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

August 95 Vol.6 No.2

Statistical COMPUTING & GRAPHICS

A WORD FROM OUR CHAIRS

Statistical Computing

Mary Ellen Bock is the 1995 Chair of the Statistical Computing Section. She welcomes your feedback and comments on any issue of interest to your section.

The June 1995 Interface meeting in Pittsburgh was a resounding success. As a rule, members of the Statistical Computing Section participate heavily in the conference and its location is the regularly scheduled site of the first meeting of the year for the section's executive officers. (The officers will have to travel a bit further than Pittsburgh, Pennsylvania, to meet at the 1996 Interface, though. It is scheduled for July 8-10, 1996, in Sydney, Australia, and is being chaired by Nicholas Fisher and Lynne Billard.) (*Editor's note: The current information about the Sydney International Statistical Congress is available on the WWW at* `http://www.dms.csiro.au/sisc/index.html` *and includes a nifty form to request further information.)* The 1995 Interface meeting in Pittsburgh was hosted by our newsletter editors Michael Meyer and James Rosenberger. This adds another item to the long list of things for which we owe them thanks.

Their newsletter has proved so popular that other publishers have expressed interest in reprinting newsletter articles. In general, if you are interested in reproducing material from the newsletter, please contact the editors who are quite happy to grant permission with proper attribution.

CONTINUED ON PAGE 3

A WORD FROM OUR CHAIRS

Statistical Graphics

The officers of the Statistical Graphics Section have only begun the process of upgrading the section services available to its members. Secretary Robert Newcomb is currently preparing a WWW site. The section has also made a substantial investment with other sections for hardware to provide Internet connectivity at ASA headquarters. On a less substantial note, in addition to the usual fare, you can enjoy an ice cream sundae at our Orlando mixer Monday evening (well, at least the first one hundred lucky folks).

David Scott is the 1995 Chair of the Statistical Graphics Section. He and the editors encourage feedback, directly to the chair, or with letters to the editors of the Newsletter. Do you have something graphical you would like to display?

The Orlando program promises to be exciting. Check out the many statistical graphics paper opportunities available, thanks to Sally Morton. Your decision to be a paying member of the section has a direct impact on the program, as the allocation of invited paper sessions is strongly related to the number of members in each section. If you have any late-blooming ideas for sessions next year, Steve Eick would like to hear from you soon.

The section is involved in many activities in addition to ASA meetings. For example, Karen Kafadar and Tom Devlin organized a number of short courses at the Interface of Computer Science and Statistics conference in Pittsburgh in June.

CONTINUED ON PAGE 2

EDITORIAL

This August 95 issue (Vol. 6 No. 2) of the Newsletter should arrive before the Joint Statistical Meetings in Orlando Florida, August 12-17, 1995. The final issue for 1995, which is also the final issue for the current editors, should arrive before the end of the year. The deadline for this next issue is October 25. Thanks again to all who have contributed.

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

In his "Unix Computing" column, see page 13, Phil Spector writes clearly about the often confusing issue of file permissions. How many times have we UNIX users been hindered by a file which was put on our systems with the wrong permissions? Sometimes by our own system administrators? And then we might not even be able to use them. The opposite problem, of leaving confidential files with open permissions on a networked system should also haunt us. This article will clear up some of the confusion, and by using group permissions, allow us to share our files with just the group of individuals we wish to share our files with. Of course if giving permissions seems like a complicated problem, one needs only to work on a system like MS-DOS or Windows, where better permissions would be much preferred to the chaos which results when inexperienced users try to change the system, and have open access to the entire disk.

In another article in this issue, see page 8, Dan Carr describes a multivariate visualization problem and a suggested tool to solve it in his column on "Topics in Scientific Visualization". The suggested plot solves a problem of visualizing 4-dimensional data about potential energy in protein molecules.

Mike Conlon returns in this issue to continue his series on topics facing many of us in charge of "departmental computing", whether as faculty members, applied statisticians, department heads, or system administrators. The topic this time is *Productivity*, see page 6, and the factors which enhance and improve our performance.

In "Net Snooping" Mike Meyer talks about the latest developments on the world wide web which are of direct interest to our members. Especially useful is the search capability through the abstracts for the upcoming Joint Statistical Meetings in Orlando (see `http://www.stat.cmu.edu/joint95/`) and a new home page for the ASA. Another home page worth surfing describes the Interface '96 meeting in Sydney Australia (see `http://www.dms.csiro.au/sisc/index.html`) which provides the facility to enter your name on the mailing list to request further information about the meeting. The Interface '96 meeting is part of the Sydney International Statistical Congress (SISC-96), being held July 8-12, 1996, which is a joint meeting of The Statistical Society of Australia, The Interface Foundation of North America, and The Institute of Mathematical Statistics.

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

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

The conference co-organizers were none other than our two highly esteemed newsletter editors, proving successful collaborations can be repeated in a different arena. Short course presenters interested in the next Interface should get their bids in early, as the 1996 meeting will be in Sydney, Australia.

A number of you have asked about research opportunities available in this field. I would like to bring to your attention a forthcoming proceedings which summarizes a two-day meeting on the topic of Massive Data Sets, an NSA/NSF/CATS-funded meeting hosted in early July at the NRC. The workshop leaders were Jon Kettenring (ASA President-Elect) and Daryl Pregibon. The majority of participants were not statisticians, but scientists working with data from remote sensing, marketing, telephone networks, population surveys, meteorology, health providers, imaging, and database services. The workshop was videotaped, and selected video snippets will appear on the NSF WWW pages. Among the problems of scaling up statistical analysis and modeling were visualization problems and opportunities covering: detection of outliers and groups of outliers caused by nonstationarity; basic EDA tools for large data sets; presentation of massive data analyses; and presentation of estimation errors in pictures of three and four dimensional spatial field estimates. A small portion of this work will be covered in a special session in Orlando.

I would like to close by congratulating those selected by you to participate as officers in the section: Sally Morton, Rand Corporation, Chair-Elect; Dianne Cook, Iowa State University, Program Chair-Elect; and Lorraine Denby, AT&T Bell Labs, Council Of Sections Representative. You can help them by sharing your ideas and your willingness to serve in the future. Good luck to our newest officers.

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

The section (and all of ASA) also owes thanks to Michael Meyer and Lorraine Denby for the key role they are playing in bringing the ASA up to date with its Internet connection. The section officers have voted to provide both moral and financial support. Think of them when you visit the ASA web page.

The strong response to the 1995 Statistical Computing Student Paper Competition means that it will be repeated next year, 1996 in Chicago. Encourage students to submit an abstract and paper on a statistical computing topic. Winners receive travel support to the 1996 annual meeting of the ASA and present their papers in a special session. Even if a student does not win, the paper can be presented in a contributed session or a poster session. See the accompanying notice in the newsletter for details associated with the January 10, 1996, deadline.

Sallie Keller-McNulty stepped down as Chair-Elect to avoid conflict with her duties as Director of the Statistics and Probability Program at National Science Foundation. She has been a super activist in many roles for the section and we will miss her enthusiasm and leadership.

Congratulations to the newly elected officers who will take up the following positions in January, 1996: Program Chair-Elect James Rosenberger, Chair-Elect Daryl Pregibon, Secretary-Treasurer Evelyn Crowley, and Council of Sections Representative Janis Hardwick.

