

References Cited

- [1] Omar Benjelloun, Hector Garcia-Molina, David Menestrina, Qi Su, Steven Euijong Whang, and Jennifer Widom. Swoosh: A generic approach to entity resolution. *VLDB Journal*, 18(1):255–276, January 2009.
- [2] J. Bertin. *Sémiologie Graphique*. Gauthier-Villars, 1967. English translation by W.J. Berg as *Semiology of Graphics*, University of Wisconsin Press: Madison, WI, 1983.
- [3] Indrajit Bhattacharya and Lise Getoor. Collective entity resolution in relational data. *ACM Trans. Knowl. Discov. Data*, 1(1):5, 2007.
- [4] Michael Bostock and Jeffrey Heer. Protovis: A graphical toolkit for visualization. *IEEE Trans. Vis. and Comp. Graphics*, 15(6):1121–1128, 2009.
- [5] Cynthia A. Brewer. Color use guidelines for data representation. In *Proceedings of the Section on Statistical Graphics, American Statistical Association*, pages 55–60, 1999.
- [6] W. C. Brinton. *Graphic presentation*. Brinton Associates, 1939.
- [7] D. M. Bruls, C. Huizing, and J. J. van Wijk. Squarified treemaps. In *Data Visualization Eurographics/IEEE Symp.*, pages 33–42, 2000.
- [8] Michael J. Cafarella, Alon Halevy, Daisy Zhe Wang, Eugene Wu, and Yang Zhang. WebTables: Exploring the power of tables on the web. *Proc. VLDB*, 1(1):538–549, 2008.
- [9] Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman, editors. *Readings in information visualization: using vision to think*. Morgan Kaufmann, San Francisco, CA, 1999.
- [10] Stuart K. Card, Bongwon Suh, Bryan Pendleton, Jeffrey Heer, and John W. Bodnar. Timetree: Exploring time changing hierarchies. In *IEEE VAST*, pages 3–10, October 2006.
- [11] H. Chafi, Z. DeVito, A. Moors, T. Rompf, A. Sujeeth, P. Hanrahan, M. Odersky, and K. Olukotun. Language virtualization for heterogeneous parallel computing. In *Onward!*, 2010.
- [12] Laura Chiticariu, Phokion G. Kolaitis, and Lucian Popa. Interactive generation of integrated schemas. In *SIGMOD '08: Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 833–846, New York, NY, USA, 2008. ACM.
- [13] W. S. Cleveland. *Visualizing Data*. Hobart Press, 1993.
- [14] W. S. Cleveland, P. Diaconis, and R. McGill. Variables on scatterplots look more highly correlated when the scales are increased. *Science*, 216:1138–1141, Jun 1982.
- [15] W. S. Cleveland and R. McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79:531–554, 1984.
- [16] The College Board. <http://www.collegeboard.com>, July 2010.
- [17] Kenneth Cukier. Show Me: New ways of visualising data. http://www.economist.com/specialreports/displaystory.cfm?story_id=15557455, February 2010.
- [18] Tamraparni Dasu and Theodore Johnson. *Exploratory Data Mining and Data Cleaning*. Wiley-Interscience, May 2003.
- [19] Tamraparni Dasu, Theodore Johnson, S. Muthukrishnan, and Vladislav Shkapenyuk. Mining database structure; or, how to build a data quality browser. In *ACM SIGMOD*, 2002.
- [20] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.
- [21] B. D. Dent. *Cartography: Thematic Map Design*. William C. Brown Publishing, 1998.
- [22] AnHai Doan, Jayant Madhavan, Robin Dhamankar, Pedro Domingos, and Alon Halevy. Learning to match ontologies on the semantic web. *The VLDB Journal*, 12(4):303–319, 2003.
- [23] D. J. Duke, R. Borgo, M. Wallace, and C. Runciman. Huge data but small programs: Visualization design via multiple embedded DSLs. In *Practical Aspects of Declarative Languages*, pages 31–45. Springer, 2009.
