William S. Cleveland

Background, Past Research, Current Research, and Publications

DEPARTMENT OF STATISTICS, PURDUE
250 N. UNIVERSITY STREET
WEST LAFAYETTE, IN 47907

EMAIL: WSC@PURDUE.EDU
OFFICE: HAAS 222
CELL: (773) 402-9625

Background and Past Research

Positions: William S. Cleveland has been the Shanti S. Gupta Distinguished Professor of Statistics and Courtesy Professor of Computer Science at Purdue University since 1/1/2004. Previous to this, he was a Distinguished Member of Technical Staff in the Statistics Research Department at Bell Labs, Murray Hill; for 12 years he was the Department Head.

Education: Cleveland received an A.B. in Mathematics from Princeton University; his senior thesis adviser was William Feller. He received his PhD in Statistics from Yale University; his PhD thesis adviser was Leonard Jimmie Savage.

Awards and Honors: In 1996 Cleveland was chosen national Statistician of the Year by the Chicago Chapter of the American Statistical Association. In 2002 he was selected as a Highly Cited Researcher by the American Society for Information Science and Technology in the newly formed mathematics category. He has twice won the Wilcoxon Prize and once won the Youden Prize from Technometrics. He is a Fellow of the American Statistical Association, the Institute of Mathematical Statistics, and the American Association of the Advancement of Science. He is an Elected Member of the International Statistical Institute.

Data Science: Cleveland defined data science as it is used today in a talk at the 1999 biennial meeting of the International Statistical Institute, and in a 2001 paper, number [25] in the coming publication list. The term had been used before, but with different meanings. See the Web page `en.wikipedia.orgwikiData_science`. The paper was republished in 2014 [1] together with a discussion and with another paper about D&R with Tessera [2], described next, which requires research in all technical areas of data science.

The technical areas of data science are those that have an impact on how a data analyst analyses data in practice: (1) Statistical theory; (2) Statistical models; (3) Statistical and machine-learning methods; (4) Algorithms for statistical and machine-learning methods, and optimization; (5) Computational environments for data analysis; (6) Live analyses of data where results are judged by the findings, not the methodology and systems that where used.

The implications for an academic department are that it is not necessary each individual be a researcher in all areas. Rather, collectively, the department needs to have research in all areas. There must be an exchange of knowledge so that all department members have at least a basic understanding of all areas.

Areas of Research: Cleveland's areas of research have been in statistics, machine learning, data visualization, data analysis for multidisciplinary studies, and high performance computing for deep data analysis.

Data Analysis Projects: Cleveland has been involved in many projects requiring the analysis and modeling of very diverse datasets from many fields, including computer networking, healthcare engineering, telecommunications, homeland security, environmental monitoring, public opinion polling, cyber security, and visual perception.

Widely-Used Methods and Their Publication: In the course of this work in data analysis, Cleveland has developed many new analytic methods and new computer systems for data analysis that are used throughout the worldwide technical community. He has published 129 papers and 3 books on this work. A chronological publication list is below. For citations to the publications, see the Web page scholar.google.com/citations?hl=en&user=ds52UHcAAAAJ/.

Data Visualization: In data visualization, Cleveland has written two books, co-authored another and one user's manual, was the Editor of two books and a special issue of the Journal of the American Statistical Association. He is the founder of the Graphics Section of the American Statistical Association, which means he led the group that successfully petitioned the ASA board of directors for approval.

His two books on data visualization have been reviewed in many journals from a wide variety of disciplines. The Elements of Graphing Data was selected for the Library of Science. J. Lodge reviewed it in Atmospheric Environment and wrote: "*certain kinds of tendency toward bad graphics could be cured if as many authors as possible would not just read, but, in the words of the Anglican Prayer Book, 'learn, mark, and inwardly digest' this volume.*" B. Gunter reviewed Visualizing Data in Technometrics and wrote: "*This is a terrific book — in my opinion, a path-breaking book. Get it. Read it. Practice what it preaches. You will improve the quality of your data analysis*."

Cleveland and colleagues developed trellis display, a powerful framework for data visualization. It has been used by a large, worldwide community of data analysts as a result of its implementation in the two software systems based on the S language for data analysis, the S-Plus commercial system and the R open source system.

Current Research:
High Performance Computing for Deep Data Analysis

"Big Data" Misses Badly: The widely used term "big data" carries with it a notion of computational performance for the analysis of big datasets. But for data analysis, computational performance depends very heavily, not just on size, but on the computational complexity of the analytic routines used in the analysis. Data small in size can be a big challenge, too. Furthermore, the hardware power available to the data analyst is an important factor.

