

# Andreas Buja

## Liem Sioe Liong/First Pacific Company Professor

### Address:

Statistic Department, The Wharton School  
University of Pennsylvania  
471 Jon M. Huntsman Hall, 3730 Walnut Street  
Philadelphia, PA 19104-6302  
Tel.: (215) 898-8236; Fax: (215) 898-1280  
<http://www-stat.wharton.upenn.edu/~buja/>

### Education:

- Ph.D. in Mathematics/Statistics (1980)  
Swiss Federal Institute of Technology (ETH, Zurich)  
Advisors: F. Hampel, P. J. Huber, H. Foellmer
- Diploma in Mathematics (ETH, 1975, won the ETH award medal)
- Systems Course (Digital Equipment Corporation, Marlboro, MA, 1977)

### Employment:

- Liem Sioe Liong/First Pacific Company Professor,  
Statistics Dept., The Wharton School, University of Pennsylvania, July 2003–.
- Professor,  
Statistics Dept., The Wharton School, University of Pennsylvania, Jan 2002–.
- Technology Consultant (= job rank; not consulting),  
AT&T Labs – Research, Florham Park, New Jersey, 1996–Jan 2002.
- Visiting Faculty (Ph.D. level summer teaching),  
Stanford University, Summer 1997.
- Member of Technical Staff,  
AT&T Bell Laboratories, Murray Hill, NJ, 1994–1996.
- Member of Technical Staff,  
Bell Communications Research, Morristown, NJ, 1987–1994.
- Associate,  
Salomon Brothers Inc, New York, NY, 1987 (first half).
- Associate Professor (with tenure),  
Department of Statistics, University of Washington, Seattle, 1987–1988.
- Consultant, Bell Communications Research,  
Morristown, NJ, 1986 (7 months).

- Consultant, Bell Laboratories,  
Murray Hill, NJ, 1985 (1 week).
- Assistant Professor, Department of Statistics,  
University of Washington, Seattle, 1982–1987.
- Visiting Assistant Professor, Department of Statistics,  
Visiting Scientist, Stanford Linear Accelerator Center,  
Stanford University, Stanford, 1981–1982.
- Research Associate,  
ETH Zurich and Children’s Hospital, Zurich, 1980–1981.
- Visiting Lecturer,  
Stanford University, Stanford, California, Summer 1980.
- Research Assistant,  
Swiss Federal Institute of Technology, Zurich, 1975–80.

**Supervised Ph.D. Students** (*as principal or co-principal advisor*):

- Kartik Ghia (2007, Wharton, UPenn),  
J.P. Morgan.
- Lisha Chen (2006, Wharton, UPenn),  
Yale University.
- Yi Shen (2005, Wharton, UPenn),  
City Group, St. Louis, MO.
- Young-Seop Lee (1999, Rutgers University, New Brunswick),  
Dongguk University, Korea,  
Co-advisor with J. Cabrera.
- Harry Leung (1998, Rutgers University, New Brunswick),  
Baptist University, Hong Kong,  
Co-advisor with J. Cabrera.
- Dianne Cook (1992, Rutgers University, New Brunswick),  
Iowa State University, Ames, IA,  
Co-advisor with J. Cabrera.
- Catherine Hurley (1987, University of Washington, Seattle),  
National University of Ireland, Maynooth, Ireland,  
Co-Advisor with J. A. McDonald.
- Deborah Donnell (1987, University of Washington, Seattle),  
Fred Hutchinson Cancer Research Center, Seattle,  
Co-advisor with W. Stuetzle.

**Service to the Profession:**

- External Review Committee, Department of Statistics, University of California, Berkeley (April 2006).

