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

December 95 Vol.6 No.3

Statistical COMPUTING & GRAPHICS

A WORD FROM OUR CHAIRS

Statistical Computing

Time goes so swiftly that it is hard to believe a year has gone by since I became a voice for the Statistical Computing Section in this column. In January, the Section's voice will be that of Sallie Keller-McNulty who will become the new chair of the Section. She is currently at the National Science Foundation serving as the director of the Statistics Program. It is a time when the budget outlook for research looks pretty grim. The Stastistics Program is part of the Division of Mathematical Sciences Division and its Director Don Lewis is looking for input from the statistical community on priorities. The power of email gives you a voice.

Mary Ellen Bock is the outgoing 1995 Chair of the Statistical Computing Section. She welcomes the new incoming Chair for 1996 and makes a call for new candidates for the section. She also endorses the Interface meeting scheduled for Sydney, Australia next July

One of the things that the Executive committee of the Section has discovered is that the Secretary-Treasurer has a steep learning curve in the year-long term of office. Just as he or she learns what is going on, it is over. That is why we have recommended that this office have a two year term. You will see this proposed change to the constitution on the ballot next time. If passed, it will affect the 1997 candidates. The list of candidates for the Section offices is outstanding and ensures the health of the Section. There is always a search for new voices though and suggestions from you for nominees for future slates are welcome.

One of the most popular activities of the Section are the short courses that are offered. In particular, a tuition waiver/scholarship is offered for these by the Section for the Joint Statistical Meetings. CONTINUED ON PAGE 3

A WORD FROM OUR CHAIRS

Statistical Graphics

David Scott, the outgoing 1995 Chair of the Statistical Graphics Section, gives his parting words as Chair. He hands the gavel to the incoming 1996 Chair, and thanks the many participants of the section during the past year.

I would like to begin this my last column by expressing my gratitude for the opportunity to serve the Statistical Graphics Section these past several years, and I am looking forward to a year as past-Chair. I am particularly proud of the accomplishments of the section, as well as the extraordinary efforts of the officers. Robert Newcomb has done a fine job putting together our web page, which you can visit through `http://www.amstat.org/`. The scientific program in Orlando was well-attended, and I would like to thank Sallie Keller-McNulty and Sally Morton for putting the program together. Steve Eick has put together an excellent program for Chicago that draws on communities not usually in attendance at our meetings. I am pleased to report that the Section's finances are in excellent order, and that we have a balanced (no surplus) budget in place that should improve the visibility and services offered.

The scientific program in Orlando was well-attended, and I would like to thank Sallie Keller-McNulty and Sally Morton for putting the program together.

The officers have tried to be particularly responsive to related activities. For example, we have offered substantial support to the effort to improve the internet connectivity of ASA headquarters; CONTINUED ON PAGE 3

EDITORIAL

Final Edition

This year end, December 1995, issue (Vol. 6 No. 3) of the Newsletter should have arrived before the end of the year. We hope it arrives in time to bring you some cheer during the holidays. But with the inclusion of color and the demands of other deadlines, and a few reluctant writers, it will be a close call. The deadline for the next issue is February 15, though read on to learn where to send your letters and articles.

Three years later—it is time for new editors

Mike and I certainly want to thank all of the contributors to the Newsletter these past three years for their interesting and informative columns and articles. We hope you have enjoyed the Newsletter, for the articles, the announcements and our attempts to bring both Computing and Graphics news that was timely and educational.

We started with lots of enthusiasm. We created a new face and new format for this joint Newsletter, and then prepared a LaTeX style file to simplify the typesetting of the Newsletter. This allowed us to embed tables and figures and prepare the final product in camera ready PostScript. We received most submissions electronically, including photos, tables and figures. The last three years' issues are also available on the WWW, in PostScript format, on my homepage `http://www.stat.psu.edu/`. In most cases the electronic version was available weeks before the printing and mailing brought a copy to your door.

Our new editors will hopefully have the same enjoyment of bringing you (and us) the latest ideas in statistical computing and graphics. The new computing editor is Mark Hansen, a member of the technical staff at AT&T Bell Laboratories, and recent Ph.D. in Statistics from UC Berkeley. His position at the Lab should be a useful source of material and people networking. I am enthusiastic about handing over the reins of the Computing editorship to him. As you can read in the article on page 4, he has already received what I could provide of the endowment from the past three years.

From Mike

Three years is a long time. When Jim and I started doing the newsletter, my life was a lot less complicated. Since then my wife and I have started to raise a family, Jim and I have hosted the Interface conference, and I have several times significantly changed my administrative work load. All of these changes have taken a toll on my ability to devote time to the newsletter and Jim has gracefully filled in a lot of void. All of the credit for getting out a timely newsletter has to go to Jim.

I'm delighted with the choice of the two new editors. The new graphics editor is Mario Peruggia of Ohio State university. Mario has a lot of experience with computers and graphics, and should bring a quite different perspective to the newsletter. I wish both Mark and Mario the best of luck. The easy thing about editing the newsletter is that, given the pace of change in the computing industry, there is never a shortage of things to write about. If there is one thing that **you** can do to make their lives easier, it is to communicate. One of my biggest disappointments as editor was the lack of feedback (and lack of spontaneous submissions) from the readers.

Jim and I made a first pass at making the newsletter available electronically, and someday I might even finish converting the back issues into HTML. I certainly hope that the new editors will consider being much more aggressive about an electronic manuscript than we were. Of course, given my own interests, I would be delighted if they chose to make the material available on either StatLib or on `amstat.org`, the American Statistical Association server.

In this issue

We again bring you some helpful articles from our regular columnists.

In his "UNIX Computing" column, see page 18, Phil Spector provides a tutorial on how to write your own UNIX commands, sometimes called scripts.

Our Geographical Information System commentator, Mark Monmonier, returns in this issue to bring us an account from his recent experience in court. He tells how the key statistical concept of "Margin of Error" played a role in trying to extract precise information from a map. Embedding and extracting information from maps can result in error, and representing this source of variability fairly provides a challenge for the cartographers and GIS specialists. Read his article on page 8 to learn how important this information can be.

In the color article in this issue, see page 11, Dan Carr displays an experimental plotting technique to summarize the data from a massive table in his column on "Topics in Scientific Visualization". The suggested plot combines several tools for embedding complex data in an uncluttered plot—but you be the judge, and send your views to the author, or better yet, to the Newsletter.

Mike Conlon provides some excellent advice on the constant challenge we face when buying hardware for our own departments in his "departmental computing" column. This advice can help faculty members, department heads, or system administrators, and also provide a useful reference if you are having trouble convincing someone of the need.

Mike Meyer unveals a new product bursting onto cyberspace, HotJava, in his "NetSnooping" column. This product will liven up your net surfing while also reducing the demands on the Internet. It accomplishes this by distributing the processing required for animations to the viewing machine. Read it and try it for yourself—and remember that you read it here first.

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, but will gladly accept plain ASCII text files as well. All submissions and correspondence should now be sent to Mario or Mark whose addresses are given on page 4 in the article below.

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 Graphics

CONTINUED FROM PAGE 1

we continue our support of the undergraduate data analysis contest; we are supporting the archival efforts of Harry Posten and the distinguished statistician video library; and seed fundings for a new award initiated by COPSS.

The Section is investigating mechanisms to help encourage graduate student research and participation in graphics. Suggestions for the creation of an award for outstanding graphics paper will also be considered. The video library has been improved and our librarian, Deborah Swayne, is the correct person to contact (and, again, visit our www page).

We have put together an outstanding slate of candidates for the next election. Please vote and suggest to your colleagues to consider joining the Section. The discounted rate for joining both the Graphics and Computing Sections is a real bargain. A slightly revised charter is also on the agenda. The Section offers many opportunities for new initiatives and general participation, and I encourage you to contact our new chair, Bill DuMouchel. After January 1, of course. It's all yours, Bill!

David W. Scott
Chair, Statistical Graphics Section
Department of Statistics
Rice University
scottdw@rice.edu

***FROM OUR CHAIRS (Cont.)*...**

Statistical Computing

CONTINUED FROM PAGE 1

It complements the Student Paper Competition which you will see described in this newsletter. Although the Computing Section will offer an exciting program of lectures and talks at the Joint Statistical meeting in Chicago in August, 1996, it also supports the Interface meeting which will be held in July, 1996, in Sydney, Australia. The meeting coincides with the Australian Statistical Conference and several others and promises to be a great one. It's time to dig out the passport and get a visa.

I would like to give special thanks to the officers who served the Statistical Computing Section this year. Particular thanks go to the Statistical Computing Section's Newsletter Editor Jim Rosenberger who served for three years. He and the Statistical Graphics Section Newsletter Editor Mike Meyer led the newsletter to new heights. Pull out your old copies if you have them and watch the transformations over the three years. There is a great new crew coming in though. Join me in welcoming them.

Mary Ellen Bock
Chair, Statistical Computing Section
Department of Statistics
Purdue University
mbock@stat.purdue.edu

NEWS CLIPPINGS

New Newsletter Editors

by Mark Hansen

Some weeks ago, a large shar file was deposited on my electronic doorstep; a gift from `jlr@stat.psu.edu`. The moment I unpacked its contents, PostScript and TeX commands spilling out onto my disc, I realized how much work Jim and Mike had put into this publication over the past three years. The attention given to this newsletter by Jim and Mike and the other editors emeritus has shaped it into an incredible resource for those interested in statistical computing and graphics.

As incoming editors, Mario and I are committed to maintaining the high quality of this publication. Not surprisingly, this depends on our ability to solicit articles from the computnik community at large. In the next few weeks, we will be establishing a website through Netlib to start collecting input from our readers. In addition, we are hoping to renew as many of the current columnists as possible, and have high hopes for establishing new regular features. Mario and I will work hard to ensure that this newsletter continues to be a timely, quality resource for the statistical community.

Mark Hansen
Incoming Editor
Statistical Computing Section
AT&T Bell Laboratories
`cocteau@research.att.com`
and
Mario Peruggia
Incoming Editor
Statistical Graphics Section Ohio State University
`peruggia@stat.mps.ohio-state.edu`

STATISTICAL COMPUTING

Student Paper Competition

by Trevor Hastie

The Statistical Computing Section of the ASA is again sponsoring a Student Paper Session at the Joint Statistical Meetings in 1996. Our 1995 competition was successful, and attracted 16 entries. The judges had a difficult task selecting the 4 winners, who all presented their papers in the special session at the 1995 ASA meeting in Orlando.

The topic of the session is *Statistical Computing*. Three to four students will be selected to participate in this session, which will include a discussant nominated by the selection committee. All fees associated with registration, accommodation and travel to the conference will be awarded to the participants in this Session.

Students at all levels (undergraduate, Masters, and Ph.D.) are encouraged to participate. To be eligible, an applicant must be a registered student in the fall of 1995. The applicant must be the first author of the paper.

To be considered for selection in the session, students must submit an abstract, a six page manuscript, a resume, and a letter of recommendation from a mentor familiar with their work. The manuscript should be single-spaced in a 10 point font with one inch margins (this is consistent with ASA's Proceedings guidelines.) In the case of joint authorships, the mentor should indicate what fraction of the contribution is attributable to the applicant.

All application materials MUST BE RECEIVED by January 10, 1996. They will be reviewed by the Statistical Computing Section Student Paper Competition Award committee. The topic of the paper should be in the area of statistical computing, and might be original methodological research, some novel application, or any other suitable contribution (for example, a software related project). Selection will be based on a variety of criteria at the discretion of the selection committee, and will include novelty and significance of contribution, amongst others.

Award announcements will be made January 22, 1996. The selection committee's decision will be final. There will be a discretionary cap of $1000 on any given award, but it is anticipated that this figure should be more than sufficient to cover the expenses.

Students not selected for inclusion in the Session may submit their abstract and a registration fee to ASA by February 1 if they plan to attend the Joint Meetings. Those abstracts must be submitted following the ASA abstract submission instructions described in AMSTAT News. Students selected for inclusion in the session will receive further information about abstract submission and fee waivers from the award committee.