High Performance Computing for Deep Data Analysis Cleveland's current research is in High Performance Computing for Deep Data Analysis (HPC-DDA). The goal is to enable both deep analysis and HPC.

HPC means computations are feasible and practical for wide ranges of dataset size, computational complexity, and hardware power. Deep analysis means analyzing data at their finest granularity, and not just summary statistics. Deep analysis also means that the analyst can apply any of 1000s of methods of statistics, machine learning, and data visualization.

The goal is achieved by work in the Divide & Recombine (D&R) statistical approach to analysis, and the Tessera D&R software implementation that makes programming D&R easy. The work ranges from statistical theory to cluster design, covering all of the areas of data science; furthermore, work in the different areas is highly integrated, one area affecting another. Integrated work in data science is necessary to succeed.

D&R: Cleveland and colleagues have been developing D&R since 2009. In D&R, the data are divided into subsets (div), analytic methods are applied to each subset independently without communication between subsets (ana), and the subsets' outputs for each method are recombined (rec).

Research in statistical theory seeks division methods and recombination methods to optimize the statistical accuracy of D&R results. While D&R is a statistical approach, it is done for the purpose of HPC-DDA. Most of the D&R computation is embarrassingly parallel, the simplest and fastest parallel computation.

Tessera Tessera D&R software runs on a cluster. The front end has R and the Tessera datadr R package that makes D&R programing easy. The back end, the Hadoop distributed file system and parallel compute engine, executes the datadr/R commands: (div), (ana), and (rec). In between, the R package RHIPE (R and Hadoop Integrated Programming Environment) provides communication between datadr and Hadoop.

PhD Students: Cleveland is currently the advisor for 6 students; since joining Purdue in 2004, he has advised 18 students. The students have made major contributions to research in D&R with Tessera.

More Information For more information. see the Web page `tessera.io`

Publications

[1] W. S. Cleveland. Data Science: An Action Plan for the Field of Statistics. *Statistical Analysis and Data Mining*, 7:414–417, 2014. reprinting of 2001 article in ISI Review, Vol 69.

[2] W. S. Cleveland and R. P. Hafen. Divide and Recombine (D&R): Data Science for Large Complex Data. *Statistical Analysis and Data Mining*, 7:425–433, 2014.

[3] R. P. Hafen, L. Gosink, W. S. Cleveland, J. McDermott, K. Rodland, and K. Van Dam. Trelliscope: A System for Detailed Visualization in the Deep Analysis of Large Complex Data. In *Proceedings, IEEE Symposium on Large-Scale Data Analysis and Visualization*, 2013.

[4] R. P. Hafen and W. S. Cleveland. The ed Method for Nonparametric Density Estimation and Diagnostic Checking. Technical report, Department of Statistics, Purdue University, 2012.

[5] D. Anderson, B. Xi, and W. S. Cleveland. Multifractal and Gaussian Fractional Sum-Difference Models for Internet Traffic. Technical report, Department of Statistics, Purdue University, 2012.

[6] D. Anderson, B. Xi, and W. S. Cleveland. Multifractal and Gaussian Fractional Sum-Difference Models for Internet Traffic. Technical report, Department of Statistics, Purdue University, 2012.

[7] S. Guha, R. Hafen, J. Rounds, J. Xia, J. Li, B. Xi, and W. S. Cleveland. Large Complex Data: Divide and Recombine (D&R) with RHIPE. *Stat*, 1:53–67, 2012.

[8] S. Guha, P. Kidwell, A. Barthur, W. S. Cleveland, J. Gerth, and Carter Bullard. A Streaming Statistical Algorithm for Detection of SSH Keystroke Packets in TCP Connections. In R. Kevin Wood and Robert F. Dell, editors, *Operations Research, Computing, and Homeland Defense*. Institute for Operations Research and Management Sciences, 2011.

[9] M. Fuentes, B. Xi, and W. S. Cleveland. Trellis Display for Modeling Data from Designed Experiments. *Statistical Analysis and Data Mining*, 4:133–145, 2011.

[10] B. Xi, H. Chen, W. S. Cleveland, and T. Telkamp. Statistical Analysis and Modeling of Internet VoIP Traffic for Network Engineering. *Electronic Journal of Statistics*, 4:58–116, 2010.

[11] R. Maciejewski, R. Hafen, S. Rudolph, W. S. Cleveland, and D. Ebert. A Visual Analytics Approach to Understanding Spatiotemporal Hotspots. *IEEE Transactions on Visualization and Computer Graphics*, 16:205–220, 2010.