- Advisory Editor, *Journal of Computational and Graphical Statistics (JCGS)* (2000–).
- Associate Editor, *Journal of the American Statistical Association (JASA)* (2005).
- Reviewer, KDD 2004.
- Managing Editor, *Journal of Computational and Graphical Statistics (JCGS)* (1997–2000).
- Co-Chair/Co-Organizer, with D.F. Swayne and M. Theus, “Workshop on Data Visualization,” sponsored by AT&T Labs, Drew University, Madison, NJ (summer 1997).
- Ad hoc committee on electronic publication of journals, American Statistical Association (1997).
- Reviewer, *Visualization*, IEEE Computer Society and ACM/SIGGRAPH, (1992–1998).
- Associate Editor, *Journal of Computational and Graphical Statistics* (1990–1997).
- Associate Editor, *The Annals of Statistics* (1989–1992).
- Program Committee, IMS, Joint Statistical Meetings, Boston (1992).
- Program Committee, *Computing Science and Statistics*, 23, Seattle (1991).
- Ad Hoc Committee, IMS representative, on founding of a new journal: *Journal of Computational and Graphical Statistics* (1990).
- Co-Chair/Co-Organizer, with P.A. Tukey, W. Stahel, IMA 1989 Summer Program, “Computational Statistics, Robustness and Diagnostics,” U. of Minnesota (1989).
- Co-Chair/Co-Organizer, with W. Stuetzle, “Graphics, Human Interfaces, and Programming Paradigms for Data Analysis,” AMS-SIAM-IMS Summer Research Conference, U. of California, Santa Cruz (1986).
- Associate Editor, *SIAM Journal on Scientific and Statistical Computing* (1983–1985).

### Honors:

- Fellow, Institute of Mathematical Statistics (2006).
- IMS Medallion lecture, Joint Statistical Meetings, New York (2002).
- Keynote speaker, European Meeting of the Psychometric Society, Leiden (1995).
- Fellow, American Statistical Association (1994).
- Award Medal for diploma thesis in mathematics, Swiss Federal Institute of Technology (ETH Zurich, 1975).

### Recent Talks:

- Invited speaker at the Berkeley-Stanford Joint Statistics Seminar in honor of J.H. Friedman’s retirement (April 2006)
- Keynote speaker, SIAM Conference on Datamining (SDM06) (April 2006).
- Joint Statistical Meetings, session on datamining, topic of talk: “Bagging” (August 2006)

**Papers Submitted or Under Revision:**

- “The Plumbing of Interactive Graphics,” Wickham, H., Lawrence, M., Cook, D., Buja, A., Hofmann, H., and Swayne, D.F. (under review)
- “Functional Principal Components Analysis via Penalized Rank One Approximation,” Huang, J.Z., Shen, H., and Buja, A. (under review)
- “Principal Component Analysis of Two-Way Functional Data Using Two-Way Regularized Singular Value Decompositions,” Huang, J.Z., Shen, H., and Buja, A. (under review)
- “Local Multidimensional Scaling for Nonlinear Dimension Reduction, Graph Layout and Proximity Analysis,” Chen, L., A. Buja (under revision)
- “Loss Functions for Binary Class Probability Estimation and Classification: Structure and Applications,” Buja, A., Stuetzle, W., and Shen, Y. (under revision)
- “Sampling/Resampling Methods for Simultaneous Inference with Applications to Function Estimation and Functional Data,” Buja, A., and Rolke, W. (under revision) ([www-stat.wharton.upenn.edu/~buja/PAPERS/paper-sim.pdf](http://www-stat.wharton.upenn.edu/~buja/PAPERS/paper-sim.pdf)).

**Special Status Papers:**

- “Quasi-Darwinian Selection in Marketing Relationships,” Eyuboglu, N., and Buja, A., *Journal of Marketing*, 17 (Oct), 48-62 (2007). (Selected as the JM blog article of the issue and finalist for JM’s 2007 Harold H. Maynard Award.)
- “Visualization Methodology for Multidimensional Scaling,” Buja, A., and Swayne, D.F., *Journal of Classification* 19, 7–43 (2002). (Special invited lead article.)
- “Grand Tour and Projection Pursuit,” Cook, D., Buja, A., and Cabrera, J., *Journal of Computational and Graphical Statistics* 4, 155–172 (1995). (*JCGS* Invited Paper at the 1995 Symposium on the Interface, Pittsburg.)
- “Prosection Views: Dimensional Inference through Sections and Projections,” Furnas, G.W., and Buja, A., *Journal of Computational and Graphical Statistics*, 3, 323–385 (1994). (*JCGS* Invited Paper at the 1994 Joint Statistical Meetings, Toronto, with discussion.)
- “Analysis of Additive Dependencies and Concurvities Using Smallest Additive Principal Components,” Donnell, D.J., Buja, A., and Stuetzle, W., *The Annals of Statistics*, 22, 1635–1673 (1994). (Discussed article.)
- “Linear Smoothers and the Additive Model,” Buja, A., Hastie, T., and Tibshirani, R., *The Annals of Statistics*, 17, 453–555 (1989). (Discussed lead article.)

