Appendix B: References

Chapter 1

Barnard, G. A. "Fisher: A retrospective." *Chance* 3.1990 (1990): <http://math.fullerton.edu/sbehseta/Fisher.PDF>

Box, Joan Fisher. "Guinness, Gosset, Fisher, and small samples." *Statistical Science* 2.1 (1987): 45-52.

Farr, W. (1852). Report on the Mortality from Cholera in England, 1848–1849. London: HMSO. (Also published as Farr, W. 1852. Registrar General’s Report on Cholera in England 1849–1850. London: W. Clowes & Son)

General Board of Health (UK). (1855). Report of the committee for scientiﬁc inquiries in relation to the cholera-epidemic of 1854. London: HMSO.

Tukey, John W. "The future of data analysis." *The Annals of Mathematical Statistics* 33.1 (1962): 1-67. Retrieved 7/13/2013 from: <http://projecteuclid.org/DPubS?service=UI&version=1.0&verb=Display&handle=euclid.aoms/1177704711>

Bingham, P., N. Q. Verlander, and M. J. Cheal. "John Snow, William Farr and the 1849 outbreak of cholera that affected London: a reworking of the data highlights the importance of the water supply." *Public Health* 118.6 (2004): 387-394.

Fisher, Ronald A. "The influence of rainfall on the yield of wheat at Rothamsted." *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character* 213 (1925): 89-142.

Tukey and the Prim-9 video: <http://flowingdata.com/2008/01/01/john-tukey-and-the-beginning-of-interactive-graphics/>

Breiman, Leo. "Statistical modeling: The two cultures (with comments and a rejoinder by the author)." *Statistical Science* 16.3 (2001): 199-231.

Watts, Duncan, “The Myth of Common Sense: Why Everything that Seems Obvious Isn't”, Speaking to the Santa Fe institute, viewed on 7/14/2013 at <http://www.youtube.com/watch?v=EF8tdXwa-AE>

Lipowski, Earlene E. "Developing great research questions." *American Journal of Health-System Pharmacy* 65.17 (2008): 1667.

“Report of JPMorgan Chase & Co. Management Task Force Regarding 2012 CIO Losses”, retrieved 7/18/2013 <http://files.shareholder.com/downloads/ONE/2532388207x0x628656/4cb574a0-0bf5-4728-9582-625e4519b5ab/Task_Force_Report.pdf>

Hubbard, Douglas W. How to measure anything: finding the value of intangibles in business. Wiley. com, 2010.

Cook, Richard. "How complex systems fail." Cognitive Technologies Laboratory, University of Chicago. Chicago IL (1998).

Morris, Robert, and Ken Thompson. "Password security: A case history." Communications of the ACM 22.11 (1979): 594-597.

Wheelan, Charles. Naked Statistics: Stripping the Dread from the Data. WW Norton, 2013.

Kahneman, Daniel, and Gary Klein. "Conditions for intuitive expertise: a failure to disagree." American Psychologist 64, no. 6 (2009): 515..

Cook, Richard. "How complex systems fail." Cognitive Technologies Laboratory, University of Chicago. Chicago IL (1998).

Wheelan, Charles. Naked Statistics: Stripping the Dread from the Data. WW Norton, 2013.

Chapter 2

Crawley, Michael J. The R book. John Wiley & Sons, 2012.

Cotton, Richard Learning R O’Reilly Media, 2013

Shaw, Zed A. "Learn Python The Hard Way." (2010).

Lutz, Mark. Learning python. O'Reilly Media, 2013.

Chapter 3

Cohen, Yosef, and Jeremiah Y. Cohen. Statistics and Data with R: An applied approach through examples. Wiley, 2008.

McKinney, Wes. Python for data analysis. O'Reilly Media, 2012.

Chapter 4

Cook, Diane J., and Lawrence B. Holder, eds. Mining graph data. Wiley-Interscience, 2006.

Højsgaard, Søren, David Edwards, and Steffen Lauritzen. Graphical models with R. Springer, 2012.

Chapter 5

Goodman, Steven. "A Dirty Dozen: Twelve P-Value Misconceptions."Seminars in hematology. Vol. 45. No. 3. WB Saunders, 2008.