[12] S. Guha, R. P. Hafen, P. Kidwell, and W. S. Cleveland. Visualization Databases for the Analysis of Large Complex Datasets. *Journal of Machine Learning Research*, 5:193–200, 2009.

[13] R. P. Hafen, D. E. Anderson, W. S. Cleveland, R. Maciejewski, D. S. Ebert, A. Abusalah, M. Yakout, and S. Ouzzani, M. andGrannis. Syndromic Surveillance: STL for Modeling, Visualizing, and Monitoring Disease Counts. *BMC Medical Informatics and Decision Making*, 9:21:doi:10.1186/1472–6947–9–21, 2009.

[14] R. Maciejewski, R. Hafen, S. Rudolph, G. Tebbetts, W. S. Cleveland, S. J. Grannis, and D. S. Ebert. Generating Synthetic Syndromic Surveillance Data for Evaluating Visual Analytics Techniques. *IEEE Computer Graphics & Applications*, 29:3:18–28, 2009.

[15] P. Kidwell, G. Lebanon, and W. S. Cleveland. Visualizing Incomplete and Partially Ranked Data. *IEEE Transactions on Visualization and Computer Graphics*, 14:1356–1363, 2008.

[16] R. Maciejewski, S. Rudolph, R. Hafen, A. A., M. Yakout, M. Ouzzani, W. S. Cleveland, S. J. Grannis, M. Wade, and D. S. Ebert. Understanding Syndromic Hotspots - A Visual Analytics Approach. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST)*, pages 35–42, 2008.

[17] R. Maciejewski, N. Glickman, G. Moore, C. Zheng, B. Tyner, W. Cleveland, D. Ebert, S. Lance, J. Horan, and L. Glickman. Companion Animals as Sentinels for Community-Exposure to Industrial Chemicals: The Fairburn, GA Propyl Mercaptan Case-Study. *Public Health Reports*, 123(3), 2008.

[18] W. S. Cleveland. Learning from Data: Unifying Statistics and Computer Science. *ISI Review*, 73(2):217–221, 2007.

[19] R. Maciejewski, B. Tyner, Y. Jang, C. Zheng, R. Nehme, D. Ebert, W. Cleveland, M. Ozzani, S. Grannis, and L. Glickman. LAHVA: Linked Animal-Human Health Visual Analytics. *IEEE Symposium on Visual Analytics Science and Technology*, pages 27 – 34, 2007.

[20] J. Cao, W. S. Cleveland, and D. X. Sun. Bandwidth Estimation for Best-Effort Internet Traffic. *Statistical Science*, 19:518–543, 2004.

[21] J. Cao, W. S. Cleveland, Y. Gao, K. Jeffay, F. D. Smith, and M. Weigle. Stochastic Models for Generating Synthetic HTTP Source Traffic. *Proc. IEEE INFOCOMM*, pages 1546–1557, 2004.

[22] J. Cao, W. S. Cleveland, and D. X. Sun. The S-Net System for Internet Packet Streams: Strategies for Stream Analysis and System Architecture. *Journal of Computational and Statistical Graphics: Special Issue on Streaming Data*, 12:865–892, 2003.

[23] J. Cao, W. S. Cleveland, D. Lin, and D. X. Sun. Internet Traffic Tends *Toward* Poisson and Independent as the Load Increases. In C. Holmes, D. Denison, M. Hansen, B. Yu, and B. Mallick, editors, *Nonlinear Estimation and Classification*, pages 83–109. Springer, New York, 2002.

[24] J. Cao, W. S. Cleveland, D. Lin, and D. X. Sun. On the Nonstationarity of Internet Traffic. *ACM SIGMETRICS*, 29(1):102–112, 2001.

[25] W. S. Cleveland. Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics. *ISI Review*, 69:21–26, 2001.

[26] W. S. Cleveland. Comments on Approache Graphique en Analyze des Données by Jean-Paul Valois. *Journal de la Société Française de Statistique*, 141:43–44, 2001.

[27] W. S. Cleveland and D. X. Sun. Internet Traffic Data. *Journal of the American Statistical Association*, 95:979–985, 2000. Reprinted in *Statistics in the 21st Century*, pp. 214-228, edited by A. E. Raftery, M. A. Tanner, and M. T. Wells, Chapman & Hall/CRC, New York, 2002.

[28] W. S. Cleveland, D. Lin, and D. X. Sun. IP Packet Generation: Statistical Models for TCP Start Times Based on Connection-Rate Superposition. *ACM SIGMETRICS*, pages 166–177, 2000.