The success of the section's program of sessions and short courses at the August, 1995, ASA Annual Meeting in Orlando, Florida, is due to the efforts of our current Program Chair John Rice and our Continuing Education Chair Thomas Devlin. Our current Program Chair-Elect Robert Tibshirani is to be thanked for organizing the round-table discussions. If there is a topic you want to hear more about or one that you feel should receive more attention, now is the time to let program officers know about it for the 1996 annual meeting. (It is hard to believe that the planning takes place this far ahead of time but it does.) Suggestions for sessions for the 1996 program should go to Robert Tishirani and suggestions for round-table discussions at the 1996 annual meeting should go to James Rosenberger. Ideas for 1996 short courses are welcomed by Thomas Devlin.

The opportunity for input on program sessions, short courses, round-table discussions and the newsletter is one of the important benefits of section membership. On a lighter note, remember that another benefit of section membership is great ice cream sundaes at the Orlanda meeting Monday night mixer.

Thank you for your comments and suggestions and send more.

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

STATISTICAL COMPUTING

Student Paper Competition

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 are all presenting 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 comittee. 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 comittee, 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 the ASA by February 1 if they plan to attend the Joint Meetings. Those abstracts must be submitted following the ASA abstract submission instructions for contributed papers 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.

All electronic submissions of papers should be in postscript. Inquires and materials should be emailed or mailed to either one of:

Trevor Hastie
Student Paper Selection Committee
Statistical Computing Section
Statistics Department, Sequoia Hall
Stanford CA 94305
`trevor@playfair.stanford.edu`
or
Daryl Pregibon
Student Paper Selection Committee
Statistical Computing Section
Room 2C264, AT&T Bell Laboratories
600 Mountain Ave
Murray Hill, NJ 07974
`daryl@research.att.com`

CONTINUING EDUCATON

Continuing Education Proposals Sought

Proposals for short courses on topics in statistical computing or statistical graphics are invited for presentation at the Continuing Education Programs of the 1996 Joint Statistical Meetings in Chicago next August, Interface 96 in Sydney, Australia, or another professional society meeting. Courses may be either one-day (6 hours of instruction) or half-day (3 hours) in length. They may be textbook-based, but that is not a requirement.

In addition to a modest honorarium and travel expenses, offering a course provides: professional recognition for yourself and the Section, an opportunity to share new ideas with colleagues, exposure to a wide audience of statisticians, a service to the profession and the Section, and an opportunity to promote statistical education and statistical thinking.

Our sections have a great track record with Continuing Education activities. Courses we have recently co-sponsored include:

Joint Statistical Meetings

- "Multivariate Density Estimation and Visual Clustering" by David W. Scott
- "Exploring Multivariate Data with Modern Software: Examples in S" by Susan Holmes
- "Visualizing Data" by William S. Cleveland
- "Nonlinear Mixed Effects Models for Clustered Data" by Douglas Bates & Mary Lindstrom

Symposium on the Interface: Computing Science and Statistics

- "Trellis Displays" by Richard A. Becker, William S. Cleveland, and Ming Shyu
- "Using tcl/tk in Building Interfaces to Statistical Software and Robust Visualization" by John Kapenga
- "Tools for Discovering Patterns in Data" by John Elder
- "HTML and Creating World Wide Web Pages" by Bob Kuszewski
- "LaTeX for Statisticians" by Mike Meyer
- "Modern Nonparametric Regression and Classification" by Trevor Hastie and Rob Tibshirani
- "Resampling-Based Multiple Testing" by Peter Westfall and Stanley Young
- "Algorithms for Estimation and Visualization of Multivariate Density Functions with Applications to Clustering" by David W. Scott
- "Data Analysis using Interactive Dynamic Graphics: An Introduction to Xgobi" by Di Cook, Martin Koschat, and Deborah Swayne

Neural Information Processing Systems (NIPS) Conference

- "Statistical Modeling" by Leo Breiman

Joint Statistical Meetings in August 1995

The Statistical Computing Section will co-sponsor three short courses at the Joint Statistical Meetings in Orlando this August. The topics and presenters are:

- "Extending the Cox Model" by Terry Therneau
- "Applied Discriminant Analysis" by Carl J. Huberty
- "Introduction to Complex Bayesian Modeling Using BUGS" by Nicole Best and David Spiegelhalter

1996 Joint Statistical Meetings

Proposals for the 1996 JSM are being accepted through October 1, 1995. Proposals need to follow Guidelines established by the ASA Advisory Committee on CE. The Guidelines are available for anonymous ftp from `mozart.montclair.edu` in the directory `/pub/asascs` and from the sections' CE chairs. Proposals, inquiries and suggestions should be sent to either:

Thomas F. Devlin
CE Chair, Statistical Computing Section
Mathematics & Computer Science Dept.
Montclair State University
U. Montclair, NJ 07043
`devlin@mozart.montclair.edu`
or
Karen Kafadar
CE Chair, Statistical Graphics Section
Department of Mathematics
University of Colorado - Denver
PO Box 173364
Denver, CO 80217-3364
`kk@tiger.denver.colorado.edu`

NEWS CLIPPINGS

News from the NSF

Statistics and Probability Program

Program Directors

For FY96, which begins October 1, 1995, the Statistics & Probability Program Directors will again be Sallie Keller-McNulty, `smcnulty@nsf.gov`, 703-306-1883, who handles Statistics, and Stephen M. Samuels, `ssamuels@nsf.gov`, 703-306-1884, who handles Probability.

New Division Director

The Statistics & Probability Program is one of eight disciplinary programs within the Division of Mathematical Sciences (DMS). The new Division Director, effective July 1, 1995, is Donald J. Lewis, Professor of Mathematics at the University of Michigan. He succeeds Frederic Y. M. Wan, who at the conclusion of his two years at DMS became vice chancellor of research and dean of graduate studies at the University of California at Irvine.

Target Date and Screening Panel

Just as in 1994, there will again be a target date for research proposals submitted for consideration for FY96 funding by the Statistics & Probability program. This year's target date is November 9, the same as last year's. For FY96, the Statistics & Probability program will again employ a Preliminary Screening Panel as part of the merit review process. The panel will convene early in December.

Proposals which miss the target date will be handled only as time permits and may miss both panel review and/or funding decisions. Priority will be given to proposals arriving on or before the target date of November 9, 1995.

As in 1994, the screening panel's review will consist of two Panelist Checklists and a Panel Summary, reflecting the panel's discussion. Verbatim copies of these are eventually sent to Principal Investigators in the same manner as mail reviews. The panel advises the Foundation that a proposal either (i) be recommended for decline without additional review; (ii) receive additional mail merit review; or (iii) be recommended for award without additional review. Last year, approximately one-third of the proposals were in category (i), and all but a very few of the remaining proposals were in category (ii). Although the Program Directors may exercise their own discretion, last year they followed the panel's advice in nearly all cases.

NSF on the World Wide Web

You can read and print NSF publications and Program Announcements as well as the Abstracts of all NSF Awards made since 1989. The National Science Foundation's Home Page is `http://www.nsf.gov` and `http://www.fastlane.nsf.gov` (for current awards lists and abstracts). You can, for example, click on "Info & Pubs" and then on "STIS Database" (Gopher) Search to get the Gopher Menu, which includes many items, among them "Search NSF Award Abstracts", "Search NSF Publications", and "MPS" — Directorate for Math and Physical Sciences. MPS includes DMS, hence Letters, Program Guidelines and Reports from the division can be found here. Another item, Cross-Directorate Programs, has announcements which may be of great interest to statisticians and probabilists.

