A joint newsletter of the Statistical Computing & Statistical Graphics Sections of the American Statistical Association.

April 94 Vol.5 No.1

Statistical COMPUTING & GRAPHICS

A WORD FROM OUR CHAIRS

Statistical Computing



Trevor Hastie is the 1994 Chair of the Statistical Computing Section. He follows the able leadership of Sandy Weisberg. Trevor is currently with Bell Laboratory.

This is my first article as Statistical Computing section chair. By the time my year as chair is up I might know enough about section issues to write an informative article about the section. In the meanwhile, let me use this opportunity to blather on about another issue that I seem to discuss with many people.

Members of the Statistical Computing section are most likely applied statisticians—they use computers for data analysis and modelling, and for developing (statistical) software for such tasks. CONTINUED ON PAGE ??

Statistical Graphics



Roy Welsch is the 1994 Chair of the Statistical Graphics Section. This is his first note to the section, and the editors encourage feedback to the section, directly to the chair, or with letters to the editors.

When asked if I would accept the nomination for chair-elect of the Statistical Graphics section, the voice (forever to remain nameless) did not mention that, if elected, I would have to write a column for the Newsletter when I became section chair the following year. Would that have changed my answer? Well, from now on if you are nominated and then elected you know what is in store for you!

CONTINUED ON PAGE ??

FEATURE ARTICLE

The Process of Reading Statistical Maps: The Effect of Color

by Linda W. Pickle and Douglas J. Herrmann

As part of the design process for a national mortality atlas at the National Center for Health Statistics, we are conducting research into how best to design statistical maps that will fairly represent the underlying patterns in the data. To this end, we have developed a cognitive model for map reading that we hope will help focus the map designer's attention on the visual elements that are most critical to successful map reading (Herrmann 1993). We will use this model as the basis for illustrating when the use of color can enhance the design of a statistical map. This article will focus on the design of a classed choropleth map (also called a categorized shaded area map) of death rates, but the concepts can be applied to maps of other statistics or to symbol or line maps as well.

A Cognitive Model for Map Reading

Our model is based on the foundation established by cartographers many years ago (see, e.g., Dent 1993), that is, that successful map reading involves three components: the map itself, the experience and skill of the map reader, and the task to be done using the map. The most common map-reading tasks are reading a rate from the map and geographic pattern recognition, examples of elementary and intermediate tasks according to Bertin's classification (Bertin 1983). Each of these components (the map, the reader, and the task) affects the nature of thought processes (often called cognition) that underly map reading. CONTINUED ON PAGE ??

EDITORIAL

This April 94 issue (Vol. 5 No. 1) of the Newsletter should arrive before the Spring Interface Meeting in Cary, North Carolina, June 15–18, 1994 (although we continue to be amazed at just how many weeks it can take for some subscribers to see their issues!). The second issue for 1994 should arrive before the annual Joint Statistical Meetings in Toronto this August, and the final issue is a year end issue. The deadline for the next issue is June 15, and for the final issue October 15, 1994. Thanks again to all who have contributed.

In this issue we introduce the new chairs of our sections, Trevor Hastie and Roy Welsch. Within days of taking office the chairs became (we hope politely) aware that one of the benefits of being section chair is a regular column in the newsletter. They have both risen to the challenge and written interesting and probably provocative challenges for the profession.

Our feature article is by Linda Pickle and Douglas Herrmann from the National Center for Health Statistics. They discuss problems with and methods for reading statistical maps with a particular focus on the use of color. We have taken this opportunity to introduce some color graphics in the newsletter. If the printed quality is good enough we might be persuaded and cajoled into publishing more color articles in a future issue. We are particularly interested in your reactions to our use of color.

Speaking of color, one of our regular and colorful columnists, Dan Carr, has jumped at the chance to finally get a color plot published. His column discusses color perception and makes some practical suggestions about using color in maps.

A second theme of this issue is hazardous waste. Mark Monmonier writes about spatial resolution in geographic information systems, with a particular emphasis on a hazardous materials survey of New York state. Al Liebetrau describes how photographs of the interiors of hazardous chemical waste tanks at Hanford are transformed and combined to give useful images of the the tanks.

Our other regular columnists have continued their suite of interesting articles. Phil Spector in his Unix Computing column continues his discussion of the Unix shell; Mike Conlon ponders the costs and benefits of uniformity in his Departmental Computing column; and Mike Meyer continues snooping the internet with Mosaic and the WWW tools.

We again wish to encourage submissions of articles to this newsletter. Letters to the editors, short informative pieces, information, and longer articles are all welcome. Send submissions and questions about appropriateness via e-mail to the editors. We encourage submissions in LaTeX(did you realize that more and more journals, including the *American Statistician* are now being set in TeX), but will accept plain ASCII.

James L. Rosenberger
Editor, Statistical Computing Section
JLR@stat.psu.edu

Mike M. Meyer
Editor, Statistical Graphics Section
mikem@stat.cmu.edu



FROM OUR CHAIRS (Cont.). . .

Statistical Computing

CONTINUED FROM PAGE 1

Many other disciplines can also be characterized this way: Pattern Recognition, Image Processing, Machine Learning, Speech and Vision Modelling, and so on. In many of these fields there is considerable overlap with the kinds of things we do as applied statisticians. This section can help to expose its members to these fields, mainly via the programs and special events at the Interface and Annual meetings, through this newsletter, and of course the Journal of Computational Graphics and Statistics, which we cosponsor.

One such related discipline is Neural Networks, which is the subject of this article. Many statisticans I have spoken with dismiss this field as a fad, and complain that they (the Netters) continually reinvent wheels discovered long ago in Statistics. To some extent these sentiments are justified, and exacerbated by the level of *hype* associated with this new field. Statisticians attending Neural Network talks find that some of the reinvented terminology grates a bit: input patterns rather than predictors, bias term rather than constant or intercept, learn rather than fit or estimate. Well, so what—in some cases I wish we had been a bit more exotic in choosing names. I have personally witnessed their rediscovery of the bias/variance tradeoff (this is the more traditional use of "bias") and regularization—concepts well known in Statistics for a long time.

But wait—don't dismiss too quickly! It is true that Netters have been learning statistics on the fly, and at an incredible rate too. They even venture into and reference our journals now (for which they get my admiration!) The more pertinent issue for us is "What can we learn from them?" I believe a great deal.

One could argue that Neural Networks started because of a void created by statisticians. The typical problems they tackle are large-scale prediction problems involving complicated systems, and tons of data and variables. Sixty thousand training observations on hundreds or even thousands of variables is not unusual. Armed with such huge data sets, they can afford to fit correspondingly huge models with thousands of parameters. I think it is safe to say that models of this size are unthinkable to most of us, and our software would not bear up to the task very well.

Their tools are different. One could argue that a feedforward neural network is no more than a highly parametrized nonlinear model. True—but along with their models come clever *online* algorithms that make fitting and pruning such large models feasible. On the negative side, it seems they are willing to use this heavy hammer indiscriminately, in cases where slightly beefed up models from *our* toolbox would do the job better.

I started off a skeptic, but now find I enjoy attending some of the Neural Network conferences. This is a young, dynamic crowd with plenty of very clever ideas, who seem to garner the *most* interesting problems to work on. So rather than be left behind, leave your critical hat at home and have a good look at what they do. You won't return empty handed, I guarantee.

There have been a recent spate of review articles by statisticians that can get you going quite quickly—notably papers by Brian Ripley (*Neural Networks and Related Methods for Classification, (with discussion)*, **JRSSB**, 1994) and Bing Cheng and Mike Titterington (*Neural Networks : a Review from a Statistical Perspective (with discussion)*, **Statistical Science**, 1994).

Enough said—now back to the section. This long rave is intended to entice you to some special Neural Network events that are planned for the coming Interface and ASA meetings.

For the Interface meeting in June, John Sall (our secretary-treasurer, and program chair of the Interface) asked Ron Gallant to organize a session on Neural Networks. Amongst the speakers is Andrew Barron of Yale, a face often seen at both Statistics and Neural networks meetings, and one of the first to study the convergence properties of Neural Net models.



Geoffrey Hinton

At the ASA meeting in Toronto we are hosting a special one-person session—a pioneer and leader of the field: Geoffrey Hinton. Geoffrey Hinton is professor of Computer Science and Psychology at the University of Toronto. He received his PhD in Artificial Intelligence from Edinburgh University in 1978.

He does research on ways of using neural networks for learning, memory, perception and symbol processing and has over 100 publications in these areas. He was one of the researchers who introduced the back-propagation algorithm that is now widely used for practical applications.

Geoff won't get off scott free—John Hartigan of Yale and Dick deVeaux of Princeton will be discussants.

Trevor Hastie
Chair, Statistical Computing Section
trevor@research.att.com

FROM OUR CHAIRS (Cont.) . . .

Statistical Graphics

CONTINUED FROM PAGE 1

I have been doing a bit of reflecting this year which is my 25th at MIT. (At this anniversary, MIT gives you a chair automatically—a wooden one.) As an undergraduate at Princeton working with John Tukey, I had been exposed to a new world of statistical graphics, but I never really was challenged to use these ideas until shortly after I arrived at MIT and was asked to participate is some research at the Harvard Medical School.

Patients in intensive care were being monitored with 20 or 30 continuous data streams. Two problems were presented to me. How to display quickly the status of the patient and how to prevent false alarms (for example, a patient might sneeze) that summoned high-level personnel needlessly? I addressed the second problem with what I knew about robust statistics, but the first problem actually proved to be much more challenging. We tried a number of ideas such as star plots (Welsch, 1976) and crude "control charts" before and after reducing dimensionality with multivariate methods. A few years later Herman Chernoff visited MIT and I saw his

"faces". If only I had that idea when working on the intensive care data! Faces are a lot better that stars for communicating a patient's condition.

What was true in intensive care then, seems to be true almost everywhere today. Business and government monitor everything they can and all of us want to know instantly the health of the patient (and we do not like false alarms). Today I spend major portions of my time in manufacturing and finance and the need to solve these problems is acute. Recently, I was asked for ways to display the "health" of a chip manufacturing line and to help a group visualize the "health" of their financial holdings.

Here are a few challenges for those who create, use, and teach statistical graphics.

We all need to work very hard to figure out how to display uncertainty and risk. Most of our customers assume that the real or implied picture that we present is certain. Those same people would demand some sort of uncertainty measure to go with a statistic. (Many newspapers now at least publish error bounds for political polling.) Most plots and graphs are fuzzy in some sense. How do we convey this? How do we help people visualize uncertainty and variation?

The quality of teaching is a hot topic. Many articles say that we should present material in a variety of ways to appeal to the diversity in our students (and audiences). Graphical communication is essential to improving the productivity (the deans will love you) and quality of our teaching as well as of business and government in general. It is also very hard to do well. We need more case studies that show us the way and explore the use of multimedia technology. For those who do research in graphics, a few case histories telling us how new graphical ideas were created would be fun.

The "information highway" presents new challenges for graphics. With transaction level data available for many financial markets, we need to continuously update the graphics that support financial analytics. Although technology is advancing rapidly, our dynamic displays (especially displays that convey information about uncertainty and risk) will need to pay close attention to the needs of data compression so that rapid updating is possible.

Perhaps you will be stimulated by attending the Interface Symposium in June and especially the tour of the University of North Carolina Graphics and Image Lab (Virtual Reality Center).

If I go on, I will have no material left for the two more Newsletters that are coming this year. That would be enough to panic a section chair and an editor. Let me know what you would like to see in this column and what you would like to see the Statistical Graphics section accomplish this year. Thanks for your help.

References

Welsch, R. E. 1976. "Graphics for Data Analysis", Computers and Graphics, 2, No. 1, pp 31-37.

Roy Welsch
Chair, Statistical Graphics Section
rwelsch@sloan.mit.edu

Toronto Joint Meeting Program

by Sallie Keller-McNulty
Program Chair, Statistical Computing Section

Between our two sections and the JCGS invited paper we have a very full invited program this year at the Joint Statistical Meetings. In addition to all of the invited sessions, there will be over 60 contributed papers and posters presented at the meetings. This includes a Special Topics Contributed Paper Session on "Classification and Clustering," organized by Kenneth Higbee, Battelle–PNL. We also have 12 roundtable luncheon discussions planned. Below you will find the Invited Paper Program and a description of the Roundtable Luncheon Discussions for both the Statistical Computing and Statistical Graphics sections.

Be sure to check the final program for time changes as well as for the times and topics of the contributed paper sessions. Also, don't miss our Sections' Joint Business Meeting and Mixer on Monday night, August 15th. It is sure to feature free food, drink, and door prizes!!

Statistical Computing and Statistical Graphics Invited Paper Program

Title: **Graphics Within Database Systems**
Time: Monday August 15, 8:30 am
Organizer: William F. Eddy, Carnegie-Mellon
Chair: William DuMouchel, Columbia
Speakers:

- An Interactive Image Index, William F. Eddy, Carnegie-Mellon
- Graphical and Query-Based Statistical Data Editing at BLS, Richard Esposito, Kenneth Fox, Dong-yow

Lin, and Kevin Tidemann, BLS
Discussant. Paul Velleman, Cornell and Data Description Inc.