- [24] Ahmed K. Elmagarmid, Panagiotis G. Ipeirotis, and Vassilios S. Verykios. Duplicate record detection: A survey. *IEEE Trans. Knowl. Data Eng.*, 19(1):1–16, 2007.
- [25] Jean-Daniel Fekete. The InfoVis Toolkit. In *IEEE InfoVis*, pages 167–174, 2004.

- [26] S. Few. *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Analytics Press, 2004.
- [27] S. Few. *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. Analytics Press, 2009.
- [28] Michel Fortin. Sim Daltonism. <http://michelf.com/projects/sim-daltonism/>.
- [29] Freebase. <http://freebase.com>, July 2010.
- [30] J.F. Gantz, C. Chute, A. Manfrediz, S. Minton, D. Reinsel, W. Schlichting, and A. Toncheva. The diverse and exploding digital universe: An updated forecast of worldwide information growth through 2011. Technical report, IDC White Paper, March 2008. http://www.emc.com/about/destination/digital_universe/.
- [31] J.F. Gantz, D. Reinsel, C. Chute, W. Schlichting, J. McArthur, S. Minton, I. Xheneti, A. Toncheva, and A. Manfrediz. The expanding digital universe: A forecast of worldwide information growth through 2010. Technical report, IDC White Paper, March 2007. http://www.emc.com/about/destination/digital_universe/.
- [32] P. P. Gilmartin. Influences of map context on circle perception. *Annals of the Association of American Geographers*, 71:253–258, 1981.
- [33] Luis Gravano, Panagiotis G. Ipeirotis, H. V. Jagadish, Nick Koudas, S. Muthukrishnan, Lauri Pietarinen, and Divesh Srivastava. Using q-grams in a dbms for approximate string processing. *IEEE Data Engineering Bulletin*, 24(4):28–34, December 2001.
- [34] T. R. G. Green. Cognitive dimensions of notations. In *Proceedings of the fifth conference of the British Computer Society, Human-Computer Interaction Specialist Group on People and computers V*, pages 443–460, New York, NY, USA, 1989. Cambridge University Press.
- [35] J. Heer and M. Agrawala. Multi-scale banking to 45°. *IEEE Trans. on Visualization and Comp. Graphics*, 12:701–708, Sep/Oct 2006.
- [36] J. Heer, N. Kong, and M. Agrawala. Sizing the horizon: the effects of chart size and layering on the graphical perception of time series visualizations. In *ACM CHI*, pages 1303–1312, 2009.
- [37] J. Heer, F. Viégas, and M. Wattenberg. Voyagers and voyeurs: Supporting asynchronous collaborative information visualization. In *ACM CHI*, pages 1029–1038, Apr 2007.
- [38] Jeffrey Heer. Exploring Enron. <http://hci.stanford.edu/jheer/projects/enron/>. Accessed: July 2010.
- [39] Jeffrey Heer. Flare. <http://flare.prefuse.org>, July 2010.
- [40] Jeffrey Heer and Maneesh Agrawala. Software design patterns for information visualization. *IEEE Trans. Vis. and Comp. Graphics*, 12(5):853–860, 2006.
- [41] Jeffrey Heer and Maneesh Agrawala. Design considerations for collaborative visual analytics. *Information Visualization*, 7(1):49–62, 2008.
- [42] Jeffrey Heer, Maneesh Agrawala, and Wesley Willett. Generalized selection via interactive query relaxation. In *ACM CHI*, pages 959–968, April 2008.
- [43] Jeffrey Heer and Michael Bostock. Crowdsourcing graphical perception: Using mechanical turk to assess visualization design. In *ACM CHI*, pages 203–212, April 2010.
- [44] Jeffrey Heer and Michael Bostock. Declarative language design for interactive visualization. *IEEE Trans. Visualization and Computer Graphics*, to appear, 2010.
- [45] Jeffrey Heer and danah boyd. Vizster: Visualizing online social networks. In *IEEE Information Visualization*, pages 32–39, October 2005.