NSF E-Mail List

If you are not already receiving NSF announcements and would like to, send the following one-line message to `listmanager@nsf.gov`:

`subscribe statdept` [followed by your full name].

Sallie Keller-McNulty
NSF Program Director, Statistics
`smcnulty@nsf.gov`
and
Stephen M. Samuels
NSF Program Director, Probability
`ssamuels@nsf.gov`

DEPARTMENTAL COMPUTING

Productivity

by Michael Conlon

You've spent the money on hardware and software and you have your machines connected to the network. You have a talented system manager. Now how can you make yourself and your department more productive?

We will consider the question in terms of the four elements of a computing system—hardware, software, communications and people.

Hardware

Productivity means getting more done. Some people confuse this with running tasks on a computer faster. It may well be that running tasks faster would help you get more done, but the real question is not how long it takes the computer to get a particular task finished, but whether you are getting more of your work done.

A faster machine could possibly be a good start. But maybe not because it could run tasks faster. Maybe the reason to have a new machine is to run the software that will truly make you more productive.

Most modern computing systems benefit greatly from having more RAM— more RAM is typically the cheapest way to improve your system's performance.

If staff are waiting for equipment to become available productivity is suffering.

The way to improve your use of the computer may be to add more disk space. Having documents on disk is a great deal faster than having to hunt through diskettes or tapes for the necessary information. Put what you need on-line.

If access to a particular piece of hardware creates waiting time, consider getting more of the hardware people are waiting for. Faster printers, additional scanners, projectors, multi-media development stations all cost money, but if staff are waiting for equipment to become available productivity is suffering.

Hardware may provide a quick fix and obvious improvement — a processor upgrade, more RAM, or more disk space, may quickly get your computer to provide snappier answers. But becoming more productive will mean changing the way you work and changing the way you use technology.

Software

You must multi-task. That is, you must work in an environment where you initiate several tasks and are notified when tasks are completed. People who manage other people are accustomed to this—they delegate tasks to others who then complete the tasks and report back. People who use mainframe systems are familiar with the notion that they can submit jobs and when the jobs are completed they are notified. If you were bound to complete one task before starting another you would be unlikely to get much work done. If your computer is bound to complete one task before starting another you most certainly will get little done. Create a computing environment where you can initiate many tasks.

Make the computer do more work. In the old days of monolithic computing and computer accounts one was tempted to use own's time to save the computer some time. This would conserve precious computing credits. Today's environments could not be more different. Your time is very valuable. The computer's time is not. A computer will work 24 hours a day, 7 days a week for years. Say a computer must be replaced after three years. That's 6,000 working hours (40 hour week) or 24,000 hours of full-time operation. A typical desktop computer costs about $2,000, or under $0.10/hour. Cheap labor. Use it.

Use software that can be automated. Graphical User Interfaces (GUIs) are great and they tremendously reduce training costs and increase productivity of first-time tool users. But some software with GUIs can be difficult to automate. Newer tools on the MacOS, such as Applescript help. UNIX systems should have the language for task automation. Windows does not have a general task automation utility in wide use.

Use software that can be integrated. Insist that your software cooperate. Use tools that can export and import data from other tools easily. Use tools that respect your operating system and allow you to work in a manner of your choosing—not the tool's. Head for software that can easily be learned and easily adapted to your tasks.

If you spend your time looking for things on your system, consider getting software to help you look. There are many available tools for helping you hunt through what you have on-line. I have about 23,000 files on-line in my personal directories. I occasionally need to hunt through them using search tools.

Improve your email environment. You should be able to store and retrieve thousands of email messages and organize them in collections. Your email system should be easy to use and you should easily be able to get, read and respond to 100 messages per day. Email provides valuable archives of technical information and work flow. Implement MIME-base mail for sending and receiving word processing documents and graphics. MIME (Multimedia Internet Messaging Extensions) is a standard way to send and receive non-ASCII information via email. It is critical that work group members be able to send each other graphics, documents and databases without having to put them on external storage media such as disk and tape and physically move copies.

Get access to the web. There is a tremendous amount of useful material in archives on the web. Get on the web, learn about search tools and indices and start using material you find there. Going to the library takes time. Clicking on the web viewer is much faster.

Use automated procedures to create on-line documentation for the web. Write scripts to automate typical repetitious tasks. If you catch yourself doing the same thing over and over, give thought to turning the task over to the computer.

Think about how you can use the web to have information available for the people you work with in a timely manner, a consistent format and with excellent remote access. Put documents on-line to save paper production time and expense and to enable people to gain access to documents that otherwise they would have not received or would have misplaced. You don't need to play telephone tag with someone about information if you can call it up on your screen when you need it.

Communications

Today's networked systems move data much faster than by modem or (gasp!) diskette. Don't move data via diskette! Get computers on the net and move data using Internet protocols such as ftp.

Use less paper. It is amazing how many people are still entering data into computers that is printed on paper that was produced by a computer. Have the computers connected and transfer the data. You'll save time and errors.

Use standard high speed networks such as 10BaseT ethernet. If you are on a slower network, work hard to have the network modernized. The time saved by working on a high speed network is substantial.

Solve the problem of remote access to your data and your place of work. The goal is to have full access to your electronic resources (data, programs, documents) from wherever you are. Certainly you want to be able to access these resources from home. You will also want full access when you are on the road.

Getting on the Internet is essential. Email is cheap and easy and saves tremendous time and effort compared to the phone. Consider email lists for notifying groups of people and for sending memos. Create automated web archives of material sent by email so people can review and catch up.

People

Involve your system administrators in your productivity efforts. It is very tempting for system administrators to confine their operations to "the system" and leave the application level work to the users. Explain that the users are part of the system and it all exists to get work done. Of course there are system tasks that must be done for security reasons, maintenance reasons and for reasons of system stability. But good system administrators should be very concerned that the system is useful to the users and willing and able to help with productivity issues.

Reduce training costs by standardizing on software that is easy to use and has the features your department needs — especially features related to automation and integration with other tools.

Learn more about the tools you already have. Most large modern complex software systems have many features that go unused. Skim through the reference manual for a tool you use everyday and consider features that you do not yet use. Some features may simplify your use of the software.

Insist on a stable environment with a high-speed network and standard software tools that interoperate. Consider how these tools can best be used to reduce your effort in performing your tasks.

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

TOPICS IN SCIENTIFIC VISUALIZATION

Scanning a 4-D Domain for Local Minima: A Protein Folding Application

by Daniel B. Carr

With Contributions From Peter J. Munson and Geetha Vasudevan* (* Analytical Biostatistics Section, LSB, DCRT, National Institutes of Health)*

1. Introduction

Methods from statistical graphics apply to a wide range of applications. At Interface '95 I conjectured to Peter Munson that while protein folding (the collapse of a protein chain into a specific compact structure) occurs in three dimensions, insights might be obtained by considering constraints and using the methodology of higher-dimensional graphics. Peter immediately had a problem for me. He and Geetha Vasudevan had computed theoretical energies of short protein chain segments, described by a 7 x 7 x 7 x 7 lattice of dihedral angles. They knew the point of the minimum energy on the lattice. Peter asked if I could provide a visualization that would shed more insight into the energy surface. In Section 2 Peter and Geetha provide more details about the data. In Section 3 I indicate my design considerations in developing a first display of this data. Peter and Geetha discuss some implications of the display in Section 4. Finally Section 5 indicates some extensions. I have been thinking about extensions because Peter and Geetha have tougher problems at hand.

