom), pp. 569–574 (2015)
283. Matsudaira, K.: The science of managing data science. Communications of the ACM **58**(6), 44–47 (2015)
284. Mattison, R.: Data warehousing and data mining for telecommunications. Artech House (1997)
285. Mattmann, C.A.: Computing: A vision for data science. Nature **493**(7433), 473–475 (2013)
286. Mattmann, C.A.: Cultivating a research agenda for data science. Journal of Big Data **1**(1), 1–8 (2014)
287. McCartney, P.R.: Big data science. The American Journal of Maternal/Child Nursing **40**(2), 130–130 (2015)
288. McKinsey: Big data: The next frontier for innovation, competition, and productivity (2011). McKinsey Global Institute
289. Microsoft: Azure. URL www.azure.microsoft.com/
290. Miller, C.C.: Data science: The numbers of our lives. New York Times (2013). URL http://www.nytimes.com/2013/04/14/education/edlife/universities-offer-courses-in-a-hot-new-field-data-science.html?pagewanted=all&_r=0
291. Miller, K., Taddeo, M.: The ethics of information technologies. In: Library of Essays on the Ethics of Emerging Technologies (ed.). NY: Routledge (2017)
292. MIT: Data analytics in urban transportation. URL http://dusp.mit.edu/transportation/project/data-analytics-urban-transportation
293. MIT: Checklist for software quality (2011). URL http://web.mit.edu/~6.170/www/quality.html
294. Mitchell, M.: Complexity: A Guided Tour. Oxford University Press (2011)
295. Mitchell, T.: Machine Learning. McGraw Hill (1997)

296. Mittelstadt, B., Floridi, L.: The ethics of big data: current and foreseeable issues in biomedical contexts. Sci. Eng. Ethics **22**, 303–341 (2015)
297. de Moraes, R.M., Martinez, L.: Computational intelligence applications for data science. Knowledge-Based Systems **87**, 1–2 (2015)
298. Morrell, A.J.H.: Information processing 68 (ed.). In: Proceedings of IFIP Congress 1968. Edinburgh, UK (1968)
299. Murray-Rust, P.: Data-Driven science: A scientist's view. In: NSF/JISC 2007 Digital Repositories Workshop (2007). URL http://www.sis.pitt.edu/repwkshop/papers/murray.pdf
300. Naur, P.: 'datalogy', the science of data and data processes. In: Proceedings of the IFIP Congress 68, pp. 1383–1387 (1968)
301. Naur, P.: Concise Survey of Computer Methods. Studentlitteratur, Lund, Sweden (1974)
302. NCSU: Institute for advanced analytics, north carolina state university (2007). URL http://analytics.ncsu.edu/
303. NCSU: Master of science in analytics, institute for advanced analytics, north carolina state university (2007). URL http://analytics.ncsu.edu/
304. Neapolitan, R.E.: Learning Bayesian Networks. Prentice-Hall, Upper Saddle River, NJ, USA (2003)
305. Nelson, M.L.: Data-driven science: A new paradigm? EDUCAUSE Review **44**(4), 6–7 (2009)
306. von Neumann, J., Kurzweil, R.: The Computer and the Brain, 3rd Edition. Yale University Press (2012)
307. Neville, J., Jensen, D.: Relational dependency networks. The Journal of Machine Learning Research **8**, 653–692 (2007)
308. NICTA: National ict Australia (2016). URL https://www.nicta.com.au/
309. NIST: NIST text retrieval conference data (2015). URL http://trec.nist.gov/data.html
310. NSB: Long-lived digital data collections: Enabling research and education in the 21st century. In: US National Science Board (2005). URL http://www.nsf.gov/pubs/2005/nsb0540/
311. NSF: US NSF07-28. In: Cyberinfrastructure Vision for 21st Century Discovery (2007). URL http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf
312. OECD: OECD principles and guidances for access to research data from public funding (2007). URL https://www.oecd.org/sti/sci-tech/38500813.pdf
313. OECD: Data-driven innovation: Big data for growth and well-being (2015). doi:http://dx.doi.org/10.1787/9789264229358-en
314. OECD: The next production revolution: Implications for governments and business (2017). doi:http://dx.doi.org/10.1787/9789264271036-en
315. O'Leary, D.E.: Ethics for big data and analytics. IEEE Intelligent Systems **31**(4), 81–84 (2016)