Title: **Estimation for Stochastic Processes through Simulation**
Time: Monday August 15, 10:30 am
Organizer: Katherine Ensor, Rice
Chair: Monnie McGee, Rice
Speakers:

SIMEST, SIMDAT and Pseudoreality, James Thompson, Rice
Monte Carlo Based Statistical Inference for Stochastic Processes, Peter Glynn, Stanford, and Katherine Ensor, Rice
Resampling Methods for Time Series Modeling, Lori Thombs, Univ. of South Carolina
Discussant: Thomas Wehrly, Texas A&M

Title: **Neural Network Models**
Time: Monday August 15, 2:00 pm
Organizer/Chair: Trevor Hastie, AT&T
Speakers:

Neural Network Models, Geoff Hinton, Univ. of Toronto
Discussants: John Hartigan, Yale; Richard deVeaux, Princeton

Title: **Self-Validating Numerical Methods in Statistical Computing**
Time: Tuesday August 16, 8:30 am
Organizer/Chair: William Kennedy, Iowa State
Speakers:

Applications of Numerical Interval Analysis to Obtain Self-Validating Results in Statistical Applications, Ouhong Wang, Iowa State
Errors-Free Least Squares Problems Based on Multiple Homomorphic Images, Sallie Keller-McNulty, Kansas State
Object Oriented Software Support for Numerical Interval Analysis in Statistics, Morgan Wang, Univ. of Central Florida

Title: **Graphical Solutions to Complex Applied Problems**
Time: Tuesday August 16, 10:30 am
Organizer/Chair: William DuMouchel, Columbia
Speakers:

Plotting DNA Sequences, P.M. Leong and S. Morgenthaler, EPFL, Lausanne
Graphical Analysis of Data on Networks, Richard Becker, Linda Clark, and Diane Lambert, AT&T Bell Laboratories
Visualizing Panel Data, Deborah Swayne, Bellcore, and Martin Koschat, Yale

Title: **Space for Statistics in Geographic Information Systems**
Time: Tuesday August 16, 2:00 pm
Organizer: Wayne L. Myers, Pennsylvania State
Chair: Daniel C. Carr, George Mason
Speakers:

Statistics of quality Control for GIS, Russell Congalton, New Hampshire
DataViewer: An Advanced Dynamic Graphics Tool for GIS and Remote Sensing, Emanuel Arbogast, Statistical Sciences
Previsual Statistical Analysis: Prospects for Exploratory Analysis of Images, Wayne L. Myers, Pennsylvania State
Discussant: Colin Goodall, Pennsylvania State

Title: **Computationally Intensive Statistical Methods**
Time: Tuesday August 16, 4:00 pm
Organizer/Chair: James Gentle, George Mason University
Speakers:

Resampling Methods for Tuning Parameter Selection, Christian Leger, Univ. de Montreal
Robust Sample Survey via Bootstrapping and Bias Calibration, Raymond Chambers, Australian National Univ., Alan Dorfman, BLS, and Philip Kokic, Australian Bureau of Agriculture and Resource Economics
Combinatorics and the Bootstrap, Susan Holmes, Unit'e de Biometrie INRA
Discussant: David Banks, Carnegie Mellon

Title: **Image Processing from an Electrical Engineering Point-of-View**
Time: Wednesday August 17, 8:30 am
Organizer/Chair: Victor Solo, Macquarie
Speakers:

Multiresolution Stochastic Models, Alan Willsky, MIT
Image Understanding Using Statistical Methods: Promises, Progress, and Payoffs, Rama Chellappe, Univ. of Maryland

Title: **What Motivates Software Development? A Panel Discussion**
Time: Wednesday August 17, 10:30 am
Organizers: Sallie Keller-McNulty, Kansas State, and Mary Ann Hill, SYSTAT
Chair: Mary Ann Hill, SYSTAT

Panelists:

John Chambers, AT&T Bell Laboratories
John Sall, SAS Inc.
Paul Velleman, Cornell University and Data Description
Leland Wilkinson, SYSTAT

Title: **Journal of Computational and Graphical Statistics Invited Paper**
Time: Wednesday August 17, 2:00 pm
Organizer/Chair: William F. Eddy, Carnegie Mellon University
Speaker:

Projection and Section Views of High-Dimensional Objects: Powers and Limitations of Two Data Visualization Techniques, George Furnas and Adreas Buja, Bell Communications Research
Discussant: Daniel B. Carr, George Mason University

Title: **The Frontiers of Maximum Entropy**
Time: Thursday August 17, 8:30 am
Organizer/Chair: Harry Martz, Los Alamos National Laboratory
Speakers:

Recent Results in Quantum Statistical Inference, Richard Silver, Los Alamos National Laboratory
Recent Results in Maximum Entropy, John Skilling, Univ. of Cambridge
Discussant: David Donoho, Stanford

Title: **Hidden Markov Models**
Time: Thursday August 17, 10:30 am
Organizer/Chair: Karen Kafadar, National Cancer Institute
Speakers:

Hidden Markov Models and their Estimation, Arthur Na'das, IBM
Topics in Hidden Markov Models and their Applications in Speech Recognition, Don X. Sun, SUNY Stony Brook
Hidden Markov Models for Markov Chains Observed with Error, Glen Satten, Center for Disease Control
Discussant: Robert Stine, Univ. of Pennsylvania

Roundtable Luncheon Discussions

Statistical Graphics Section

1. Plotting with Students, Doug Nychka, North Carolina State University.
   By adding a small set of user friendly functions to S one has a powerful tool for teaching students about graphics. We have created a series of labs that introduces the student to modern methods of statistical graphics and simulation. We will share some of our experiences using these labs and describe the advantages of a statistics teaching laboratory.
2. Picture Yourself in School: Getting Statisticians Involved in their School Districts, Jerry Moreno, John Carroll University.
   As the NCTM Standards continue to be implemented, elementary and secondary teachers will be wanting ideas on teaching statistics. Ideas will be shared on how we can help our district's teachers teach statistics, particularly through statistical graphing and by promoting the K–12 American Statistics Poster Competition.
3. Statistical Graphics Section: Spending Section Funds and New Initiatives, David Scott, Rice University.
   The reorganization of ASA has brought new power to the sections together with new income. The discussion will be open for new ideas and priorities and long term direction.
4. Statistical Analysis of Graphical Data: Interface '95, James L. Rosenberger, Pennsylvania State University.
   This roundtable will provide an opportunity to discuss topics of interest to be considered for inclusion in the technical sessions for the 1995 Interface Meeting. Discussion will focus on the new statistical challenges which arise in the statistical analysis of graphical data, deriving inference from graphical displays, and the graphical presentation of inferential data.
5. Statistical Graphics in the Popular Press, Abbe Herzig, Consumers Union.
   Statisticians have devoted much attention to the success of graphics in communicating technical information. What are the elements of a "successful" graph in other environments such as newspaper or magazine publishing, where the graphs serve additional functions such as entertainment, enticement, and decoration?
6. Graphical Displays for Multiple Comparisons, Mario Peruggia, Ohio State University.
   What are the pros and cons of the techniques currently implemented in statistical packages? What inferential and perceptual tasks do they perform? What are suitable criteria to measure their effectiveness and is there a need for new, improved representations?

Statistical Computing Section

1. Symbolic Computation in Statistics, David Andrews, University of Toronto.

   Participants will be encouraged to discuss current work in symbolic and hybrid (symbolic and numerical) computing. The completeness of the current collection of fundamental algorithms useful in statistics will be reviewed; gaps will be noted. The potential and requirements of coordinated algorithm libraries will be explored.

2. Smoothing Multidimensional and Spatial Data, Katherine Hansen, Sandia.

   The smoothing of spatial and other multidimensional data will be discussed. Topics may include algorithms for smoothing various types of data, applications of multidimensional smoothers in a variety of fields, and the performances of different smoothers in comparative studies.

3. Parallel Computing - Is is Worth the Effort?, Janice Hartwick, University of Michigan.

   Discussion will focus on the software costs associated with making effective use of parallel architectures. Problem areas include matching the parallelism and communication characteristics of an algorithm to those of the architecture, and the difficulties in porting code without sacrificing performance.

4. Neural Networks, Trevor Hastie, AT&T Bell Laboratories.

   There is this huge, very active and volatile area of research called Neural Network Modelling. Neural Networks are used, mostly by engineers, to solve every problem under the sun, typically using, but not restricted to, huge amounts of data, and often with great success. Many statisticians dismiss this field as a fad, especially because they appear to be reinventing many old statistical wheels. This dismissal is dangerous — both fields can learn a lot from each other. Lets talk about it!

5. Computational Issues in Time Series Analysis, Joseph Newton, Texas A&M.

   Among the topics that could be discussed are: 1) monte Carlo methods for time series analysis (e.g. the Gibbs sampler), 2) computing for nonlinear time series, 3) graphical methods for time series analysis, 4) resampling methods (bootstrap, jackknife), 5) time series analysis in statistical packages.

6. Hidden Markov Models, Robert Stine, University of Pennsylvania.

   This roundtable focuses on the use of Hidden Markov Models (HMM) in statistical applications. HMM provide an effective methodology for tackling a variety of difficult computational tasks and unify methods from various disciplines. Some applications to be discussed include speech recognition, dynamic time warping, missing data models, and switching regression models.



DEPARTMENTAL COMPUTING

Departmental Computing: Cost/Benefit of Uniformity

by Michael Conlon

Multicultural computing?

There was a golden age of desktop computing (5 years ago?) when users could choose their own computing platform—UNIX, PC or Macintosh and be blithely unaware of what was going on in the next room. Each user had their own little dot matrix printer and had no need to computationally interact with the other people in their own department. Sharing of documents was done on diskette, and as long as your floppy could read my diskettes, we could work together.

Those days are gone.

We now expect computers to communicate with each otherand software to operate across platforms

Now we expect that all users in the department can send and receive e-mail, transparently access each other's files across the network, access data on an enterprise mainframe, step in to run data analysis software and work on each other's documents.

These developments have led many departments to standardize on one computing platform. Some departments have had good consensus for standardization without much discussion. Others have had pitched battles over appropriate platforms.

Each platform has merits

I am currently the Director of Information Resources for my College of Liberal Arts and Sciences. We have 22 academic departments and an additional 20 centers and institutes. In this capacity I advise departments on networking and desktop computing. I try to help departments design systems that will help them get their

work done. I see systems built around UNIX, around Macs and around PCs (Windows and Netware).

The UNIX people have obvious advantages. They have powerful machines, built from the ground up for multi-processing and networking. They have powerful statistical tools, excellent graphics and access to a huge array of free software of the highest quality. UNIX systems require system management, the more talented the better.

Macs have always been very strong in user interface—both ease of use and consistency. Macs are trivial to set-up, trivial to network and do not require local management. With the recent price reductions and the announcement of the Power-Mac series, Macintosh computers may now be the most cost effective computers around. They are excellent for multi-media work.

PCs have strength because of their installed base. It is easy to find people who have PC experience and can install and operate PC networks and use PC software. Recent advances in Netware promise the ability to manage accounts and resources in much the same way that UNIX people have done for years. PCs have advantages in commercial software.

When I think about platform preferences (and I have my own preferences), I try to keep four things in mind:

- No platform uniformly dominates the others. Where one has strength the others may have comparative weaknesses.
- Many people approach platform choice with the fervor of religion. Can we all just get along?
- Software is the key. Power is coming to all platforms. The PowerPC and other advances in RISC architecture will boost UNIX and Mac performance. Pentium advances will recharge PCs. Multiprocessing is coming to all platforms.
- People can change platforms. I see it all the time. The argument "I have to have X, because that's what I know" just doesn't make sense.

Why Uniformity

So if all the platforms have merit, can't each user just have whatever they want? To help answer this let's consider the benefits of uniformity.

Easier to provide basic services. Unlike those good old days of desktop computing, modern office systems are expected to interoperate. People need to work together, using their computers. We expect to be able to provide network printing, transparent file access, electronic mail, cooperative scheduling, work group document production, work group data management and analysis. It is far easier to provide these basic services if there is just one platform to support.

Easier to manage. Less manpower is needed to keep a uniform system operational. There are fewer vendors to deal with, fewer software systems and fewer different kinds of equipment to be installed and repaired. Managers often seize on this as the single most important reason for imposing uniformity. Imposing uniformity can be a big mistake. Users must feel a part of the decision making process if their computing platform is to be changed.

Uniform computer systems are easier to manage but that same uniformity might stifle innovation.

Reduce training. If there's one system, there can be one training program. One set of manuals, one set of short courses. Users can help each other, since they are all using the same kind of system. In the stand-alone desktop days, users were often thwarted from helping each other because different software or different versions of the same software was loaded on each machine.

Improve interaction. A good reason (besides reducing costs) for uniformity is to foster interaction between the users of the system. When people are using different kinds of systems it may be difficult for them to work together on projects and difficult for one person to step into another person's project.

Why not uniformity?

There are three consistent arguments against uniformity. I have some sympathy with each.

Requires retraining. If a department adopts a standard, uniform computing environment (and that means not only a standard hardware platform, but standard software tools as well) people will need to be retrained. Data analysts typically have little trouble adapting to a new environment. Clerical people may require more attention (and the supervisor may require the most attention).

No platform lasts forever. There was a time before DOS and there will be a time after DOS and after Windows. There will be a time after System 7 and a time after X windows. In my computing career I have used dozens of different computing environments, some more productive (and more productively) than others. Change is possible.

To change requires more than retraining. It requires "buying in" to the process. People must understand the benefits of the change (and not just that it will save a

manager a few bucks in a maintenance budget). There should be real human benefits.

Change also requires conversion of documents and data. Adequate resources must be planned for such conversions which are often not trivial. The situation is improving for such activities. Most systems provide ways of getting data from proprietary internal formats to standard interchangeable formats which can then be reloaded in new systems.

Benefits overstated. I often hear, in response to discussions of department systems, "I don't need X." Where X may be network printing, backup, cooperative computing, electronic mail or Internet access. Some people have worked alone for a long time and they see no reason to invest in retraining for the dubious pleasure of having the ability to work with others electronically. These voices are becoming increasingly rare. The benefits to the organization of having its members on-line are becoming overwhelming. The access to information (both local and world-wide) provided by the network is particularly compelling.

Stifles innovation. A uniform system, once in place, may stifle innovation. I have often heard "She can't have X, because it's different from what we have." Real innovation must proceed in computing environments. We didn't invest in desktop computer networks only to repeat the mainframe mistakes of the past. There must be opportunities for innovation. Users who want to try new things must be encouraged and supported. But careful consideration must be made of whether the changes are innovations, or regressions. Do users want something different because "that's what they know?" Such requests are not innovation.

Innovation comes in the form of hardware and software. Someone may want to try a new generation processor that offers significant speed increases. Another may wish to try new data analysis software that enables the development of dynamic graphics. The system staff may wish to try new networking gear that can increase the speed of the network by a factor of 10. Such efforts must be coordinated. Productivity is lost when each goes off to do their own thing. But massive opportunities for department improvement can be lost when innovation is discouraged under the heading of uniformity.

Convergence?

There are signs of convergence in the desktop world. We've seen Macs running Windows, UNIX machines running Mac programs, and all systems running TCP/IP networking. The situation with regards to support of multi-cultural computing is improving.