2. The Computed Protein Folding Data

This data set represents a theoretical potential energy as a function of the shape of a small segment of a protein molecule. The protein molecule is a long chain of residues which collapse into a very specific shape. Predicting this shape is known as the protein folding problem. We have modeled only four links of the protein's polypeptide chain which form a "reverse turn", basically a U-turn in the naturally occurring protein backbone. There are four main types of turns, designated I, I', II, and II' (Schulz 1979). The shape of such turns can be described by four torsion angles in the links of the chain. In an attempt to describe the complex energy landscape surrounding each of these turn types, we used a well-established potential energy function (CHARMm - see

Brooks, 1983). We expect naturally occurring proteins to take conformations which are close to their energy-minimal shape. Type II' turns have canonical values for the torsion angles as follows: phi1=60, psi1=-120, phi2=-80 and psi2=0 degrees. We calculated all the energies in a 4-dimensional, 60 degree window around these canonical values, and hoped to find a well-defined energy-well containing the unique energy-minimal conformation somewhere within. We also expected to see some very high energy conformations corresponding to "impossible" twists of the protein backbone.

The visualization of such energy landscapes is commonly done in two dimensions, phi and psi, (known as the Ramachandran plot). Here the challenge is to visualize the energy surface when there are 4 dimensions.

3. Graphical Design Considerations in Representing the Energy Data

My problem (Dan) was which visualization approach to select. The multivariate arsenal of statistical graphics tools continues to grow. For example cone plots (Dawkins 1995) are a recent addition. Given my historical bias toward ray glyphs in relative low dimensions, I still had the problem of which method to apply first: a 5-D display (e.g. stereo + ray angle + length as in Carr et al., 1986), a one-factor conditioned plot sequence of 4-D displays (stereo + ray), or a two-factor conditioned plot sequence of 3-D plots (rays or stereo). My first choice was the two-factor conditioned plot as illustrated in Figure 1.

Figure 1 conditions on the 7 x 7 levels of Phi1 and Psi1 to produce 49 small plots. In terms of recent history, the layout for Figure 1 (with margins added) dates back to Tukey and Tukey (1983) who called it an X3, X4 plot windowed by X1, X2. Cleveland (1993 and earlier) uses the word coplot to label the collection of conditioned plots. Tufte (1983) refers to the collection as small multiples. Here I use the word coplot and refer to individual conditioned plots as panels. Whatever the label, coplots have proved effective in breaking visual problems down into visually manageable pieces. A coplot seemed a good first choice.

> ***Here the challenge is to visualize the energy surface when there are 4 dimensions.***

Figure 1 differs from the Tukey and Tukey plots because individual plots (or panels) represent three variables: Phi2, Psi2, and energy . Given that the two angles are represented with x and y position (as shown in the legend) the question is how to represent energy? The numerous choices include ray angle, stereo depth, framed rectangles (Cleveland 1985) , circle area, colored contours, perspective views of fitted surfaces and colored dots. For a monochrome static view I chose ray angle to represent energy. Carr, Olsen and White (1992) discuss some merits of this choice.

> ***Coplots have proved effective in breaking visual problems down into visually manageable pieces.***

Both ray-glyphs (dot at the base) and arrows (arrow head at the tip) can represent angle. Arrow heads have line terminators and introduce additional visual angles. I use rays because reducing the number of line terminators simplifies the plot appearance, and eliminating extra angles makes it easier to focus attention on the information encoded as angles. Without explanation or experience, ray-glyphs are ambiguous as to direction. In my informal checks, some people have preferred arrows because their direction is not ambiguous. Careful cognitive testing may be required to shed more insight into the relative merits of the two representations for angle. Those who want to see the arrows' variation can obtain the plot by `ftp` as indicated below.

> ***Careful cognitive testing may be required to shed more insight into the relative merits of the two representations for angle.***

The dependent variable is energy. The units, Kilocalories/mole, can be negative and are to be interpreted relative to the minimum value. A simple Boltzmann formula converts the difference from the minimum into a probability that the system will appear with specified dihedral angles given by (phi1, psi1, phi2, psi2). I subtracted the minimum value before producing Figure 1. This simplifies the scale labels in the legend. The translated energy values cover a large range, [0–1097], relative to the region of primary interest which is only a few units from zero. To provide resolution in the region of interest, Figure 1 masks values above 20 and uses a nested scale. As the legend shows, white rays on a dark gray background encode values in the interval [0–10] and black rays on a light gray background encode values in the interval [10–20]. For Figure 1, I chose 20 as the upper limit to show a substantial portion of the data. My first picture for Peter and Geetha narrowed attention to values below 10. Using gray backgrounds rather than black and white backgrounds reduces the contrast and makes the figure easier to study. The two dark-gray regions in the Figure 1 make it immediately obvious that there are at least two local minima. This was not expected as described in Section 4.