**Refereed and Solicited Papers:**

- “Data Visualization with Multidimensional Scaling,” Buja, A., Swayne, D.F., Littman, M., Hofmann, H., and Chen, L., *Journal of Computational and Graphical Statistics* (to appear 2008).
- “Boosted Classification Trees and Class Probability/Quantile Estimation,” Mease, D., Wyner, A.J., and Buja, A., *Journal of Machine Learning Research*, 8 (Mar), 409–439 (2007).
- “Observations on Bagging,” Buja, A., and Stuetzle, W., *Statistica Sinica* (special issue on machine learning), 16 (2), 323–352 (2006).
- “Computational Methods for High-Dimensional Rotations in Data Visualization,” Buja, A., Cook, D., Asimov, D., and Hurley, C., *Handbook of Statistics*, eds. E.J. Wegman and C.R. Rao (2005).
- “Visual Comparison of Datasets Using Mixture Distributions,” Gous, A., and Buja, B., *Journal of Computational and Graphical Statistics*, 13 (1) 1–19 (2004).
- “Exploratory Visual Analysis of Graphs in GGobi,” Swayne, D.F., Buja, A., and Temple-Lang, D., refereed proceedings of the *Third Annual Workshop on Distributed Statistical Computing* (DSC 2003), Vienna.
- “GGobi: Evolving from XGobi into an Extensible Framework for Interactive Data Visualization,” Swayne, D.F., Temple Lang, D., Buja, A., and Cook, D., *Journal of Computational Statistics and Data Analysis* (2002)
- “Robust Phoneme Discrimination Using Acoustic Waveforms,” Cvetkovic, Z., Beferull-Lozano, B., and Buja, A., *Proceedings of ICASSP 2002* (May 2002).
- “Data Mining Criteria for Tree-Based Regression and Classification,” Buja, A., and Lee, Y.-S. *Proceedings of KDD 2001*, 27–36 (2001).
- “Sampling Schemes for Model Visualization,” Smith, A.J., Nelder, J. Buja, A., Mallick, Z., Tweedie, L., and Spence, R., *Journal of Computational and Graphical Statistics*, 10, 545–554 (2001).
- “XGobi: Interactive Dynamic Data Visualization in the X Window System,” Swayne, D.F., Cook, D., and Buja, A., *Journal of Computational and Graphical Statistics*, 7, 113–130 (1998).
- “Missing Data in Interactive High-Dimensional Data Visualization,” Swayne, D.F., and Buja, A., *Computational Statistics*, 13 (1998).
- “Manual Controls for High-Dimensional Data Projections,” Cook, D., and Buja, A., *Journal of Computational and Graphical Statistics*, 6, 464–480 (1997).
- “Interactive High-Dimensional Data Visualization,” Buja, A., Cook, D., and Swayne, D., *Journal of Computational and Graphical Statistics*, 5, 78–99 (1996).