- [46] Jeffrey Heer and Stuart K. Card. Doitrees revisited: Scalable, space-constrained visualization of hierarchical data. In *Advanced Visual Interfaces*, pages 412–424, May 2004.
- [47] Jeffrey Heer, Stuart K. Card, and James A. Landay. prefuse: a toolkit for interactive information visualization. In *Proc. ACM CHI*, pages 421–430, 2005.
- [48] Jeffrey Heer and George G. Robertson. Animated transitions in statistical data graphics. *IEEE Trans. Vis. and Comp. Graphics*, 13(6):1240–1247, 2007.
- [49] Joseph M. Hellerstein. Quantitative data cleaning for large databases, Feb 2008. White Paper, United Nations Economic Commission for Europe.

- [50] Joseph M. Hellerstein, Peter J. Haas, and Helen J. Wang. Online aggregation. In *ACM SIGMOD*, pages 171–182, May 1997.
- [51] Mauricio A. Hernández, Renée J. Miller, and Laura M. Haas. Clio: a semi-automatic tool for schema mapping. In *SIGMOD '01: Proceedings of the 2001 ACM SIGMOD international conference on Management of data*, page 607, New York, NY, USA, 2001. ACM.
- [52] Melanie Herschel, Mauricio A. Hernández, and Wang-Chiew Tan. Artemis: a system for analyzing missing answers. *Proc. VLDB Endow.*, 2(2):1550–1553, 2009.
- [53] Victoria Hodge and Jim Austin. A survey of outlier detection methodologies. *Artif. Intell. Rev.*, 22(2):85–126, 2004.
- [54] E.R.S. Hodges. *The Guild Handbook of Scientific Illustration*. Van Nostrand Reinhold, 1989.
- [55] N. Holmes. *The Best in Diagrammatic Graphics*. Quarto Publishing, 1993.
- [56] P. Hudak. Building domain-specific embedded languages. *ACM Comput. Surv.*, 28(4), 1996.
- [57] David Huynh and Stefano Mazzocchi. Freebase gridworks. <http://code.google.com/p/freebase-gridworks/>. Accessed: July 2010.
- [58] David F. Huynh, Robert C. Miller, and David R. Karger. Potluck: semi-ontology alignment for casual users. In *ISWC'07/ASWC'07: Proceedings of the 6th international The semantic web and 2nd Asian conference on Asian semantic web conference*, pages 903–910, Berlin, Heidelberg, 2007. Springer-Verlag.
- [59] P. Indyk, N. Koudas, and S. Muthukrishnan. Trends in massive time series data sets using sketches. In *VLDB*, 2000.
- [60] L. Itti and C. Koch. Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2(3):194–203, 2001.
- [61] Laurent Itti. Visual salience. http://www.scholarpedia.org/article/Visual_salience, March 2010.
- [62] Hyunmo Kang, Lise Getoor, Ben Shneiderman, Mustafa Bilgic, and Louis Licamele. Interactive entity resolution in relational data: A visual analytic tool and its evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 14(5):999–1014, 2008.
- [63] Juho Kim, Ranjitha Kumar, and Scott R Klemmer. Crowdsourcing interface for collecting correspondences of web pages. In *ACM UIST Posters*, 2009.
- [64] Nam Wook Kim, Stuart K. Card, and Jeffrey Heer. Tracing genealogical data with timenets. In *Advanced Visual Interfaces*, pages 241–248, May 2010.
- [65] R. Kohavi, R. M. Henne, and D. Sommerfield. Practical guide to controlled experiments on the web: listen to your customers not to the hippo. In *ACM KDD*, pages 959–967, 2007.
- [66] Nicholas Kong, Jeffrey Heer, and Maneesh Agrawala. Perceptual guidelines for creating rectangular treemaps. *IEEE Trans. Visualization and Computer Graphics*, to appear, 2010.
- [67] S. M. Kosslyn. Understanding charts and graphs. *Applied Cognitive Psychology*, 3:185–225, Jul-Sep 1989.
- [68] Klaus Krippendorff. *Content Analysis: An Introduction to Its Methodology*. Sage, 2004.