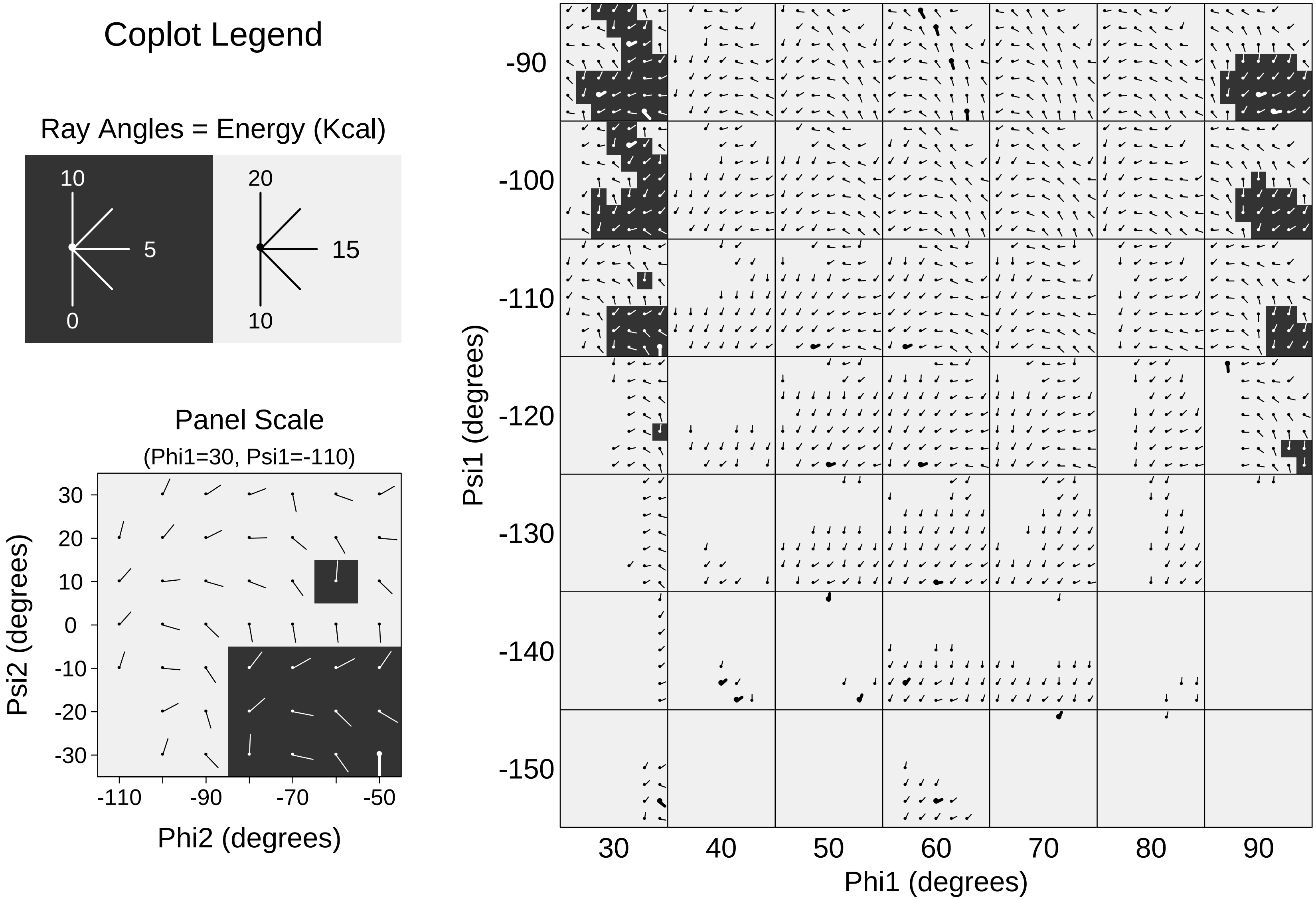


Figure 1. A Coplot of an Energy Surface.

Like coplots, nested scales allow readers to focus attention while limiting the mental burden through the use of identical structures for the scales. Typically color distinguishes the different scales. The monochrome Figure 1 shows two scales. This could be extended using shades of gray. Full color plots provide more options for distinguishing scales. The use of ordered colors (saturation and lightness) helps in mentally gluing the pieces together. The use of distinctive hues brings preattentive vision into play and promotes rapid evaluation within individual scales. A single plot cannot optimize both for overview and detail.

For a single plot, understanding the patterns across the nested scale transitions involves extra work. In Figure 1 white rays pointing up have energies similar to black rays pointing down. White rays pointing down have energies much less than black rays point up. Keeping track of this distinction is an extra mental burden. Rather than leave all the work to the reader in the search for local minima, it helps to algorithmically flag candidate values. Figure 1 shows the candidate local minima as enlarged ray-glyphs. Surprisingly there are 28 such local minima in the plot.

Like coplots, nested scales allow readers to focus attention while limiting the mental burden through the use of identical structures for the scales.

The local minimum values found depend on the definition of a local neighborhood. The square, cube, and hypercube lattices are awkward because “neighboring” points fall into two distinct classes, non-diagonals and diagonals. Non-diagonal neighbors are much closer than diagonals in the following sense. If one looks at the near-neighbor regions (hypercubes) about the lattice points, the hypercubes about non-diagonal points will share “cube faces” with the hypercube about the center point. The non-diagonal hypercubes barely touch the hypercube about the center point. Given a choice, I prefer to use the body-centered hypercube lattice in 4-D because the near neighbor regions (24-cells) for both non-diagonals and diagonals share “octahedron faces” with the 24-cell about the center point. A local minimum is established by comparison against the 24 neighboring points. For the current hypercube lattice, I choose to use non-diagonal neighbors to define the local neighborhood. Each point has 2 x 4 = 8 neighbors except for points on the edge of the domain that have fewer neighbors. A consequence of this definition is that diagonal troughs of local minima can appear. The four points in the (60,-90) panel are part of such a trough. The five points in the (50, 60) x (-130, -120, -110) set of panels form another diagonal trough. The two points in the top right panel are diagonally connected.

The coplot layout is important because it attempts to keep points in multivariate space close to each other in the plot. The layout represents a compromise since closeness in the plot is not equivalent to closeness in the 4-D space. The non-diagonal neighbors with a panel are much closer together than non-diagonal neighbors that have the same position in adjacent panels. At first glance, one might think that the point just to the left of the bottom right point in the (70, -100) panel is a local minimum. However, the values get lower as one goes up a panel and then left a panel. Reversing the roles of (phi1 psi1) and (phi2 psi2) would facilitate the study of energy as a function of phi1 or psi1 given the other values.

The coplot layout is important because it attempts to keep points in multivariate space close to each other in the plot.

Scanning the plot for more local minima is easier with interactive tools. Consider coloring the local regions dark gray or light gray depending on the whether or not the energy is above or below a particular cutoff value. When this cutoff value is under slider control, the user can increase the energy cutoff and immediately see when dark gray appears in visually disjoint regions. Slight slider oscillation will blink the new regions. For complicated surfaces many disjoint regions may appear and spotting new ones gets progressively harder. Switching to a new distinctive color can help. Interactive visualization can convey information about the energy well depth and shape.

Lattice plots like Figure 1 have some interpretational dangers. The dependent variable may change radically between lattice points. The plot does not reveal the exact location of the apparent local minima. Further, the plot provides no indication of a local minimum if an energy well is completely contained in a region between lattice points. Unless the lattice point separation is known to be smaller that the scale of all energy wells, the possibility of missing local minima remains.

4. Implications of Figure 1

Figure 1 reveals several important facts about our problem. First, the global energy minimum does not lie right at the center, but is just on the edge of the domain. In fact, it is on the boundary in three out of four dimensions, suggesting that the real minimum lies outside the window and more energy calculations are needed with a shifted domain. These energy calculations may require several hours on a workstation, so we can't expect to modify the domain interactively. We also notice that there is a second energy well (phi1=90) which is completely disconnected from the first well (phi1=30). Such multiple minima are a common feature in molecular modeling. If the energy barriers separating multiple minima are high enough, molecules can be completely trapped in local minima even though a nearby, lower-energy states exist. Likewise, numerical energy-minimization algorithms can become trapped in these local minima which makes use of simple gradient descent or Newton-Gauss type minimizers unreliable. We did not expect to see significant, multiple energy minima in this very simplified system (full proteins have hundreds of times the complexity of our modeled fragment), and indeed, until we visualized our data set, we had no idea that they existed here. One should not forget that the mathematical model employed here (CHARMm potential energy) may not be a good description of the actual forces influencing real molecules in its natural, watery environment. More complex, and time-consuming, but more realistic calculations would include the effects of water on the intramolecular forces and energies.

5. Extensions

Peter and Geetha indicate that there are many extensions of the protein folding problem to consider. I have started to think about them. For example they may compute at higher resolution. If need be, I will answer with a three feet wide plot of up to 9 feet in length. While static views cannot provide progressive disclosure, good old human pan and zoom is not bad. When I first spoke to Peter I had naively anticipated that 5-D plots would do the trick. For starters he tells me about domain dimensions in steps of 2 from 2-D to 30-D. Figure 1 suggests that 4-D domains are quite manageable. At first consideration I think that 6-D domains are within the limits of divide and conquer comprehension. I suspect that understanding a function on a 6-D domain in any kind of overview sense will take a lot of mental energy even with the best of visualization methods. Right now thinking about 8-D domains is too hard.