Uniformity remains an important concept in the design and deployment of departmental systems, but uniformity at what price?

Michael Conlon
Department of Statistics
University of Florida
mconlon@stat.ufl.edu

GEOGRAPHIC INFORMATION SYSTEMS

Spatial Resolution, Hazardous Waste Siting, and Freedom of Information

by Mark Monmonier

Geographic information systems support a formal process of site selection based on overlays of electronic maps. The process lets a locational analyst consider a variety of geographic factors, some essential for the planned facility and others merely desirable. As described more fully in my first GIS column (*Statistical Computing and Statistical Graphics Newsletter*, November 1990), the analyst divides the region into a grid of square cells and then rates each cell for every factor; the grid overlay of ratings for a specific factor is called a *coverage*. For a binary coverage each cell is coded either 1 (to indicate the factor's presence) or 0 (to represent its absence). Some coverages have three or more favorability categories. Soil permeability, for example, might have five categories represented by an integer scale ranging from 1 (very permeable) for loose sandy soil and alluvial gravel to 5 (highly impermeable) for most clays. Site selection typically proceeds in two stages: an *exclusionary screening*, in which the GIS identifies cells unacceptable because of state environmental laws or other reasons, followed by *candidate area identification*, in which the system computes a summary favorability score for each remaining cell by accumulating for all coverages the product of the cell's rating and the coverage's weight. Each coverage is assigned a weight representing its relative importance so that cells with the highest scores are arguably better sites than cells with lower scores. When the facility requires a site larger than a single cell, the analyst typically estab-

lishes a *minimum favorability score* and focuses further attention only on sufficiently large groups of contiguous cells that meet or surpass this threshold. GIS-based siting can be rigorously systematic, but because of subjective judgments about categories, ratings, weights, and thresholds, the technique is hardly objective.

Finding a Disposal Site

However subjective, GIS is an appealing strategy for an agency seeking a rational, unbiased location for a locally objectionable facility such as an incinerator or a landfill. In 1986, under pressure from Congress to choose an in-state disposal site for low-level radioactive waste (LLRW) generated within its borders, the New York legislature set up an LLRW siting commission to choose a disposal site after careful analysis and public hearings. The siting commission apparently got the idea for using a GIS from a ten-page Nuclear Regulatory Commission pamphlet that noted a GIS could be helpful but said nothing about cell size or data quality.

Rather than hire a large number of temporary workers, the commission advertised for an environmental consulting firm to design a suitable GIS siting process, develop coverages, select 10 comparatively large candidate areas for staff review and public comment, and narrow this selection to no more than four potential sites. Several firms submitted bids, and in accord with state law the commission gave the job to the qualified contractor who submitted the lowest bid. Lacking adequate expertise in spatial analysis and apparently blinded by expediency, the LLRW siting commission accepted the contractor's proposal to use a grid of 640-acre cells (one-mile squares) for statewide exclusionary screening and candidate area identification and a more detailed GIS with 40-acre cells (squares a quarter mile on each side) for selecting potential sites within the ten candidate areas. Although 640- and 40-acre cells minimized data acquisition costs and helped the commission meet deadlines, these grids were too coarse to capture significant local variations in soils, land use, and surficial geology.

Pitfalls and Challenges

Inadequate spatial resolution was not the only flaw in the LLRW siting commission's GIS: several of the coverages were irrelevant or erroneous. One spurious factor was precipitation, which had been blamed for the horizontal migration beyond the dumpsite of hazardous materials buried in shallow trenches that filled with water and overflowed. This "bathtub effect" is a relevant siting factor only in states like Washington, where the climate ranges from rain forest to desert. All of New York State is fairly humid, and even places with a relatively low average rainfall have occasional wet years. Yet the contractor dutifully digitized a precipitation coverage, for which class breaks at 40 and 50 inches divided the state into low, medium, and high categories. Equally outrageous was the highly generalized source map, on which smooth, nicely rounded isohyets based on 105 weather stations conveniently ignored the regional and microclimatic influences of lakes, cities, and rough terrain. Another faulty coverage, intended to minimize transport costs and reduce the risk of highway or rail accidents, measured the relative accessibility of each 640-acre cell to the centers of the state's 62 counties. Although weighting county-to-cell distances by the amount of low-level radwaste generated in each county seems reasonable—calculations were based on presumably precise reports by individual generators to state regulators—straight-line distance calculations ignored such significant obstacles to direct transport as the Adirondacks, the Catskills, and the Finger Lakes. Despite the appearance of technical precision, the GIS had the integrity of snake oil.

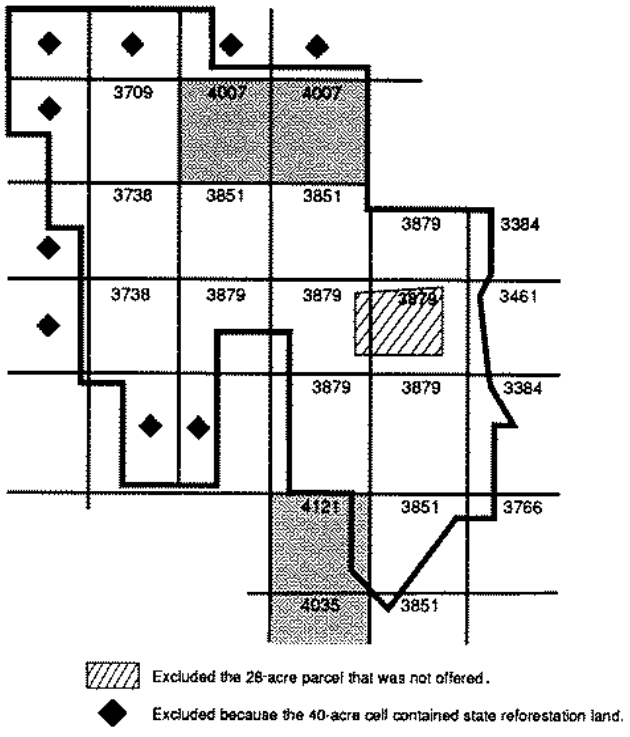


Figure 1: The volunteered Taylor North Site.

An LLRW dump raises fears based on uncertainty over the safe disposal of radwaste as well as widespread suspicion about government bureaucracy and a discredited nuclear establishment. At public hearings in counties with candidate areas, local residents voiced the skepticism and hostility typical of NIMBY (not in my back-

yard) controversies. After the search narrowed to five potential dumpsites, concentrated in two rural counties, protest groups and local officials not only challenged the findings but deconstructed the entire siting process. In Cortland County, after detailed maps prepared by town officials, the county health department, and local activists suggested that the siting commission had violated its own guidelines by accepting as one of the five finalists an otherwise unacceptable "volunteered site." Senator Alfonse D'Amato asked the U.S. General Accounting Office to audit the GIS analysis.

What the GAO saw, it didn't like. In an August 1992 report euphemistically titled "Nuclear Waste: New York's Adherence to Site Selection Procedures is Unclear," federal auditors documented a number of irregularities, including a potential site (the volunteered Taylor North site) that clearly failed to meet the siting commission's ostensibly objective criteria. According to the GAO,

> For two reasons, this decision was inconsistent with the commission's requirement that an offered site must be 'at least as good as' other sites under consideration at the time. First, Taylor North contained more than 5 acres of agricultural land in active production, contrary to state regulations that prohibited a low-level radioactive waste disposal site from containing more than 5 acres of such land. Second, the site scored below the minimum score that the commission's staff had established to identify promising sites for further consideration.

According to the siting commission's guidelines, an acceptable site must have at least five contiguous cells scoring 3900 or higher, but as a map in the GAO report illustrated, only two pairs of contiguous cells met this criterion (shaded in Figure 1). Moreover, the map revealed another irregularity: the so-called volunteered site surrounded a 28-acre tract owned by another landowner not eager to sell.

The Public's Right to Know

New York's LLRW siting controversy also raised issues of freedom of information, openness in government, and public access to data. The siting commission rejected the request of Concerned Citizens of Allegany County (CCAC) for a copy of the GIS data. CCAC sued, and a state judge ordered the commission to hand over data tapes. Although the siting commission and its contractor complied with the court order, CCAC's science advisors at a local college soon discovered the tapes were unreadable. Developed with the contractor's proprietary software, the geographic data were incompatible with ArcInfo or any other widely available GIS software. If local people wanted to interrogate the database, they would have to travel to the contractor's offices in Seattle. Embarrassed by this discovery, the New York legislature revised the radioactive waste siting law to ensure effective citizen access. Although the siting commission eventually acquired its own in-house GIS and a small support staff, by that time residents of the affected areas were no longer interested—in responding to the public outcry, the legislature had told the commission to overhaul the entire siting process, and begin anew, this time by choosing an appropriate disposal method (including above-ground monitored retrievable storage) before selecting a site.

Who was fooling whom? A lot is written these days about empowerment and biased representations. According to the postmodern deconstructionist's would-be orthodoxy, nontechnical people are powerless—or at least weak and poorly prepared to defend themselves—against the scientific-technical experts and propaganda of government, corporations, and other elites. Rubbish! Although the "little people" don't always win, they know what they don't like and when they are being manipulated. In land-use controversies with high uncertainty and grim consequences, local people will either find their own experts or develop their own expertise—if they don't already have it. Attentive citizens and pubic officials in two of New York's more rural counties successfully fought an attempt to impose by artifice, not negotiation, a potentially hazardous facility they didn't want. They didn't like the GIS consultant's data or the siting commission's conclusions, and in challenging the analysis, they won largely by embarrassing a high-level state commission. Moreover, these farmers, homemakers, school teachers, factory workers, and local officials demonstrated convincingly that graphic and numerical representations are weapons much like grenades—quite lethal when tossed back at an unwary invader. If anyone was fooled by the GIS, it was the appointed members and staff of the LLRW siting commission, who trusted their contractor and a technically correct methodology, and New York's governor and legislature, who trusted the commission and its self-proclaimed objectivity.

Mark Monmonier
Geography Department
Maxwell School of Citizenship and Public Affairs
Syracuse University
mon2ier@mailbox.syr.edu

The Process of Reading Statistical Maps: The Effect of Color

CONTINUED FROM PAGE 1

Unfortunately, while this traditional framework specifies the general variables that affect map reading, it does not specify just how these variables exert their effect. The model proposed here does specify the impact of these variables and thus can provide practical guidance for map design.

In order to move beyond this basic foundation, we need to better understand what it means to "read a map". First, we assume that attainment of map reading goals requires successful completion of four cognitive stages:

- Map Orientation—What geographic area does the map represent? What is the map designed to communicate?
- Legend Comprehension—How are the values of the mapped statistic represented?
- Map/Legend Integration—Match the legend scale with the map itself.
- Extracting Information from the Map—Answer the questions asked of the map. Among others, these could include: What are the (approximate) data values associated with specific locations? How does the value for an area compare to a referent value (e.g., U.S. rate)? Which locations have the highest rates? Is there any apparent spatial trend of the mapped statistic?

Each of these cognitive stages can be shown to depend on a series of processes, including sensation, perception, comprehension, encoding of information into memory, memory retrieval, and reasoning. The particular processes involved differ from stage to stage. In the case of map orientation, the visual characteristics of different areas must be resolved into patterns of figure and background corresponding to geographic units with different rates. This stage is dominated by sensory, especially visual, and perceptual processes. To understand the legend, one must be able to identify how many categories are being used as well as to remember the ordering of the categories. This stage is dominated by sensory and perceptual processes as well, but also by linguistic comprehension of the category labels and encoding of this information into memory. Integrating the map with the legend involves applying the symbolization of the legend to the map itself. This requires retrieval of what was remembered about the legend and perceptual matching of the visual image of particular category scales with the visual images perceived on the map.

Finally, to extract information from the map, one must compare visual properties of the map, making use of prior knowledge about the legend symbology and the geographic units depicted on the map. This final stage makes use of cognitive processes that vary from simply reading a rate from the legend or memory of the legend corresponding to a particular geographic unit, to the more complex process of establishing rules for clustering neighboring geographic units for geographic pattern recognition.

Graphical Elements of Map Design

Graphical elements available to the designer of a statistical map are: hue, saturation and brightness (or intensity) of color; symbol type and size; and texture (e.g., hatch patterns). This is an adaptation of Bertin's list of "visual variables" for graphics, with shape and orientation variables eliminated from the original scheme as inappropriate for statistical maps (Bertin 1983). Note that color can be represented in several ways; we use the hue/saturation/brightness scale to illustrate the concepts. The reader is referred to Monmonier's recent books (1991, 1993) for an in-depth discussion of how each of these variables may be applied to map design.

The Conventional Approach

Cartographers and others have long recommended using an ordered scale for quantitative maps (Dent 1993; Monmonier 1991, 1993). Our cognitive model outlined above suggests that this has been a successful strategy because the ordered legend is readily understood and remembered, and can be integrated with the map to extract the necessary information more quickly than a less intuitive scale which requires the map reader to repeatedly refer back to the legend. Of the visual variables listed above, saturation and brightness of color, symbol size, and some choices of texture (e.g., increasingly dense cross- hatched lines) are generally recognized as having a natural ordering. Although hues technically can be ordered by wavelength, studies have shown a great deal of variability in how people order different hues, so hue is not recommended when the ordering of the underlying mapped statistic is important (Kosslyn 1994).

The New Cognitive Approach

The cognitive model identifies the importance of visual discriminability of the rate categories. That is, the

shades or patterns that represent each rate class must be clearly identifiable, not only on the legend, where the category boxes are ordered, but also on the map, where there is no predetermined order. Shades that are too similar lead to difficulty integrating the legend and map, even when separately the legend and the map appear to be clear.

We will illustrate problems that can occur when the map design interferes with the map/legend integration stage by using two maps (Figures 1 and 2). These examples demonstrate the impact of visual variable ordering and discriminability on map/legend integration with and without the use of color. Each example is a map of colon cancer death rates among white males during 1980-1989, age-adjusted to the 1960 U.S. population. The basic geographic unit is the Health Service Area, 802 areas defined according to where residents obtained their routine hospital care (Makuc, 1991). Rates were categorized into seven classes defined by the following percentile cut-points: 10%, 20%, 40%, 60%, 80%, 90%. This is a simple quintile classification with the two extreme categories further split to identify outliers in the rate distribution. Maps were produced using AtlasPro (Strategic Mapping, Inc.) under DOS on a 386 PC. The percent saturations were chosen to give the impression of equal differences in saturation among the categories (Kimerling 1975, 1979).