- [69] Jean Lave and Etienne Wenger. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press, 1991.
- [70] S. Lewandowsky and I. Spence. Discriminating strata in scatterplots. *Journal of the American Statistical Association*, 84:682–688, Sep 1989.
- [71] H. W. Lie. *Cascading Style Sheets*. PhD thesis, University of Oslo, 2005.
- [72] G. D. Logan. The code theory of visual attention: An integration of space-based and object-based attention. *Psychological Review*, 103(4):603–649, 1996.
- [73] J. A. Lohse. Cognitive model for the perception and understanding of graphs. In *ACM CHI*, pages 137–144, Apr/May 1991.
- [74] B. Lucas, G. D. Abram, N. S. Collins, D. A. Epstein, D. L. Gresh, and K. P. McAuliffe. An architecture for a scientific visualization system. In *IEEE Visualization*, pages 107–114. IEEE Computer Society Press, 1992.

- [75] P. Lyman and H. Varian. How much information? <http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/>, 2003.
- [76] J. D. Mackinlay, P. Hanrahan, and C. Stolte. Show Me: Automatic presentation for visual analysis. *IEEE TVCG*, 13:1137–1144, Nov/Dec 2007.
- [77] Jock Mackinlay. Automating the design of graphical presentations of relational information. *ACM Trans. Graph.*, 5(2):110–141, 1986.
- [78] M. Mernik, J. Heering, and A. M. Sloane. When and how to develop domain-specific languages. *ACM Comput. Surv.*, 37(4):316–344, 2005.
- [79] P. Mijksenaar and P. Westendorp. *Open Here: The Art of Instructional Design*. Joost Elffers Books, New York, 1999.
- [80] Robert C. Miller and Brad A. Myers. Interactive simultaneous editing of multiple text regions. In *USENIX 2001 Annual Technical Conference*, pages 161–174, June 2001.
- [81] Donald A. Norman. *The Psychology of Everyday Things*. Basic Books, New York, NY, 1988.
- [82] S. Pinker. A theory of graph comprehension. In *Artificial intelligence and the future of testing*, pages 73–126. Lawrence Erlbaum Associates, 1990.
- [83] Processing. <http://processing.org>, July 2010.
- [84] Erhard Rahm and Philip A. Bernstein. A survey of approaches to automatic schema matching. *The VLDB Journal*, 10(4):334–350, 2001.
- [85] Vijayshankar Raman and Joseph M. Hellerstein. Potter’s wheel: An interactive data cleaning system. In *VLDB ’01: Proceedings of the 27th International Conference on Very Large Data Bases*, pages 381–390, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.
- [86] George G. Robertson, Mary P. Czerwinski, and John E. Churchill. Visualization of mappings between schemas. In *ACM CHI*, pages 431–439, 2005.
- [87] Ruth Rosenholtz, Yuanzhen Li, Jonathan Mansfield, and Zhenlan Jin. Feature congestion: a measure of display clutter. In *ACM CHI*, pages 761–770, New York, NY, USA, 2005. ACM.
- [88] Ruth Rosenholtz, Nathaniel R. Twarog, Nadja Schinkel-Bielefeld, and Martin Wattenberg. An intuitive model of perceptual grouping for HCI design. In *ACM CHI*, pages 1331–1340, New York, NY, USA, 2009. ACM.
- [89] S.F. Roth, J. Kolojejchick, J. Mattis, and J. Goldstein. Interactive graphic design using automatic presentation knowledge. In *ACM CHI*, pages 112–117, Apr 1994.
- [90] S.F. Roth and J. Mattis. Automating the presentation of information. In *IEEE Artificial Intelligence Applications*, pages 90–97, Feb 1991.
- [91] Nathan Sakunkoo and Pattie Sakunkoo. Gliflix: Using movie subtitles for language learning. In *ACM UIST Demos*, 2009.
- [92] S. Sarawagi and A. Bhamidipaty. Interactive deduplication using active learning. In *Proc. of ACM SIGKDD*, Edmonton, Alberta, 2002.