- “What Criterion for a Power Algorithm?” in: *Robust Statistics, Data Analysis and Computer Intensive Methods; Festschrift on the Occasion of Peter J. Huber’s 60th Birthday*, H. Rieder (ed.), Springer Lecture Notes in Statistics, 109 (1996).
- “Projection Pursuit Indices Based on Orthogonal Function Expansions,” Cook, D., Buja, A., Cabrera, J., and Hurley, C., *Journal of Computational and Graphical Statistics* 2, 225–250 (1995).
- “Penalized Discriminant Analysis,” Hastie, T., Buja, A., and Tibshirani, R., *The Annals of Statistics*, 23, 73–102 (1995).
- “Flexible Discriminant Analysis,” Hastie, T., Tibshirani, R., and Buja, A., *Journal of the American Statistical Association*, 89, 1255–1270 (1994).
- “Inequalities and Positive-Definite Functions Arising from a Problem in Multidimensional Scaling,” Buja, A., Logan, B.F., Reeds, J.R., and Shepp, L.A., *The Annals of Statistics*, 22, 406–438 (1994).
- “Remarks on Parallel Analysis,” Buja, A., and Eyuboglu, N., *Multivariate Behavioral Research* 27, 509–540 (1993).
- “Dynamics of Channel Negotiations: Tactics, Outcomes, and Contention Level,” Eyuboglu, N., and Buja, A., *Psychology and Marketing* 10, 47–65 (1992).
- “Structural Power in Channel Relationships,” Eyuboglu, N., Buja, A., and Didow, N., *Advances in Distribution Channel Research*, 1, 81–113 (1992).
- “Interactive Data Visualization using Focusing and Linking,” Buja, A., McDonald, J.A., and Stuetzle, W., *Visualization ’91* (IEEE), 156–163 (1991).
- “Painting Multiple Views of Complex Objects,” McDonald, J.A., Stuetzle, W., and Buja, A., *OOPSLA/ECOOP ’90 Proceedings*, 245–257 (1990).
- “Analyzing High-Dimensional Data with Motion Graphics,” Hurley, C.B., and Buja, A., *SIAM Journal on Scientific and Statistical Computing*, 11, 1193–1211 (1990).
- “Remarks on Functional Canonical Variates, Alternating Least Squares Methods, and ACE,” Buja, A., *The Annals of Statistics*, 18, 1032–1069 (1990).
- “Elements of a Viewing Pipeline for Data Analysis,” Buja, A., Asimov, D., Hurley, C.B., and McDonald, J.A., in: *Dynamic Graphics for Statistics*, W.S. Cleveland, M.E. McGill (eds.), Wadsworth, Statistics/Probability Series, 277–297 (1988).
- “On the Huber-Strassen Theorem,” Buja, A., *Probability Theory and Related Fields*, 73, 149–152 (1986).
- “Simultaneously Least Favorable Experiments, Part II: Standard loss functions and their applications,” Buja, A., *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 69, 387–420 (1985).
- “Simultaneously Least Favorable Experiments, Part I: Upper standard functionals and sufficiency,” Buja, A., *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 65, 367–384 (1984).

**Invited Discussions:**

- With W. Stuetzle: On “Evidence Contrary to the Statistical View of Boosting,” by Wyner, A., and Mease, D., *Journal of Machine Learning Research*, 9, 41-46 (2008).
- On “Boosting Algorithms: Regularization, Prediction and Model Fitting,” by Bühlmann, P., and Hothorn, T., *Statistical Science*, to appear (2008).
- On “Tukey’s Paper After 40 Years,” by Mallows, C., *Technometrics*, 48 (3), 327-330 (2006).
- On “Exploratory Data Analysis for Complex Models,” by Gelman, A., *Journal of Computational and Graphical Statistics*, 13 (4), 780–784 (2004).
- On “Additive Logistic Regression: A Statistical View of Boosting,” by Friedman, J.H., Hastie, T., and Tibshirani, R., *The Annals of Statistics*, 28, 387–391 (2000).
- On “Interactive Graphical Methods in the Analysis of Customer Panel Data,” by Koschat, M., and Swayne, D., *Journal of Business and Economics Statistics* (1996).
- On “The Use of Polynomial Splines and their Tensor Products in Multivariate Function Estimation,” by Stone, Ch.J., *The Annals of Statistics*, 22, 171–176 (1994).
- On “Multivariate Adaptive Regression Splines,” by J.H. Friedman, (Buja, A., Duffy, D., Hastie, T., and Tibshirani, R.) *The Annals of Statistics*, 19, 93–98 (1991).
- On “OMEGA: Online Multivariate Exploratory Graphical Analysis,” by Weihs, C., and Schmidli, H., (Buja, A., and Hurley, C.B.), *Statistical Science*, 5, 208–211 (1990).
- On “Graphical Perception: The Visual Decoding of Quantitative Information on Graphical Displays of Data,” by Cleveland, W.S., and McGill, R., *Journal of the Royal Statistical Society, Series A*, 150, 215–216 (1987).
- On “Computers in Statistical Research,” by W. Eddy et al. (Buja, A., Fowlkes, E.B., Kettenring, J.R.), *Statistical Sciences*, 1, 440–442 (1986).
- On “Projection Pursuit,” by Huber, P.J., (Buja, A., and Stuetzle, W.), *The Annals of Statistics*, 13, 484–490 (1985).
- On “Estimating Optimal Transformations for Multiple Regression and Correlation,” by Breiman, L., and Friedman, J.H., (Buja, A., and Kass, R.E.), *Journal of the American Statistical Society*, 80, 602–607 (1985).