Inquires and materials should be emailed or mailed to either one of:

Trevor Hastie, Statistics Department, Sequoia Hall, Stanford CA 94305.
trevor@playfair.stanford.edu

Daryl Pregibon, Room 2C264, AT&T Bell Laboratories, 600 Mountain Ave, Murray Hill, NJ 07974.
daryl@research.att.com

All electronic submissions of papers should be in postscript.

BUSINESS MEETING MINUTES

Statistical Computing Section

Minutes of the June 23, 1995 Section Executive Committee Meeting, Interface Conference, Pittsburgh, PA

Present: Mary Ellen Bock, Michael Myer, Thomas Devlin, Sallie Keller- McNulty, Karen Kafadar, James Rosenberger, Deborah Swayne, Terry Therneau

Newsletter

Jim discussed the issue of a replacement newsletter editor, to begin on January 1, 1996. He has approached a small number of people without success. A suggestion was made that a good choice for this post would be someone who has just gotten tenure; the pressure of publication has lifted somewhat, and this person would benefit most from the exposure, in terms of promotion to the next level. Several good names were then suggested by members of the Committee, along with rankings . Jim will contact these people sequentially and report on his success at the August meeting.

Request to Republish Articles

The Statistical Software Newsletter, of the International Journal of Computation and Data Analysis, has forwarded a request to republish some of the articles found in our own newsletter. It was agreed that on a whole this is a good idea, since the purpose of the newsletter is to distribute information widely. A final decision on this, however, was felt to be the purview of the editors of the newsletter. The following policy statement is made as a suggestion for their consideration: The republished information should attribute both the source (our newsletter) as well as the original author; permission should be gained from both the newsletter editor and the original author for anything that will be republished. We noted that the newsletter already contains an editorial statement of similar effect.

Internet connection at the ASA office

Mike Myer gave a report on the status of an Internet connection for the American Statistical Association office. Per the electronic mail discussion held a few months ago, both our section and the computational graphics section agreed to underwrite a portion of the cost to create the infrastructure necessary for the ASA office to have an Internet connection. By the ASA meetings in August, Mike intends to have a World–Wide Web page in place; the mail service will likely remain on Compuserve for several more months.

His overall plan, once this is set up, is to encourage the members of the ASA office to view ASA membership as a resource. The ASA staff are enthusiastic but inexperienced, and do not have extra resources (people) to take on this new role as a primary portion of their job.

The assigned name of the ASA office is amstat.org, (“asa” was already taken). Mike has made plans to reserve the name “ims.or” as well, in anticipation of that society’s desire to join the network at a later date. Many WWW pages are nearly ready to go, so the service should be useful as soon as it comes up. Dan Horowitz is very supportive; he intends to allow searching of the member directory through the WWW interface.

Many sections offered to contribute money, so we have $15,000 to work with at the present moment. Access charges will be about $600 a month; Mike feels that this should be sufficient to cover the first year’s cost. The Statistical Computing Section’s original funding was $4,000 or $5,000; it was moved and seconded to approve additional funding in this amount, if necessary to complete the project.

Secretarial term

The change of Secretary/Treasurer to a two-year term, discussed in prior meetings, requires a change to the bylaws. The issue will appear on the ballot for the 1996 election and will take effect, if passed, for the 1997 electee.

Continuing Education Report

The Section sponsored Leo Breiman’s course in partitioning models at the January conference on neural nets. There were 75 attendees out of a pool of 220 conference attendees. Tom Devlin felt that this was a very successful outreach, both from a business aspect and as a way of introducing our Section’s activities to people outside of the statistical profession.

At the Joint Statistical meeting in August, three of the four courses proposed by the Section have been accepted for presentation. They are the course on Bayesian Methods (BUGS), Applied Discriminant Analysis, and Extending the Cox Model. We expect an enthusiastic turnout.

The Section sponsors a tuition waiver/scholarship program for the short courses offered at the joint meetings. The announcement of that program was sent in too late to appear in the Amstat Newsletter. It was suggested that the announcement be mailed to the Statlib distribution list, as well as the list of department heads of statistical sections.

Proposals are due in October for courses at the next American Statistical Association meetings. Tom solicited suggestions from anyone who had them. A discussion ensured about which courses would be most appropriate for our Section to sponsor. One suggested input to this process is to create an article for the newsletter describing what courses have been given in years past. This could serve to both avoid unnecessary duplication, as well as periodic representation of our most popular offerings. Given the success of our foray into the neural nets conference this winter, we should also consider which other organizations we might cooperate with for future offerings.

Report on meetings

Sallie presented to the group the discussion and findings of the ASA discussions on improvements and suggestions for the content of the statistical meetings. Particular issues were:

- how to improve presentations
- make posters more exciting or tantalizing to presenters
- invited poster sessions

She pointed out that the Statistical Computing Section has fewer sponsored sessions than it has in the past years.

Several ideas were suggested as input to this process including the following:

1. In many other societies, the equipment needed for presentation is marked on the abstract form. This might help coordination of appropriate equipment for presentation.
2. Network connectivity and specialized computing equipment could be much more useful, reliable, and effective in a poster setting. It was suggested that the Section fund an invited poster section, including some of the equipment costs. This special poster session might be displayed for a longer period of time, in a more prominent location, or with other features to make it more of an attraction.

The meeting adjourned at 1:30pm.

Respectfully submitted,

Terry M. Therneau
Secretary/Treasurer
Statistical Computing Section

DEPARTMENTAL COMPUTING

Departmental Computing: Buying Hardware

by Michael Conlon

In this column I'll assume that you have the job of buying more hardware and that you have decided which platform is appropriate for your department. I'll also assume that you have made this decision keeping in mind the software you need to run, the volume you need to handle and the integration you need to achieve with the rest of your enterprise.

The Big Buy

Making a major purchase—enough machines to replace the hardware in your department—provides great opportunities to achieve higher productivity. A well-designed system, achieving pre-planned goals of integration, networking and support of key application programs will lower maintenance costs, lower training costs and increase productivity. In some cases, a departmental make-over will enable the department to standardize on a hardware platform and an operating system. Standardizing is a desirable goal in small to medium-sized departments with common operational characteristics. Such a purchase should be made with as much input as possible from the members of the department. Few people embrace change. Department members should be convinced that the change will result in simplifying their jobs or enabling them to do tasks that were previously difficult or impossible.

You may require assistance in planning your major purchase. Don't be shy. Ask for help from knowledgeable members of your department, from other departments, and from your computer center. Enlist the help of vendors. Compare the claims of various vendors. Consider how your purchase will use existing equipment, interface with your existing network, hook to existing printers, storage devices, phone systems. Many vendors have trade-in or trade-up programs. These may enable you to take your purchase further than you might have thought at first. You may be able to combine funds planned for the major purchase with funds earmarked for maintenance of existing equipment. The maintenance may no longer be needed on new equipment (warranty service may suffice) enabling you to stretch your dollars. Achieving a department-wide make-over is highly desirable to standardize, support interoperability and lower training, maintenance and support costs.

A large purchase may enable you to negotiate with the supplier regarding price. Single unit and small count purchases must typically be done at published price. Larger purchases, especially those that enable a vendor to make a major market expansion in your enterprise, may put you in a very enviable bargaining position. Take advantage of the opportunity to drive down prices. A bidding process may work to your advantage on a major purchase.

Give people input into the design and purchase process. But don't confuse input with decision-making responsibility. If you, and your committee are designing the system, solicit input, but don't abdicate final decision-making. One path to disaster is to give each department member a "budget" and ask them to make their own single-machine purchase decisions. The result will be incompatible hardware that cannot be assembled into an inter-operable departmental computing system.

The Single Machine

Individual machines must be purchased in keeping with a well-organized and understood departmental plan for computing. The department should purchase computers from a single vendor. Read Deming. Fostering a relationship with a vendor gains you access to product announcements and discounts. You should purchase computers, even one at a time, from a reputable major vendor. Stay informed. Read the industry magazines. You need to be informed when talking to your vendor. The industry moves very quickly and you may have heard something about a system that your vendor has not.

Getting the configuration you want is very important. You don't want to pay for parts you will not use. Sound cards, slow modems and small capacity tape drives are often added as "features" by vendors to make their systems more attractive. Think of how the machines will be used and if any of these things have value. An office computer should not need a modem—a network card should be standard equipment. Office computers should not need tape drives—they should be connected to a server via an office network and should be backed up across the network using a daily automated process.

Consider the lifetime of the computer you are purchasing. It should begin with an easy return policy. If there are problems "out of the box" the vendor should have easy to use procedures for returning defective parts and/or systems and getting immediate replacement. Your computer should be covered by a three year warranty. Your computer should use industry standard memory, disks and other parts that can easily be replaced or upgraded. If your computer is a PC, it should support the industry "plug and play" standards. While this feature may not be important to you today, it will be in the near future as operating systems recognize plug and play components.

Larger purchases, especially those that enable a vendor to make a major market expansion in your enterprise, may put you in a very enviable bargaining position.

Price cannot be the determining factor in making a computer purchase. Saving $100 on a $3,000 computer can be penny-wise and pound-foolish when your discount computer turns out to be incompatible with your networking standards, or your maintenance provider. Investigate the contracts your enterprise may already have in place with major vendors. These may save you considerable amounts over "magazine" prices.

Speed cannot be the determining factor in making a computer purchase. If computer A is 5% faster than computer B at the same price, but computer A is requires lengthy system management procedures to maintain and upgrade, the advantage is easily lost.

"Free" software that comes bundled with systems must be evaluated carefully. Games, obscure utilities, office productivity software that you will not use have no value. You may be able to delete the software bundle from a major vendor's machine and lower the price.

Upgrading Existing Machines

Upgrading existing machines may be a cost effective way of achieving a collection of machines that can use a standard operating system and common software. Adding RAM is the easiest and safest way to improve performance of existing computers. Investing in processor upgrades can also work well. Small upgrades that enable marginal machines to resume "front-line" status running the current applications and operating systems is money well spent.

In general, upgrades to machines that are more than four years old will not work well. These machines require such major improvements and the money to do the upgrades will approach the cost of a new system. A $399 processor upgrade can jump to $1,600 when I/O upgrades, video upgrades, drive upgrades, RAM upgrades and others must be added to support the new processor. The result may still cost less than a new system but will be a collection of miscellaneous parts rather than a well-designed, integrated system. When this process is repeated on several machines in a department over a period of years, the result is an unmanageable collection of parts rather than a collection of systems. Avoid upgrades to older machines.

Trickle Down Computing

As new computers replace older computers, the older computers must find new homes. In some cases, they will have been traded in or upgraded, and no net gain in computers will have resulted from the purchasing. In other cases, new computers will replace old and the older will replace still older computers. Such is trickle-down computing, where someone's "new" computer is actually someone else's old computer. This can work well in larger organizations with diverse computing needs. Research scientists with demanding computing applications may outgrow computers that are still perfectly suitable for clerical computing. Make sure the department member receiving the "new" computer is benefiting from the change—the change should move them toward the department's standard environment.

Draw the line when the older computers will require disproportionate share of the maintenance or training effort of the department. If your department has, for example, standardized on OS/2, but your oldest computers can't run it, it may be better to move the older computers entirely out of the department rather than trying to live with them as DOS machines and requiring an entire additional set of application programs, networking software and operating procedures. In some cases, older computers may begin new careers as data entry terminals, terminal servers, service gateways or other dedicated applications supporting your department.

Computers which can no longer be used productively within your department may be traded across departments for some additional resources. Or you may be able to trade the computers to a used computer dealer. Or your organization may have standard means for compensating your department for such surplus equipment. At some point, the cost of maintaining these older systems is larger than their operational value and they must be removed from service, even if they continue to function.