- [93] Eric Saund. Finding perceptually closed paths in sketches and drawings. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25(4):475–491, April 2003.
- [94] Christopher Scaffidi, Brad Myers, and Mary Shaw. Topes: reusable abstractions for validating data. In *ICSE ’08: Proceedings of the 30th international conference on Software engineering*, pages 1–10, New York, NY, USA, 2008. ACM.
- [95] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.
- [96] B. Shneiderman. Tree visualization with tree-maps: 2-d space-filling approach. *ACM TOG*, 11:92–99, 1992.
- [97] B. G. Shortridge. Stimulus processing models from psychology: Can we use them in cartography? *The American Cartographer*, 9:155–167, 1982.
- [98] D. Simkin and R. Hastie. An information-processing analysis of graph perception. *Journal of the American Statistical Association*, 82:454–465, Jun 1987.

- [99] A. Slingsby, J. Dykes, and J. Wood. Configuring hierarchical layouts to address research questions. *IEEE Trans. Vis. and Comp. Graphics*, 15(6):977–984, 2009.
- [100] M. Southworth and S. Southworth. *Maps: A visual survey and design guide*. Little Brown, 1982.
- [101] I. Spence and S. Lewandowsky. Displaying proportions and percentages. *Applied Cog. Psych.*, 5:61–77, 1991.
- [102] Chris Stolte, Diane Tang, and Pat Hanrahan. Polaris: A system for query, analysis, and visualization of multi-dimensional relational databases. *IEEE Trans. Vis. and Comp. Graphics*, 8:52–65, 2002.
- [103] A. M. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136, 1980.
- [104] L. Tremmel. The visual separability of plotting symbols in scatterplots. *Journal of Computational and Graphical Statistics*, 4:101–112, Jun 1995.
- [105] E. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, 1983.
- [106] E. Tufte. *Envisioning Information*. Graphics Press, 1990.
- [107] E. Tufte. *Visual Explanations*. Graphics Press, 1997.
- [108] E. Tufte. *Beautiful Evidence*. Graphics Press, 2006.
- [109] Hal Varian. Hal Varian on how the web challenges managers. *McKinsey Quarterly*, January 2009.
- [110] Fernanda B. Viégas and Martin Wattenberg. Communication-minded visualization: A call to action. *IBM Systems Journal*, 45(4):801–812, 2006.
- [111] Fernanda B. Viégas, Martin Wattenberg, Frank van Ham, Jesse Kriss, and Matt McKeon. Many Eyes: a site for visualization at internet scale. *IEEE Trans. Vis. and Comp. Graphics*, 13(6):1121–1128, 2007.
- [112] H. Wainer. *Visual revelations: Graphical tales of fate and deception from Napoleon Bonaparte to Ross Perot*. Copernicus New York, NY, USA, 1997.
- [113] C. Ware. *Information Visualization: Perception for Design*. Morgan-Kaufmann, 2nd edition, 2004.
- [114] M. Wattenberg and D. Fisher. A model of multi-scale perceptual organization in information graphics. In *IEEE InfoVis*, pages 123–133, 2003.
- [115] M. Wattenberg and J. Kriss. Designing for social data analysis. *IEEE Trans. on Visualization and Comp. Graphics*, 12:549–557, Jul/Aug 2005.
- [116] Christopher E. Weaver. Building highly-coordinated visualizations in Improvise. In *Proc. IEEE InfoVis*, pages 159–166, 2004.
- [117] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer, 2009.
- [118] P. Wildbur and M. Burke. *Information graphics: Innovative solutions in contemporary design*. Thames & Hudson, 1999.
- [119] Leland Wilkinson. *The Grammar of Graphics (Statistics and Computing)*. Springer-Verlag, Secaucus, NJ, 2005.
- [120] Wesley Willett, Jeffrey Heer, and Maneesh Agrawala. Scented widgets: Improving navigation cues with embedded visualizations. *IEEE Trans. Visualization and Computer Graphics*, 13(6):1129–1136, 2007.