[29] W. S. Cleveland, L. Denby, and C. Liu. Random Location and Scale Effects: Model Building Methods for a General Class of Models. In E. J. Wegman and Y. M. Martinez, editors, *Computer Science and Statistics*, volume 32, pages 3–10. Interface Foundations of North America, 2000.

[30] L. Clark, W. S. Cleveland, L. Denby, and C. Liu. Modeling Customer Survey Data. In C. Gatsonis, R.E. Kass, A. Carriquiry, A. Gelman, I. Verdinelli, and M. West, editors, *Case Studies in Bayesian Statistics IV*, pages 3–57. Springer, New York, 1999.

[31] L. Clark, W. S. Cleveland, L. Denby, and C. Liu. Competitive Profiling Displays: Multivariate Graphs for Customer Satisfaction Survey Data. *Marketing Research*, 11:25–33, 1999.

[32] J. D. LeGrange, S. A. Carter, M. Fuentes, J. Boo, A. E. Freeny, W. Cleveland, and T. M. Miller. The Dependence of the Electro-Optical Properties of Polymer Dispersed Liquid Crystals on the Photopolymerication Process. *Journal of Applied Physics*, 81:5984–5991, 1997.

[33] R. A. Becker and W. S. Cleveland. *Trellis Graphics User's Manual*. Insightful, Inc., 1996.

[34] W. S. Cleveland and C. L. Loader. Rejoinder to Discussion of "Smoothing by Local Regression: Principles and Methods". In W. Härdle and M. G. Schimek, editors, *Statistical Theory and Computational Aspects of Smoothing*, pages 113–120. Springer, New York, 1996.

[35] W. S. Cleveland and C. L. Loader. Smoothing by Local Regression: Principles and Methods. In W. Härdle and M. G. Schimek, editors, *Statistical Theory and Computational Aspects of Smoothing*, pages 10–49. Springer, New York, 1996.

[36] R. A. Becker, W. S. Cleveland, and M. J. Shyu. The Design and Control of Trellis Display. *Journal of Computational and Statistical Graphics*, 5:123–155, 1996.

[37] W. S. Cleveland. *The Elements of Graphing Data*. Hobart Press, Chicago, 1994.

[38] W. S. Cleveland. Coplots, Nonparametric Regression, and Conditionally Parametric Fits. In T. W. Anderson, K. T. Fang, and I. Olkin, editors, *Multivariate Analysis and Its Applications*, pages 21–36. IMS Lecture Notes — Monograph Series, Vol. 24, Hayward, California, 1994.

[39] W. S. Cleveland. *Visualizing Data*. Hobart Press, Chicago, 1993.

[40] W. S. Cleveland. Discussion of "Varying Coefficients Models," by T. Hastie and R. Tibshirani. *Journal of The Royal Statistical Society*, Series B:789–790, 1993.

[41] W. S. Cleveland and R. A. Becker. Discussion of "Graphic Comparisons of Several Linked Aspects" by J. W. Tukey. *Journal of Computational and Statistical Graphics*, 2:41–48, 1993.

[42] W. S. Cleveland, C. L. Mallows, and J. E. McRae. ATS Methods: Nonparametric Regression for Nongaussian Data. *Journal of the American Statistical Association*, 88:821–835, 1993.

[43] W. S. Cleveland. A Model for Studying Display Methods of Statistical Graphics (with Discussion). *Journal of Computational and Statistical Graphics*, 2:323–364, 1993.

[44] W. S. Cleveland, E. Grosse, and M. J. Shyu. Local Regression Models. In J. M. Chambers and T. Hastie, editors, *Statistical Models in S*, pages 309–376. Chapman and Hall, New York, 1992.

[45] W. S. Cleveland. A Science of Graphical Data Display. In R. Penman and D. Sless, editors, *Designing Iinformation for People*, pages 111–121. Communication Research Institute of Australia, Canberra, 1992.

[46] W. S. Cleveland and R. A. Becker. Visualizing Multidimensional Data. *Pixel*, 2(2):36–41, 1991.

[47] W. S. Cleveland and R. A. Becker. Take a Broader View of Scientific Visualization. *Pixel*, 2(2):42–44, 1991.

[48] W. S. Cleveland and E. Grosse. Computational Methods for Local Regression. *Statistics and Computing*, 1:47–62, 1991.

[49] W. S. Cleveland. A Model for Graphical Perception. In *1990 Proceedings of the Section on Statistical Graphics*, pages 1–24. American Statistical Association, Fairfax, VA, 1990.

[50] W. S. Cleveland. Statistical Graphics Software: Core Methods of Exploratory Graphics. Technical report, Bell Labs, 1990.