Michael Conlon
Department of Statistics
Box 100212 HSC
University of Florida
Gainesville, FL 32610–0212
`mconlon@stat.ufl.edu`
Home page:
`http://www.clas.ufl.edu/~mconlon`

GEOGRAPHIC INFORMATION SYSTEMS

Margin of Error

by Mark Monmonier

For me, the O.J. trail was not the only legal proceeding to drag on last summer: I was an expert witness in an environmental protection hearing that stretched eight days of testimony from early June through mid-August. The folks on my side had hoped the hearing could be wrapped up in two days. But the other side extended the process by hammering relentlessly on a concept near (if not always dear) to the hearts of statisticians and GIS experts: error.

Background

I testified for only a few hours each on days 1 and 3, and for less than an hour on day 8, as the final rebuttal witness. I had been hired as a expert in cartography and geographic analysis by the Onondaga County Resource Recovery Agency (OCRRA), which planned to open a new solid-waste landfill in the Town of Van Buren, a comparatively rural area in the northwest corner of Onondaga County in upstate New York. OCRRA had recently built a "resource-to-energy facility"—a trash-burning electrical generating station—and was trucking the incinerator ash to a commercial landfill about eighty miles away. The new landfill would not only be closer and cheaper but also guarantee a disposal site for several decades.

OCRRA's other experts were a civil engineer employed by the firm that had designed the landfill and a consulting soil scientist, who until a years few ago had directed the state office of the federal Soil Conservation Service

(SCS). Our position was bolstered by a team from the regional office of the New York State Department of Environmental Conservation (DEC): a lawyer and several environmental scientists, who were satisfied that the trash agency had met the requirements for a permit. The other side also consisted of two parties: the Town of Van Buren ("the Town") and People United for Rural Environment (PURE), a citizens group. In the jargon of environmental law, OCRRA was "the applicant," the Town and PURE were "the intervenors," and the DEC team was "staff." Directing the proceedings was an administrative law judge ("the ALJ"), from DEC headquarters in Albany.

The Hearing

The hearing focused on a section of the New York Code of Rules and Regulations governing the siting of landfills in farming areas. Referred to as the "agricultural prohibition," NYCRR 360-1.7(a)(2)(i) prohibits DEC from issuing a permit if the land is to be "taken through eminent domain; consists predominantly of agricultural soil group 1 or 2 ([according to the] Land Classification System as certified by the New York State Commissioner of Agriculture and Markets); and is within an agriculture district formed pursuant to the Agriculture and Markets Law." This provision applied because the owner of one of the eight parcels had refused to sell. OCRRA thus had the burden of demonstrating that the better soils did not cover more than half of that portion of the site within the agricultural district. Since the state land classification system is based on published SCS soil survey reports, applying the provision might seem a simple matter of measuring areas and calculating the percentage of prime soils: if the result is greater than 50 percent, the permit is denied, and if it's less than or equal to 50 percent, the permit is approved—assuming the applicant meets all other requirements.

Seems simple, but it wasn't. After dismissing seven other objections filed by the intervenors, the ALJ accepted a single issue for adjudication: the predominance of prime agricultural soils. Several weeks before the hearing, he divided the issue into two separate yet related questions: (1) "In interpreting the Onondaga County Soil Map, what is the percentage of group 1 and 2 soils comprising that portion of the facility within the Agricultural District?" and (2) "Is the percentage of group 1 and 2 soils, as indicated on the Map, within the Map's margin of error (if any), such that a predominance determination based solely upon the Map is not capable of being made in this instance?" The first question directed the parties' attention toward the measurement of area, whereas the second question focused on the reliability of the map.

The OCRRA team was confident the facts were on our side. Using both AutoCad (a CAD/CAM system) and a polar planimeter, our engineering expert had measured the area of the site within the agricultural district (the denominator in the predominance calculation) and the area within that of group 2 soils (the numerator)—there were no group 1 soils on the site. In digitizing soil boundary lines for the AutoCad calculation, he had followed soil boundaries "conservatively" by striving consciously not to underestimate the extent of group 2 soils. He plotted his digitized soil boundaries, compared them with the original lines on the soils map, and made appropriate adjustments to ensure an upper-bound estimate, which he calculated as 43.45%. And as a check, he ran three trials with a polar planimeter, which returned results between 43.24% and 43.38%. Our soils expert also estimated the percentage: using a grid sampling method described in the National Soil Survey Manual, he run nine trials, with results between 38.3% and 40.1%. (Later in the hearing, during DEC's presentation, a staff scientist who had applied AutoCad using two different methods and a combined total of five trials, reported results between 43.45% and 44.45%.)

Recruited only a week before the hearings began, I had reviewed the engineer's and soil scientist's reports, and concluded that their methods were suitable for a result safely below the 50-percent threshold for predominance. When OCRRA's counsel asked me under oath for my opinion on the second question, I replied: "The evidence that we have here and the testimony concerned with a variety of different cartographic measurements is clearly adequate to reach a determination that Group 2 soils are substantially less than half the area of the site within the agricultural district. There is absolutely no doubt whatever in my mind on that point." And when asked to address the first question, I opined with equal conviction that "the calculations were performed correctly," and that "the percentage of [prime] soils is less than half," with "a strong likelihood that it is somewhere around 43 and possibly, or quite likely, less than 43 [percent]."

My cross-examination was not as painful as that of OCRRA's engineering expert. The Town's attorney challenged his choice and use of the planimeter (a mechanical device), his experience with AutoCad, his ability to apply consistently his conservative digitizing strategy, and his use of all the decimal digits reported by AutoCad as well as the accuracy of the property survey, the propriety of using the center line of the "described right of way" instead of the center line of the road or the edge of the road as the outer boundary of the site, and his techniques for joining four map sheets from the soils survey into a single map and for relating the resulting soils map to the property survey.

PURE's representative, an instructor in engineering technology at the local community college, was more inventive—and frustratingly slow—in repeatedly suggesting that the polygons of group 2 soils should be digitized along the outer edge of the cartographic boundary line, rather than along the center of printed boundary. Disclosures shortly after the first week of testimony indicated that this calculation had rewarded PURE with a tenuous estimate of 50.30%—hardly a comfortable margin, though, when compared to its companion inner-edge estimate of 42.84%. Our engineer looked forward to feeding OCRRA's attorney embarrassing questions about the implications of these numbers as well as the reliability of PURE's jagged, decidedly biased digitizing—the reason, perhaps, why PURE decided not to introduce the report in its direct testimony.

Cross-Examination

I held up nicely during my cross-examination, which began on both days 1 and 3 in mid-afternoon, when my biological system demands, and normally receives, a caffeinated beverage. On day 3, though, I was taken aback when the Town's attorney asked me to read a paragraph from my book How to Lie with Maps. Having no choice, I read: "That crisp, definitive lines on soils maps mark inherently fuzzy boundaries is unfortunate. More appalling, though, is the uncritical use in computerized geographic information systems of soil boundaries plotted on 'unrectified' aerial photos subject to the relief-displacement error described in figure 3.6. Like quoting a public figure out of context, extracting soils data from a photomap invites misinterpretation. When placed in a database with more precise information, these data readily acquire a false aura of accuracy." (Monmonier, 1991, p. 38). The Town attorney's strategy was obvious: AutoCad could be used as a GIS, and OCRRA's engineering expert had, indeed, overlaid an electronic version of the property survey map for the site onto the digitized soils boundaries. When I tried to explain, I was interrupted, thanked for my contribution, and dismissed. Fortunately, OCRRA's attorney provided the opportunity, on re-direct examination, to point out that the passage referred not to the comparatively nonproblematic registration of the site's external boundary onto a single overlay, but to the overlay of a detailed soils map, say, onto a detailed vegetation map compiled from different sources and likely to yield spurious 'sliver' polygons where lines that should correspond don't.

(Never heard of *How to Lie with Maps?* The book with the tantalizing title borrowed from Darnel Huff's classic *How to Lie with Statistics?* Don't blame the University of Chicago Press, which sent separate review copies, albeit without result, to JAMA, the American Statistician, and this newsletter-the current editors of which will, I trust, pardon this shameless plug for the expanded second edition, due out this spring with a few enhancements, including full-color illustrations for an enlarged color chapter. The first edition, by the way, has been translated into French and Japanese.)

As alluded to in the aforequoted paragraph from *How to Lie with Maps*, another source of uncertainty broached by the Town's attorney was the geometric distortion found on aerial photos of hilly areas like central New York. Simply put, because an areal photograph is a perspective view, horizontal positions are displaced along lines radiating from the center point. High features are displaced outward, relative to low features, which are displace inward, and the resulting "relief displacement" is generally greater near the edges of the photo than near the center. This displacement could be removed by a differential rectification, a process that yields geometrically accurate orthophotos and orthophotoquad maps, which are now available for parts of New York State. But the Onondaga County soil survey, published in 1977, was started in 1945, and the SCS plotted soil boundaries on a mosaic of unrectified air photos. While a potentially severe source of error for distance measurements, relief displacement would seem to have little discernible effect on the percentage calculation. Besides, because the steeper, more easily eroded land is less agriculturally productive, area measurements on air photos of any type would tend to overestimate the relative surface area of group 2 soils.

Inclusions

The intervenors, who also attacked the published soil boundaries, were the first to use the I-word: inclusion. As environmental scientists and authors of environmental regulations are well aware, soils maps, like all other maps, are not exact representations of reality but generalizations, filtered and smoothed for legible presentation at a standardized map scale. One kind of generalization common (yet not unique) to soils maps is the suppression of small patches of another kind of soil within the boundary of a larger, easily mappable zone. A mapped patch labeled soil X thus might include minute amounts of soils Y and Z. Not easily shown on the map—or identified in their entirety in the field—these inclusions are an inconvenient reality that soil scientists deal with in three ways: by suppressing known inclusions smaller than the head of pencil on the finished map, by observing an SCS guideline that "limiting" and "non-limiting" inclusions ought not exceed 15 and 25 percent, respectively, of the area of a mapped soils unit, and by listing in the detailed description of each soil category the estimated overall percentage of inclusions and the names of included soils recognized in the field. More precise mapping is possible, to be sure, but the map sheets would be enormous, the survey could consume several lifetimes, and the field work might require more digging than landowners would tolerate.

Margin of Error

Throughout the hearing the Town and PURE brought up, again and again, the phrase "margin of error." Each intervenor presented a retired SCS soil scientist as an expert witness, and both experts testified that the maps were not sufficiently accurate to calculate the area of prime soils to within 7 percent. One testified that the map's margin of error was "15 to 25 percent," whereas the other considered the margin narrower, "in the neighborhood of 10 percent." But neither witness addressed error as a task-specific concept.

Another term treated with equal obscurity was "accuracy"—the intervenors, and at times the ALJ, seemed unwilling to acknowledge that "accuracy" is meaningful only in reference to a specific measurement or problem. It was frustrating to hear adults fail to differentiate the internal homogeneity of soil mapping units from the positional accuracy of the delineated boundaries, or to distinguish the accuracy of a calculated percentage from the accuracy of determining preponderance. Judicial proceedings don't allow someone not testifying to inject a comment or raise a hand. So you roll your eyeballs, wring your hands, and think of nasty things to say about lawyers in a GIS column.

The Outcome

OCRRA is confident the state ultimately will award the landfill permit, and that the courts will reject any subsequent legal challenge. After all, numerous complementary measurements indicate without exception that group 2 soils are not even close to predominant. And no evidence suggests that the Onondaga County soil survey is significantly less accurate than similar maps for other New York counties. Moreover, had the soil survey been demonstrably flawed or some plausibly appropriate measurements fallen on the other side of the 50-percent threshold, OCRRA's case would still be solid. The agricultural provision is based directly on the delineated land classification—making the existing soils map an "official map," which need not be a scientifically accurate map, and the standard of proof in civil proceedings is preponderance of evidence, not reasonable doubt. Even so, the Town and PURE were able to thwart the timely issuance of a permit, and with persistent legal maneuvering the intervenors might delay the process further, until OCRRA's options-to-purchase expire. Should the matter reach state Supreme Court and even the Court of Appeals, a legal precedent useful in similar adjudications might emerge. If soil scientists, statisticians, and GIS experts fail to develop well-defined, widely accepted procedures for addressing simple questions involving uncertain data, lawyers, judges, and jurors will develop the standards for them.