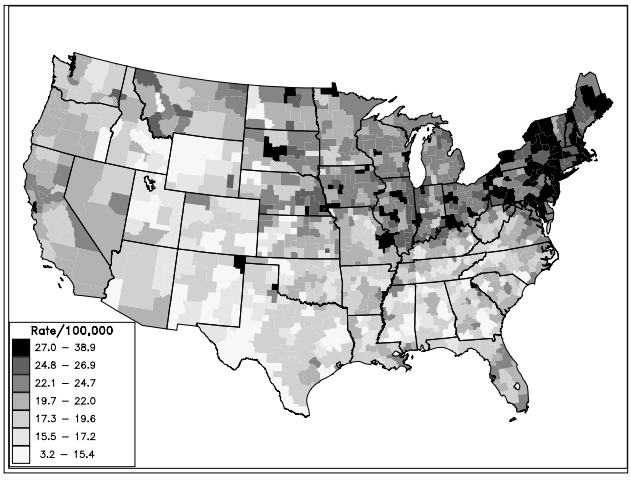


Figure 1: Monochromatic (gray scale) map.

The Monochromatic Map

Tufte (1983) and others have argued that a monochromatic (gray) scale map is adequate for choropleth statistical maps and avoids the expense of color printing for map publication. As shown in Figure 1, there is a clear progression of dark to light shades corresponding to high to low rates. Unfortunately, the discriminability of the gray shades is poor. Because white shading appears so much brighter than any of the gray shades, its use as a category shade can confuse some readers, who are unsure whether black or white represents the highest rate. Therefore choices of saturation in Figure 1 are restricted to roughly 5% to 100%, which permits an average of 15% difference in saturation for a seven-category classification. Even with the best output device, the human eye has trouble distinguishing shades that are this close in saturation.

To test yourself, use Figure 1 for both a rate-readout and a pattern recognition task. First, find your home town on the map. Can you tell what the range of rates is in this place? You probably found that the ease of determining the exact shade of gray depended on the surrounding area shades. This problem of map/legend integration for a rate-readout task can be partly alleviated by providing a "legend slider", a legend on a separate page that can be moved across the map surface, like a small ruler, to help match an area's shade with the legend shades. Now look at Figure 1 again and identify the region of the country with the highest rates. Is there an approximate spatial trend in rates across the country? You probably found that this was easier than the rate-readout task because broad geographic patterns of dark, middle, and light shades can be identified quickly without determining exactly which of the seven shades is in a specific area. This nearly instantaneous "visual impression" of broad patterns can be a drawback, as Tukey pointed out (Tukey 1979); if the geographic units are of greatly varying size, the eye is drawn to the larger areas, although often the small areas have greater population. A recent experiment suggests that the size of the geographic unit affects the perception of clusters of areas with similar rates (Lewandowsky 1993). An advantage to using the Health Service Areas for our death rate maps is that they are more homogeneous in area than counties or other proposed county aggregations.

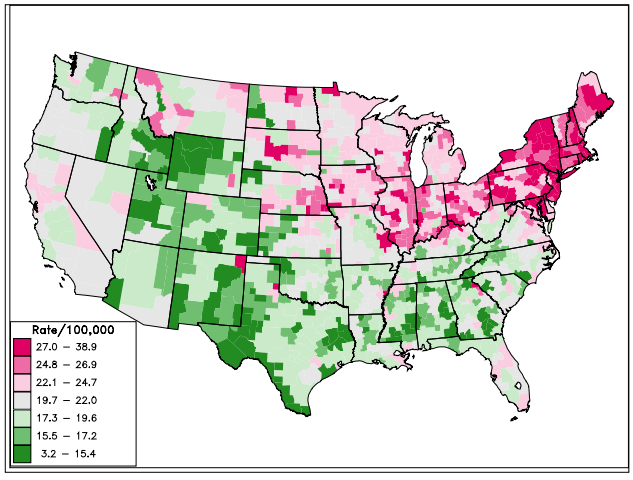


Figure 2: Double-ended chromatic map.

The Chromatic Map

Figure 2 shows how the use of color can alleviate the discriminability problem. Try to answer the same questions that you did for Figure 1. In this map, areas with rates above the 60th percentile are red, areas with rates between the 40th and 60th percentiles are gray, and areas with rates below the 40th percentile are green. For each hue, there are no more than three levels of saturation so these can be chosen to be clearly different (18%, 55%, and 100% in Fig. 1). This example uses a "double-ended" color scheme, where the saturation for the two hues is greatest for rates at the extremes of the rate distribution. A hue/saturation combination where the saturation is greater for higher rates (e.g., less to highly saturated green, followed by less to highly saturated brown as rates increase from lowest to highest) has been used successfully by the Bureau of the Census (U.S. Dept. of Commerce 1990). Of course, the problem of the perceived order of the hues remains, but the reader needs only to remember two hues, not seven.

Two-dimensional Coding: Hue and Saturation

Next we will illustrate the effects of map design on the stage of extracting information from the map. Color has long been used for bivariate mapping but experimentation has suggested that most readers struggle to understand this format (Wainer and Francolini 1980). We would argue from our cognitive model that the problem has not been the use of color per se, but the way in which it has been used. Specifically, the use of different hues to convey an ordering on a map is not recommended because of the reader's uncertainty about the intended order of the hues. However, hues can be selected that are highly discriminable and so can be used to convey a qualitative (or equivalently a dichotomized quantitative) characteristic of the mapped statistic (Antes and Chang 1990). Note that we are proposing to superimpose on a single map information about two characteristics of the same variable (e.g., a death rate), rather than information about two different variables.

Most readers struggle to understand the use of color in bivariate mapping . . . especially when different hues are used for an ordering

We noted above that of the visual variables available for statistical mapping, hue, symbol type, and some textures were not generally perceived as having a natural order and so are poor choices for representing a quantitative variable if rank is important. These visual variables can, however, be useful as indicators of qualitative information. For example, the double-ended color scheme in Figure 2 conveys two pieces of information about each place: its approximate rate (within a range), and whether that rate is above, below, or about the same as the median rate. This design allows the reader to perceive general patterns by mentally grouping similar saturations for one hue, but also permits rate readout because of the highly discriminable hue/saturation combinations. Figure 3 at first seems at odds with our recommendation that hues should not be used for a quantitative ordering. However, if our primary task for a map is rate readout, we want the most distinct shades/patterns, which can be accomplished by using a multi-hue design. Saturation, dichotomized to two distinct levels, can then be used to indicate whether the rate is unreliable. This style was used in the Canadian cancer atlas (Statistics Canada, 1991). Because no variance estimates were available for our colon cancer data, we arbitrarily considered rates based on fewer than 50 deaths to be "unreliable" in Figure 3. Category cutpoints are quintiles of the rate distribution. This design works well for rate readout tasks because of the clear discriminability of the 10 hue/saturation combinations. However, the memorability of the legend hues is poor, which leads to problems integrating the legend and map for clustering and pattern recognition tasks.

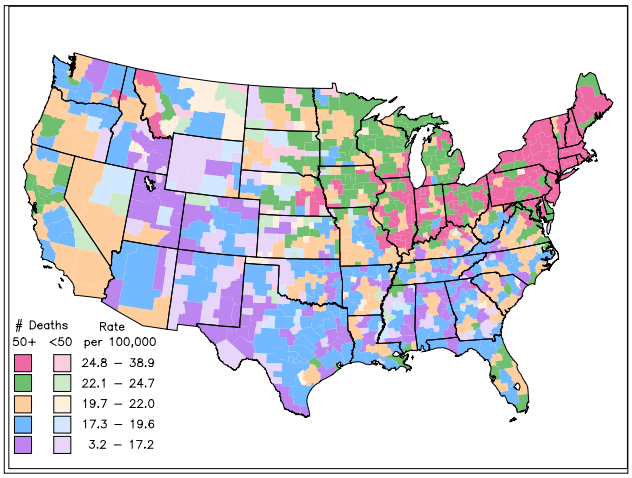


Figure 3: Multi-hue map with reduced saturation for "unreliable" rates, defined to be rates based on < 50 deaths.

A related design often suggested for a map of rates and their statistical significance is to use increasing saturation to represent increasing rates, then to overlay a single symbol (e.g., *) on those places where the rate is significantly different from the referent rate (e.g., the U.S. rate). As with any symbol map, this design breaks down when there are hundreds of small geographic units on the map, as is the case in Figure 3. For this situation, the use of color to show significance or rate reliability

is an improvement over the use of symbols, which can completely mask the underlying shading of small areas.

Discussion

This article has shown how the effects of color can be analyzed by the use of a cognitive model of the map reading task. The model discussed here may be seen as a response to calls for the development of systematic research programs in graphic communication (Kosslyn, 1985; Cleveland, 1993; Simkin & Hastie, 1989). The model has also been influenced by research on the cognitive approach to cartography (Barrell & Cooper, 1986; Blades & Spencer, 1986; Eastman, 1985; Gilmartin, 1981; Griffin, 1983; Lloyd & Steinke, 1984; Muller, 1991; Olson, 1979, 1983; Petchenik, 1983; and Peterson, 1987). As mapping has become a tool for data analysis and not just the means for representing location, questions asked of a map have become more complex. With recent technological advances, we have the ability to tailor the map style to the needs of the task. For example, maps can be easily produced in color today whereas a decade ago such production would have been difficult and costly. In order to use color most effectively with current technology, we need to understand the ways in which readers mentally process the map. Map reading speed and accuracy is a function of many variables affecting one or more cognitive stages. As illustrated by the examples we have presented in this article, a single map design may be optimal for some, but not all, tasks. The cognitive model points to the importance of experimentation or pretesting. It also suggests certain principles of map design. First, refinements of a map must affect the stage that has been identified as impaired by experimentation or pretesting. Second, experimental/pretest design should isolate the various stages. With the challenge that color poses to map design, understanding the map reading process is paramount to account for the effects of color.

References

Antes, J. R., and Chang, K. 1990. "An empirical analysis of the design principles for quantitative and qualitative area symbols." *Cartography and Geographic information Systems*, 17, 271-277.

Bertin, J. 1983 "Semiology of Graphics: Diagrams, Networks, Maps." Madison, WI: University of Wisconsin.

Barrell, G. V. and Cooper, P. J. 1986. "Cognitive processes in orienteering: The interpretation of contours and responses to the map as a whole." *Scientific Journal of Orienteering*, 2, 25-46.

Blades, M. and Spencer, C. 1986. "The implications of psychological theory and methodology for cognitive cartography." *Cartographica*, 23, 1-13.

Cleveland, W. S., and McGill, R. 1984. "Graphical perception: Theory, experimentation, and application to the development of graphical methods." *Journal of the American Statistical Association*, 79, 531-534.

Cleveland, W. S., 1993. "A model for studying display methods of statistical graphics." *Journal of Computational and Graphical Statistics*, 2, 323-343.

Dent, B.D. 1993. "Cartography: Thematic Map Design." Dubuque, Iowa: Wm. C. Brown.

Eastman, J. R. 1985. "Graphic organization and memory structures for map learning." *Cartographica*, 22, 1-20.

Gilmartin, P. P. 1981. "The interface of cognitive and psychological research in cartography." *Cartographica*, 18, 19-20.

Griffin, J. L. C. 1983. "Problem solving on maps: The importance of user strategies." *The Cartographic Journal*, 20, 1010-109.

Herrmann D, Kerwin J, Pickle LW, Croner C, White A, Jones G, Jobe J. 1994. "Cognitive processes in statistical map reading." *Proceedings of the Statistical Graphics Section*, American Statistical Association 1993 Meeting, San Francisco, CA. In press.

Kimerling, A. J. 1975. "A cartographic study of equal value gray scales for use with screened gray areas." *American Cartographer*, 2, 119-127.

Kimerling, A. J. 1979. "Visual value as a function of percent area inked for the cross- screening technique." *American Cartographer*, 6, 141-148.

Kosslyn, S. M. 1985. "Graphics and human information processing." *Journal of the American Statistical Association*, 80, 499-512.

Kosslyn, S.M. 1994. "Elements of Graph Design." New York: W. H. Freeman and Co.

Lewandowsky, S., Herrmann, D., Behrens, J.T., Li, S., Pickle, L.W., Jobe, J.B. 1993. "Perception of clusters in statistical maps." *Applied Cognitive Psychology*, 7,533-551, 1993.

Lloyd, R. & Steinke, T. 1984. "Recognition of disoriented maps: The cognitive process." *The Cartographic Journal*, 21, 55-59.

Makuc, D.M., Haglund, B., Ingram, D.D., Kleinman, J.C., Feldman, J.J. 1991. "Health service areas for the United States." National Center for Health Statistics. Vital Health Stat (2) 112.

Monmonier, M. 1991. "How to Lie with Maps." Chicago: University of Chicago Press.

Monmonier, M. 1993. "Mapping It Out." Chicago: University of Chicago Press.

Muller, J. C. (Ed). 1991. "Cognitive cartography: A new heart for a lost soul." In *Advances in Cartography*. International Cartography Association, pp. 201-221.

Olson, J. M. 1979. "Cognitive cartographic experimentation." The Canadian Cartographer, 16, 34-44.

Olson, J. M. 1983. "Future research directions in cartographic communication and design." In D. R. F. Taylor (Ed.) *Graphic Communication and Design in contemporary Cartography*. New York: Wiley.

Petchenik, B. B. 1983. "A map maker's perspective on map design research 1950-1980". In D. R. F. Taylor (Ed.) *Graphic Communication and Design in contemporary Cartography*. New York: Wiley.

Peterson, M. R. 1987. "The mental image in cartographic communication." *The Cartographic Journal*, 24, 35-51.

Simkin, D., and Hastie, R. 1987. "An information-processing analysis of graph perception." *Journal of the American Statistical Association*, 82, 454-465.

Statistics Canada 1991. *Mortality Atlas of Canada*. Ottawa, Canada: Canada Communications Group.

Tufte, E.R. 1983. "The Visual Display of Quantitative Information." Cheshire, CT: Graphics Press.