316. O'Neil, C., Schutt, R.: Doing data science: Straight talk from the frontline. O'Reilly Media, Sebastopol, CA (2013)
317. OpenCV: Open source computer vision library (2016). URL www.opencv.org/
318. OPENedX: OPENedX online education platform (2016). URL https://open.edx.org/
319. O'Reilly, T.: What is web 2.0 (2005). URL http://oreilly.com/pub/a/web2/archive/what-is-web-20.html?page=3
320. Pal, S.K., Meher, S.K., Skowron, A.: Data science, big data and granular mining. Pattern Recognition Letters **67**(2), 109–112 (2015)
321. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Trans. on Knowl. and Data Eng. **22**(10), 1345–1359 (2010)
322. Patil, D.: Building Data Science Teams. O'Reilly Media (2011)
323. Patterson, K., Grenny, J.: Crucial Conversations Tools for Talking When Stakes Are High (Second Edition). McGraw-Hill Education (2011)
324. Paul, R., Elder, L.: The Thinker's Guide to Scientific Thinking Based on Critical Thinking Concepts & Principles. Foundation for Critical Thinking (2008)
325. Paulk, M.C., Curtis, B., Chrissis, M.B., Weber, C.V.: Capability maturity model version 1.1. IEEE Software **10**(4), 18–27 (1993)

326. Pearson, K.: Report on certain enteric fever inoculation statistics. Br Med J. **2**(2288), 1243–1246 (1904)
327. Peter, F., James, H.: The science of data science. Big Data **2**(2), 68–70 (2014)
328. Philip, J.C.: Computer Generated Artificial Life: A Biblical And Logical Analysis (Integrated Apologetics), 10th edition. Philip Communications (2015)
329. Pike, J.: Global command and control system (2003). URL https://fas.org/nuke/guide/usa/c3i/gccs.htm
330. Press, G.: A very short history of data science (2013). URL http://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/#61ae3ebb69fd
331. Priebe, T., Markus, S.: Business information modeling: A methodology for data-intensive projects, data science and big data governance. In: 2015 IEEE International Conference on Big Data (Big Data), pp. 2056–2065 (2015)
332. Provost, F., Fawcett, T.: Data science and its relationship to big data and Data-Driven decision making. Big Data **1**(1), 51–59 (2013)
333. Qian, X.: Revisiting issues on open complex giant systems. Pattern Recognit. Artif. Intell. **4**(1), 5–8 (1991)
334. Qian, X.: Building Systematism. ShanXi Sci. Technol Press, Taiyuan, China (2001)
335. Qian, X., Yu, J., Dai, R.: A new discipline of science-the study of open complex giant system and its methodology. Chin. J. Syst. Eng. Electron. **4**(2), 2–12 (1993)
336. Raghavan, S.N.: Data mining in e-commerce: A survey. Sadhana **30**(2 & 3), 275–289 (2005)
337. RapidMiner: Rapidminer (2016). URL https://rapidminer.com/
338. Redman, T.: Data Quality: The Field Guide. Digital Press (2001)
339. Renae, S.: Data analytics: Crunching the future. Bloomberg Businessweek (2011). September 8
340. Review, S.: Data integration and application integration solutions directory (2016). URL http://solutionsreview.com/data-integration/data-integration-solutions-directory/
341. Rifkin, J.: The Third Industrial Revolution: How Lateral Power is Transforming Energy, the Economy, and the World. Palgrave MacMillan (2011)
342. Rowley, J.: The wisdom hierarchy: representations of the DIKW hierarchy. Journal of Information and Communication Science **33**(2), 163–180 (2007)
343. Rudin, C., Dunson, D., Irizarry, R., Ji, H., Laber, E., Leek, J., McCormick, T., Rose, S., Schafer, C., van der Laan, M., Wasserman, L., Xue, L.: Discovery with data: Leveraging statistics with computer science to transform science and society (2014). URL http://www.amstat.org/policy/pdfs/BigDataStatisticsJune2014.pdf. A Working Group of the American Statistical Association
344. Russell, S.J., Norvig, P.: Artificial Intelligence: A Modern Approach, 2 edn. Pearson Education (2003)
345. SAS: Big data analytics: An assessment of demand for labour and skills, 2012-2017 (2013). URL https://www.thetechpartnership.com/globalassets/pdfs/research-2014/bigdata_report_nov14.pdf. Report. SAS/The Tech Partnership