References

Huff, D. 1954. *How to Lie with Statistics.* W.W. Norton. New York.

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

Mark Monmonier
Syracuse University
`mon2ier@syr.edu`

TOPICS IN SCIENTIFIC VISUALIZATION

Converting Tables To Plots: A Challenge From Iowa State

by Daniel B. Carr and Sarah M. Nusser

1. Introduction

In October 1995 I gave a seminar for the Iowa State Statistics Department. My topic was converting tables into plots. Sarah Nusser, noting that Iowa is awfully close to Missouri (the show me state), suggested that I try my hand at converting one of her summary tables for water and wind erosion. I agreed, provided that she would help me write this article. This article shows my first attempt. Most of the water erosion information appears in two plots. Figure 1 shows the water erosion land classes and erosion rates while Figure 2 shows the corresponding acreage. For brevity, this article omits the similarly designed wind erosion plots. The plots are not trivial but with modest explanation many will be able to see the patterns. I conjecture that (1) more readers will be willing to study the plots than the corresponding tables and that (2) most readers will be more confident about not missing basic patterns if they use the plots rather than the tables.

In what follows Sarah provides a description of the table and indicates some of the guidance she gave me. Then I describe my efforts to develop the plots. I generally follow the guiding principles of writers such as Cleveland (1985), Tufte (1990) and Kosslyn (1994). However the guidance does not capture all the interactions and special cases that arise in increasingly complex table conversions. (In fact Kosslyn suggests avoiding complex graphics.) While some design decisions come easily, others are struggles. Hopefully readers will pick up a useful idea or two while looking over my shoulder. Quite possibly ideas for better designs will emerge.

2. The National Resources Inventory Land Class and Erosion Table

The National Resources Inventory (NRI) is a longitudinal survey conducted by the USDA Natural Resources Conservation Service (NRCS) in cooperation with the Iowa State University Statistical Laboratory. The purpose of the survey is to assess the status and trends at 5-year intervals for natural resources on nonfederal lands of the United States. The agriculture variables of interest include land use patterns, soil types and properties, wind and water erosion, conservation practices, rangeland quality, and conversion of farmlands to non-farm uses. Environmental concerns include wetlands, earth cover, and habitat measures.

The table in question displays changes over time in soil loss due to water and wind erosion as influenced in relation to land use dynamics. The data used to create the table contains a set of points with a time series of 1982, 1987 and 1992 data on land use, long-term average annual rates of soil erosion induced by water and wind, and NRCS erodibility indices. The six land use categories are cultivated, highly erodible cropland; cultivated, not highly erodible cropland; noncultivated, highly erodible cropland; noncultivated, not highly erodible cropland; Conservation Reserve Program (CRP) cropland; and other land use.

The typical change table considers two years at a time. The earlier year's land use categories define rows. The later year's categories define columns. The table has a two way layout with six entries per cell. The cell entries are the number of points belonging to the cell, the estimated number of acres that follow the pattern defined by the grid cell, and the estimated average water erosion rates and wind erosion rates for the two years. Considering t time points involves examining up to t choose 2 tables. It doesn't take long to become saturated even for those with a high table-reading IQ.

The challenge was to take the information contained in the 3 change tables (82-87, 87-92 and 82-92) and display it on one plot. Ideally it is desirable to look at the grid defined by initial condition (rows of land use categories for the earliest year) crossed with the most recent conditions (columns of land use categories for the most recent year) with some representation of the erosion rates for the different land use classes that were present during the interim years. The grid would be five rows by six columns because the Conservation Reserve Program was not in place in 1982. Since this was a pretty tall order, we decided that a good starting place would be to consider displaying information for only one type of erosion (wind or water) and the acreage associated with the particular land use pattern over time. The acreage is an important measure of the significance of the land use pattern in the dynamics of soil loss.

3. Design Thoughts for Figure 1

As Sarah posed the problem, it seemed natural to keep the 5 x 6 matrix for the beginning (1982) and ending (1992) classes. When focusing on water erosion, the challenge was to represent the land classification and erosion values for the time series within each cell. Since I had little space I immediately thought of using parallel coordinates plots to represent the erosion values.

Inselberg (1985) and Wegman (1990) introduced parallel coordinate plots. Parallel coordinate plots are natural to consider when plotting space is at a premium. Parallel coordinate plots are particularly useful for representing two variables (see Carr and Olsen 1995 for a map legend application) and for representing a small number of time series. In comparison to scatterplot matrices that show all possible pairs of variables, parallel coordinate plots weakly express the relationship between non-adjacent variables. However, for time series the relationships between non-adjacent times are of lesser importance than those between adjacent times, so the compact representation will often suffice.

The next question was how to represent the land class information. My immediate ideas included use of different symbols, line textures and colors. (See Kosslyn 1994 for good symbol and line patterns.) Fortunately there are only six classes. This is crucial in the design considerations because most people can only remember seven plus or minus two chunks of information. With many more classes, creating subclasses and layering the information becomes important. To go for a simple appearance, I chose to use round dots and to avoid angles and end-points associated with other symbols and line textures. Thus puts all the burden on color to represent the six classes. As discussed later this choice returns to haunt me because the symbols are small and small areas make it difficult to see color differences.

With a design in mind, the next step was to try it out. In the S-Plus context I chose my own row-labeled plot functions (Carr 1994a) to handle the layout and labeling but Trellis (see Cleveland 1993 for examples) might have been easier to use. A little effort went into splitting the data set into cells, some effort went into making a rudimentary function to plot the data in a cell, and a great deal of effort went into attending to details.

Developing the plotting function was not too bad since I had previously developed a parallel coordinate function. Attending to details meant providing global scaling so all plots had the same vertical scale, using a gray background with white grid lines to increase the perceptual accuracy of extraction, scaling the data to keep points inside each cell, and plotting axes labels on the right. Noting that color perception is difficult for small areas, I chose thick lines to connect dots over time. The dot-connecting line color used the most recent class assignment. Thus a line from a 1982 value to a 1987 value would use the class assignment for 1982.

Consider the color-driven class interpretation for the parallel coordinate panels show in Figure 1. The color of the row and column labels provide the color legend. By construction, the color of the row label matches the color of the all 1982 dots in that row. The lines to the 1987 time period also have the same color. The color of the column label matches the color of all the 1992 dots in the column. For example the top row second column dots all start in red and the time series all end in green. The 1987 dot color and subsequent line gives the intermediate transition class.

The first experimental plot (that I care to mention) had overplotted dots. Within a matrix cell, overplotting of 1982 dots did not cause an interpretation problem. For hidden dots the class membership and water erosion values were the same as that of the overplotting dots. Similarly the 1992 dot overplotting did not cause an interpretation problem. However the 1987 overplotted dots were problematic. In some cases I could not tell the class membership. Further I could not pair a line entering an overplotted point with the corresponding line leaving the point so two or more time series became confused with each other. To address this my second try changed the line color half way between the time periods. This paired the corresponding lines entering and leaving a dot since they were all the same color. The lines were often sufficiently visible to establish the color of hidden points. Unfortunately the color change between axes made the lines look more complicated and the strategy did not resolve the problem of overplotted lines near zero.

For my third try, I simply resolved the overplotting of six or fewer dots on the 1987 time axis. Carr (1994a) describes a point nudging algorithm that move dots slightly to avoid overplotting. The algorithm computes a small positive increment for a point if its nearest neighbor on the left (or below) is too close. Similarly the algorithm computes a small negative increment for a point if its nearest neighbor on the right (or above) is too close. At each iteration each point changes by the sum of its two increments. The increments cancel when neighbors on both sides are too close. The nudging spreads out the points that are too close. The “too close” criterion can stop moving points when they just touch. This approach almost works.

A point nudging algorithm moves dots slightly to avoid overplotting.

Nudging can change small positive values to negative values. Some people find the shift from positive to negative erosion rates unacceptable. To avoid this interpretation shift, a further step increments small values so that the smallest value becomes zero. This step does not alter larger values that are separated from the lowest values by a sufficiently large gap.

Figure 1 shows the result of applying the modified nudging algorithm to all water erosion rates. While adjusting the 1987 values is sufficient for interpretation purposes, plots often look simpler without overplotting and Figure 2 exploits the lack of overplotting for the 1982 values. For those concerned about the graphical license taken, the most extreme rate adjustment for 1987 (green-cyan-cyan) increased by 1.49 tons per acre per year. The adjustment for 1992 provided more distortion since more values were small. The worst case was the (gray-red-cyan) increase of 4.1 tons per acre per year. This is not acceptable and such discrepancies needed to be flagged or fixed even though the stacked dots suggest the existence of a problem. The obvious solution for this plot is to drop the nudging for 1992. In future years, rescaling, partial dot overplotting and better nudging can help the approximation. Better nudging moves points horizontally and vertically toward a staggered or hexagonal pattern centered on the axis with considerations for line overplotting.

National Water Erosion Patterns In Relation to Land Use

Water Erosion Rates In Tons/Acre/Year (Right Margin)