The statistical graphics community has much to offer in developing higher-dimensional visualizations for protein folding applications and a host of other applications. The opportunity seems so great that I felt compelled to write about it even though I have only been working on this problem for two days. If there are existing methods in the protein folding field that I don't know about, there is still a good chance that our community can produce something better. Figure 1 represents my first effort. The variations cited in Section 2 and those that occur to readers still need to be evaluated. I anticipate that my notion of preferred graphics will evolve as I try different variations and as Peter and Geetha guide my efforts. I hope others take up the challenge. Those that want to try their hand at visualizing the results of protein folding computation experiments can contact Peter at munson@helix.nih.gov. Readers can obtain my data sets and Splus script files by anonymous `ftp` to `galaxy.gmu.edu`. The files will be under `/pub/submissions/protein`.

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

Brooks, B. et al, 1983. *Journal of Computational Chemistry*, Vol 4, No. 187.

Carr, D. B., A. R. Olsen, and D. White. 1992. "Hexagon Mosaic Maps for Display of Univariate and Bivariate Geographical Data." *Cartography and Geographic Information Systems*, Vol. 19, No. 4, pp. 228-236, 271.

Carr, D. B., W. L. Nicholson, R. J. Littlefield, and D. L. Hall. 1986. "Interactive Color Display Methods for Multivariate Data." *Statistical Image Processing and Graphics*, eds. E. J. Wegman and D. J. DePriest,, New York NY: Marcel Decker, pp 215-250.

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.

Dawkins, B. P. 1995. "Investigating the Geometry of a P-Dimensional Data Sets," *Journal of the American Statistical Association*, Vol 90, No. 429, pp. 350-359.

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

Schulz, G. E. and Schirmer, R. H. 1979. *Principles of Protein Structure*, New York: Springer-Verlag.

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

File Permissions

by Phil Spector

Introduction

Since most computers running UNIX have many users accessing the same file systems, it is imperative that there are safeguards to prevent users from accidentally or maliciously overwriting or removing important system files or the files of other users. The first line of defense against such unauthorized access is a system of file permissions. All of the routines in the UNIX kernel which access files (which include devices like tape drives, CDs and diskettes) are governed by these permissions. On a well-administered UNIX system, it's possible that you could never run into any problems related to file permissions, but if you want to provide a little more security for some of your files, or understanding the workings of the UNIX system more thoroughly, it's a good idea to know something about file permissions.

In this article, I'll cover the basics of what file permissions are and how to manipulate them and touch on a few more advanced permission issues which you might encounter.

The Basics

Every file or directory on a UNIX system is associated with an owner(usually the creator of the file), as well as a group. The group associated with a file is the default group of it's owner. You can see what groups you are in by typing the command `groups`; the first group listed is your default group.

Thus, permissions can be controlled separately for three classes of users: the owner (usually referred to as "user" in UNIX jargon) members of the same group as is associated with the file, and everyone else ("other" in UNIX jargon). For each of these classes there are three categories of permissions, namely read, write and execute. So, for example, you could have a file which was only readable by yourself and members of the group you are in, but not others, or which was writable only by you, executable by others in your group, and restricted completely for other users.

One point about file permissions which should always be kept in mind is that the superuser (root) or anyone who has access to the root password, or to your password, can do whatever they want to any of your files. So if you have files which absolutely should not be seen or manipulated by anyone else, you should either encrypt them in some fashion, or store them somewhere other than a computer system which is accessible by others.

To display information about permissions, the `-l` and `-g` options of the unix command `ls` can be used. The `-l` option specifies the long form of a file or directory listing; it includes all the information about the three forms of permission for the three classes of users, as well as the name of the owner of the file, and the size of the file in bytes. The `-g` option shows the group association of the file. Here's an example, listing a few files from the `/usr/bin` directory, using the `ls -lg` command:

```
-rwxr-xr-x  1 root  staff   5504 Jan 20  1994 cmp
-rwxr-xr-x  1 root  staff   8136 Jan 20  1994 col
-rwsr-xr-x  1 root  staff  24576 Jan 20  1994 crontab
```

The first column contains an indication of the type of the file. For a regular file, this column will contain a hyphen (`-`), as in the example above. For a directory, the letter `d` will appear, while symbolic links are designated with the letter `l`. (A symbolic link is a filename which really represents a file at a different location in the file system. When trying to access these files, the permissions which are used are those of the actual file, not of the link. See the manual page for the `ln` command for more information.). There are one or two other letters which may appear in the first column of the `ls -l` listing to indicate special files such as storage devices and terminal sessions.

It's a good idea to remove write permission in a directory if you don't want the files in that directory deleted.

The next nine columns display the permissions of the file. These nine columns can be thought of as three sets, each of three columns. The first set of columns applies to the owner (user) of the file, followed by columns for the group and finally those for everyone else (other). For each of the three sets of columns, the individual columns refer to read, write and execute permissions in that order. So in the example above, for the file `cmp`, the user (`root`) has read, write and execute access, users in the group `staff` have read and execute permissions, as do all other users. Usually, these permissions are represented by a `r`, `w`, or `x`, but as can be seen for the `crontab` entry, an `s` may appear. This is explained in the section on Special Permissions.

The next column following the block of permission columns shows the number of links to the file, followed by the owner, the group, and the size of the file in bytes. The next few columns display the date on which the file was last modified, finally followed by the name of the file.

What the Permissions Mean

The actual meaning of read, write and execute are different depending on whether they are referring to a normal file or to a directory. When you use the `ls` command on a directory, it will not normally show you information about the directory itself, but on the files found in the directory. To get specific information about the directory, you should use the `-ld` or `-ldg` flag of the `ls` command.

For a normal file, read permission allows opening the file so, for example, it can be viewed with an editor or pager, or read by an executing program. It also allows copying of the file to a different location (provided, of course, that you have permission to write files to that location). For a directory, read permission allows listing of the files in the directory, usually using the `ls` command. However, since individual file permissions override the permission of the directory, there is no guarantee that you can read all the files in a directory for which you have read access.

For a normal file, write access allows changing or deleting the file. To create a file, however, you must have write access to the directory in which you want the file to reside. Write access in a directory also allows removal of files in that directory. If a file is located in a directory for which you have write permission, but you don't have write permission for the individual file, you will be given the option of overriding the file's write protection, so it's a good idea to remove write permission in a directory if you don't want the files in that directory deleted.

You can only modify permissions on the files which you own.

Finally, execute permission for a file allows you to have the commands in the file executed by typing the name of the file. When you use a compiler to create an executable binary file, it automatically gives the file execute permission. However, if you create a file of executable commands like a shell script or a `perl` program, you'll have to set the execute permission for the file as described in the next section, or the system will respond with "`Permission denied`" when you try to execute the file. For a directory, execute permission allows you to use that directory as your current directory, as well as allowing you to open and execute files in the directory. If another user reports seeing lots of error messages when trying to access files in one of your directories, even though they can eventually access the files individually, check the execute permission of the directory in question.