346. SAS: Sas enterprise miner (2016). URL http://www.sas.com
347. SAS: SAS insights (2016). URL http://www.sas.com/en_us/insights.html
348. Sayama, H.: Introduction to the Modeling and Analysis of Complex Systems. Open SUNY Textbooks (2015)
349. Schadt, E., Chilukuri, S.: The role of big data in medicine (2015). URL http://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/the-role-of-big-data-in-medicine
350. Schoenherr, T., Speier-Pero, C.: Data science, predictive analytics, and big data in supply chain management: Current state and future potential. Journal of Business Logistics **36**(1), 120–132 (2015)
351. Schulmeyer, G.G., Mcmanus, J.I.: Handbook of Software Quality Assurance, 3rd Edition. Prentice Hall PTR (1998)
352. SCJ: Science council of Japan - code of conduct for scientists (2017). URL www.scj.go.jp/en/report/code.html

353. Scott, J.: Social Network Analysis (4th Edition). SAGE Publications (2017)
354. SDS: Social data science lab. URL http://socialdatalab.net/
355. Sebastian-Coleman, L.: Measuring Data Quality for Ongoing Improvement: A Data Quality Assessment Framework. Morgan Kaufmann (2013)
356. Security, C.I.: Big data strategies and actions in major countries (2015). URL http://www.cac.gov.cn/2015-07/03/c_1115812491.htm
357. Shi, C., Yu, P.S.: Heterogeneous Information Network Analysis and Applications. Springer (2017)
358. SIAM: Siam career center (2016). URL http://jobs.siam.org/home/
359. Siart, C., Kopp, S., Apel, J.: The interface between data science, research assessment and science support - highlights from the German perspective and examples from Heidelberg university. In: 2015 IIAI 4th International Congress on Advanced Applied Informatics (IIAI-AAI), pp. 472–476 (2015)
360. Silk: Data science university programs (2016). URL http://data-science-university-programs.silk.co/
361. Simovici, D.A., Djeraba, C.: Mathematical Tools for Data Mining: Set Theory, Partial Orders, Combinatorics. Springer Publishing Company (2008)
362. Siroker, D., Koomen, P.: A / B Testing: The Most Powerful Way to Turn Clicks Into Customers. Wiley (2015)
363. Smarr, L.: Quantifying your body: A how-to guide from a systems biology perspective. Biotechnology Journal **7**(8), 980–991 (2012). doi:10.1002/biot.201100495. URL http://dx.doi.org/10.1002/biot.201100495
364. Smith, F.J.: Data science as an academic discipline. Data Science Journal **5**, 163–164 (2006)
365. SMU: Living analytics research centre (2017). URL https://larc.smu.edu.sg/
366. Sobel, C., Li, P.: The Cognitive Sciences: An Interdisciplinary Approach (2nd Edition). SAGE Publications (2013)
367. Society, B.R.A.: Astronomical databases and archives. URL https://www.ras.org.uk/education-and-careers/for-everyone/126-astronomical-databases-and-archives
368. Sonnenburg, S., Raetsch, G.: Shogun (2016). URL http://www.shogun-toolbox.org/
369. SSDS: Springer series in the data sciences (2015). URL http://www.springer.com/series/13852
370. Stanford: Stanford data science initiatives, Stanford university (2014). URL https://sdsi.stanford.edu/
371. Stanton, J.: An introduction to data science (2012). URL http://surface.syr.edu/istpub/165/
372. Stevens, M.L.: An ethically ambitious higher education data science. Research & Practice in Assessment **9**, 96–97 (2014)
373. Stewart, T.R., McMillan, J.C.: Descriptive and prescriptive models for judgment and decision making: Implications for knowledge engineering. In: J.L. Mumpower, O. Renn, L.D. Phillips, V.R.R.U. (Eds.) (eds.) Expert Judgment and Expert Systems, pp. 305–320. Springer-Verlag, London (1987)
374. Stonebraker, M., Madden, S., Dubey, P.: Intel 'big data' science and technology center vision and execution plan. SIGMOD Record **42**(1), 44–49 (2013)
375. Suchma, L.: Human-Machine Reconfigurations: Plans and Situated Actions. Cambridge University Press (2006)