Tukey, J. W. 1979. "Statistical mapping: what should not be plotted", in *Proceedings of the 1976 Workshop on Automated Cartography and Epidemiology*, Hyattsville, MD, DHEW Publ. No.(PHS) 79-1254, 18-26.

U.S. Department of Commerce 1990. *Agricultural Atlas of the United States: 1987*, Vol 2 (Subject Series), Part 1, AC87-S-1, Bureau of the Census, U.S. GPO, Washington, D.C.

Wainer, H., Francolini, C.M. 1980. "An empirical inquiry concerning human understanding of two-variable color maps." *American Statistician*, 34, 81-93.

Linda W. Pickle
and
Douglas J. Herrmann
National Center for Health Statistics
CDC, Hyattsville, MD
LWP0@NCH09A.EM.CDC.GOV



TOPICS IN SCIENTIFIC VISUALIZATION

Color Perception, the Importance of Gray and Residuals, on a Choropleth Map

By Dan Carr

Note: This article is the third in the sequence on choropleth maps. The previous articles addressed smooths and legends, respectively. This article shows the residual map for the continuing example of white male colon cancer mortality rates. As will be discussed below, the residuals are of interest in the context of generating epidemiological hypotheses. Since the plot is in color I have chosen to begin the article with a discussion of color and color perception.

The color discussion below provides a physics-based description and comments on perception. Many aspects of color use are not covered such as distinguishing foreground from background, adding additional layers of information, creating perceptual groups, providing depth cues, achieving balance in terms of complementary colors, and highlighting in terms of complementary colors. Further, little is said about color systems. Readers wanting more information can refer to the newsletter article on color systems (Eddy 1990) and a host of additional resources such as Foley et al (1990), Durrett (1987), Friedhoff and Benzon (1989), Tufte (1990), and Mante (1972).

Color

A physics-based description of visible light yields an intensity versus wavelength curve over the spectrum of 400 to 700 nanometers. Figure 1 provides an example for one color. Light consists of photons with characteristic wavelengths and wavelength corresponds inversely with frequency and energy. A physics-based intensity (energy) measure for a specific wavelength counts the number of photons striking a given area in a given length of time and scales the count by the energy for that particular wavelength. The perception of color is based on a weighted visual response to the intensity curve. Common descriptions refer to three different weightings, one for each of the three basic photon responsive components of the eye: rods, green cones and red cones. The construction of a brightness measure, called luminance, proceeds by settling on a single luminous efficiency curve (see Figure 1) as being representative of the composite visual response. The integrated product

of the intensity function with the efficiency curve (or weighting function) produces a measure with units such as foot lumens. This measure is a rough approximation to perceived brightness.

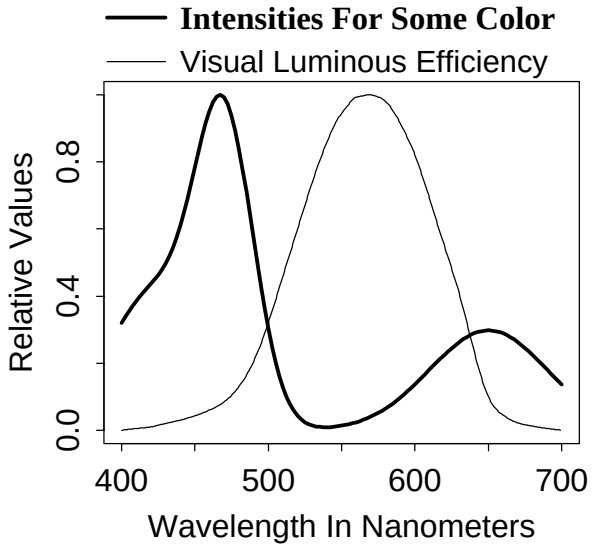


Figure 1. Intensity and Luminance Curves.

Many different intensity curves produce the same perceived color. Such colors are called monomers. Descriptive systems for characterizing perceptually discriminable colors typically involve three facets or dimensions. The HLS system describes color using the concepts of hue, lightness and saturation.

For monochromatic light, hue corresponds directly to the wavelength of the light. For example wavelengths of 470, 490, and 572 nanometers correspond to what is generally perceived to be pure blue, green and yellow, respectively. Most light that we see is not monochromatic but rather is composed of wavelength mixtures over regions of the visible spectrum. The hue of a mixture can be found by perceptually matching the color with mixtures of a monochromatic light and achromatic light. The monochromatic light in the match gives the hue. Hue is conveniently thought of as the dominant wavelength of a color. However in some cases the color actually has zero intensity for that particular wavelength. For example the combination of monochromatic red and green light does not contain yellow but appears yellow because photons with those two wavelengths jointly stimulate the red and green cones in the eye in a fashion similar to that of yellow light photons. Multidimensional scaling of judgements concerning hue closeness places hues on a “C” shaped curve in a plane (Shepard and Carroll 1966). The HLS descriptive system orders hues around a circle. Consequently hues are not good for representing values of an ordered variable.

A second aspect of color, saturation, refers to excitation purity. Monochromatic light appears fully saturated in color (provided the photon rate is high enough). Light containing a mixture of different wavelengths appears less saturated. Sunlight and other mixtures that are of roughly uniform intensity across the visible spectrum appear as totally unsaturated or achromatic. Saturation can be thought of as an ordered scale that involves a mixture of a pure chromatic color (such as red) and an achromatic color (such as some shade of gray). Representing an ordered variable using saturation is reasonable. The main limitation is that humans cannot discriminate many levels of saturation.

When the wavelength specific intensities are roughly the same across the visible spectrum, the color is perceived as being achromatic. Achromatic light can be thought of as gray levels from white through black. We perceive the mixtures produced by the sun and many electric bulbs as white light. We perceive surfaces reflecting white light equally across the visible spectrum as white when the reflectance is above 80% A surface reflecting 3% or less light appears to be black (Murch 1987). We perceive intermediate reflectances as levels of gray. In printing, the half-toning process produces grey-levels by putting patterns of black ink on white paper to reduce the reflectance. For CRT displays, “equal” mixtures of red, green and blue light emitted by the CRT phosphors produce gray. The apparent gray-level from white to black, depends on the intensity of the light from the three phosphors and the surrounding brightness.

Different authors refer to the notion of composite light intensity (or intensity contrast) using such words as brightness, lightness, and value. Whatever the word chosen, humans are very responsive to the intensity dimension of light. Humans can discriminate many levels of brightness, so brightness is the best color dimension for representing an ordered variable. The use of gray-levels is particularly convenient since it provides a brightness dimension without the risk of suggesting other competing orderings based on hue and saturation.

The above physics-based description of brightness follows that of Foley et al (1990). However, luminance as a measure of brightness does not tell the whole story. Murch (1987) uses the word brightness to refer to perceptual color changes that are a function of light intensity or roughly speaking, photon rates. As intensity increases it first becomes possible to see something, then it becomes possible to see dark unsaturated colors, next it is possible to see more and more saturated col-

ors, and eventually it is possible to see saturated reds and yellows. At greater extremes of light intensity, the hues appear more desaturated and the colors can appear dazzling. A further phenomena is the apparent change in hue that accompanies a change in luminance. This is known as the Bezold-Brucke Effect and can necessitate some adjustments except at the three pure colors indicated above.

In controlled settings, colors can be ordered in terms of saturation and brightness but commonly encountered situations are much more complicated because color perception is relative. Our eye-brain systems diminish in response to steady state stimulation. When a small colored disk is surrounded by an annulus of a different color, adaptation (or diminished response) to the annulus color shifts the apparent color of the small disk toward the complementary color of the annulus. (Complementary colors in an additive light model are color pairs that produce white light.) If the annulus is green, the visual response to the green rapidly diminishes both for the annulus and the disk. For the disk, this adaptation produces a perceptual shift toward magenta. Land (see Friedhoff and Benzon 1991) discovered that he could produce richly colored images using two lights sources, such as white and red. He produced two gray-level slides using red and green filters. He then projected red light through the red filter slide and white light through the green filter slide. The result was not pink as in pigment mixing but a full color image with the color of regions relatively low in red saturation being shifted toward cyan. Apparent gray levels also relative. A brighter surround makes the inner disk appear darker. The relativity of color perception can complicate interpretation when color represents an ordered variable in choropleth maps and other varying-local-background applications.

Visual Subchannels

According to Friedhoff and Benzon (1991), visual processing involves separate pathways or channels that emphasize different facets of a scene. Their description gives three channels, a color channel that does not see the objects in great detail, a channel associated with movement and binocular depth information, and a channel that carries high-resolution information about shapes. The high-resolution shape channel processes changes in value (or brightness). This shape channel also carries much of the monocular information about depth.

Readers can perform a simple experiment to gain insight into the importance of brightness contrast in monocular depth perception. Start with a perspective view such that in Figure 2. Then make a similar image using a background color such as green and lines in another color such as red. Modify the intensities until the colors are brightness matched. In other words the lines should be as difficult as possible to see. Then observe how flat the image appears compared to the black and white version.

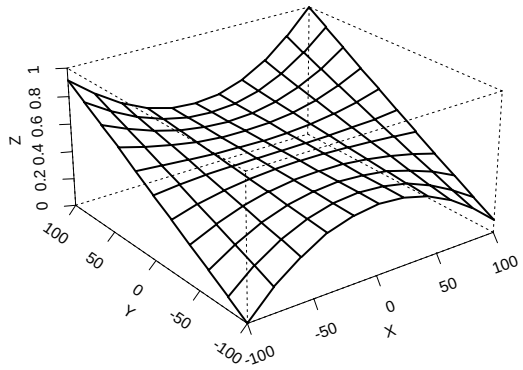


Figure 2. A simple pserspective plot.

Many scientific images use a spectral color order (red, orange, yellow, green, blue, violet) to represent an ordered variable. This can produce unintended visual effects because the colors are not ordered in brightness. The shape-channel picks out edges between colors of contrasting brightness and extracts monocular depth information based on local brightness comparisons. While purposeful jumps in brightness may emphasize selected contours and create perceptual groups that are relatively close in brightness, all to often people make color selections without regard to the shape and depth information that the color selection generates.

A gray-level plot can be used as a rough consistency check for a plot that uses color to encode an ordered variable. When colors have been encoded as RGB values and the RGB values are based on the standard NTSC phosphors, the Y component of the YIQ transformation, Y= .299R+.587G+.114B (see Eddy 1990) corresponds to luminance. This Y value can be equated to gray-level for production of the plot on a CRT display or on paper. (The paper version can have much better spatial resolution but may not be as effective because of halftone edge problems.) If the gray-level version appears inconsistent with the intended message, then the color selection should be reconsidered.

Controlling Contrast Noise

Tufte's (1990) description of laying and separation provides an introduction the elegant use of color. In his

description, wild local variations in brightness (value) appear as noise. This phenomena is likely to occur in a choropleth map that represents residuals from spatial smooths. Figure 3 addresses the noise problem by plotting HSAs in gray when their residuals are small in absolute value. (See Carr and Pickle (1993) and Carr (1993) for details concerning the data and smoothing.) The legend in the noisy residual plot (not shown) prior to Figure 3, gave approximate values for the 5th and 95th population-based percentiles. The class interval from -3 to 3 is a convenient, symmetric rounding of those percentiles. The values of -8 and 8 were chosen heuristically based on the residual extrema. The double-ended color scale with darker saturated colors at the extremes is consistent with the separate representation of positive and negative residuals. This color scheme uses saturated color sparingly in accordance with Tufte's advice.

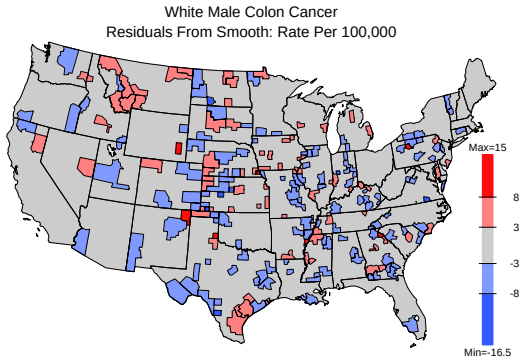


Figure 3. Residuals from a smooth fit.

Hypothesis Generation

Plotting residuals from spatial smooths of mortality rates is a useful device for generating epidemiological hypotheses. Knowing the location of a large mortality rate residual triggers the mental search for possible cancer related variables that are also unusual for the location. A match with locally unusual values provides a hypothesis for investigation. The failure to find a match among the immediately known variables suggests checking on additional variables. More generally, spatial smoothing provides a surrogate adjustment for cancer related variables that vary smoothly over space. The residuals can call attention to cancer related variables that do not vary smoothly over space.

In the residual plot, the red HSA in eastern Nebraska comes as no surprise to epidemiologists. In past decades epidemiologists noticed high rates in eastern Nebraska compared to other midwest rates. Further study found that the rates were associated with people of Czechoslovakian background and their dietary patterns (see Pickle et al 1987). Based on viewing cancer mortality maps for the three previous decades, it is possible that the other dark red HSAs in Figure 3 have not yet been studied as having locally high colon cancer rates. In Pennsylvania, the red HSA draws additional attention because of the light blue (low) neighboring HSA rates. Readers having hypotheses about any of the dark red regions are encouraged to send them to me. The clusters of light red HSAs in Montana and Texas are roughly consistent with the three decade trend map shown in Pickle et al 1987. However most of the other significant trend regions do not appear in this residual plot. The Montana cluster is evident in the 1970-1979 map but the Texas cluster is new. The dark blue HSAs and the light blue clusters can also be consider for possible associations.

Cautious Interpretation

Interpretation of this residual plot should proceed cautiously for several reasons. Unexplained variability remains. The smoothing fraction may be too large and the use of Poisson-based weights needs to be reviewed. The large residuals call for robust residual weights. In general smoothing methods are suspect at map edges and most methods are not designed to track discontinuities, should they be present in the data. In addition, the use of class intervals can hide important detail. As examples of this, the apparent Texas cluster is the result of smoothing two locally high HSA rates and the single dark red Nebraska HSA is not the only high rate HSA in the area. Smoothing methods accomplish their purpose when the examination of the smooth and residuals helps the analyst focus attention. The results are not to be interpreted in isolation from data, diagnostics and related knowledge.