Neter, John, William Wasserman, and Michael H. Kutner. Applied linear statistical models. Vol. 4. Chicago: Irwin, 1996.

Yau, Nathan Data Points: Visualization That Means Something. Wiley, 2013

Chang, Winston. R graphics cookbook. O'Reilly, 2012.

Wheelan, Charles. Naked Statistics: Stripping the Dread from the Data. WW Norton, 2013.

Chapter 6

Kosara, Robert, “In Defense of Pie Charts” <http://eagereyes.org/criticism/in-defense-of-pie-charts> *Retrieved August 27, (2013); 2011*

Cleveland, William S., and Robert McGill. "Graphical perception: Theory, experimentation, and application to the development of graphical methods."Journal of the American Statistical Association 79.387 (1984): 531-554.

Cleveland, William S., and Robert McGill. "Graphical perception and graphical methods for analyzing scientific data." Science 229.4716 (1985): 828-833.

Cairo, Alberto. The Functional Art: An introduction to information graphics and visualization. New Riders, 2012.

Card, Stuart K., and Jock Mackinlay. "The structure of the information visualization design space." Information Visualization, 1997. Proceedings., IEEE Symposium on. IEEE, 1997.

Healey, Christopher G., Kellogg S. Booth, and James T. Enns. "High-speed visual estimation using preattentive processing." ACM Transactions on Computer-Human Interaction (TOCHI) 3.2 (1996): 107-135.

Ware, Colin. Information visualization. Vol. 2. San Francisco: Morgan Kaufmann, 2000.

Stone, Maureen. "Choosing colors for data visualization." Business Intelligence Network (2006). Retrieved Sep, 2013 <http://www.perceptualedge.com/articles/b-eye/choosing_colors.pdf>

Tufte, Edward. *Envisioning Information.* Graphics Press, 1990

Simkin, David, and Reid Hastie. "An information-processing analysis of graph perception." Journal of the American Statistical Association 82.398 (1987): 454-465.

Yau, Nathan. Data Points: Visualization That Means Something. Wiley, 2013.

Few, Stephen. Show me the numbers: Designing Tables and Graphs to Enlighten. Vol. 1. No. 1. Oakland, CA: Analytics Press, 2004.

Tufte, Edward. Envisioning Information. Graphics Press, 1990

Ware, Colin. Information visualization. Vol. 2. San Francisco: Morgan Kaufmann, 2000.

Chapter 7

Verizon RISK Team. "VERIS Community Database." *Available:* [https://github.com/vz-risk/VCDB](http://www.verizonenterprise.com/DBIR).

Open Security Foundation. "Data Loss DB." *Available:* [http://datalossdb.org](http://www.verizonenterprise.com/DBIR).

Privacy Rights Clearinghouse. "Chronology of Data Breaches." *Available:* http://www.privacyrights.org/data-breach.

Verizon RISK Team. "2013 data breach investigations report." *Available:* <http://www.verizonenterprise.com/DBIR>.

Verizon RISK Team. "VERIS Community." *Available:* http://veriscommunity.net .

Chapter 8

Harrington, Jan L. Relational database design and implementation: clearly explained. Morgan Kaufmann, 2009.

Lublinsky, Boris, Kevin T. Smith, and Alexey Yakubovich. Professional Hadoop Solutions. John Wiley & Sons, 2013.

Tiwari, Shashank. Professional NoSQL. John Wiley & Sons, 2011.

Chapter 9

Mitchell, Tom M. "Machine learning. 1997." Burr Ridge, IL: McGraw Hill 45 (1997).

Bilge, Leyla, et al. "Disclosure: detecting botnet command and control servers through large-scale NetFlow analysis." Proceedings of the 28th Annual Computer Security Applications Conference. ACM, 2012.

Genuer, Robin, Jean-Michel Poggi, and Christine Tuleau-Malot. "Variable selection using random forests." Pattern Recognition Letters 31.14 (2010): 2225-2236.

Emran, Syed Masum, and Nong Ye. "Robustness of canberra metric in computer intrusion detection." Proc. IEEE Workshop on Information Assurance and Security, West Point, NY, USA. 2001.

