Appendix B: References

Chapter 1

Barnard, G. A. "Fisher: A retrospective." *Chance* 3.1990 (1990): l. retrieved from <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>

Kahneman, Daniel, and Gary Klein. "Conditions for intuitive expertise: a failure to disagree." *American Psychologist* 64.6 (2009): 515.

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.

*Conditions for intuitive expertise: a failure to disagree by Kahneman and Klein*.

*How complex systems fail by Richard Cook.*

*Naked Statistics by Charles Wheelan*.

Chapter 2

*The R Book* by Michael J. Crawley

*Learning R* by Richard Cotton

*Learn Python the Hard Way* by Zed A. Shaw

*Learning Python* by Mark Lutz

Chapter 3

*Statistics and Data with R: An Applied Approach Through Examples* by Yosef Cohen and Jeremiah Y. Cohen (John Wiley & Sons, Ltd. ISBN: 9-780-470758052)

*Python for Data Analysis* by Wes McKinney (O’Reilly Media, Inc. ISBN: 9-798-1-4493-1979-3)

Chapter 4

*Mining Graph Data* by Diane J. Cook and Lawrence B. Holder (John Wiley & Sons, 2006). ISBN: 9780471731900

*Graphical Models with R* by Højsgaard, Søren, David Edwards, and Steffen Lauritzen (Springer, 2012). ISBN: 978-1461422983

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.

*Data Points: Visualization That Means Something* by Nathan Yau

*R Graphics Cookbook* by Winston Chang

*Naked Statistics: Stripping the Dread from the Data* by Charles Wheelan

Chapter 6

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

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

Few, Stephen. "Save the pies for dessert." <http://perceptualedge.com/articles/visual_business_intelligence/save_the_pies_for_dessert.pdf> *Retrieved August 27* (2013): 2007.

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

Chapter 8

*Relational Database Design Clearly Explained, Second Edition,* by Jan L. Harrington

*Professional Hadoop Solutions* by Boris Lublinsky, Kevin T. Smith, and Alexey Yakubovich

*Professional NoSQL* by Shashank Tiwari

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. com, 2007.

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

Chapter 10

*Information Dashboard Design: The Effective Visual Communication of Data* by Stephen Few

*Security Metrics: Replacing Fear, Uncertainty, and Doubt* by Andrew Jaquith

*The Visual Display of Quantitative Information* by Edward R. Tufte, and P. R. Graves-Morris

*Design, Evaluation, and Analysis of Questionnaires for Survey Research* by Willem E. Saris and Irmtraud N. Gallhofer

*Modern Analysis of Customer Surveys: with Applications Using R* by Ron Kenett and Silvia Salini

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.

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

Leland Wilkinson. *The Grammar of Graphics.* Springer, 2005.

*The Design of Everyday Things* by Donald A. Norman

*Interactive Data Visualization for the Web* by Scott Murray

*The Functional Art* by Alberto Cairo

“VisAlert: From Idea to Product” by Stefano Foresti and James Agutter—This is found in *VizSEC 2007*, pp. 159-174. Springer Berlin Heidelberg, 2008

*D3 Tips and Tricks: Interactive Data Visualization in a Web Browser* by Malcolm Maclean—See http://leanpub.com/D3-Tips-and-Tricks

Chapter 12

James, Bill. *Battling Expertise with the Power of Ignorance.* 2010.retrieved October 26, 2013 from http://crllearns.kucrl.org/events/battling-expertise-with-the-power-of-ignorance