[51] W. S. Cleveland. The Collected Works of John W. Tukey, 1990. Transcript of talk, *Newsletter, Northern New Jersey Chapter — ASA*.

[52] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning. STL: A Seasonal-Trend Decomposition Procedure Based on Loess (with Discussion). *Journal of Official Statistics*, 6:3–73, 1990.

[53] W. S. Cleveland and D. C. Hoaglin. Challenges for the 90's in Statistical Graphics. In *Challenges for the 90's*, pages 26–27. American Statistical Association, Fairfax, VA, 1989.

[54] R. A. Becker, W. S. Cleveland, and G. Weil. An Interactive System for Multivariate Data Display. In *Proceedings of the 11th Conference on Probability and Statistics*, pages J1–J9. American Meteorological Society, Boston, 1989.

[55] W. S. Cleveland and J. E. McRae. The Use of Loess and STL in the Analysis of Atmospheric CO_2 and Related Data. In W. P. Elliot, editor, *The Statistical Treatment of CO_2 Data Records*, pages 61–81. National Oceanic and Atmospheric Administration, Silver Springs, Maryland, 1989.

[56] W. S. Cleveland. Comment on a Paper of Carlin and Dempster. *Journal of the American Statistical Association*, 84:24–27, 1989.

[57] W. S. Cleveland, R. A. Becker, and G. Weil. The Use of Brushing and Rotation for Data Analysis. In W. S. Cleveland and M. E. McGill, editors, *Dynamic Graphics for Statistics*. Wadsworth and Brooks/Cole, Pacific Grove, CA, 1988. Reprinted in *First IASC World Conference on Computational Statistics and Data Analysis*, 114–147, International Statistical Institute, Voorburg, Netherlands, (1987).

[58] W. S. Cleveland, M. E. McGill, and R. McGill. The shape parameter of a two-variable graph. *Journal of the American Statistical Association*, 83:289–300, 1988.

[59] W. S. Cleveland and S. J. Devlin. Locally-Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association*, 83:596–610, 1988.

[60] W. S. Cleveland, S. J. Devlin, and E. Grosse. Regression by Local Fitting: Methods, Properties, and Computing. *Journal of Econometrics*, 37:87–114, 1988.

[61] W. S. Cleveland. The Graphics Papers of John W. Tukey. In W. S. Cleveland, editor, *The Collected Works of John W. Tukey*, pages 10–49. Chapman and Hall, New York, 1987.

[62] W. S. Cleveland, R. A. Becker, and G. Weil. The Use of Brushing and Rotation for Data Analysis (Graphs Only and with Discussion). In *Bulletin of the International Statistical Institute: Proceedings of the 46th Session*, Book 3, pages 369–386, 1987.

[63] R. A. Becker, W. S. Cleveland, and A. R. Wilks. Dynamic Graphics for Data Analysis (with Discussion). *Statistical Science*, 2:355–395, 1987. Reprinted in *Dynamic Graphics for Data Analysis*, edited by W. S. Cleveland and M. E. McGill, Wadsworth and Brooks/Cole, Pacific Grove, CA.

[64] W. S. Cleveland. Research in Statistical Graphics. *Journal of the American Statistical Association*, 82:419–423, 1987.

[65] W. S. Cleveland and R. McGill. Graphical Perception: The Visual Decoding of Quantitative Information on Statistical Graphs (with Discussion). *Journal of the Royal Statistical Society Series A*, 150:192–229, 1987.

[66] R. A. Becker and W. S. Cleveland. Brushing scatterplots. *Technometrics*, 29:127–142, 1987. Reprinted in *Dynamic Graphics for Data Analysis*, edited by W. S. Cleveland and M. E. McGill, Chapman and Hall, New York, 1988.

[67] W. S. Cleveland, M. E. McGill, and R. McGill. The Shape Parameter of a Two-Variable Graph. In *Proceedings of the Section on Statistical Graphics*, pages 1–10. American Statistical Association, Washington, D.C., 1986.

[68] W. S. Cleveland and R. McGill. An Experiment in Graphical Perception. *International Journal of Man-Machine Studies*, 25:491–500, 1986.

[69] W. S. Cleveland. *The Elements of Graphing Data*. Chapman and Hall, New York, 1985, 1994.

[70] R. A. Becker and W. S. Cleveland. Brushing a Scatterplot Matrix: High-Interaction Graphical Methods for Analyzing Multidimensional Data. Technical report, Bell Labs, 1985.

[71] W. S. Cleveland, P. Grambsch, and I. J. Terpenning. Concurrent and Forecast Seasonal Adjustment and DeForest Extension. Technical report, Bell Labs, 1985.