Rows: Beginning Class, Columns: Ending Class

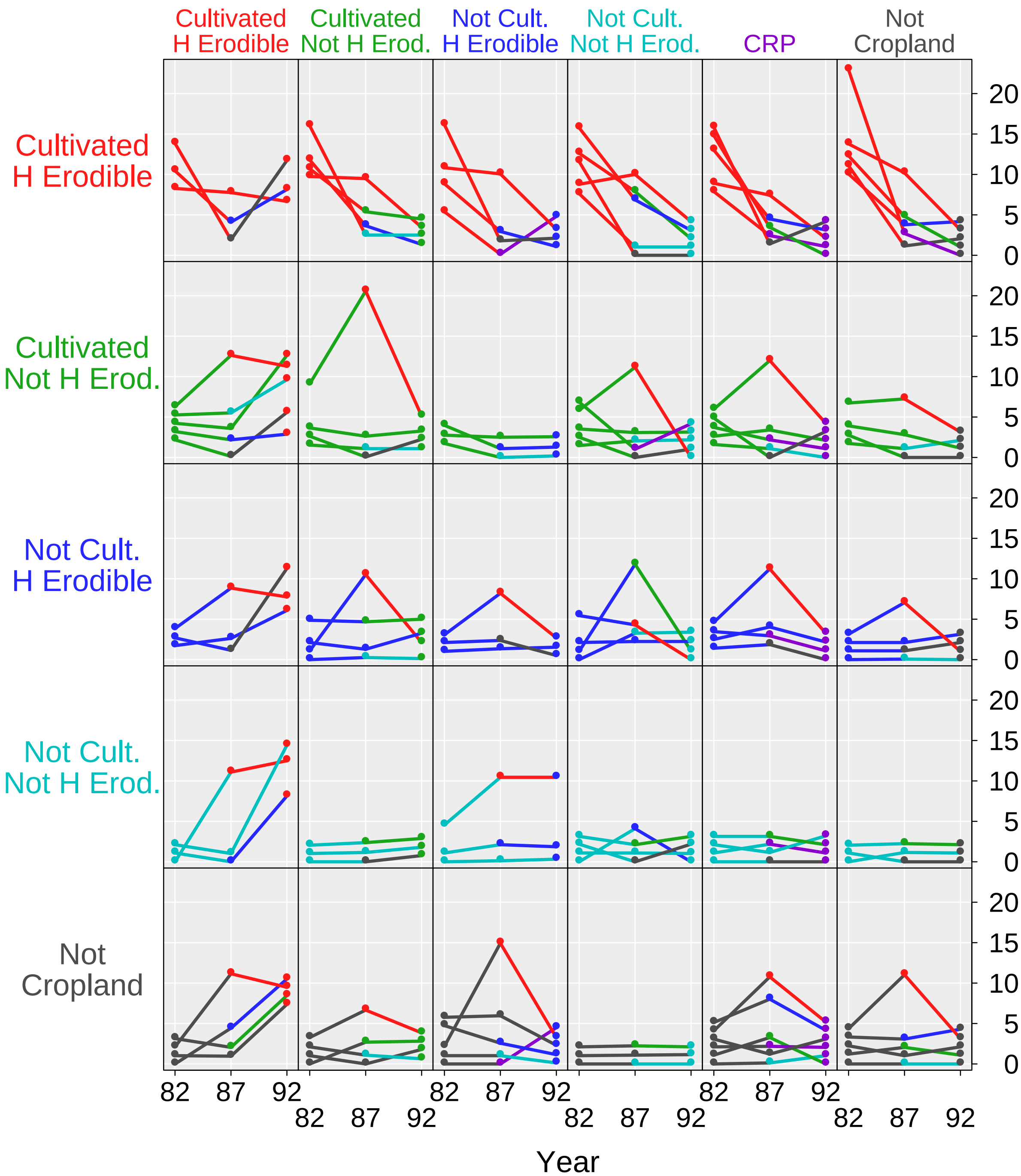

Figure 1: Land Class By Color, Erosion Rate by Parallel Coordinates

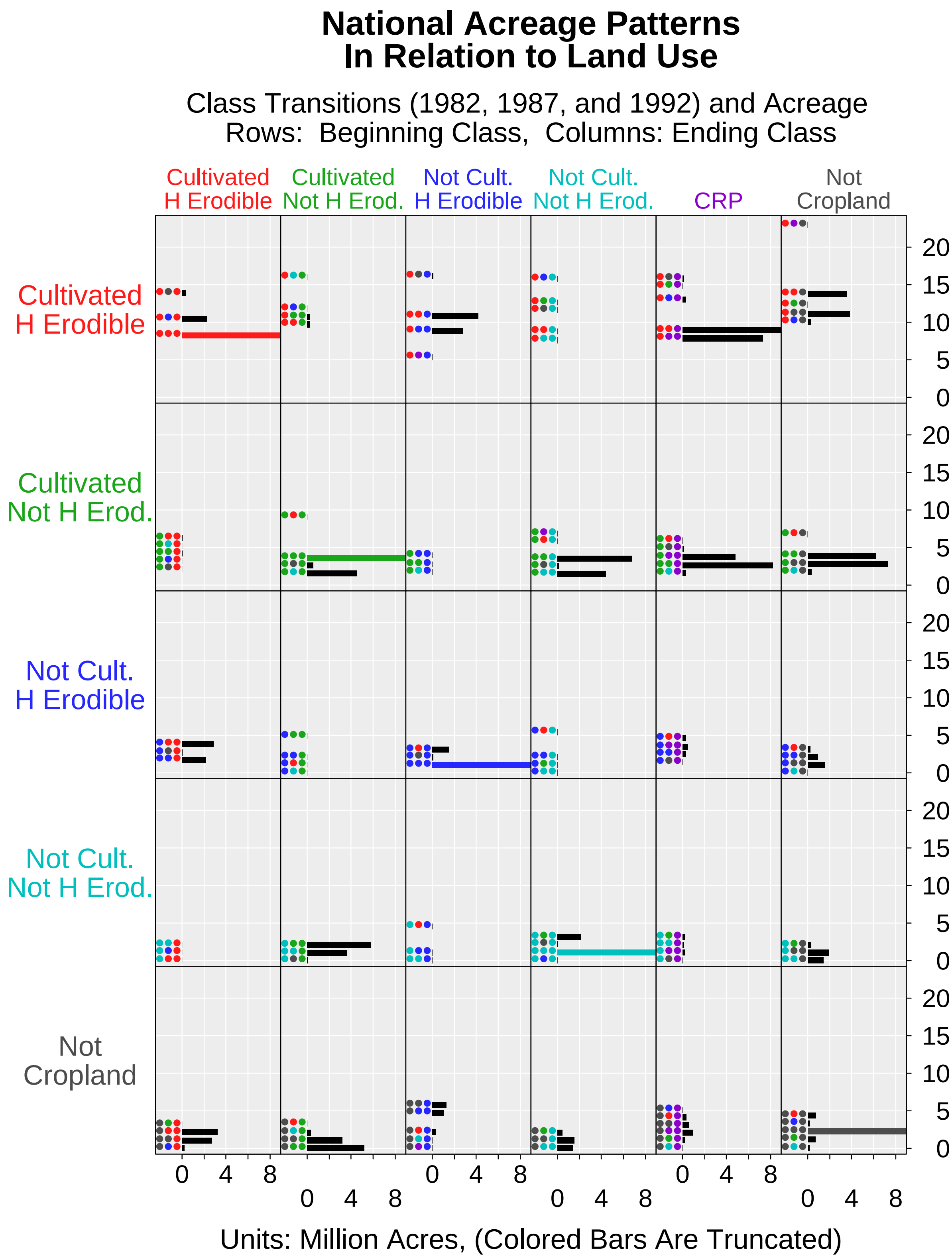

Figure 2: Linked Plot Using 1982 Water Erosion Values

Next steps included matrix labeling, plot titling, sizing of the plot to maximize use of the available space, and staggering the time axis labels to avoid overplotting. As indicated above, Figure 1 avoids the need for a separate classification legend by using color for the row and column labels. This keeps the labels very close to the data, minimizes memory burdens and reminds the reader about the matrix construction. However, plotting labels in color raises the issue of unequal contrast with the background. Outlined fonts with background-contrasting outlines increase the colors that can be used in this way but such fonts were not available. (I tried overplotting a font on top of a font outline but the outline wasn't quite big enough. Multiple drawing of the background text in a tight circular pattern and then overplotting will work.) The available software handled titling, color, resizing and staggering of axis tic labels but had to be modified to support multiple line row labels.

Detailed logical labeling conflicts with space constraints. Selecting a compromise involved some iteration. Putting the label for the water erosion axis in the title is not entirely satisfactory. (Later I may try a vertical label that uses horizontal letters.)

Having the option of color is wonderful but attempting to use color well can be frustrating. Every device I used produced different colors from the same postscript file. Hopefully, the newsletter version will be close to my preferred version. Since the task is to represent different classes, hue differences are appropriate. Some have suggested choosing color from among cyan, magenta, yellow, black, red, green and blue to avoid half-toning on many color printers. I hate to admit it but my hue choices ended up coming from this set. I lightened and darkened some colors in the attempt to make them more distinguishable. The color assignment uses red to call attention to the high risk cultivated highly erodible class. Blue, the darkest color on the light background, calls attention to the other highly erodible class. Gray represents the class of least interest, the not cropland class.

In looking for color guidance. Tufte (1990) says to use natural colors and minimize on highly saturated colors. Kosslyn (1994) warns about the different focal lengths for blue and red, and the fact that 8% of men aren't going to see the red-green distinction. He recommends selecting colors that follow the conventions for the specific problem and audience. Brewer (1994) provides guidance for several one variable and two variable color schemes. Her 2 x 3 layout with two shades of green, blue and magenta is one of several elegant examples. This problem is a 2 x 2 layout plus 2 classes. I would like to combine all of the above advice. Perhaps two shades of brown could convey cultivated land and two shades of green could convey uncultivated land. Guidance is easier to find than to apply.

The current plot problem is tougher than common mapping problems. Brushing aside issues associated with two different backgrounds (white and light gray) there are two basic differences. First, the dots and lines have small area. Color perception is dependent on having adequate area. In this problem subtle color distinctions are not going to work. Second, the plot has crossing lines while map regions do not typically overplot. Mapping techniques, such as putting black lines around regions, can mitigate surround induced variation in color perception. If one uses black outlines to make, for example, yellow lines visible, crossing lines generate new, unwanted patterns. The monochrome Figure 3 illustrates the problem. The line crossings in the corresponding (green row, cyan column) cell in Figure 1 draw much less attention. Since the objective is to follow the lines in a flat plot and not through space, outlining is undesirable.

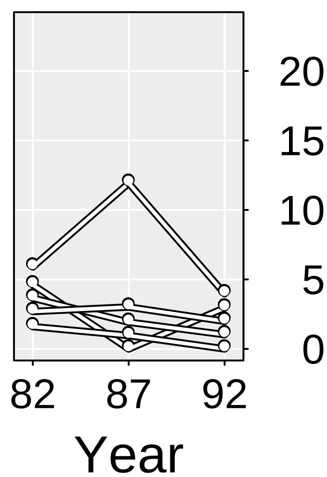

Figure 3. Outlines

Graphical design is often open ended. The thick lines in Figure 1 allow me to readily distinguish among all six colors in my prints. In Figure 2 there are just dots. I can still make red/magenta and green/cyan distinctions but have to pay attention. It may be safer to add small white (or black) dots to the magenta and cyan dots or to use larger areas and different symbol shapes to designate class membership. The design for Figure 2 may not have stopped at the best point and that has implications for Figure 1 due to the linkage.

4. Interpretation and Acreage Plots

Patterns in Figure 1 are easy to spot. High erosion rates typically correspond to the cultivated highly erodible class. High positive slope lines typically end in a red dot. That is, a marked increase in erosion generally goes with a transition to the cultivated highly erodible status. Most of the high erosion rates (above 15 tons/acre/per year) appear in 1982. Tracking the (red, red, red) and (green, green, green) dots over time suggests a slight reduction in erosion rates for land staying in the same class. Trends are flat for the other non-transition lands. Figure 2 aids the interpretation by providing the acreage associated with the land classes.

Figure 2 is basically a horizontal bar plot showing acres. The design is similar to that in Carr (1994b). New variations include the addition of dots to indicate the class transitions, the placement of the bars to match the position of the 1982 values in Figure 1, and the truncation of large acreages for non-transition land. Another choice would be to link erosion rates at the last time period. In either case the dot nudging in Figure 1 provides the space for the horizontal bars in Figure 2.

Scanning Figure 2 for the black transition acreage bars reveals eight substantial bars at the top right. These are transitions to CRP and not cropland classes from the cultivated classes. Somewhat smaller not highly erodible acreages make the transitions between being cultivated and not cultivated. Knowing the acreage helps in assessing the importance of the transitions.

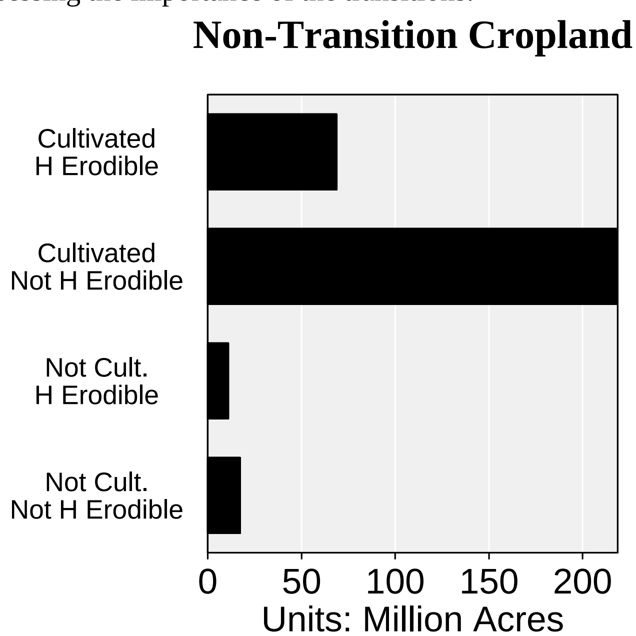

Figure 4. Bar Plot

Figure 4 shows the non-transition cropland acreage truncated in Figure 2. To provide resolution this figure omits the not cropland class with its 1.49 billion acres. The plot interpretation is straight forward.

5. Closing Remarks

Soil erosion affects long-term agricultural production and historically has concerned farm policy makers. More recently, effects of the degradation of soil resources is being understood in the broader context of environmental concerns for water quality. Degradation of soil function in natural and agroecosystems is an emerging policy issue. Improved abilities to understand soil erosion dynamics are an important part of assessing soil function in ecosystems..