The data are available and readers are encouraged to produce their own smoothed plots.

The impact of inverse variance weights should be considered in the interpretation of ordinary residuals. For example, Long Island has a large population and correspondingly the mortality rate has a small variance estimate. The inverse variance weight is relatively large, so the Long Island rate strongly influences its own smooth estimate and the estimates for the neighboring HSAs. This is evident if one examines the maps in the two previous articles. Seeing a large residual for Long Island would be a surprise. The anticipated result is that extreme residuals will tend to belong to small population HSAs. With a Poisson model for the number of deaths, the small population rates are more variable and the

HSA's exert less pull on the local smooth. Clusters of small population HSAs with extreme residuals are not particularly surprising. Since counties have been aggregated into HSAs to reduce the population size imbalance among analysis units, the situation is not as extreme as in county residual maps. Still, patterns in Figure 3 may be predominantly due to the "random variation" in small population HSA rates so plotting studentized residuals is an obvious next step.

Readers are encouraged to improve on the smoothing, to make their own choropleth residual maps, to generate their own hypotheses and reach their own conclusions. The data structures, modeling and plotting scripts, and postscript files are available by anonymous ftp to `galaxy.gmu.edu` under `submission/eda/maps`.

Acknowledgements

Research related to this article was supported by NSF under grant no. DMS-9107188 and by EPA under cooperative agreement no. CR8280820-01-0. The article has not been subjected to the review of the EPA and thus does not necessarily reflect the view of the agency and no official endorsement should be inferred.

References

Carr, D. B. (1993), "Constructing Legends for Classed Choropleth Maps." *Statistical Computing & Statistical Graphics Newsletter*, Vol. 1. No. 1, pp. 15-19.

Carr, D. B. and L. W. Pickle (1993), "Plot Production Issues and Details." *Statistical Computing & Statistical Graphics Newsletter*, Vol. 1. No. 1, pp. 16-20.

Durrent, J. H. (ed.). (1987), *Color and the Computer*, Academic Press Inc., New York

Eddy, W. F. (1990), "An Introduction to Color Systems." *Statistical Computing & Statistical Graphics Newsletter*, Vol. 1. No. 1, pp. 7-10.

Foley, J. D., A. van Dam, S. K. Feiner, and J. F. Hughes. (1990), *Computer Graphics, Principles and Practice*, Second edition, Addison-Wesley Publishing Co. New York.

Friedhoff, R. M. and W. Benzon. (1989), *the Second computer revolution VISUALIZATION*, W. H. Freeman and Company, New York.

Mante, H. (1972), *COLOR DESIGN in Photography*, Van Nostrand Rienhold Company, New York.

Murch, G. (1987), "Color Displays and Color Science." Color and the Computer, (H. J. Durrett ed.) Academic Press Inc., New York, pp. 1-25.

Pickle, L. W., T. J. Mason, N. Howard, R. Hoover, and J. F. Fraumeni, Jr, (1987), Atlas of U.S. Cancer Mortality Among White: 1950-1980. Washington, D.C.: USGPO, DHHS Publ. No. (NIH) 90-1582.

Shepard, R. N. and J. D. Carroll. (1966), Parametric Representation of Nonlinear Data Structures in Multivariate Data Structures. Multivariate Analysis (P. R. Krishnaiah, ed.) Academic Press, New York, pp. 261-292.

Tufte, E. R. (1990), Envisioning Information, Graphics Press, Cheshire CN.

Daniel B. Carr
George Mason University
`dcarr@galaxy.gmu.edu`

UNIX COMPUTING

What is a Shell, Part Two

by Phil Spector

Note: This article is a continuation of the article appearing in the December, 1993 issue of this newsletter.

Job Control

The UNIX operating system accommodates the execution of programs in two modes: foreground mode, where the terminal or window being used does not prompt for a response until the program is completed; and background mode, which executes the program without tying up the terminal or window from which the command invoking the program was executed. By default, the commands you type are carried out in the foreground. In an environment with multiple windows, these distinctions are not so critical, since additional windows can be opened if the one in which a command was entered is not available.

Background commands are particularly useful in non-windowed environments.

When using a terminal with just one window, it can be very useful to run one or more commands in the background while an interactive session is being carried out. Even in a windowed environment, it can often be useful to run a program in the background if you know you will be logging off before the program has completed executing, since jobs in the foreground will terminate when you logoff. To make it known the the UNIX operating system that you wish to run a job in the background,

terminate the command with an ampersand (&). (Under some versions of the Bourne and Korn shells, you may need to use the nohup commands if you want to make sure that a background job will not terminate when you log off.) When running a job in the background, output from the background process will still be displayed on your computer screen, so it's usually a very good idea to use redirection (described in the first part of this article) to redirect both the standard output *and* the error output to a file. This is especially important if you will be logging out before the job is completed, because in that case there will be no way to retrieve output which has not been redirected. Remember that the ampersand should be the very last character of the command you wish to place in the background.

When a job is running in the foreground, you can communicate with it by sending a signal to it. The most commonly used signal is an interrupt signal, which is sent when you hit control-C on the keyboard (that is, type a "c" while holding down the control key). This will stop many programs, but some programs (notably many interactive statistical programs) will trap the interrupt signal and, for example, terminate the currently executing command and return to the program's prompt.

Signals allow you to communicate to both foreground and background jobs.

A more powerful signal, known as the quit signal, can be sent to an executing program using control-\. Most programs will be terminated after receiving the quit signal, but sometimes you will need to use the UNIX kill command if the program does not respond to a quit signal or if it is running in the background. For foreground jobs, you will also need to find the process id (PID) of the job, which can be obtained through the use of the UNIX ps command. For background jobs, you can use the job number (printed in square brackets when a job is put in the background or available through the jobs UNIX command) preceded by a percent sign (%) in place of the PID. You can find out more about these commands by accessing the online documentation through the UNIX man command.

An additional form of job control, available in most shells except the Bourne shell, allows you to temporarily suspend a running job. This can be very useful if you are running a job in the foreground, but then decide to put it in the background. To do this in a shell which supports it, you first suspend the running job by sending it a suspend signal, control-Z. Then you can place the job in the background with the UNIX command bg. The shell will respond with a job number, just as if the program was initiated in background mode, and the terminal or window will again be available for interactive commands. Remember that a job suspended with control-Z does not go away—it still uses some computer resources, and is available for restarting, either in the background (with the bg) command or in the foreground (with the fg) command.

Aliases and Startup Files

An alias is a way of abbreviating or redefining a command to customize it to your personal preferences. (Aliases are not available in many versions of the Bourne shell.) One very common example of an alias involves the UNIX rm command. By default, rm silently removes files without asking for your approval. (Any user who has accidentally typed rm * knows that this default may not be very desirable.) There is an option to rm, namely -i which instructs rm to prompt you with the name of each file it would otherwise remove by default, giving you a chance to confirm that you actually want it deleted. To make sure that every time you use the rm command it is invoked with the -i option, you could issue the following command to the shell:

```
alias rm rm -i        # in c-shell and tcsh
alias rm="rm -i"      # in Korn shell
```

If you wish to override the meaning of an alias, and have it revert back to its usual default, precede the alias name with a backslash (\). So if you wanted to use rm without being prompted for verification of each file, you could use the command \rm instead of rm.

Naturally, you may not remember to type this every time you start up a shell, so each of the shells reads and executes a file of commands from your home directory every time they are started up. For the Bourne and Korn shells, the name of the file is .profile, while for csh and tcsh, the name of the file is .cshrc.

Shell Variables

Within each of the shells, you have an option of setting variables to values of your choice. These variables can be set locally, so that only the currently executing shell is aware of them, or they can be set so that they are transmitted to all new shells that are created from the current shell. In the Bourne and Korn shells, you can set a local variable with the syntax variable=value. Under the Bourne shell, you need to issue a separate export command to make the variable available to any newly created shells, while the Korn shell allows you to both set and export a variable using just a single export command with the syntax export variable=value. In csh and tcsh, a local variable is set using the syntax set variable=value, while global variables are set with the

setenv command, omitting the equal sign. In all of the shells, a reference to a shell variable can be recognized by a dollar sign ($) at the beginning of the variable's name, although as you can see from the above description of how to set a variable, you do not use the dollar sign when defining the variables.

There are two main functions of shell variables. First, you can use a shell variable in a fashion similar to an alias, that is as an abbreviation for a longer or more complex piece of text. Unlike an alias, however, shell variables are not limited to just commands. For example, suppose you are accessing data sets in a directory with a complicated name, such as /usr/data/survey/jan/rawdata. If you entered the following command (or placed it in your .cshrc file):

```
survey=/usr/data/survey/jan/rawdata
     # in Bourne and Korn shells
set survey=/usr/data/survey/jan/rawdata
     # in c-shell and tcsh
```

then you could refer to the file
/usr/data/survey/jan/rawdata/datafile
as simply $survey/datafile.

A more common use of shell variables is to communicate to other programs about the location of resources or personal preferences. For example, many programs will present output and invoke a UNIX pager to pause when the text has filled your screen. If there is a particular pager program you prefer, then you can inform these programs by setting the global variable PAGER to the name of the command you prefer. Of course there is no guarantee that all programs will look for this variable, but it is a fairly common practice. So you could set the PAGER variable to the value less, a popular public domain paging program with the commands

```
PAGER=less export PAGER  # in Bourne shell
export PAGER=less        # in Korn shell
setenv PAGER less        # in c-shell and tcsh
```

Note that under the Bourne shell, the command which sets the variable and the one which exports it to other shells can be typed on the same line. Other common variables which are used this way include PRINTER and EDITOR. In addition, the documentation of many UNIX programs includes information about shell variables specific to those programs. Most users find that setting these variables in their startup files (.cshrc or .profile) provides the most effective way of communicating their preferences to the programs they use.

History and Command Editing

One of the main motivations of developing new shells after the original Bourne shell was to add capability to "re-use" old commands. Much of the work which is done on a computer is repetitive, and it was felt that some mechanism to access and modify previously entered commands would be very useful. The C-shell (csh) provides this functionality through a history mechanism. First, the built-in command history will display recently entered commands, each proceeded by a number. (The number of commands displayed is controlled by the shell variable history which can be set in your .cshrc file.) In the C-shell, to access a previous command, you can enter its history number preceded by an exclamation point (!). In addition, the command !! repeats the last command, !-1 repeats the next to last command, and so on. There are also ways of modifying previous commands, which are explained in the UNIX manual pages for csh, most likely under the heading "History Substitution".

Rather than this form of access to previous commands, many users prefer to edit their previous commands using the editor of their choice, and then resubmit the commands to the shell. This functionality is built into the Korn shell (ksh) as well as the tcsh shell. These shells create a single line editing environment in which you can scroll your previous commands using the same key strokes as you would use to move up and down through a file, and to modify commands just as you would edit a line in a file. Upon hitting the return key, your edited command is submitted to the shell. These shells allow editing in either the style of emacs or vi, the two most popular UNIX editors. In the Korn shell you can choose between these two styles by setting an appropriate option with one of the following statements:

```
set -o emacs   # for emacs
set -o vi      # for vi
```

while in tcsh, you can use the bind command:

```
bind emacs     # for emacs
bind vi        # for vi
```

Once again, these statements should be placed in the appropriate startup file, so that the shell will always respond in the way you expect.

Phil Spector
Applications Manager
Department of Statistics
UC at Berkeley
spector@stat.Berkeley.EDU

NET SNOOPING

URLs for Fun, and Profit

by Mike Meyer

Many software tools such as the Alex virtual filesystem, Andrew File System (AFS) and its derivative Distributed File System, have tried to unify diverse filesystems so that they appear unified. Each of these tools has to some extent succeeded, but none has had the phenominal success of the very simple, Universal Resource Locator (URL).

I introduced the concept of a URL last issue, and this time I intend to point out a few URLs which I find I often return to. One reason to simply list and describe interesting URLs is the obvious difficulty of searching the WWW for information that is interesting to you. There are some tools evolving tools for searching through the WWW, and there are many enterprising souls who attempt to organize the WWW into (what appears to them to be) a logical structure. But the difficulties are enormous. If you have ever forgotten the name of a file in your own computer system and had to use tree traversing and searching software to find the file, then you have a first approximation of what it is like to search for something on the internet.

In order to organise my own internet browsing I have taken to creating my own Mosaic "Home Page" that contains a short description of interesting sites and the corresponding hyper-link.

With those few words of introduction, I'll jump into some of my more interesting haunts.

A recent revelation to me was that the United States Census Bureau (`http://www.census.gov/`) is on the internet and provides a great deal of useful information. The server contains the expected (and very interesting) sections about Census history and legislation, but also provides access to unique resources like SIPP On Call. That service allows you to define variables and perform custom extracts from the Survey of Income and Program Participation and selected Current Population Survey data sets.

A less statistical, but still interesting service, is the Unix Help system at the Univeristy of Edinburgh (`http://www.ucs.ed.ac.uk/Unixhelp/TOP_html`). This provides quite a good introduction to unix for novice to medium level users. The material is not as good as Phil Spector's columns, but it is close.

Another favorite of mine is information about how to incorporate in the U.S. state of Delaware (`http://www.service.com/tcc/home.html`). Every year at tax time this looks more and more interesting. By the time you read this most of you will have already submitted your taxes. For those of you who are still looking for U.S. tax forms, all the current forms are available in Adobe's Portable Document Format (PDF) at (`http:inept.scubed.com:8001/tax/fed/fed_index.html`). Many forms are also available in TIFF and Postscript formats.

If you are interested in how the U.S. portrays itself to the rest of the world, you can obtain transcripts of Voice of America broadcasts at (`gopher://gopher.voa.gov/`).

There are also convenient hooks into a geographic name service (`http://www.mit.edu:8001/geo`), for chemists, an interactive periodic table (`http://www.cs.ubc.ca/elements/periodic-table`), and a really cool map of Australian WWW sites (`http://life.anu.edu.au/links/ozmap.html`).

Finally, for all us struggling to obtain NSF grants, (`gopher://stis.nsf.gov/1`) the Science and Technology Information System provides convenient access to the NSF's documents.