[72] W. S. Cleveland. Reply to Letter of Stellman. *The American Statistician*, 39:239, 1985.

[73] W. S. Cleveland and R. McGill. Graphical Perception and Graphical Methods for Analyzing Scientific Data. *Science*, 229:828–833, 1985.

[74] W. S. Cleveland and R. McGill. The Many Faces of a Scatterplot. *Journal of the American Statistical Association*, 79:807–822, 1984.

[75] 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.

[76] W. S. Cleveland. Graphical Methods for Data Presentation: Dot Charts, Full Scale Breaks, and Multi-based Logging. *American Statistician*, 38:270–280, 1984.

[77] W. S. Cleveland. Graphs in Scientifc Publications. *American Statistician*, 38:261–269, 1984.

[78] J. M. Chambers, W. S. Cleveland, B. Kleiner, and P. A. Tukey. *Graphical Methods for Data Analysis*. Chapman and Hall, New York, 1983.

[79] W. S. Cleveland. Comments on 'Modeling Considerations in the Seasonal, Adjustment of Economic Time Series' by Hillmer, Bell, and Tiao. In A. Zellner, editor, *Applied Time Series Analysis of Economic Data*, pages 106–119. U.S. Census Bureau, Washington, D.C., 1983.

[80] W. S. Cleveland, A. E Freeny, and T. E. Graedel. The Seasonal Component of Atmospheric CO_2: Information From New Approaches to the Decomposition of Seasonal Time Series. *Journal of Geophysical Research*, 88(C15):10934–10946, 1983.

[81] W. S. Cleveland. Seasonal and calendar adjustment. In D. R. Brillinger and P. R. Krishnaiah, editors, *Time Series in the Frequency Domain, Handbook of Statistics*, volume 3, pages 39–72. North-Holland, New York, 1983.

[82] W. S. Cleveland. Lowess: Robust Locally Weighted Regression for Smoothing and Graphing Data in Two or More Dimensions. *Graphical Techniques for Exploring Data*, 1983. Association for Computing Machinery SIGGRAPH Tutorial 23.

[83] W. S. Cleveland and R. McGill. The Human Graph Interaction. *Graphical Techniques for Exploring Data*, 1983. Association for Computing Machinery SIGGRAPH Tutorial 23.

[84] W. S. Cleveland and R. McGill. A Color-Caused Optical Illusion on a Statistical Graph. *The American Statistician*, 37:101–105, 1983.

[85] W. S. Cleveland, C. S. Harris, and R. McGill. Experiments on Graphical Perception. *Bell System Technical Journal*, 62:1659–1674, 1983.

[86] W. S. Cleveland. Comments on 'Estimating Historical Heights' by Wachter and Trussell. *Journal of the American Statistical Association*, 77:294–295, 1982.

[87] W. S. Cleveland and S. J. Devlin. Calendar Effects in Monthly Time Series: Modeling and Adjustment. *Journal of the American Statistical Association*, 77:520–528, 1982.

[88] W. S. Cleveland, S. J. Devlin, and I. J. Terpenning. The sabl seasonal and calendar ajustment procedures. In O. D. Anderson, editor, *Time Series Analysis: Theory and Practice 1*, pages 539–564. North-Holland, Amsterdam, 1982.

[89] W. S. Cleveland and R. McGill. Communication by Scatterplot. *The Harvard Library of Computer Graphics*, 1982. Harvard Laboratory for Computer Graphics, Cambridge, Mass.

[90] W. S. Cleveland and R. McGill. How to Make a Scatterplot for Data Analysis. In *Proceedings of the Third Annual Conference and Exposition of the National Computer Graphics Association*, pages 1044–1050. National Computer Graphics Association, Washington, D.C., 1982.

[91] W. S. Cleveland, P. Diaconis, and R. McGill. Variables on Scatterplots Look More Highly Correlated When the Scales are Increased. *Science*, 216:1138–1140, 1982.

[92] W. S. Cleveland, C. S. Harris, and R. McGill. Judgements of Circle Sizes on Statistical Maps. *Journal of the American Statistical Association*, 77:541–547, 1982.

[93] W. S. Cleveland and I. J. Terpenning. Graphical Methods for Seasonal Adjustment. *Journal of the American Statistical Association*, 77:52–62, 1982.

[94] W. S. Cleveland. A Reader's Guide to Smoothing Scatterplots and Graphical Methods for Regression. In Launer and Siegel, editors, *Modern Data Analysis*, pages 37–43. Academic Press, New York, 1982.