376. Swan, A., Brown, S.: The skills, role & career structure of data scientists & curators: Assessment of current practice & future needs. In: UK Joint Information Systems Committee (2008). Technical Report. University of Southampton
377. Swan, M.: The quantified self: Fundamental disruption in big data science and biological discovery. Big Data **1**(2), 85–99 (2013)
378. Taddeo, M., (eds.), L.F.: The ethical impact of data science. Phil. Trans. R. Soc. A **374** (2016). URL http://rsta.royalsocietypublishing.org/content/374/2083
379. Taleb, N.N.: The Black Swan: The Impact of the Highly Improbable. Random House, New York (2007)

380. Tang, L., Liu, H.: Community Detection and Mining in Social Media. Morgan and Claypool (2010)
381. Technavio: Top 10 healthcare data analytics companies (2016). URL http://www.technavio.com/blog/top-10-healthcare-data-analytics-companies
382. TFDSAA: IEEE task force on data science and advanced analytics (2013). URL http://dsaatf.dsaa.co/
383. Thrun, S., Pratt, L.e.: Learning to learn. Boston, Mass.: Kluwer Academic (1998)
384. Tilburg: Msc specialization data science: Business and governance (2017). URL https://www.tilburguniversity.edu/education/masters-programmes/data-science-business-and-governance/
385. TOBD: IEEE transactions on big data (2015). URL https://www.computer.org/web/tbd
386. Today, P.A.: 29 data preparation tools and platforms (2016). URL http://www.predictiveanalyticstoday.com/data-preparation-tools-and-platforms/
387. Tukey, J.W.: The future of data analysis. Ann. Math. Statist. **33**(1), 1–67 (1962)
388. Tukey, J.W.: Exploratory Data Analysis. Pearson (1977)
389. Tutiempo: Global climate data (2016). URL http://en.tutiempo.net/climate
390. U-Waikato: Weka (2016). URL http://www.cs.ubc.ca/labs/beta/Projects/autoweka/
391. UCI: UCI machine learning repository (2016). URL www.archive.ics.uci.edu/ml/
392. UCL: Msin105p: Critical analytical thinking (2015). URL https://www.mgmt.ucl.ac.uk/module/msin105p-critical-analytical-thinking
393. Udacity: Udacity courses (2016). URL https://www.udacity.com/courses/data-science
394. Udemy: Udemy courses (2016). URL https://www.udemy.com/courses/search/?ref=home&\src=ukw&q=data+science&lang=en
395. UK: Uk big data (2016). URL https://www.ukri.org
396. UK-HM: Uk hm government. In: Open Data White Paper: Unleashing the Potential (2012). URL http://data.gov.uk/sites/default/files/Open_data_\White_Paper.pdf
397. UK-OD: UK open data (2016). URL http://data.gov.uk/
398. UMichi: Michigan institute for data science, university of Michigan (2015). URL http://midas.umich.edu/
399. UN: United nation global pulse projects (2010). URL http://www.unglobalpulse.org/
400. Uprichard, E.: Big data, little questions? (2013). URL http://discoversociety.org/2013/10/01/focus-big-data-little-questions/
401. US National Science Foundation: Critical techniques and technologies for advancing foundations and applications of big data science & engineering (bigdata) (2015). URL https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504767
402. US National Science Foundation: Computational and data-enabled science and engineering (cds&e) (2017). URL https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504813&org=CISE&sel_org=CISE&from=fund
403. US-OD: US government open data (2016). URL https://www.data.gov/
404. USAID: Usaid recommended data quality assessment (dqa) checklist (2016). URL https://usaidlearninglab.org/sites/default/files/resource/files/201sae.pdf
405. USD2D: US national consortium for data science (2016). URL www.data2discovery.org
406. USDSC: US degree programs in analytics and data science (2016). URL http://analytics.ncsu.edu/?page_id=4184
407. USNSF: US big data research initiative (2012). URL http://www.nsf.gov/cise/news/bigdata.jsp
408. UTS: Master of analytics (research) and doctor of philosophy thesis: Analytics, Advanced Analytics Institute, University of Technology Sydney (2011). URL http://www.uts.edu.au/research-and-teaching/our-research/advanced-analytics-institute/education-and-research-opportuniti-1