Mike Meyer
Carnegie Mellon University
`mikem@stat.cmu.edu`



by Andrejs Dunkels

Statistical Computing And Graphics In Science And Industry

by Al Liebetrau

This column features statistical computing and statistical graphics activities in science and industry. Please send your comments and suggestions for future columns, and inquiries to Albert M. Liebetrau, Analytic Sciences Department, Battelle-Northwest, MS K7-34, P.O. Box 999, Richland, WA 99352, AM_Liebetrau@pnlg.pnl.gov, *509-375-2694.*

Introduction

Since its inception during the second world war, the U.S. nuclear program has generated large volumes of radioactive and hazardous chemical wastes. At the Hanford Site, much of the waste is stored in large underground tanks. There are 149 single-shell tanks and 28 double-shell tanks, each with a capacity of approximately 750,000 gallons. In the past, various campaigns have been conducted to reduce the volume of liquid in the single-shell tank wastes. The remaining contents vary from one tank to another, but generally the tanks contain some mixture of liquids and solids. The typical solid material is a salt cake that was formed by evaporation. Because of past pumping and evaporation activity, the contents of a tank may be quite heterogeneous spatially; in other words, the concentration of any given constituent may vary widely from one location to another within the tank.

Surface maps are used to measure spatial variability at the surface of the waste tank.

It is now a national priority to find suitable means for permanent storage of the wastes in these tanks. A crucial first step is to characterize the tanks, or to determine the radionuclide and chemical composition of their present contents. Characterization is based primarily on the analysis, both chemical and statistical, of physical samples of the tank contents. However, sampling is costly (labor intensive, material is hazardous and/or radioactive, etc.) and the spatial resolution that can be achieved with sample sizes that are affordable is limited. Moreover, there are physical limits on the extent of spatial coverage because samples can be taken only at riser locations; i.e., at locations of access ports to the tank.

Surface Maps

Recently, surface maps have been employed to help understand the composition of waste in a tank and to interpret analytical results. In particular, surface maps are used to measure spatial variability at the surface of the waste tank. Repeated surface maps provide a convenient time history, so it is hoped that these maps can also be used to detect tank leaks and other important changes.

A surface map is created from a series of photographs of the waste surface taken with a camera inserted into the tank at one of the risers. The camera is manipulated remotely to obtain a set of photos that cover the waste surface (with considerable overlap). Photos are taken in three stages or tiers. First, the camera is set at -80° to horizontal to obtain a series of photos that give "close in" coverage near the foot of the riser. Next, the camera is set at -50° to horizontal to obtain medium range coverage, and finally, the camera is set at -20° to horizontal to give “far field” coverage. At each stage, the camera is rotated to produce a series of photos that give full circular coverage. A total of 22 photographs are taken that cover the entire surface of the tank. The original images are taken on 2.25-inch color negative film. A typical photo, taken at -20° to horizontal, is shown in Figure 1.



Figure 1. A typical photo of the waste surface.

After the photos are available, the first step is to select those that will actually be used to create the surface map. Those with identifiable “land masses,” i.e., identifiable

regions of different intensity than their surroundings, are selected because their features can be used to piece the photos together into a composite image. The selected photos are enlarged and then converted to electronic form by scanning them at 400 dpi to create a 24-bit true color image. The resulting image is stored in a TIFF file that is approximately 3200 by 3200 by 3 pixels in size. The three layers in the z direction contain intensities for the three primary colors; red, blue and green. The second step is to correct the TIFF file images for differences in perspective. Perspective correction is required for two reasons: First, the camera is usually inserted at a riser far from the center of the tank and, second, photos are taken from three different viewing angles. The images are transformed from the perspective of the camera to a vertical perspective, as if the surface were viewed from directly above (at a great distance) with the top of the tank removed. The perspective change is accomplished by means of appropriate equations that describe the position of the camera relative and the portion of the waste surface covered by the image. Separate transformations are required for each tier because the camera position changes from one tier to the next. Figure 2 shows the image of Figure 1 after the perspective correction has been applied. Since the camera is operated remotely, its position relative to a fixed set of coordinates is known only approximately. This uncertainty in the position of the camera is reflected in the results of the perspective correction and in all subsequent steps in the creation of a surface map.

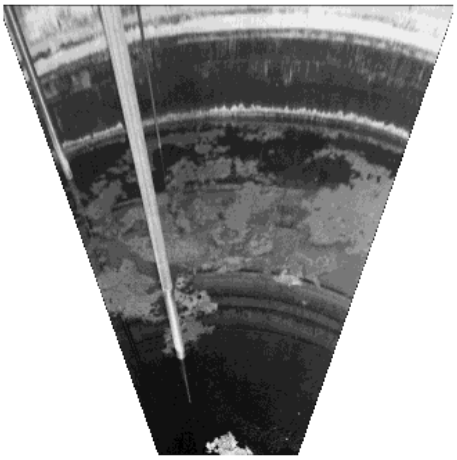


Figure 2. The result of applying a perspective correction transformation to the image of Figure 1.

Next, individual images are placed on an outline of the tank (a circle) in the positions they will occupy in the composite image. Placement is accomplished by rotating and translating the images: Specific transformations are determined from information about the position of the camera relative to known reference points and structural features of the tank (such as pipe locations). Matching the arc in the image that represents the interface between the waste surface and the tank wall with the corresponding arc in the underlying circular tank outline is a key aspect of this positioning step. Figure 3 shows the image of Figure 2 after it has been positioned on the composite.

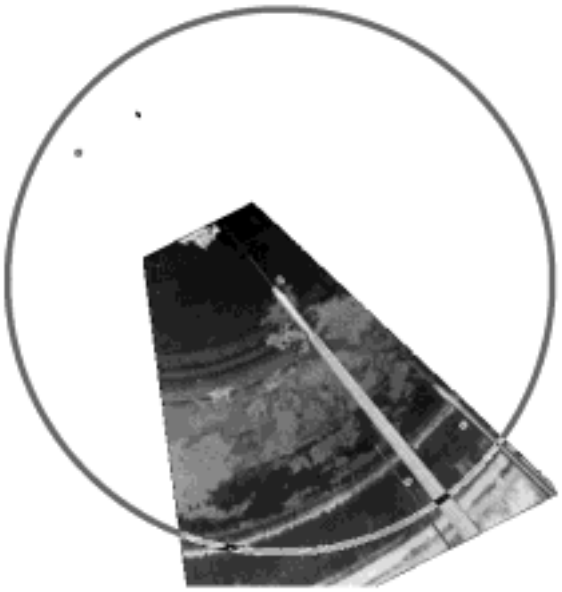


Figure 3. The image of Figure 2 after initial positioning on the composite image.

For the images in the "far field" tier (those taken at -20° to horizontal), perspective distortion is large because of distance from the camera and uncertainty about its position. It is evident in Figure 3, for example, that the arc in the image representing the interface between the waste surface and the tank wall does not match the corresponding arc of the underlying tank outline. Thus, for "far field" images where distortion is extreme, an additional positioning step is required to match corresponding arcs in the image and the tank outline. This step involves transforming the polygonal image in Figure 3 to that shown in Figure 4. This polygon-to-polygon transformation is derived from information about the location of the wall and other structural features of the tank (e.g., pipes), together with the location of previously-positioned images. Figure 4 shows the image in its final

position in the composite. The image has been cropped to minimize overlap with other images.

Successive images are placed on the map using the sequence of steps described above: They are corrected for perspective, and rotated and translated into their proper positions in the composite. Images in the "far field" tier undergo an additional transformation to match the arc of the waste surface in the image with the tank wall in the underlying composite. The outline of the tank, its structural features, and previously-placed images are used to guide the placement of successive images on the map.

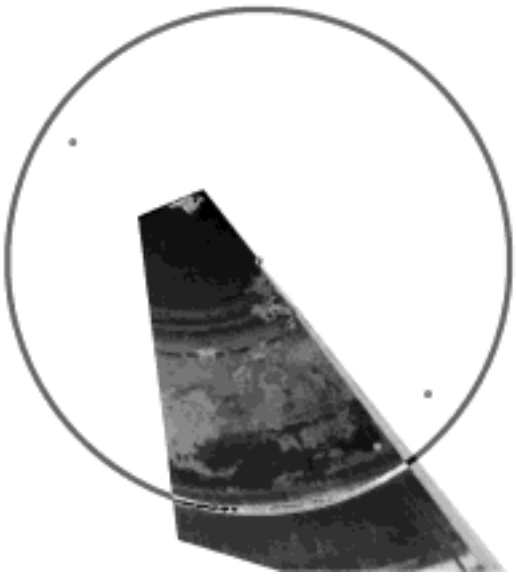


Figure 4. The image of Figure 3 in its final position on the composite image.

The final step in creating the composite image is to resolve discrepancies in the corresponding pixels of overlapping images. Where images overlap and multiple values are available for a given pixel, the highest value is selected. This procedure is equivalent to selecting the brightest of the available images at each pixel. Thus, where two images overlap, border discrepancies are resolved by enlarging the brightest areas of surface (which generally correspond to solid waste). The completed surface map for one tank is shown in Figure 5.

Evaluation and Future Enhancements

Inspection of Figure 5 reveals that seams exist between some pairs of adjacent images. This is due in part to the difference in perspective associated with the three camera orientations. Seams are also due in part to limitations in the camera positioning apparatus. The ideal way to handle this problem is to develop an improved system to control the camera and determine its location. Another source of error is our inability in some areas to determine whether high intensities solids in the tank or light reflections off of the liquid surface. This problem can be mitigated by better positioning of the camera. Finally, the seams between adjacent images can be reduced by employing a suitable spatial averaging scheme whereby intensities in a given pixel are weighted by the intensities in adjacent pixels. However, it must be kept in mind that a primary goal is to clearly identify the surface features of the waste surface. Spatial averaging has not been employed because it reduces the sharpness of the distinctions we are trying to identify.

Not all tanks are good candidates for mapping. Extremely homogeneous wastes generally do not generate enough contrast to produce useful surface maps. Likewise, it is difficult to superimpose images for tanks that lack enough easily identifiable structural features. Thus, surface maps are most accurate and useful for those tanks with easily recognizable structural features and whose contents are a mixture of liquids and floating solids.

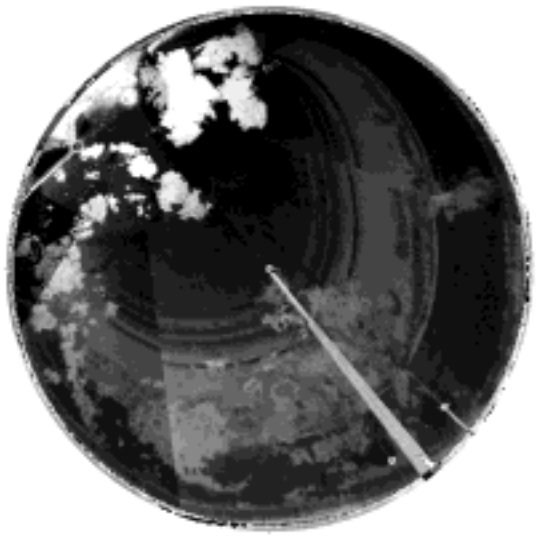


Figure 5. The completed surface map for a waste tank.

Work continues to improve this pioneering effort at creating tank surface maps. However, the current maps are of sufficient quality to increase our understanding of the nature of the waste in a tank, and thereby to help us interpret the corresponding analytical results.

Acknowledgments

I wish to thank my colleagues Cheryl Vorvick, Daryl Baird and Pat Heasler for permitting me to describe their work in this column.

Albert M. Liebetrau
Battelle Pacific Northwest Laboratories
AM_Liebetrau@pnlg.pnl.gov

CONFERENCE NOTICES

Interface '94

26th Symposium on the Interface:
Computing Science and Statistics
Computationally Intensive Statistical Methods
June 15-18, 1994,
Research Triangle Park, NC,
Sheraton Imperial Hotel



The Interface Conference is the premier annual conference on the interface of computing and statistics. It is sponsored by the non-profit Interface Foundation of North America, and will be hosted in 1994 by SAS Institute, with John Sall as program chair. The 1994 conference will be in Research Triangle Park, the home of many top research centers, and anchored by three major universities.

Contact: Email: interface94@sas.com phone: 919-677-4499, fax: 919-677-8224, mail: Interface 94, SAS Campus Drive, Cary NC 27513 USA

Call for Papers:

The deadline for submitting a contributed paper is Jan 31. It may be a presentation, or poster session. Submit an abstract either by email (if you don't have messy formulas) or hardcopy formatted to 6.5 inches wide by at most 4 inches tall, with centered lines for title, author(s), and address.

Financial Support:

Limited funds may be available to support travel and per diem expenses of young researchers and graduate students. Preference will be given to those who will be presenting papers. Please apply to John Sall, program chair [sall@sas.com].

Adjoining Meetings in the Triangle:

Before the Interface is the Spring Research Conference on Statistics in Industry and Technology. After the Interface is the Third World Congress of the Bernoulli Society and the 55th Annual IMS meeting.

Hotel:

For reservations, call or write Sheraton Imperial Hotel, I40 Exit 282 at Page Rd, P O Box 13099, Research Triangle Park, NC 27709. Phone 919-941-5050.

Exhibits:

If you would like to exhibit, please contact Armistead Sapp, SAS Institute, Cary NC, 919-677-4499, sasaws@unx.sas.com.

Tours:

Several tours will be available on Thursday evening, one to SAS Institute, and one to the UNC Graphics and Image Lab (Virtual Reality Center).

Keynote:

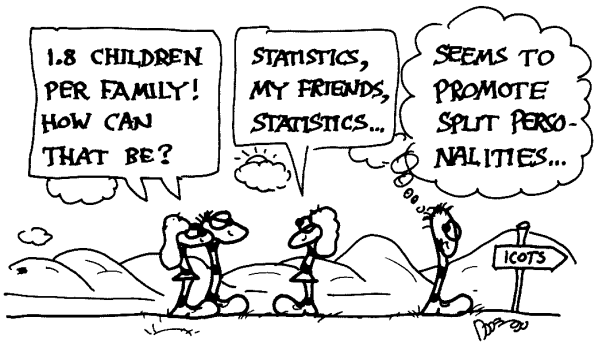
- G.W. Stewart, Institute for Advanced Computer Studies, University of Maryland, a well-known authority in the field of numerical linear algebra.