Weston, Steven and Calaway, Rich. “Getting Started with doParallel and foreach”, retrieved 10/2013 from <http://cran.r-project.org/web/packages/doParallel/vignettes/gettingstartedParallel.pdf>

Conway, Drew, and John White. Machine Learning for Hackers. O'Reilly Media, Inc., 2012.

Cherkassky, Vladimir, and Filip M. Mulier. Learning from data: concepts, theory, and methods. Wiley, 2007.

James, Gareth, et al. An Introduction to Statistical Learning with Applications in R. Springer, 2013.

Chapter 10

Few, Stephen. Information dashboard design. Analytics Press, Second edition, 2013.

Jaquith, Andrew. Security metrics: replacing fear, uncertainty, and doubt. Upper Saddle River: Addison-Wesley, 2007.

Tufte, Edward R., and P. R. Graves-Morris. The visual display of quantitative information. Vol. 2. Cheshire, CT: Graphics press, 1983.

Saris, Willem E., and Irmtraud N. Gallhofer. Design, evaluation, and analysis of questionnaires for survey research. Vol. 548. John Wiley & Sons, 2007.

Kenett, Ron, and Silvia Salini. Modern analysis of customer surveys: with applications using R. Vol. 117. Wiley. com, 2011.

Chapter 11

Stefano Foresti, James Agutter, Yarden Livnat, Shaun Moon, and Robert Erbacher, “Visual Correlation of Network Alerts,” IEEE Computer Graphics and Applications, vol. 26, no. 2, pp. 48-59, March/April, 2006.

Schwartz, Barry. “The Paradox of Choice: Why Less Is More.” New York: Ecco, 2004.

Wilkinson, Leland. The grammar of graphics. Springer Berlin Heidelberg, 2012.

Norman, Donald A. The design of everyday things. Basic books, 2002.

Murray, Scott. Interactive Data Visualization for the Web. O'Reilly Media, 2013.

Cairo, Alberto. The Functional Art. Peachpit Press, 2012.

Foresti, Stefano, and James Agutter. "VisAlert: From Idea to Product." InVizSEC 2007, pp. 159-174. Springer Berlin Heidelberg, 2008.

Maclean , Malcom. D3 Tips and Tricks: Interactive Data Visualization in a Web Browser. http://leanpub.com/D3-Tips-and-Tricks

Chapter 12

James, Bill. *Battling Expertise with the Power of Ignorance.* 2010.<http://crllearns.kucrl.org/events/battling-expertise-with-the-power-of-ignorance>

R Packages Used

aplpack: Peter Wolf and Uni Bielefeld (2013). aplpack: Another Plot PACKage: stem.leaf, bagplot, faces, spin3R, plotsummary, plothulls, and some slider functions. R package version 1.2.9. http://CRAN.R-project.org/package=aplpack