**Invited Talks with Proceedings Papers (*unrefereed*):**

- “GGobi: XGobi Redesigned and Extended,” Swayne, D.F., Temple Lang, D., Buja, A., and Cook, D., in: *Computing Science and Statistics*, 33 (2001).
- “Flexible Discriminant and Mixture Models,” Hastie, T., Tibshirani, R., and Buja, A., in: “Neural Networks and Statistics,” J. Kay, D. Titterton (eds.), Oxford University Press (1997).

- “Visualizing the Embedding of Objects in Euclidean Space,” Littman, M.L., Swayne, D.F., Dean, N., and Buja, A., in: *Computing Science and Statistics*, 24, 208–217 (1992).
- “XGobi: Interactive Dynamic Graphics in the X Window System with a Link to S,” Swayne, D.F., Cook, D., and Buja, A., in: *Proceedings of the 1991 Joint Statistical Meetings, American Statistical Association* (1992).
- “XGobi Meets S: Integrating Software for Data Analysis,” Swayne, D.F., Hubbell, N., and Buja, A., in: *Computing Science and Statistics*, 23, 430–434 (1991).
- “A Data Viewer for Multivariate Data,” Buja, A., Hurley, C.B., and McDonald, J.A., in: *Computing Science and Statistics*, 18, 171–174 (1986).
- “Grand Tour Methods: An Outline,” Buja, A., and Asimov, D., in: *Computing Science and Statistics*, 17, 63–67 (1985).

#### Proceedings Papers (*unrefereed*):

- “The Grand Tour via Geodesic Interpolation of 2-Frames,” Asimov, D., and Buja, A., in: *Visual Data Exploration and Analysis, Symposium on Electronic Imaging Science and Technology, IS&T/SPIE* (The Society for Imaging Science and Technology/ International Society for Optical Engineering) (1994).
- “Empirical Comparisons of Neural Networks and Statistical Methods for Classification and Regression,” Buja, A., Duffy, D., Yuhas, B., and Jain, A., in: *Neural Networks in Telecommunications*, B. Yuhas, N. Ansari (eds.), Kluwer Academic Publishers, Norwell MA (1993).
- “An Analysis of Polynomial-based Projection Pursuit,” Cook, D., Buja, A., and Cabrera, J., *Computing Science and Statistics*, 24, 478–482 (1992).
- “Direction and Motion Control in the Grand Tour,” Cook, D., Buja, A., and Cabrera, J., *Computing Science and Statistics*, 23, 180–183 (1991).
- “Discovering Features of multivariate Data through Statistical Graphics,” Buja, A., Fowlkes, E.B., Keramidas, E.M., Kettenring, J.R., Lee, J.C., Swayne, D.F., and Tukey, P.A., *ASA Proceedings in Statistical Graphics*, 98–103 (1986).
- “Designing an Intelligent System for Spectral Analysis,” Percival, D.B., Buja, A., Martin, R.D., Belcher, E.O., Kerr, R.K., Yee, S.D., and Hurley, C.B., *Computing Science and Statistics*, 17, 29–37 (1985).