Hopefully the design comments above suggest some new possibilities for displaying time series in relation to a matrix of transition states. We can represent rather complicated relationships graphically and have them understood by scientific audiences. As the relationships get more complicated and detailed, graphical design becomes more crucial and the need for graphical approximation increases. Cartographers and geographers have long recognized that communication requires hiding and simplifying information as a function of map scale. They are taking up the challenge of representing data quality (see Van Der Well, Hootsman and Ormeling 1994, MacEachren and Pickle 1995, and Howard and MacEachren 1995). Similarly the statistical graphics community could address graphical approximations as a function of information complexity and work further in such areas such as representing estimate variability and sampling plan adequacy. Figure 1 is not the full story because nothing has been said about estimate quality.

Figure 1 will generalize to at least one more time period. Adding new time axes is straight forward. A potential problem is that new transition combinations could generate many new lines. With two intermediate time periods there are 36 possible classes per cell, Fortunately future NRI data are likely to add only a few new combinations. I have a few more years before having to face the too many combinations challenge.

This article continues the theme of converting tables into plots. Further challenges are welcome, as are gentle constructive suggestions. Those wanting to obtain the data or adapt the software can do so by anonymous `ftp` to `galaxy.gmu.edu`. The directory is `/pub/submissions/erosion`.

Acknowledgments

Graphics research related to this article was supported by EPA under cooperative agreement No. CR8280820-01-0. The article has not been subject 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

Brewer, C. A. 1994. “Color Use Guidelines for Mapping and Visualization,” *Visualization in Modern Cartography*, Eds. MacEachren and Taylor, Pergamon/Elsevier Science Ltd., pp. 123-147.

Carr, D. B. and A. R. Olsen. 1995. "Parallel Coordinate Plots For Representing Distribution Summaries in Map Legends." Proceedings 1 of the 17th International Cartography Association Conference 10th General Assembly of the ICA. pp. 733-742.

Carr, D. B. 1994a. "Converting Plots to Tables," Technical Report No. 101, Center for Computational Statistics, George Mason University, Fairfax, VA. 22030.

Carr, D. B. 1994b. "Using Gray in Plots." *Statistical Computing & Graphics Newsletter*, Vol. 5, No. 2, pp. 11-14.

Cleveland, W. S. 1985. *The Elements of Graphing Data*, Monterey CA: Wadsworth Advanced Books and Software.

Cleveland, W. S. 1993. *Visualizing Data*, Summit NJ: Hobart Press.

Inselberg, A. 1985. The Plane With Parallel Coordinates, *The Visual Computer*, 1, pp. 69-91.

Howard, D and A. M. MacEachren. 1995. "Constructing and Evaluating an Interactive Interface For Visualizing Reliability," Proceedings 1 of the 17th International Cartography Association Conference 10th General Assembly of the ICA. pp. 320-329.

Kosslyn, S. M. 1994. *Elements of Graphic Design*, New York, NY: W. H. Freeman and Company.

MacEachren, A. M. and L. Pickle. 1995. "Mapping Health Statistics, Representing Data Quality," Proceedings 1 of the 17th International Cartography Association Conference 10th General Assembly of the ICA. pp. 311-319.

Van Der Well, F. J. M., R. M. Hootsman, and F. Ormeling, 1994. "Visualization of Data Quality," *Visualization in Modern Cartography*, Eds. MacEachren and Taylor, Pergamon/Elsevier Science Ltd., pp. 313-331.

Tufte, E. R. 1990. *Envisioning Information*, Cheshire, CT: Graphics Press.

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

UNIX COMPUTING

UNIX Commands

by Phil Spector

One of the most attractive features of UNIX is its extensibility. Even a novice UNIX user can add commands to the operating system, and those commands will be immediately accessible, and can be accessed in the same way system commands are accessed, namely by typing the name of the command. This is because when the UNIX shell encounters a command name, it searches through a set of directories known as the search path until it encounters a file with the name of the command it encountered. If that file has execute permissions appropriately set, then it is executed. (Recall that to set the execute permission for a file, the appropriate UNIX command is `chmod +x filename`).

Two types of files can serve as commands on UNIX systems. First, a text file which has appropriate execute permissions is interpreted as a collection of commands. By default, these commands are executed by the Bourne shell, `/bin/sh`. A curious mechanism is used to inform the operating system of an alternative interpreter for commands within an executable file; if the first line of the file is of the form

```
#! program_name
```

then `program_name` will be used to interpret the commands instead of the Bourne shell. Since that line would have to be ignored under other circumstances, this technique will work for any program which uses the pound sign (#) as a comment delimiter. Such programs include all the UNIX shells (`/bin/sh`, `/bin/csh`, `/bin/ksh`, etc.) as well as the perl (usually `/usr/local/bin/perl`) text processing language.

The second type of file which is executable is a binary file created as the result of compiling a program in some low-level language like C or Fortran. The operating system recognizes such files by the so-called "magic number" which is a bit pattern inserted at the beginning of the file when it is first created, and which is different for each different type of file which is supported on the system. (See the man page for the `file` command or the `magic` file format for more information about magic numbers.) Since the compiler or loader deals with this detail, you will rarely have to worry about it. Executable text and binary files can be freely mixed in directories without causing any confusion.

If you write a command of your own, and you want it to automatically execute when you type its name, it must be found in one of the directories included in your search path. Usually your search path will be set in the start-up file for your shell (either `.cshrc` for the C shell or `.profile` for the Bourne or Korn shells). Under the Bourne and Korn shells, the search path is stored in the shell variable `PATH`, while in the C shell it is stored in the shell variable `path`. You can see your search path by typing the UNIX command `echo $PATH`. The `bin` subdirectory of your home directory is usually on your search path, and is a good choice for storing commands which you write. Remember that there are some built-in commands which the shell will always find before it goes to the search list; one common example which sometimes causes problems is the `test` command. You would be well advised to not name your command files `test`, since the shell will generally find the built-in command first. Alternatively, you can always refer to a UNIX command by its full pathname; in particular, to execute a file called `commands` in the current directory, you can type

```
./commands
```

at the UNIX prompt, and there will be no possibility of confusion as to which file you want to execute. One word of caution if you use the C shell; the C shell doesn't rescan commands in its search path after it is first started unless you type the command `rehash`. So if you are using the C shell, and you add a command to a directory in your search path, you need to type `rehash` before the C shell will recognize it in an already running shell.

The simplest command files consist of a sequence of commands exactly like you would type them in at the command line. When the command is executed, the individual commands will be executed, one after the other. This can be a good way to execute commands which need to be done in a specific order, or to execute multiple runs of another command, because it is inefficient to run several invocations of a program at once, as opposed to running them serially.

Command Line Arguments

Although creating a command file to execute a series of commands can be useful, the real utility of UNIX commands is that you can pass information to them easily through the command line. UNIX commands generally follow a common format for the command line. While some UNIX commands require their arguments in a specified order (like the commands `cp` or `mv`), most UNIX commands offer more flexiblity by observing the convention that options are single letters, preceded by hyphens (`-`), and may or may not be followed by optional arguments. Even common commands, like `ls`, which lists the files in a directory, may have many options, without interfering with their simple default behavior. If you're writing your own UNIX commands, it's easy to access the information on the command line in a way similar to most standard commands. The rest of this article will provide ways of reading the command line arguments for shell scripts, perl programs and C and Fortran programs.

For the sake of example, suppose we have a program which will extract one or more lines from a file. We wish to accept a starting line number, an ending line number, a flag indicating whether to print the line number along with the line, and finally a file name. Thus there are two options with required arguments, which we'll call `-s` and `-f` for start and finish, respectively, and an option which simply serves as a flag, which we'll call `-n` for numbered.

Arguments in Shell Scripts

The examples provided here are for Bourne or Korn shell scripts; similar techniques are available for C shell scripts. Individual arguments to shell scripts are denoted as `$0` (the name of the command), and `$1`,. . .,`$9`, and the entire set of arguments is denoted `$*`. The simplest way to process arguments in a shell script is to loop over the set of arguments and use a `case` statement along with the `shift` statement, which removes arguments from the `$*` variable. The shell variables `start`, `finish`, `num` and `file` will be used to hold the extracted information.

```
#!/bin/sh
  start=1
  finish=1
  num=0
  file=""
for i in $*; do
  case $i in
  -s ) start=$2
       shift
       shift ;;
  -f ) finish=$2
       shift
       shift ;;
  -n ) num=1
       shift ;;
   * ) filename=$1 ;;
  esac
done
if test "$filename" = ""
then
  echo "No filename given. Exiting"
  exit
fi
```

The two semicolons `;;` end each clause of the case statement. For options with required arguments, the value is extracted from the next argument on the list (`$2`), and the argument list is shifted twice. For options serving as flags, only one shift is needed. In the example above, only the filename was checked; it general, it's a good idea to check to make sure all the arguments were presented properly.

Arguments in Perl Programs

There are several perl subroutines which are available to parse command line options. One convenient one is `&Getopts`, found in the file `getopts.pl`, which should be part of your distribution. You pass `&Getopts` a string which contains the letters you wish to recognize as valid options, following those which have required arguments with a colon (`:`). `&Getopts` removes the options from the argument list (which is stored inside perl as the `@ARGV` array), and sets variables of the form `$opt_letter` to appropriate values when the flag represented by `letter` is found on the command line. If a particular option is not given on the command line, the corresponding `$opt_letter` value is undefined. Thus, the previous example could be written as follows in perl:

```
#!/usr/local/bin/perl

require 'getopts.pl';

&Getopts("s:f:n");

$start = $opt_s ? $opt_s : 1;
$finish = $opt_f ? $opt_f : 1;
$num = $opt_n ? 1 : 0;

die "No filename given" if (! $ARGV[0]);
$filename = $ARGV[0];
```

Since `&Getopts` removes arguments from the `@ARGV` array, the filename will be the first element (index 0) of that array after the arguments are processed.

Arguments in C Programs

Command line arguments are passed to C programs through two optional arguments to the `main()` function; `argc` is an integer giving the number of arguments (using the convention that the 0th argument is the command name), and `argv` is an array of character strings containing the arguments themselves. The memory allocated to the strings in `argv` will not be reused, so their addresses can be safely copied into local variables. Here is one example of how to process the command line arguments of the previous example:

```
#include <stdio.h>

main(int argc, char *argv[])
{
  int i,start=1,finish=1,num=0;
  char *filename=NULL;

  i = 1;
  while(i < argc){
    if(argv[i][0] == '-')goto doopt;
    filename = argv[i];
    i++;
    continue;

  doopt:
    if(argv[i][1] == 's')
      sscanf(argv[++i],"%d",&start);
    else if(argv[i][1] == 'f')
      sscanf(argv[++i],"%d",&finish);
    else if(argv[i][1] == 'n')
      num = 1;
    else fprintf(stderr,
         "(%s)Unrecognized option, %s\n",
      argv[0], argv[i]);
    i++;
    }

  if(filename == NULL){
    fprintf(stderr,
              "(%s)No filename given\n");
    exit(1);
    }
    . . .
```

Arguments in Fortran Programs

As a final example, the fortran subroutine `getarg` and function `iargc` will be used to read command line arguments inside a fortran program. The fortran `read` statement is used to convert the character string read by `getarg` to a number where needed.

```
   character*20 filename
   character*20 rdarg
   character brkarg(20)
   equivalence (rdarg,brkarg)
   integer start,finish,num

   start = 1
   finish = 1
   num = 0
   narg = iargc()
   do while (i.le.narg)
     call getarg(i,rdarg)
     if(brkarg(1) .ne. "-")goto 5
     if(brkarg(2) .eq. "s")then
```

```
      call getarg(i + 1,rdarg)
      read(rdarg,*)start
      i = i + 2
      goto 10
      end if
   if(brkarg(2) .eq. "f")then
      call getarg(i + 1,rdarg)
      read(rdarg,*)finish
      i = i + 2
      goto 10
      end if
    if(brkarg(2) .eq. "n")then
      num = 1
      i = i + 1
      goto 10
      end if
 5     filename = rdarg
       i = i + 1
10     end do
       . . .
```

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

NET SNOOPING

Java—Is it hot or not?

by Mike Meyer

The computing trade rags and internet news groups are all agog again over HotJava, which can be viewed as another way of distributing information over networks, or just another extension to Netscape. In this article I give a short introduction to what HotJava is and I muse about what influence it might have on the way we do statistics. For starters, Java and HotJava are respectively languages and browsers developed by Sun Microsystems. In some ways one can view them as logical extensions of the current WWW/HTML protocols. There are other similar extensions and proposals, but Java has at least some momentum behind it.