* binom: Sundar Dorai-Raj (2009). binom: Binomial Confidence Intervals For Several Parameterizations. R package version 1.0-5. http://CRAN.R-project.org/package=binom
* car: John Fox and Sanford Weisberg (2011). An {R} Companion to Applied Regression, Second Edition. Thousand Oaks CA: Sage. URL: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion
* colorspace: Ross Ihaka, Paul Murrell, Kurt Hornik, Jason C. Fisher, Achim Zeileis (2013). colorspace: Color Space Manipulation. R package version 1.2-2. URL http://CRAN.R-project.org/package=colorspace
* colorspace: Achim Zeileis, Kurt Hornik, Paul Murrell (2009). Escaping RGBland: Selecting Colors for Statistical Graphics. Computational Statistics & Data Analysis, 53, 3259-3270. doi:10.1016/j.csda.2008.11.033
* devtools: Hadley Wickham and Winston Chang (2013). devtools: Tools to make developing R code easier. R package version 1.3. http://CRAN.R-project.org/package=devtools
* effects: John Fox (2003). Effect Displays in R for Generalised Linear Models. Journal of Statistical Software, 8(15), 1-27. URL http://www.jstatsoft.org/v08/i15/.
* effects: John Fox, Jangman Hong (2009). Effect Displays in R for Multinomial and Proportional-Odds Logit Models: Extensions to the effects Package. Journal of Statistical Software, 32(1), 1-24. URL http://www.jstatsoft.org/v32/i01/.
* gdata: Gregory R. Warnes, Ben Bolker, Gregor Gorjanc, Gabor Grothendieck, Ales Korosec, Thomas Lumley, Don MacQueen, Arni Magnusson, Jim Rogers and others (2013). gdata: Various R programming tools for data manipulation. R package version 2.13.2. http://CRAN.R-project.org/package=gdata
* ggdendro: Andrie de Vries and Brian D. Ripley (2013). ggdendro: Tools for extracting dendrogram and tree diagram plot data for use with ggplot.. R package version 0.1-14. http://CRAN.R-project.org/package=ggdendro
* ggmap: David Kahle and Hadley Wickham (2013). ggmap: A package for spatial visualization with Google Maps and OpenStreetMap. R package version 2.3. http://CRAN.R-project.org/package=ggmap
* ggplot2: H. Wickham. ggplot2: elegant graphics for data analysis. Springer New York, 2009.
* ggthemes: Jeffrey B. Arnold (2013). ggthemes: Extra themes, scales and geoms for ggplot. R package version 1.5.1. http://CRAN.R-project.org/package=ggthemes
* grid: R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
* gridExtra: Baptiste Auguie (2012). gridExtra: functions in Grid graphics. R package version 0.9.1. http://CRAN.R-project.org/package=gridExtra
* lattice: Sarkar, Deepayan (2008) Lattice: Multivariate Data Visualization with R. Springer, New York. ISBN 978-0-387-75968-5
* maps: Original S code by Richard A. Becker and Allan R. Wilks. R version by Ray Brownrigg. Enhancements by Thomas P Minka <tpminka@media.mit.edu> (2013). maps: Draw Geographical Maps. R package version 2.3-6. http://CRAN.R-project.org/package=maps
* maptools: Roger Bivand and Nicholas Lewin-Koh (2013). maptools: Tools for reading and handling spatial objects. R package version 0.8-27. http://CRAN.R-project.org/package=maptools
* plyr: Hadley Wickham (2011). The Split-Apply-Combine Strategy for Data Analysis. Journal of Statistical Software, 40(1), 1-29. URL http://www.jstatsoft.org/v40/i01/.
* portfolio: Jeff Enos, David Kane, with contributions from Daniel Gerlanc and Kyle Campbell (2013). portfolio: Analysing equity portfolios. R package version 0.4-6. http://CRAN.R-project.org/package=portfolio
* RColorBrewer: Erich Neuwirth (2011). RColorBrewer: ColorBrewer palettes. R package version 1.0-5. http://CRAN.R-project.org/package=RColorBrewer
* rgdal: Roger Bivand, Tim Keitt and Barry Rowlingson (2013). rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 0.8-11. http://CRAN.R-project.org/package=rgdal
* rjson: Alex Couture-Beil (2013). rjson: JSON for R. R package version 0.2.13. http://CRAN.R-project.org/package=rjson
* RJSONIO: Duncan Temple Lang (2013). RJSONIO: Serialize R objects to JSON, JavaScript Object Notation. R package version 1.0-3. http://CRAN.R-project.org/package=RJSONIO
* scales: Hadley Wickham (2012). scales: Scale functions for graphics.. R package version 0.2.3. http://CRAN.R-project.org/package=scales
* splines: R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
* vcd: David Meyer, Achim Zeileis, and Kurt Hornik (2013). vcd: Visualizing Categorical Data. R package version 1.3-1.
* vcd: David Meyer, Achim Zeileis, and Kurt Hornik (2006). The Strucplot Framework: Visualizing Multi-Way Contingency Tables with vcd. Journal of Statistical Software, 17(3), 1-48. URL http://www.jstatsoft.org/v17/i03/
* vcd: Achim Zeileis, David Meyer, and Kurt Hornik (2007). Residual-based Shadings for Visualizing (Conditional) Independence. Journal of Computational and Graphical Statistics, 16(3), 507-525.
* verisr: Jay Jacobs (2013). verisr: Tools for working with VERIS objects. R package version 0.1.
* zoo: Achim Zeileis and Gabor Grothendieck (2005). zoo: S3 Infrastructure for Regular and Irregular Time Series. Journal of Statistical Software, 14(6), 1-27. URL http://www.jstatsoft.org/v14/i06/