[95] W. S. Cleveland, S. J. Devlin, D. R. Shapira, and I. J. Terpenning. *The SABL Computer Package: Users' Manual, and The SABL Computer Package Specialized Usage*. Bell Labs Memoranda, 1981.

[96] W. S. Cleveland, S. J. Devlin, and I. J. Terpenning. *The SABL Statistical and Graphical Methods, The Details of the SABL Transformation, Decomposition, and Calendar Methods, The SABL Seasonal and Calendar Adjustment Package, and The SABL Computer Package: Installation Manual, The SABL Computer Package: Dictionary of Routines, and The X-11, X-11 ARIMA, and SABL Seasonal Adjustment Procedures*. Bell Labs Memoranda, 1981.

[97] W. S. Cleveland. LOWESS: A Program for Smoothing Scatterplots by Robust Locally Weighted Regression. *The American Statistician*, 35:54, 1981.

[98] W. S. Cleveland. Author Guidelines for Figures, 1981. Prepared for the style guide of the American Statistical Association journals.

[99] W. S. Cleveland. Challenging Assumptions in Time Series Models. *Technometrics*, 75:479–481, 1980. Comments on a paper of J. Horowitz.

[100] W. S. Cleveland and S. J. Devlin. Calendar Effects in Monthly Time Series: Identification by Spectrum Analysis and Graphical Methods. *Journal of the American Statistical Association*, 75:487–496, 1980.

[101] W. S. Cleveland and R. McGill. Color Induced Distortion in Graphical Displays. *The Harvard Library of Computer Graphics*, 17:7–10, 1980. Harvard Laboratory for Computer Graphics Cambridge, Mass.

[102] W. S. Cleveland and T. E. Graedel. Photochemical air pollution in the northeast united states. *Science*, 204:1273–1278, 1979.

[103] W. S. Cleveland. Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association*, 74:829–836, 1979.

[104] W. S. Cleveland, C. S. Harris, and R. McGill. Circle Sizes for Thematic Maps. *The Harvard Library of Computer Graphics*, 11:40–47, 1979. Harvard Laboratory for Computer Graphics, Cambridge, Mass.

[105] W. S. Cleveland and J. E. McRae. Weekday-Weekend Ozone Concentrations in the Northeast United States. *Environmental Science and Technology*, 12:558–563, 1978. Reprinted in *Preprint Volume*: *Fifth Conference on Probability and Statistics*, American Meteorological Society, Boston, (1977), 112–117. Reprinted in *Air Quality Meteorology and Atmospheric Ozone*, American Society for Testing Materials, (1978), 407–420.

[106] W. S. Cleveland, D. M. Dunn, and I. J. Terpenning. SABL — A Resistant Seasonal Adjustment Procedure with Graphical Methods for Interpretation and Diagnosis. In A. Zellner, editor, *Seasonal Analysis of Economic Time Series*, pages 201–231. U.S. Department of Commerce, Bureau of the Census, Washington, D.C., 1978.

[107] W. S. Cleveland and R. Guarino. The Use of Numerical and Graphical Statistical Methods in the Analysis of Data on Learning to See Complex Random Dot Stereograms. *Perception*, 7:113–118, 1978.

[108] W. S. Cleveland and D. M. Dunn. Some Remarks on Time series Graphics. In D. R. Brillinger and G. C. Tiao, editors, *Directions in Time Series*, pages 112–122. Institute of Mathematical Statistics, Hayward, California, 1978.

[109] W. S. Cleveland. Visual and Computational Considerations in Smoothing Scatterplots by Locally Weighted Regression. In *Computer Science and Statistics: Eleventh Annual Symposium on the Interface*, pages 96–100. Institute of Statistics, North Carolina State University, Raleigh, North Carolina, 1978.

[110] W. S. Cleveland, B. Kleiner, J. E. McRae, R. E. Pasceri, and J. L. Warner. Geographical Properties of Ozone Concentrations in the Northeastern United States. *Journal of the Air Pollution Control Association*, 27:325–328, 1977.

[111] W. S. Cleveland, T. E. Graedel, and B. Kleiner. Urban Formaldehyde: Observed Correlation with Source Emissions and Photochemistry. *Atmospheric Environment*, 11:357–360, 1977.

[112] W. S. Cleveland and R. Guarino. Some Robust Statistical Procedures and Their Application to Air Pollution Data. *Technometrics*, 18:401–409, 1976.

[113] W. S. Cleveland, B. Kleiner, J. E. McRae, R. E. Pasceri, and J. L. Warner. The Analysis of Ground-Level Ozone Data from New Jersey, New York, Connecticut, and Massachusetts: Data Quality Assessment and Temporal and Geographical Properties. In *Proceedings of the International Ozone Conference*. Research Triangle Park, North Carolina, 1976. U.S. Environmental Protection Agency (September 13–16).