Changing Permissions

The command which is used to control permissions is `chmod`, short for changing the "mode" of the file. First, keep in mind that you can only modify permissions on the files which you own, that is, those where the `ls -l` command shows you as the owner of the file. If you accidentally remove your own permissions to a file, don't despair; as the owner, you can always reinstate them. There are two forms of syntax for using the `chmod` command: symbolic and absolute. With symbolic syntax, you specify requests as a combination of a class of users, an action and the permissions involved; multiple requests are separated by a comma. The characters used with symbolic syntax and their meanings are displayed in the table below:

Class		Action		Permission	
value	meaning	value	meaning	value	meaning
`u`	owner	`+`	add	`r`	read
`g`	group	`-`	remove	`w`	write
`o`	other	`=`	set	`x`	execute
`a`	all				

A common example mentioned in the previous section is allowing a file containing shell commands to be made executable. The command `chmod +x myscript` would allow you to type `myscript` and have the commands in the file `myscript` executed by the shell. If you wanted to remove read and write access to a file called `secret` for other users (not in the file's group), you could use the command `chmod o-rw secret`. To give yourself read, write and execute access to a file called `myfile`, while giving everyone else only execute access, you could use the command `chmod a=x,u=rwx myfile` since the comma-separated requests are processed in they order they appear.

As an alternative to symbolic syntax, you can use absolute syntax, where the permissions of a file are represented as three octal digits, one each for user, group and other. Each of the digits is composed of the sum of 4 (for read access), 2 (for write access) and 1 (for execute access). So in the above example for `myfile`, you could use the command `chmod 711 myfile`; `7=4+2+1`, representing full access for user, and `1` representing only execute access for group and other. The choice between symbolic and absolute syntax is strictly a personal one, although symbolic syntax is usually easier if all you want to do is add or delete one class of permissions, as in the `myscript` example above.

Default Permissions

When you create a file or directory through a UNIX command (like an editor, or the `cat` or `mkdir` commands, the operating system provides default permissions of read and write permission for everyone (666 in absolute syntax) for files, or read, write and execute permission for everyone (777) for directories or executable files. These system defaults, however, are always modified by the shell using a value known as a umask. The umask value, which is generally set in the `.profile` or `.cshrc` file, is an octal number which is used to mask the system default value as described above, effectively removing those permissions which are represented by the value of umask. A commonly used value for umask is 022; this removes write permission for group and other, resulting in absolute permission 644 (`rw-r--r--`) for files and 755 (`rwxr-xr-x`) for directories and executables. (You can check your value of umask by typing `umask` with no arguments.) If you wish to provide more security for your files, you could set your umask to a more restrictive value, for example 077, which would provide no permissions for anyone but you. This would be done with the command `umask 077`, which you could either enter in your `.cshrc` or `.profile` file to have it in effect every time you log in, or type as a normal UNIX command to have it in effect for a single shell session. (Keep in mind, however, the point raised in the Introduction, namely that anyone with the root password or your password will be able to override the permissions you set.)

Special Permissions

In the example of the output of the `ls` command, the entry for `crontab` had an `s` in place of the `x` which would otherwise by expected for execute permission. This indicates that a the file has a special type of execute permission known as setuid permission. This means that when this file is executed, it is executed utilizing the file permissions of the owner of the file instead of the usual file permissions of the user executing the file. By using this technique, system files can be modified by ordinary users, without giving those users direct access to the files. A closely related form of permission, known as setgid, executes the file with the permissions of the group which owns the file, instead of the user who is executing the file. Setgid permissions are indicated by an `s` in place of the expected `x` for the group permission in an `ls -l` display. This description of setuid and setgid assumes that the usual execute permissions for owner or group are in effect; if not, or if the file in question is a directory, the meaning of these permissions varies from system to system; consult your system's manual page for the `chmod` command for more information.

Since the execute bit can be modified for user and group, it is not surprising that a similar modification exists for other. This permission is known as the "sticky bit", and its meaning varies from system to system. Once again the manual page for the `chmod` command should explain the particulars on your system.

In absolute mode, these permissions are controlled using a leading octal digit formed in a similar way to the three digits following it, with setuid = 4, setgid = 2 and the sticky bit = 1.

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

NET SNOOPING

Internet Access and the American Statistical Association

by Mike Meyer

In this short column I wish to describe two unrelated (but soon to be related) aspects of how the American Statistical Association is making use of the Internet. The two examples are the abstracts for the Joint Statistical Meetings which are available on StatLib, and the soon to be announced ASA WWW server.

Joint Meetings Abstracts

For the past two years the ASA office has provided StatLib with the full abstracts of papers to be presented at the annual meetings. The current batch of abstracts are in `http://www.stat.cmu.edu/joint95/`, or available by sending the one-line e-mail message `send index from joint95` to `statlib@lib.stat.cmu.edu`. Using either the e-mail or WWW interface it is possible to search the abstracts by author, keyword, or title. So far this year, there have been over 8,250 transactions (that is searches or subsequent lookup of data) to the database. The vast majority (7,800) have come via the WWW.

Just from looking at the logs, it becomes pretty clear how people are using this information. There seem to be two classes of use. Many people are interested in an area of methodology or application and will do searches for particular keywords or phrases. Others are interested in particular authors (often their friends and colleagues).

What are these log files?

Some readers may be surprised to know what information StatLib collects about its clients. For people who use StatLib via e-mail I maintain a log record that shows the e-mail address of the user and what information they retrieved from StatLib. A particular contributor to StatLib will sometimes ask about who has retrieved their material, or just the number of people who have accessed it, and hence parts of the log file are sometimes shared. Otherwise the information is just securely archived.

For WWW, FTP, and Gopher users I mostly have much less information. The only reliable data I have is the host name or IP address of the machine that established a connection to StatLib and then the files that were retrieved. For some hosts (like the various America On Line computers), this doesn't come close to identifying a unique user. For other hosts—particularly where I have some meta knowledge about who owns the machine—knowing the host name is very close to knowing the user name. Again this information is not public and only occasionally divulged to the the StatLib contributors.

Some of you may be concerned about the privacy of this information, especially with the recent media firestorm about the so-called Carnegie Mellon study of Internet pornography. I can honestly assure you that the information kept by the primary StatLib server is kept quite securely and not used for anything other than "statistical" purposes.

How is the JSM information collected?

The information available in the `joint95` and `joint94` areas of StatLib was provided directly by the ASA office, specifically with the help of the Tim Gill, the ASA computer specialist. In 1994 Tim provided me with a flat file containing all the abstracts in a more or less uniform format. I spend some considerable time repairing minor problems with the data (via a few emacs macros, and a PERL script or two, and occasional manual intervention) and eventually write another short PERL program to actually search the database.

Based on the 1994 experience, I gave Tim much clearer directions in 1995 and we were able to get the database from him into StatLib with about a days work. I've been very happy with the results and encourage everyone to explore the sessions before the meetings.

ASA, the Internet, and you

By the time you read this, I hope that `http://www.amstat.org/` will be a valid URL! As I write this I am working on the final plans for getting a direct Internet connect for the ASA main office. The connection is being sponsored by the ASA Council of Sections with money donated by several of the ASA sections. Many people have contributed to this effort. Rather than list the people and the amounts that the various sections contributed (which I would almost surely get wrong!), I do need to compliment a small team of dedicated ASA members who have kept the project running: Mike Conlon, University of Florida; Lorraine Denby, AT&T Bell Laboratories; Dan Jacobs, Maryland Sea Grant College; Tim Gill, ASA; Dan Horvitz, ASA; David Morganstein, Westat; Dan Solomon, North Carolina State University; and myself.