409. UTSAAI: Advanced analytics institute, university of technology Sydney (2011). URL https://analytics.uts.edu.au/
410. Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer-Verlag New York, New York, USA (2000)

411. Vast: Visual analytics community (2016). URL http://vacommunity.org/HomePage
412. Veaux, R.D.D., Agarwal, M., Averett, M., Baumer, B.S., Bray, A., Bressoud, T.C., Bryant, L., Cheng, L.Z., Francis, A., Gould, R., Kim, A.Y., Kretchmar, M., Lu, Q., Moskol, A., Nolan, D., Pelayo, R., Raleigh, S., Sethi, R.J., Sondjaja, M., Tiruviluamala, N., Uhlig, P.X., Washington, T.M., Wesley, C.L., White, D., Ye, P.: Curriculum guidelines for undergraduate programs in data science. Annu. Rev. Stat. Appl. **4**(2), 1–16 (2017). URL https://www.amstat.org/asa/files/pdfs/EDU-DataScienceGuidelines.pdf
413. Vesset, D., Woo, B., Morris, H.D., Villars, R.L., Little, G., Bozman, J.S., Borovick, L., Olofson, C.W., Feldman, S., Conway, S., Eastwood, M., Yezhkova, N.: Worldwide big data technology and services 2012-2015 forecast (2012). IDC
414. Viseu, A., Suchman, L.: Wearable Augmentations: Imaginaries of the Informed Body, pp. 161–184. Berghahn Books, New York (2010)
415. Walker, M.A.: The professionalisation of data science. Int. J. of Data Science **1**(1), 7–16 (2015)
416. Wang, C., Cao, L., Chi, C.: Formalization and verification of group behavior interactions. IEEE Trans. Systems, Man, and Cybernetics: Systems **45**(8), 1109–1124 (2015)
417. WEF: The global competitiveness report 2011-2012: An initiative of the world economic forum (2011)
418. Wei, W.: Copula-based high dimensional dependence modelling. Ph.D. thesis, University of Technology Sydney (2014)
419. Wei Wei Junfu Yin, J.L., Cao, L.: Modeling asymmetry and tail dependence among multiple variables by using partial regular vine. In: SDM2014 (2014)
420. Weiss, K., Khoshgoftaar, T.M., Wang, D.: A survey of transfer learning. Journal of Big Data **3**(1) (2016)
421. Whitehouse: The white house names dr. DJ patil as the first U.S. chief data scientist (2015). URL https://www.whitehouse.gov/blog/2015/02/18/white-house-names-dr-dj-patil-first-us-chief-data-scientist
422. Wikipedia: Bioinformatics. URL https://en.wikipedia.org/wiki/Bioinformatics
423. Wikipedia: Computational trust. URL https://en.wikipedia.org/wiki/Computational_trust
424. Wikipedia: Computing. URL https://en.wikipedia.org/wiki/Computing
425. Wikipedia: Dikw pyramid. URL https://en.wikipedia.org/wiki/DIKW_Pyramid
426. Wikipedia: Genetic linkage. URL https://en.wikipedia.org/wiki/Genetic_linkage
427. Wikipedia: Health care & analytics. URL http://analytics-magazine.org/health-care-a-analytics/
428. Wikipedia: Intelligent transportation system. URL https://en.wikipedia.org/wiki/Intelligent_transportation_system
429. Wikipedia: Social influence. URL https://en.wikipedia.org/wiki/Social_influence
430. Wikipedia: Social network analysis. URL https://en.wikipedia.org/wiki/Social_network_analysis
431. Wikipedia: Statistical relational learning. URL https://en.wikipedia.org/wiki/Statistical_relational_learning
432. Wikipedia: Sustainability. URL https://en.wikipedia.org/wiki/Sustainability
433. Wikipedia: Targeted advertising. URL https://en.wikipedia.org/wiki/Targeted_advertising
434. Wikipedia: Comparison of cluster software (2016). URL https://en.wikipedia.org/wiki/Comparison_of_cluster_software
435. Wikipedia: General data protection regulation (2016). URL https://en.wikipedia.org/wiki/General_Data_Protection_Regulation
436. Wikipedia: Informatics (2016). URL https://en.wikipedia.org/wiki/Informatics
437. Wikipedia: List of reporting software (2016). URL https://en.wikipedia.org/wiki/List_of_reporting_software
438. Wikipedia: National data protection authority (2016). URL https://en.wikipedia.org/wiki/National_data_protection_authority
439. Wikipedia: Sports analytics (2016). URL https://en.wikipedia.org/wiki/Sports_analytics