[114] W. S. Cleveland, B. Kleiner, J. E. McRae, and J. L Warner. The Analysis of Ground-Level Ozone Data from New Jersey, New York, Connecticut, and Massachusetts: Transport from the New York City Metropolitan Area. In *Proceedings of the 4th Symposium on Statistics and the Environment*, pages 70–90. American Statistical Association, Washington, D.C., 1976.

[115] W. S. Cleveland, R. Guarino, B. Kleiner, J. McRae, and J. L. Warner. The Analysis of the Ozone Problem in the Northeast United States. In *Proceedings of the APCA Speciality Conference on Ozone/Oxidants*, pages 109–120. Air Pollution Control Association, Dallas, Texas, March 11–12 1976.

[116] W. S. Cleveland, B. Kleiner, J. E. McRae, and J. L. Warner. Photochemical Air Pollution: Transport from the New York City Area into Connecticut and Massachusetts. *Science*, 191:179–181, 1976.

[117] W. S. Cleveland, B. Kleiner, and J. L. Warner. Robust Statistical Methods and Photochemical Air Pollution Data. *Journal of the Air Pollution Control Association*, 26:36–38, 1976.

[118] W. S. Cleveland and B. Kleiner. The Use of Some Graphical Statistical Methods in the Analysis of Photochemical Air Pollution Data. In *Proceedings of the 9th International Biometric Conference*, pages 15–34. The Biometric Society, Boston, 1976.

[119] W. S. Cleveland, L. A. Farrow, T. E. Graedel, and B. Kleiner. Chemical Kinetic and Data Analytic Studies of the Troposphere. In *Scientific Seminar on Automative Pollution*. U.S. Environmental Protection Agency Publication, 1975. EPA-600/9-75-003.

[120] W. S. Cleveland and B. Kleiner. Transport of Photochemical Air Pollution from the Camden-Philadelphia Urban Complex. *Environmental Science and Technology*, 9:869–872, 1975.

[121] W. S. Cleveland and D. A. Relles. Clustering by Identification with Special Application to Two-Way Tables of Counts. *Journal of the American Statistical Association*, 70:626–630, 1975.

[122] W. S. Cleveland and E. Parzen. Estimation of Coherence, Frequency Response, and Envelope Delay. *Technometrics*, 17:167–172, 1975.

[123] W. S. Cleveland and B. Kleiner. A Graphical Technique for Enhancing Scatterplots with Moving Statistics. *Technometrics*, 17:447–454, 1975.

[124] W. S. Cleveland and B. Kleiner. Two Graphical Statistical Methods and Their Application to Air Pollution Data. In *Transactions of the Thirty-First Annual Quality Control Conference*, pages 19–28. Rochester Society for Quality Control, 1975.

[125] W. S. Cleveland, T. E. Graedel, B. Kleiner, and J. L. Warner. Sunday and Workday Behavior of Photochemical Air Pollutants in New Jersey and New York. *Science*, 186:1037–1038, 1974.

[126] Bruntz, S. M. and Cleveland, W. S. and Graedel, T. E. and Kleiner, B. and Warner, J. L. Ozone concentrations in new jersey and new york: Statistical association with related variables. *Science*, 186:257–259, 1974.

[127] S. M. Bruntz, W. S. Cleveland, B. Kleiner, and J. L. Warner. The Dependence of Ambient Ozone on Solar Radiation, Wind, Temperature, and Mixing Height. In *Proceedings of the Symposium on Atmospheric Diffusion and Air Pollution*, pages 125–128. American Meteorological Society, Boston, 1974.

[128] W. S. Cleveland. Estimation of Parameters in Distributed Lag Economic Models. In S. E. Fienberg and A. Zellner, editors, *Studies in Bayesian Econometrics and Statistics*, pages 325–340. North-Holland, Amsterdam, 1974.

[129] W. S. Cleveland and P. A Lachenbruch. A Measure of Divergence Among Several Populations. *Communications in Statistics*, 3:201–211, 1974.

[130] W. S. Cleveland. The Inverse Autocorrelations of a Time Series and Their Applications (with discussion). *Technometrics*, 14:277–293, 1972.

[131] W. S. Cleveland. Fitting Time Series Models for Prediction. *Technometrics*, 13:713–723, 1971.

[132] W. S. Cleveland. Projection with the Wrong Inner Product and its Application to Regression with Correlated Errors and Linear Filtering of Time Series. *The Annals of Mathematical Statistics*, 42:616–624, 1971.