We have purchased an Internet 56k leased line Internet connection from PSI, and a Pentium PC to act as the server and firewall for the rest of the ASA office (it will run some flavor of Unix—probably Linux). Mike Conlon has made an excellent first pass at a set of ASA WWW pages. We view the WWW pages as just the start of an effort to provide many member services via the Internet. I'm sure you will hear more about his project at the annual meetings and later.

So, why don't we have the names `asa.org`, and `www.asa.org`? The answer is simple. Someone else already "owns" that mane. In this case a company called the Antarctic Support Organization. The only other name that we could agree on (and obtain!) was `amstat.org`, and that is what we will have.

Please explore the ASA home pages, and please be tolerant of the relatively low bandwidth we are able to afford. As we gain experience with the service we will attempt to improve all aspects of it.

I could write a *lot* more details about the hardware and Internet service, but this is probably not very interesting to anyone other than me, and way too fresh in my mind for me to even consider committing it to paper.

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

NEWS CLIPPINGS

ASA Election Results

President-Elect, 1996 (President 1997): Jon R. Kettenring, Bellcore.

Vice President (1996-98): David C. Hoaglin, Abt Associates Inc.

Results for the 1996 elections for officers of the Statistical Computing and Statistical Graphics Sections.

Statistical Computing Section

Chair Elect: Daryl Pregibon, AT&T Bell Laboratories

Program Chair-Elect: James L. Rosenberger, Penn State University

Secretary/Treasurer: Evelyn M. Crowley, Purdue University

Council of Sections Representative: Janis P. Hardwick, University of Michigan (1996-1998)

Statistical Graphics Section

Chair Elect: Sally C. Morton, The Rand Corporation

Program Chair-Elect: Dianne H. Cook, Iowa State University

Council of Sections Representative: Lorraine Denby, AT&T Bell Laboratories (1996-1998)

Congratulations!

NEWS CLIPPINGS

Interface '95 Report

The 27th Symposium on the Interface, Computing Science and Statistics, was held in Pittsburgh, Pennsylvania, June 21-24 at the Vista Hotel and David L. Lawrence Convention Center.

Raj Reddy, Dean of School of Computer Science and Robotics at Carnegie Mellon, delivered the keynote address on "Statistics, Computation, and Artificial Intelligence". He described and showed with futuristic video clips, human-computer interaction, demonstrating the important role artificial intelligence plays for the successful completion of the many tasks we might wish to assign to our silicon, robotic "gofor".

Interface '95

PITTSBURGH, PENNSYLVANIA

About 225 attendees mingled in the spacious David L. Lawrence Convention Center listening to talks centered on the theme "Statistics and Manufacturing with Subthemes in Environmental Statistics, Graphics and Imaging". Mike, Jim, and the CMU staff worked hard and were always around (and easily visible in their Interface shirts) to answer questions and point people in the right direction. They even printed a pocket sized "schedule-at-a-glance" to help us decide which talks to attend.

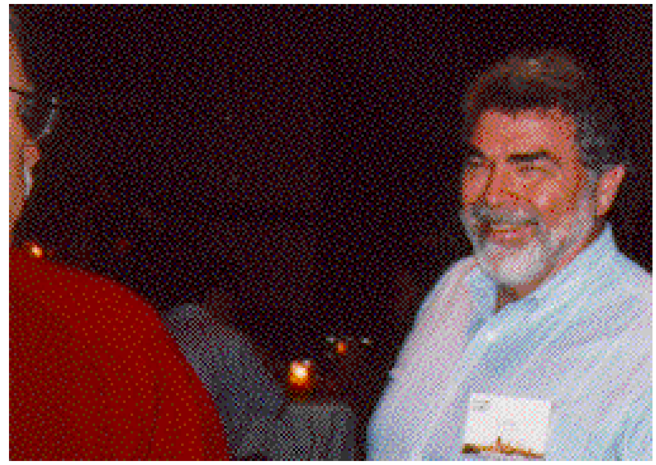
Carnegie Mellon staffers Mary Beth Meiser, directing, and Mari Alice McShane, first violinist, orchestrated the able support staff which kept the conference flowing smoothly.

Five short courses were offered on Wednesday before the conference began. They were: "Tools for Discovering Patterns in Data", "Using tcl/tk in Building Interfaces to Statistical Software and Robust Visualization", "Trellis Displays", "LaTeX for Statisticians", and "HTML and Creating World Wide Web Pages". Participants were enthusiastic about the offerings and appreciated the efforts of Tom Devlin and Karen Kafadar, who organized them.

John Elder, short course presenter on “Tools for Discovering Patterns in Data” and Sally Morton enjoying the Wednesday evening mixer with conference co-chair Jim Rosenberger.

After a shaky start on Thursday (the keynote speaker arrived late) the rest of the conference went well. Sessions were well attended and break times offered great opportunities to talk to the speakers or catch up with friends. Thursday night’s banquet was a definite highlight of the conference. Dinner on a riverboat cruising the three rivers of Pittsburgh on a beautiful warm evening — a chance to relax and mingle with the rest of the group — great idea Mike and Jim!! (And really, those who drove to the boat don’t blame you for those directions...)

Saturday was sure a good time to be clearing out of the convention center. We were being followed by the National Holstein Convention, and a bunch of cows would not have been fun to contend with. Thanks again to Jim and Mike for their hard work. See you next year in Sydney.



Larry Cox, program committee member and session organizer, discussing the environmental sessions with Mike Tarter.

Linda Clark
AT&T Bell Laboratories
lac@research.att.edu

Reader’s Notes

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) to 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) to Council of Sections
312-702-8332/8333
The University of Chicago
r-thisted@uchicago.edu

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

INSIDE

A WORD FROM OUR CHAIRS
Statistical Computing . . . 1
A WORD FROM OUR CHAIRS
Statistical Graphics . . . 1
EDITORIAL . . . 2
FROM OUR CHAIRS (Cont.). . .
Statistical Graphics . . . 2
FROM OUR CHAIRS (Cont.). . .
Statistical Computing . . . 3
STATISTICAL COMPUTING
Student Paper Competition . . . 4
CONTINUING EDUCATON
Continuing Education Proposals Sought . . . 4
NEWS CLIPPINGS
News from the NSF . . . 5
DEPARTMENTAL COMPUTING
Productivity . . . 6
TOPICS IN SCIENTIFIC VISUALIZATION
Scanning a 4-D Domain for Local Minima: A Protein Folding Application . . . 8
UNIX COMPUTING
File Permissions . . . 13
NET SNOOPING
Internet Access and the American Statistical Association . . . 15
NEWS CLIPPINGS
ASA Election Results . . . 17
NEWS CLIPPINGS
Interface ’95 Report . . . 17
SECTION OFFICERS . . . 19

Statistical COMPUTING & GRAPHICS

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

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

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

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

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

PENNSTATE

1855

Department of Statistics
University Park, PA 16802-2111

Nonprofit Organization
U. S. POSTAGE
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.

Penn State is an affirmative action, equal opportunity university.
U.Ed.SCI 96-4