#### Reports for Company Projects:

- “Exploratory Analysis of a Large Telecom Network,” Abney, S., Buja, A., and Volinsky, C.; AT&T Labs internal report (2004).
- “Finding Clusters in Common Channel Signaling Links with Correlated Failure Patterns,” in: Bellcore Common Channel Signaling Network Notes (1994).

- “An Analysis of Common Channel Signaling Link Outages and Retransmission Data,” Buja, A., and Dunlap, K., in: Bellcore Common Channel Signaling Network Notes (1994).
- “An Empirical Analysis of Link Oscillations in a Common Channel Signaling Network,” Buja, A., and Hung, N., in: Bellcore Common Channel Signaling Network Notes (1992).
- “Methodology for Relational Database Analysis — a Study in Digital Switch Memory,” Buja, A., Duffy, D., and Kagan, J., Bellcore Technical Memorandum (1990).
- “Analysis of Data on Customer Satisfaction with Residence Billing Inquiries,” Buja, A., and Duffy, D., Bellcore Technical Memorandum (1989).

### Data Visualization Software:

- “GGvis,” a plug-in for “GGobi” for graph-layout and multidimensional scaling, successor to “XGvis”.
- “GGobi,” a platform-independent system for interactive graphical data analysis, successor of XGobi (co-designer). Swayne, D.F., Temple-Lang, D., Cook, D., Hofmann, H., and Buja, A. (2000–).
- “XGvis,” a system for visualizing graphs and proximity data with multidimensional scaling (co-designer and co-implementor). Buja, A. Swayne, D.F., Littman, M.L., Dean, N., and Hofmann, H. (1992–).
- “XGobi,” a system for interactive graphical data analysis based on the X window system (co-designer). Swayne, D.F., Cook, D., and Buja, A. (1990–2000).
- “Data Viewer,” a Lisp machine-based prototype system for interactive graphical data analysis (sole designer and implementor) (1985–90).

### Films and Videos:

The first, second and last item are available from the Video Lending Library of the ASA Graphics Section.

- “XGobi: Dynamic graphics for data analysis,” Swayne, D.F., Cook, D., and Buja, A. Video, 20 min., Bellcore (1991).
- “Visualization of quantitative data,” Buja, A., Stuetzle, W., McDonald, J.A., Michalak, J., and Willis S. Video, 30 min., Bellcore and Univ. of Washington (1990).
- “Grand tour methods,” Buja, A., and McDonald, J.A.. Film, 5 min., Stanford University (1984).
- “Finding structure in unstructured data,” Buja, A., and Asimov D. Film, 12 min., Stanford Linear Accelerator Center (1983).
- “Multidimensional scaling in a new environment,” Buja, A. Film, 30 min., Stanford Linear Accelerator Center (1983).

**Edited Book:**

- “Computing and Graphics in Statistics,” Buja, A., and Tukey, P.A. (eds.), *The IMA Volumes in Mathematics and its Applications*, vol. 36, Springer-Verlag (1991).

**Grants** (*while at the University of Washington, 1982–1988*):

- NSF, DMS-8504359, 1985–1988,  
\$33,400 in FY85, \$36,000 in FY86, \$38,000 in FY87,  
“Non-parametric Techniques in Multiple Regression,” with W. Stuetzle.
- DoE, DE-FG06-85-ER25006, July 1985–June 1988,  
\$250,000 in FY85, \$260,000 in FY86, \$270,000 in FY87,  
“Non-parametric Techniques in Multiple Regression,” with W. Stuetzle.
- DoD, FY84/FY85 \$245,920  
“DoD-University Research Instrumentation Program,”  
with R.D. Martin, E. Belcher, D. Percival.
- ONR, N00014-81-k-0095, Jan 1984–Dec 1986, \$139,044  
“Expert Systems and Real-Time Graphics for Time Series Analysis,”  
with R.D. Martin, E. Belcher, D. Percival.
- NSF, MCS-8304484, Aug 1983–Jan 1985, \$30,000  
“Mathematical Sciences Research Equipment,” with R.D. Martin.
- NSF, MCS-8304234, June 1983–Nov 1985, \$24,350  
“Mathematical Sciences: Statistical Methods Based on Real-Time Graphics.”