What is HotJava?

The best way to understand what HotJava is, is to try to use it. If you are using a Sun workstation, Windows NT, or Windows 95, then there is a native HotJava application available from `http://java.sun.com`. (Of course the easy way to find these sort of things is to use one of the internet search engines. Lycos `http://www.lycos.com/` is still my favorite.) That same server also has information about writing Java applets, but more on that in a moment. The rest of us (and maybe even some of those who could use the Sun alpha-release HotJava viewer) will have to live with a version of Netscape or some other WWW viewer that also support Java support. The current beta release of Netscape for HP-UX, IRIX, Sun Solaris, and SunOS supports viewing Java applets.

But what are Java and HotJava?

One way of answering that question is to quote from the Sun advertising/technical materials.

> Java(tm) is a simple, object-oriented, distributed, interpreted, robust, secure, architecture-neutral, portable, high-performance, multithreaded, dynamic, buzzword-compliant, general-purpose programming language. Java supports programming for the Internet in the form of platform-independent Java applets.
>
> HotJava(tm) is a modular, applet-aware, extensible World-Wide Web browser written entirely in the Java programming language. The current version of the browser is Java release 1.0alpha3. The Java Developer's Kit includes an applet viewer for previewing Java applets, but not a browser.

If you are really lucky, that description answered everything for you. Otherwise, it probably still requires some explanation.

The current crop of WWW viewers allows you to view material served from remote machines, and even run programs on remote machines. It does not in general allow you to run something on your machine. So, assume, that I wanted to distribute an interactive crossword puzzle where you could fill in the crossword and get hints on the answers. Using the current WWW model someone would have to develop a program that ran on their server so that every time you entered a letter into the crossword (or maybe every time you hit the submit button), the data would be transmitted to the server, the server would look at it, and then give feedback. The idea behind Java and applets is to down-load the crossword puzzle *program* to your computer, so that all the hints and interaction are handled locally. They are called applet because the idea is that the programs are small, and self contained. The crossword example is a good one, because there is a really nice crossword puzzle applet available from `http://home.mcom.com/comprod/products/`

navigator/version_2.0/java_applets/Crossword/index.html
and an interesting Solitaire game at
http://w3.gwis.com/~thorn/BetaSol.html.

Thus, in the crudest terms, Java is an interpreted programming language, and HotJava and Netscape are capable in running Java programs. This doesn't sound like a particularly new idea. Interpreted languages are very familiar to us—for example almost all statistical packages. Java is a derivative of C++ with lots of powerful graphics extensions.

So, what can I do with it?

Using Java, it is quite easy to include animations within a WWW document. Using existing WWW technology an animation is only possible by down-loading a static "movie" to an external viewer. With Java one could program the animation right in the WWW viewer. So, for example, the Java icon is a steaming cup of coffee that sits there steaming away on many of the Java pages. There are several examples of clocks that run in your WWW browser. The code to describe what the clock looks like is a Java applet which itself queries the clock on your computer for the real time. There are tic-tac-toe games and moving backgrounds, and all sort of things that turn the WWW into an advertisers dream land. It is perfectly reasonable to expect that someone (with more time than I have!) will program at least a rudimentary statistics package in Java. I expect that we will see a growing range of instructional material, like course notes, and interactive books, that embed both static text material and dynamic calculations in the same document. So, instead of telling a student to go and run a simulation in program X, one might say "click here" to run a simulation of a random process.

Is it secure?

My first reaction on hearing about Java was to be concerned about security. Arbitrarily down-loading someone else's program (or virus, or whatever) must give every computer user the shudders. The authors of Java have a lot of experience with similar issues (the Andrew Toolkit and display postscript are two examples that come to mind), but as far as I am concerned the jury on Java security has not yet even convened.

Who cares?

The big interest in applications like Java is amongst the people who are trying to sell you things. Imagine a statistical package where every time you ran the regression command you were charged a few pennies. Java will get you close to that. Imagine a statistics package where anyone could add their own modules and charge you to get a copy. So, Fred could write a regression module and Wilma could too. You would have the choice of who you pay for your regression work. Java gets you close. Of course there are lots of non-commercial uses, but I strongly suspect that we will see something like Java as one of the forces behind an additional commercialization of the internet and of software and other intellectual property.

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

BITS FROM THE PITS

Dynamic Linear Models for Tank Level Monitoring

by Albert M. Liebetrau

This column features statistical computing and statistical graphics activities in science and industry. I invite your comments and suggestions for future columns. Please send comments, inquiries, and suggestions to Albert M. Liebetrau, Analytic Sciences and Engineering Department, Battelle-Northwest, MS K5-12, P.O. Box 999, Richland, WA 99352, 509-375-2694, AM_Liebetrau@pnl.gov

Introduction

The Hanford site in south central Washington state is home to 177 underground storage tanks. These tanks contain millions of gallons of wastes that were generated in the production of plutonium at the site following World War II and during the Cold War. These tanks were designed only for temporary storage and a program to permanently dispose of these wastes is currently underway. In the meantime, however, there are concerns about the integrity and safety of these tanks. Some tanks, for example, are known to have leaked; others are of concern because hydrogen generated by their contents can build up to dangerous levels.

In-tank instrumentation to monitor liquid level is a primary source of information about the tank. Changes in liquid level are routinely monitored to determine leaks (a decrease) or hydrogen buildup (an increase). These methods generate large quantities of data that in turn require the development of efficient and timely methods of analysis. In this article, our use of multi-state dynamic linear models (DLMs) to monitor tank liquid levels for leaks and intrusions is described. The specific purpose of this work is to identify tanks that are leaking.

Data Features

A multi-state DLM incorporates a model of the "steady state", along with models of various alternative states (features) that are of interest for some reason. The basic premise is that changes in the underlying system can be identified by means of their effects on the liquid level monitoring data. For each new observation, the strategy is to determine whether a state change has occurred and, if so, what is the likely (i.e., most probable) new state of the system. Implementation of this strategy requires a catalog of features that can be associated with changes of interest in the process. Subsets of actual tank level data that show features of interest are shown in Figure 1. The features shown are:

Steady State (SS) The steady state typically refers to behavior that is constant over time. For this application, the steady state is taken to be a change that is linear with time. Thus, the constant in this case is the rate of decrease in liquid level. Note that a rate of zero is allowed, in which case the liquid level is constant. The decrease in level over time is attributed to evaporation. The plot in the upper left quadrant of Figure 1 labeled "steady" shows typical steady state behavior for the tanks in this study.

Level Change (LC) This term is used to refer to an abrupt jump in a series of tank level measurements. Events that have produced level changes in Hanford waste tanks include

- additions of water. Water is routinely added to some tanks to control the temperature of the contents.
- a slippage in reference tape for a measurement instrument. The level measurements are referenced to an arbitrary point by means of a tape. The tape may slip during measurement or it may not be identically placed for successive measurements. Both events produce an apparent change in liquid level.
- changes in instrumentation. Several different types of instruments are used to measure liquid levels. Moreover, instruments are periodically replaced, often with instruments of a different type. These changes also produce apparent changes in liquid level.

The plot in the upper right quadrant of Figure 1 shows data for a tank that has undergone a level change.

Slope Change (SC) A change in slope indicates that the rate of level change is changing. Events that have produced slope changes include leaks and changes in the rate of evaporation, the latter are primarily due to change in the tank's venting system.

Transient (T) A transient, or outlier, refers to a single measurement that has a value far from other measurements taken at nearly the same time. Transcription errors are one source of transients. The plot in the lower right quadrant of Figure 1 shows data record that contains a transient.

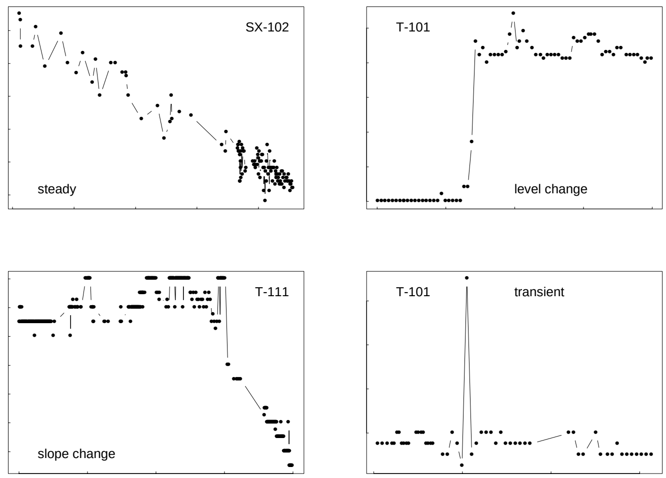

Figure 1. Tank level data that exhibit features (states) of interest. The four features shown here are "steady state", "level change", "slope change", and "transient".

Theoretical analogues of the features shown in Figure 1 are shown in Figure 2 (see Gordon and Smith (1988)). The strategy is to model the features (states) of interest and then use the data to discriminate among these states.

Methodology

The linear growth model that represents the steady state in this application can be written as follows:

$$\begin{aligned} \mu_t &= \mu_{t-1} + \beta_t + \delta_{\mu t} \\ \beta_t &= \beta_{t-1} + \delta_{\beta t} \end{aligned} \qquad (1)$$

The first of these equations describes a time varying linear trend. The μ terms indicate the level at a given time and the βs correspond to the rate of change. The $\delta_{\mu t}$ and $\delta_{\beta t}$ terms represent "noise" in the μs and βs, respectively. These terms are parameterized by their variances σ_μ^2 and σ_β^2. When both variances are set to zero, the model describes an exact linear trend.

The actual observations are modeled as

$$y_t = \mu_t + \epsilon_t \qquad (2)$$

In Equation (2), y_t represents the recorded measurement and ϵ_t represents noise in this measurement. The noise is parameterized by its variance σ_ϵ^2. Instrument resolution is a major contribution to measurement noise.

The four states described above can all be obtained from the model described by Equations (1) and (2) by making suitable choices of the model parameters σ_μ^2, σ_β^2, and σ_ϵ^2. The correspondence is as follows:

$$\text{Steady State}: \quad \sigma_\mu^2 = 0,\ \sigma_\beta^2 = 0,\ \text{and}\ \sigma_\epsilon^2 = \text{Nominal}$$

$$\text{Level Change}: \quad \sigma_\mu^2 > 0,\ \sigma_\beta^2 = 0,\ \text{and}\ \sigma_\epsilon^2 = \text{Nominal}$$

$$\text{Slope Change}: \quad \sigma_\mu^2 = 0,\ \sigma_\beta^2 > 0,\ \text{and}\ \sigma_\epsilon^2 = \text{Nominal}$$

$$\text{Transient}: \quad \sigma_\mu^2 = 0,\ \sigma_\beta^2 = 0,\ \text{and}\ \sigma_\epsilon^2 = \text{Large}$$

Additional details and discussion can be found in Gordon and Smith (1988) and West and Harrison (1989).

As waste level observations are obtained, the state definitions given above are used in a modified Kalman filter to update the description of the waste level. The description consists of (a) the probability of being in each of the four states and (b) current "best" estimates of the parameters μ_t and β_t for each of the four models.

Updating equations are derived and presented in Gordon and Smith (1988) and West and Harrison (1989).