Invited Sessions:

- *Space Filling Experimental Designs,* S. Stanley Young, Glaxo
- *Fast Implementations of Smoothers,* Steve Marron, University of North Carolina
- *Smart Monte Carlo Methods for Conditional Inference in Exponential Families,* Cyrus R. Mehta, Harvard Univ.
- *Nonparametric Regression for Edge and Peak Preserving* Alex Georgiev, Ethyl Corporation
- *Stochastic Modeling In Carcinogenesis,* Chris Portier, NIEHS
- *Convergence of Markov Chain Samplers,* Richard Smith, University of North Carolina
- *Panel of Editors of Journals for Statistical Computing,* Ed Wegman, George Mason University
- *Computational Techniques in Genetics and Molecular Biology,* Francoise Seillier-Moiseiwitsch, University of North Carolina

- *Efficient Bootstrap Computations,* Tim Hesterberg, Franklin and Marshall College
- *Neural Net Tutorials,* organized by Ron Gallant, University of North Carolina
- *Wavelets Tutorial,* Mary Ellen Bock, Purdue University
- *Gibbs Tutorial,* Adrian Smith, University of Nottingham
- *Computing for MetaAnalysis,* Bill DuMouchel, Harvard
- *Robust Regression and Multivariate Analysis,* David Rocke, University of California at Davis
- *Green Thumbs: Extensions and Applications of Tree Modeling Methods,* Sally Morton, Rand
- *Applications of Wavelets,* Iain Johnstone, Stanford
- *Bayesian Curve Fitting,* Mike West, Duke University
- *Longitudinal and Mixed Models,* Russ Wolfinger, SAS Institute
- *Statistics of Protein and Macromolecular Structures,* Peter Munson, National Institutes of Health
- *Panel: Statistics Education in the Computer Age,* Raoul LePage, Michigan State University
- *Issues in Software* Stephen Eick, AT&T Bell Labs
- Additional Tutorial Sessions. *Internet Facilities* Tim Arnold, N C State University, *Perl programming* Phil Spector, University of California at Berkeley.

John Sall
SAS Institute
sall@sas.com



by Andrejs Dunkels

Interface '95

27th Symposium on the Interface:
Computing Science and Statistics
April 13-16, 1995,
Pittsburgh PA,
Vista Hotel

The Interface Conference is the premier annual conference on the interface of computing and statistics. It is sponsored by the non-profit Interface Foundation of North America, and will be hosted in 1995 by Carnegie Mellon University and the Pennsylvania State University with Michael Meyer and James Rosenberger as joint program chairs. The 1995 conference will be in Pittsburgh Pennsylvania, home of many excellent academic and industrial statistics and computer science research progams, and one of America's most livable cities.

Contact: Email: interface95@stat.cmu.edu
Phone: (412) 268-3108 Facsimile: (412) 268-7828
Mail: Interface 95, Department of Statistics, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA.

Artificial Intelligence and Statistics

Call For Papers

Fifth International Workshop on
Artificial Intelligence and Statistics
January 4-7, 1995
Ft. Lauderdale, Florida, USA

Purpose: This is the fifth in a series of workshops that has brought together researchers in Artificial Intelligence and in Statistics to discuss problems of mutual interest. The exchange has broadened research in both fields and has strongly encouraged interdisciplinary work.

This workshop will have as its primary theme: "Learning from data". Papers on other topics at the interface of AI & Statistics are *strongly* encouraged as well (see Topics below).

Format: To encourage interaction and a broad exchange of ideas, the presentations will be limited to about 18 discussion papers in single session meetings over the three days of the technical portion of the workshop (Jan. 5-7). Focussed poster sessions will provide

the means for presenting and discussing the remaining research papers. Papers for poster sessions will be treated equally with papers for presentation in publications. Attendance at the workshop is not limited to presenters. The three days of research presentations will be preceded by a day of tutorials. These are intended to expose researchers in each field to the methodology used in the other field.

Language: The language will be English.

Topics Of Interest: We strongly encourage research papers in the areas of selecting models from data, integrated man-machine modelling methods, empirical discovery and statistical methods for knowledge acquisition, probability and search, uncertainty propagation, combined statistical and qualitative reasoning, inferring causation, quantitative programming tools and integrated software for data analysis and modelling, discovery in databases, meta data and design of statistical data bases, automated data analysis and knowledge representation for statistics, machine learning, and clustering and concept formation.

This list is not exhaustive and we encourage submissions in other areas at the interface of artificial intelligence and statistics as well.

Submission Requirements: Submissions will be extended abstracts (up to four pages). Submissions for discussion papers (and poster presentations) will be considered if postmarked by June 30, 1994. Abstracts postmarked after this date but before July 31, 1994, will be considered for poster presentation only. Please indicate the topic(s) addressed by your abstract and include an electronic mail address for correspondence. Acceptance notices will be mailed by September 1, 1994. Preliminary papers (up to 20 pages) must be returned by November 1, 1994. These preliminary papers will be copied and distributed at the workshop.

Submissions may be sent by air mail or email (latex documents preferred) to either chair:

Doug Fisher, General Chair
5th Int'l Workshop on AI & Stats
Department of Computer Science
Box 1679, Station B
Vanderbilt University
Nashville, TN 37235 USA
dfisher@vuse.vanderbilt.edu

or

Hans Lenz, Programme Chair
5th Int'l Workshop on AI & Stats
Free University of Berlin
Department of Economics
Institute for Statistics and Econometrics
14185 Berlin, Garystr 21 GERMANY
HJLENZ@fubvm.wiwiss.fu-berlin.de

Continuing Education in Statistical Computing

by Thomas F. Devlin
Statistical Computing CE Chair

CE Scholarships to Interface 94

The Statistical Computing Section will offer ten scholarships to short courses at Interface 94. These awards inaugurate the Continuing Education (CE) Scholarship Program established last August by the Executive Committee.

The short courses, which the Section is co-sponsoring are:

- "Modern Nonparametric Regression and Classification" by Trevor Hastie and Robert Tibshirani,
- "Algorithms for Estimation and Visualization of Multivariate Density Functions with Applications to Clustering" by David W. Scott,
- "Resampling-Based Multiple Testing" by Stanley Young and Peter Westfall, and
- "Data Analysis using Interactive Dynamic Graphics" by Deborah Swayne, Martin Koschat and Dianne Cook.

The scholarships, which cover the cost of a short course, will be awarded on the basis of need with preference given to beginning researchers. Recipients will be responsible for their own travel expenses. Applicants should identify the short course of interest, explain the professional importance of the course to them, and include a resume. Applications should be sent to Thomas F. Devlin, Mathematics and Computer Science Department, Montclair State College, Upper Montclair, NJ 07043, or via email to devlin@mozart.montclair.edu.

Call for Proposals for CE at 1995 Joint Statistical Meetings

It's never too soon to begin planning an ASA continuing education event. This coming September 22, 1994, is

the deadline for submission of proposals of short courses and other CE activities that the Statistical Computing or Statistical Graphics Section co-sponsor at the Joint Statistical Meetings to be held in Orlando, Florida. Co-sponsorship can benefit both you and your Section. For a copy of ASA proposal guidelines, please send email to me.

Thomas F. Devlin
Montclair State College
201-655-7244
devlin@mozart.montclair.edu



Second Annual Meeting of the International Society for Bayesian Analysis

Announcement and Call for Papers

The second annual meeting of the International Society for Bayesian Analysis (ISBA) will be held on June 10 and 11, 1994, in the beautiful setting of Alicante, Spain, immediately following the Fifth Valencia meeting. The ISBA is a newly created international scientific society with the objective of interfacing between methods of Bayesian statistics and areas such as engineering, chemistry, physics, economics, business, astronomy, education, psychology, medicine and sociology. Researchers from all fields of science are invited to submit Bayesian papers for presentation by sending a short abstract to the chairman of the Program Committee:

Professor Michel Mouchart
Institute of Statistics
34 Voie du Roman Pays
1348 Louvain-la-Neuve
Belgium
Fax: +32-10-47 3032
mouchart@stat.ucl.ac.be

In addition, a cash prize of $1000, named after Toby J. Mitchell, a Senior Research Staff Member at Oak Ridge Laboratory, who died on June 5, 1993, will be awarded to a paper describing how a Bayesian analysis has solved a truly applied problem. Entries for this award should be directed to:

Professor James S. Press
Department of Statistics
University of California
Riverside, CA 92521-0138
VXPRESS@UCRVMS.BITNET

Registration forms for this meeting can be obtained from:

Professor M.J. Bayarri
Department of Statistics and O.R.
Universitat de Valencia
Av. Dr. Moliner 50
46100 Burjasot, Valencia, Spain
Fax: +34-6-386 4735
Bayarri@vm.ci.uv.es



Reader's Notes

SECTION OFFICERS

Statistical Graphics Section - 1994

Roy E. Welsch, Chair
617-253-6601
Massachusetts Institute of Technology
rwelsch@sloan.mit.edu

David W. Scott, Chair-Elect
713-527-6037
Rice University
scottdw@rice.edu

Richard A. Becker, Past-Chair
908-582-5512
AT&T Bell Laboratories
rab@research.att.com

William DuMouchel, Program Chair
617-489-2631
dumouche@jimmy.harvard.edu

Sallie Keller-McNulty, Program Chair-Elect
913-532-6883
Kansas State University
sallie@cecil.stat.ksu.edu

Michael M. Meyer, Newsletter Editor (93-96)
412-268-3108
Carnegie Mellon University
mikem@stat.cmu.edu

Linda A. Clark, Secretary/Treasurer
908-582-4807
AT&T Bell Laboratories
lac@research.att.com

Howard Wainer, Publications Officer
609-734-5623
Educational Testing Service
Princeton, NJ 08541
hwainer@rosedale.org

Jane F. Gentleman, Rep.(94-96) to Council of Sections
613-951-8213
Canadian Centre for Health Information
Ottawa, ON K1A OT6, CANADA
GENTLEJF.NRCVM01.bitnet

Sally C. Morton, Rep.(94-95) to Council of Sections
310-393-0411
The Rand Corporation
Santa Monica, CA 90407-2138
Sally_Morton@rand.org

Statistical Computing Section - 1994

Trevor J. Hastie, Chair
908-582-5647
AT&T Bell Labs
trevor@research.att.com

Mary Ellen Bock, Chair-Elect
317-494-6053
Purdue University
mbock@stat.purdue.edu

Sanford Weisberg, Past Chair
612-625-8777
University of Minnesota
sandy@stat.umn.edu

Sallie Keller-McNulty, Program Chair
913-532-6883
Kansas State University
sallie@cecil.stat.ksu.edu

John A. Rice, Program Chair-Elect
510-642-6930
University of California at Berkeley
rice@stat.berkeley.edu

James L. Rosenberger, Newsletter Editor (93-96)
814-865-1348
The Pennsylvania State University
jlr@stat.psu.edu

John Sall, Secretary-Treasurer
Sas Institute
sall@sas.com

Karen Kafadar, Publications Liaison Officer
301-496-8556
National Cancer Institute
kk@helix.nih.gov

Ronald Thisted, Rep.(94-96) to Council of Sections
312-702-8332/8333
The University of Chicago
r-thisted@uchicago.edu

Russell Lenth, Rep.(93-95) to Council of Sections
319-335-0814
University of Iowa
rlenth@stat.uiowa.edu

Daryl Pregibon, Rep.(92-94) to Council of Sections
908-582-3193
AT&T Bell Labs
daryl@research.att.com

INSIDE

A WORD FROM OUR CHAIRS
- Statistical Computing . . . 1
- Statistical Graphics . . . 1

FEATURE ARTICLE
- The Process of Reading Statistical Maps: The Effect of Color . . . 1

EDITORIAL . . . 1

FROM OUR CHAIRS (Cont.) . . .
- Statistical Computing . . . 2
- Statistical Graphics . . . 3
- Toronto Joint Meeting Program . . . 4

DEPARTMENTAL COMPUTING
- Cost/Benefit of Uniformity . . . 7

GEOGRAPHIC INFORMATION SYSTEMS
- Spatial Resolution, Hazardous Waste Siting, and Freedom of Information . . . 9

FEATURE ARTICLE (Cont.) . . . 12

TOPICS IN SCIENTIFIC VISUALIZATION
- Color Perception, the Importance of Gray and Residuals, on a Choropleth Map . . . 16

UNIX COMPUTING
- What is a Shell, Part Two . . . 20

NET SNOOPING
- URLs for Fun, and Profit . . . 23

BITS FROM THE PITS
- Statistical Computing And Graphics In Science And Industry . . . 24

CONFERENCE NOTICES
- Interface '94 . . . 27
- Interface '95 . . . 28
- Artificial Intelligence and Statistics . . . 28
- Continuing Education in Statistical Computing . . . 29
- Second Annual Meeting of the International Society for Bayesian Analysis . . . 30

SECTION OFFICERS . . . 31

Statistical COMPUTING & GRAPHICS

The *Statistical Computing & Statistical Graphics Newsletter* is a publication of the Statistical Computing and Statistical Graphics Sections of the ASA. All communications regarding this publication should be addressed to:

James L. Rosenberger
Editor, Statistical Computing Section
Department of Statistics
The Pennsylvania State University
University Park, PA 16802-2111
(814) 865-1348
JLR@stat.psu.edu

Michael M. Meyer
Editor, Statistical Graphics Section
Department of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213-1380
(412) 268-3108
mikem@stat.cmu.edu

All communications regarding membership in the ASA and the Statistical Computing or Statistical Graphics Sections, including change of address, should be sent to:

American Statistical Association
1429 Duke Street
Alexandria, VA 22314-3402 USA
(703) 684-1221 • FAX (703) 684-2036
ASAINFO@ASA.MHS.COMPUSERVE.COM

PENNSTATE

1855

Department of Statistics
University Park, PA 16802-2111

Nonprofit Organization
U. S. POSTAGE
P A I D
Permit No. 1
University Park, PA 16802

Published by the Penn State Department of Statistics
326 Classroom Building, University Park, PA 16802-2111

This publication is available in alternative media on request.

Penn State is an affirmative action, equal opportunity university.
U.Ed.SCI 94-49