440. Wikipedia: Accuracy, precision, recall and specificity (2017). URL https://en.wikipedia.org/wiki/Precision_and_recall

441. Wikipedia: Capability maturity model (cmm) (2017). URL https://en.wikipedia.org/wiki/Capability_Maturity_Model
442. Wikipedia: Complexity (2017). URL https://en.wikipedia.org/wiki/Complexity
443. Wikipedia: Data quality (2017). URL https://en.wikipedia.org/wiki/Data_quality
444. Wikipedia: Industrial revolution (2017). URL https://en.wikipedia.org/wiki/Industrial_Revolution
445. Wikipedia: List of statistical packages (2017). URL https://en.wikipedia.org/wiki/List_of_statistical_packages
446. Wikipedia: Second industrial revolution (2017). URL https://en.wikipedia.org/wiki/Second_Industrial_Revolution
447. Wikipedia: Timeline of machine learning. retrieved 21 march 2017 (2017). URL https://en.wikipedia.org/wiki/Timeline_of_machine_learning
448. Wikipedia: Agile software development (2018). URL https://en.wikipedia.org/wiki/Agile_software_development
449. Wikipedia: Industry 4.0 (2018). URL https://en.wikipedia.org/wiki/Industry_4.0
450. Wikipedia: Internet of things (2018). URL https://en.wikipedia.org/wiki/Internet_of_things
451. Wikipedia: Open access (2018). URL https://en.wikipedia.org/wiki/Open_access
452. Wikipedia: Open data (2018). URL https://en.wikipedia.org/wiki/Open_data
453. Wikipedia: Open education (2018). URL https://en.wikipedia.org/wiki/Open_education
454. Wikipedia: Open peer review (2018). URL https://en.wikipedia.org/wiki/Open_peer_review
455. Wikipedia: Open science (2018). URL https://en.wikipedia.org/wiki/Open_science
456. Wikipedia: Open source (2018). URL https://en.wikipedia.org/wiki/Open-source_software
457. Wikipedia: Smart manufacturing (2018). URL https://en.wikipedia.org/wiki/Smart_manufacturing
458. Wikipedia: Waterfall model (2018). URL https://en.wikipedia.org/wiki/Waterfall_model
459. Williamson, J.: Big data analytics is transforming manufacturing (2016). URL http://www.themanufacturer.com/articles/big-data-analytics-is-transforming-manufacturing/
460. WIRED: How europe can seize the starring role in big data (2014). URL www.wired.com/insights/2014/09/europe-big-data/
461. Wladawsky-Berger, I.: Why do we need data science when we've had statistics for centuries? The Wall Street Journal (2014). URL http://blogs.wsj.com/cio/2014/05/02/why-do-we-need-data-science-when-weve-had-statistics-for-centuries/
462. Wolf, G.: The data-driven life. New York Times (2012). URL www.nytimes.com/2010/05/02/magazine/02self-measurement-t.html
463. Woodall P., B.A., Parlikad, A.: Data quality assessment: The hybrid approach. Information & Management **50**(7), 369–382 (2013)
464. Woodall P., O.M., A., B.: A classification of data quality assessment and improvement methods. International Journal of Information Quality **3**(4), 298–321 (2014)
465. Works, B.: Burtch works flash survey (2014). URL http://www.burtchworks.com/category/flash-survey/
466. WTTC: Big data - the impact on travel & tourism (2014). URL https://www.wttc.org/research/other-research/big-data-the-impact-on-travel-tourism/
467. Wu, J.: Statistics = data science? (1997). URL http://www2.isye.gatech.edu/~jeffwu/presentations/datascience.pdf
468. Xie, T., Thummalapenta, S., Lo, D., Liu, C.: Data mining for software engineering. Computer **42**(8) (2009)
469. Yahoo: Yahoo finance (2016). URL www.finance.yahoo.com
470. Yau, N.: Rise of the data scientist (2009). URL http://flowingdata.com/2009/06/04/rise-of-the-data-scientist/
471. Yin, J., Zheng, Z., Cao, L.: Uspan: An efficient algorithm for mining high utility sequential patterns. In: KDD 2012, pp. 660–668 (2012)
472. Yiu, C.: The big data opportunity (2012). URL http://www.policyexchange.org.uk/images/publications/the%20big%20data%20opportunity.pdf
473. Yu, B.: IMS presidential address: Let us own data science. IMS Bulletin Online (2014). 1 Oct 2014

Index

L. Cao, *Data Science Thinking*, Data Analytics,
https://doi.org/10.1007/978-3-319-95092-1