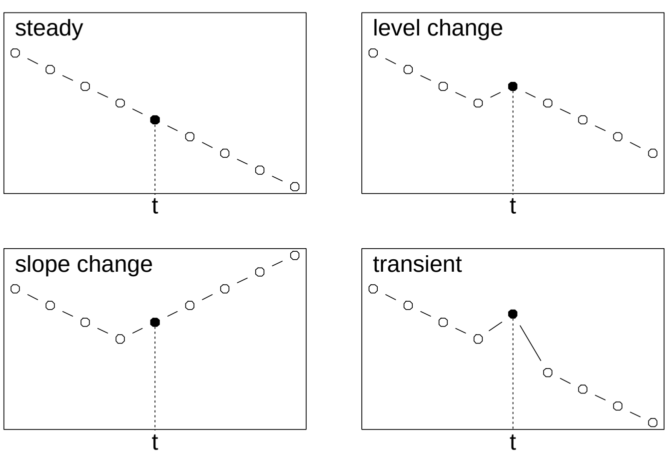

Figure 2. States for a linear growth model. These states are idealized analogues of corresponding states shown in Figure 1.

The sets of probabilities for the four states of the model are used to evaluate the state of the system after each new observation. An observation is "flagged" whenever the probability that the system is in steady state falls below some threshold. The observation y_t might be flagged, for example, whenever

$$\text{Prob}\,\{SS \text{ at time } t\} < 0.8$$

To see how this diagnostic works, refer to Figure 2 and consider the decision to be made after observing the value y_t at time t. In this case, y_t does not fit the steady state model. However, at time t, it isn't possible to identify the new state of the system. The next observation y_{t+1} is needed to make this call. Accordingly, the probability that the system is in each of the four states must be updated in light of the new observation y_{t+1}. Formula for computing these one-step-back probabilities are given in Gordon and Smith (1988) and West and Harrison (1989). A graphical summary of the one-step-back probabilities is shown in Figure 3 for a typical series of liquid level observations.

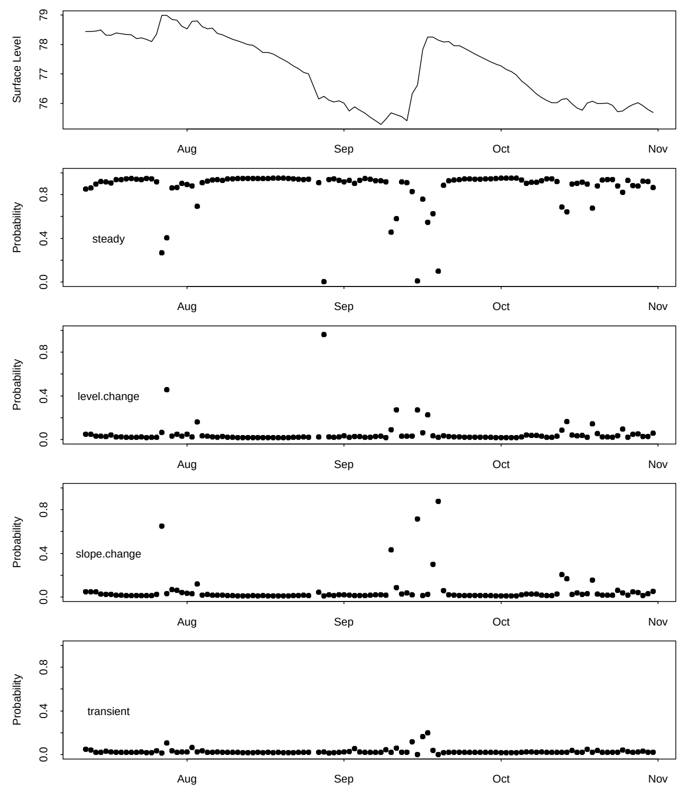

Figure 3. A one-year sequence of surface level measurements and state transition probabilities for a typical tank.

The inputs required to drive the model are (a) the variances σ_μ^2, σ_β^2, and σ_ϵ^2 for each of the states of the system at all times t, (b) initial values of μ and β for each of the states of the model, and (c) probabilities for each of the states of the model. These probabilities should be positive and sum to one. The following probabilities were used in the application described below:
Prob {Steady State} = 0.85
Prob {Level change} = 0.05
Prob {Slope Change} = 0.05
Prob {Transient} = 0.05

Applications

The DLM described here was used to evaluate the surface level data for each of 149 single shell tanks and 22 double shell tanks at the Hanford site. For a number of tanks, the analysis confirmed the determination made with other methods that it is a leaker. This modeling approach has been so effective that it is now accepted as an important tool for monitoring the status of waste tanks at the Hanford site.

The identification of level and slope changes by the DLM analysis also turned out to be an important preprocessing step in analysis of tank level data to identify trapped gas. Instrument maintenance records and, more importantly, historical documentation of transfers involving tank contents are not readily accessible in electronic form. Moreover, it would be very costly and time consuming to convert all of this information into a suitable electronic form. Hence, the breaks identified by the DLM analysis were used as proxies for that information in the trapped gas analysis.

The trapped gas analysis itself involved looking for an inverse relationship between waste level and ambient atmospheric pressure. The underlying premise is that when trapped gas is present in a tank, the liquid level in the tank will decrease with an increase in atmospheric pressure. A complicating factor not mentioned above is that in the full study, it was necessary to accommodate data obtained from four different tank level measurement systems. Details of the trapped gas analysis are presented in Whitney 1995.

Acknowledgments

I thank my colleagues Paul Whitney, Steve Sain, and Terri Miley for allowing me to share this description of their work.

References

Gordon, K. and A.F.M. Smith. 1988. “Modeling and Monitoring Discontinuous Changes in Time Series.” In *Bayesian Analysis of Time Series and Dynamic Models*, JC Spall, editor. Marcel Dekker, New York.

West, M. and J. Harrison. 1989. *Bayesian Forecasting and Dynamic Models.* Springer-Verlag, New York.

Whitney, P.D. 1995. *Screening the Hanford Tanks for Trapped Gas*, PNL-10821. Pacific Northwest Laboratory, Richland, Washington.

Albert M. Liebetrau
Staff Scientist
Battelle Pacific Northwest National Labs
AM_Liebetrau@pnl.gov

MIXER NOTES

Mixer at the Orlando Meetings

by Mario Peruggia

The Combined Business Meeting and Mixer of the Sections on Statistical Graphics & Statistical Computing was held Monday evening, August 14. Judging by the number of people who attended and the vitality of their participation, our two sections are in very good health. Throughout the evening, members were given ample opportunity to catch up with old friends and meet new ones. The chairpersons of the sections, Mary Ellen Bock (Computing) and David W. Scott (Graphics), briefed the participants on the state of the sections and on the executive meetings held earlier in the day. Their reports were followed by the raffle drawings with door prizes contributed by the following exhibitors:

- Academic Press
- Edward Arnold Publishers
- ASA Souvenirs
- BMDP Statistical Software, Inc.
- Bureau of Transportation Statistics
- Conceptual Software, Inc
- Current Index to Statistics
- Data Description, Inc
- John Wiley & Sons Inc, Publishers
- NCSS Statistical Software
- Oxford University Press
- Palisade Corporation
- Prentice Hall Publishers
- Research Triangle Institute
- Sage Publications, Inc
- SAS Institute Inc
- SIAM
- SPSS Inc
- STATSCI, a division of MATHSOFT
- W. H. Freeman Publishers

SECTION OFFICERS

Statistical Graphics Section - 1995

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

William DuMouchel, Chair–Elect
212-305-7736
Columbia University
dumouch@bayes.cpmc.columbia.edu

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

Sally C. Morton, Program Chair (appointed to replace Sallie Keller-McNulty)
310-393-0411 ext 7360
The Rand Corporation
Sally_Morton@rand.org

Stephen G. Eick, Program Chair–Elect
708-713-5169
AT&T Research Laboratories
eick@research.att.com

Michael M. Meyer, Newsletter Editor (93-95)
412-268-3108
Carnegie Mellon University
mikem@stat.cmu.edu

Robert L. Newcomb, Secretary/Treasurer (95-96)
714-824-5366
University of California, Irvine
rnewcomb@uci.edu

Deborah J. Donnell, Publications Officer (94-96)
206-283-8802 ext 258
MathSoft, Seattle, WA

Colin R. Goodall, Rep.(95-97) to Council of Sections
814-865-3993
The Pennsylvania State University
colin@stat.psu.edu

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

Sally C. Morton, Rep.(94-95) to Council of Sections
310-393-0411 ext 7360
The Rand Corporation
Sally_Morton@rand.org

Statistical Computing Section - 1995

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

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

Trevor J. Hastie, Past Chair
Stanford University
415-725-2231
trevor@playfair.stanford.edu

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

Robert J. Tibshirani, Program Chair–Elect
416-978-4642
University of Toronto
tibs@utstat.toronto.edu

James L. Rosenberger, Newsletter Editor (93-95)
814-865-1348
The Pennsylvania State University
JLR@stat.psu.edu

Terry M. Therneau, Secretary-Treasurer
507-284-1817
Mayo Clinic
therneau@mayo.edu

Karen Kafadar, Publications Liaison Officer
303-556-2547
University of Colorado-Denver
kk@tiger.cudenver.edu

MaryAnn H. Hill, Rep.(95-97) Council of Sections
312-329-2400 SPSS
hill@spss.com

Michael M. Meyer, Rep.(95-96) Council of Sections
412-268-3108
Carnegie Mellon University
MikeM@stat.cmu.edu

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

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

SECTION OFFICERS

Statistical Graphics Section - 1996

William DuMouchel, Chair
212-305-7736
Columbia University
dumouch@bayes.cpmc.columbia.edu

Sally C. Morton, Chair–Elect
310-393-0411 ext 7360
The Rand Corporation
Sally_Morton@rand.org

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

Stephen G. Eick, Program Chair
708-713-5169
AT&T Research Laboratories
eick@research.att.com

Dianne H. Cook, Program Chair–Elect
515-294-8865
Iowa State University
dicook@iastate.edu

Mario Peruggia, Newsletter Editor (96-98)
614-292-0963
Ohio State University
peruggia@stat.mps.ohio-state.edu

Robert L. Newcomb, Secretary/Treasurer (95-96)
714-824-5366
University of California, Irvine
rnewcomb@uci.edu

Deborah J. Donnell, Publications Officer (94-96)
206-283-8802 ext 258
MathSoft, Seattle, WA

Lorraine Denby, Rep.(96-98) to Council of Sections
phone
AT&T Bell Laboratories
ld@research.att.com

Colin R. Goodall, Rep.(95-97) Council of Sections
814-865-3993
The Pennsylvania State University
colin@stat.psu.edu

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

Statistical Computing Section - 1996

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

Daryl Pregibon, Chair–Elect
908-582-3193
AT&T Bell Laboratories
daryl@research.att.com

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

Robert J. Tibshirani, Program Chair
416-978-4642
University of Toronto
tibs@utstat.toronto.edu

James L. Rosenberger, Program Chair–Elect
814-865-1348
The Pennsylvania State University
JLR@stat.psu.edu

Mark Hansen, Newsletter Editor (96-97)
908-582-3869
AT&T Bell Laboratories
cocteau@research.att.com

Evelyn M. Crowley, Secretary–Treasurer
317-494-6030
Purdue University
crowley@purdue.edu

Karen Kafadar, Publications Liaison Officer
303-556-2547
University of Colorado-Denver
kk@tiger.cudenver.edu

MaryAnn H. Hill, Rep.(95-97) Council of Sections
312-329-2400 SPSS
hill@spss.com

Michael M. Meyer, Rep.(95-96) Council of Sections
412-268-3108
Carnegie Mellon University
MikeM@stat.cmu.edu

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

Janis P. Hardwick, Rep.(96-98) Council of Sections
313-769-3211
University of Michigan
jphard@umich.edu

INSIDE

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

1 8 5 5

Department of Statistics
University Park, PA 16802-2111

Nonprofit Organization
U. S. POSTAGE
PAID
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.