Statistical
COMPUTING & GRAPHICS

A WORD FROM OUR CHAIRS

Statistical Computing

Our thanks to Trevor Hastie, the outgoing 1994 Chair of the Statistical Computing Section. He recently moved to Stanford University, Department of Statistics, from Bell Laboratories.

It was nice to see our sessions at the ASA meetings in Toronto so well attended. I counted over 250 people in the neural network session featuring Geoff Hinton. Our business meeting and mixer, joint with graphics, was also a great success. I think the fact that we have improved the spread over the years has helped!

Let me summarize the highlights of the executive meeting of our section, which took place during the joint statistical meetings. One of our main missions is to promote statistical computing to students and new ASA members.

Statistical Graphics

Roy Welsch is the outgoing 1994 Chair of the Statistical Graphics Section. The editors wish to thank him for his leadership and help in providing this column during the past year.

It seems like just yesterday that I was writing my first column as Chair of the Statistical Graphics Section. Given my record at meeting Newsletter deadlines it probably was just yesterday! The time is here (and the deadline has passed) to write my final column before giving the pen to David Scott, next year's Chair. Next year I, too, will be able to enjoy my Thanksgiving vacation.

FEATURE ARTICLE

Teaching Statistics and Computing via Multimedia through the World Wide Web

by A.J. Rossini and J.L. Rosenberger

Summary

Hypertext is growing in use as a method for information retrieval and dissemination. The introduction of Mosaic has led to a rapid expansion of information services on the Internet. We propose the abstraction of statistical data, information, and procedures for dissemination and teaching through this medium. The power of hypertext unleashes a new approach for constructing and distributing learning materials for statistical courses. A paradigm of statistical viewers is discussed, and an implementation using Xlisp-Stat and Hyper-Text Markup Language (HTML) is described for UNIX and the X11 Window System. Technical details are presented, and a demonstration is available on-line on the Internet to experience first hand the potential of this medium.

Introduction

In attempting to understand statistics, a typical student spends the majority of time outside the classroom learning and studying the material. The popularization of the Internet provides students with an additional means for obtaining assistance for learning and practicing statistical methodology. This article describes a variety of instructional mechanisms using the World Wide Web (WWW), a network or virtual web of documents and information sources accessible from any computer on the Internet world-wide. Mechanisms are described and examples given for a variety of possible instructional uses.

This December 94 issue (Vol. 5, No. 3) of the Newsletter is the last issue for many of the current section officers. As the new officers take their positions, you should be thinking about how you might contribute to our sections. Also we challenge you to suggest colleagues who may be able to contribute for their own professional development, and to enhance the sections' activities. John Rice, program chair, is requesting volunteers for chairing statistical computing sessions at the 1995 Orlando Meeting of the ASA. Please contact him if you wish to nominate a junior colleague, or if you wish volunteer to chair a session yourself.

Finally, we are still looking for suggestions of persons who would like to edit this Newsletter for the next 2–3 year term. It is not a lonely task—the sections have been cooperating—with a co-editor from each section.

This issue continues the recent multimedia blitz about the multitude of uses of the Internet. Our feature article by Rossini and Rosenberger is but one example of such articles, which describes an application of the Internet which should change the face of distance education forever. *Syllabus* magazine also featured "Education and the Internet" in their Nov/Dec 1994 issue. Although we are sensitive to the need for balance in this Newsletter, these two articles on XlispStat converged at the same time, and seemed to fit together nicely. Since XlispStat is in the public domain, we hope the attention promotes the profession, and not just the developers.

The Statistical Computing Section of the ASA has approved a student paper competition for the annual joint statistical meetings, to be held in Orlando next August. Advisors should encourage their graduate students to submit an abstract and paper on a topic of interest to the statistical computing section. The winners will receive registration and travel funds, and the rest will still have the opportunity to enter their paper as a contributed paper or poster session—winners all. If you are unable to make the deadline for this year, then this reminder is for long-range planning for next year, since the intention is to make this an annual event. See details in our chair's column and on page 5.

Our regular columnists have continued their run of interesting articles. Mike Conlon discusses the issues arising in establishing a computer network server in his Departmental Computing column; Dan Carr investigates the graphical issues involved in producing boxplots, with a redesign of several features and the introduction of color to assist the eye in making comparisons. Is this evolutional shift an improvement? I'm sure he would

appreciate feedback, and we will gladly publish any letter to the editor on the topic. In the Net Snooping column, Mike Meyer describes a new product, Netscape, which has challenged Mosaic as the interface of choice to the Internet. The landscape is moving very quickly, and it should be fun to watch the rapidly emerging products available as the marketplace pushes it's way into cyberspace.

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 or TeX, but will gladly accept plain ASCII text files as well.

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

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

ⓒ

Statistical Graphics

I do have something to be thankful for and that is all the help that I have received as from many volunteers. Linda Clark has superbly managed our Section finances (without derivatives!), created logical minutes from meetings where the discussion was fast and furious, and co-produced the joint mixer at the Annual Meetings in Toronto.

Bill DuMouchel put together the excellent program this year while moving from Boston to New York to take a new job. Rick Becker has been a fountainhead of knowledge as past chair and was always reminding me of the things I needed to do to carry on the initiatives that he started. (Got that, David.)

Jane Gentleman and Sally Morton have carried our views to the Council of Sections and endured the organizational machinations of the ASA. Sally also continued her excellent management of our video library and volunteered to take over as Section Program Chair-Elect from Sallie Keller-McNulty when Sallie needed to

resign because of her new responsibilities at NSF. Sallie and Sally have already assembled an exciting program for Orlando. Deborah Swayne will assume the video library position so that Sally Morton will have time for something other than Section business. Although not a Section officer, Lorraine Denby is always ready to help and has informally served as Section archivist and historian while encouraging us to strike out in new directions as well. Lorraine also becomes the new member of the Section Committee to Nominate Fellows replacing Luke Tierney. Congratulations to Andreas Buja who was nominated by the Graphics Section and became a Fellow this year.

Jerry Moreno deserves to be thanked again for managing the very successful Graphics Poster Competitions. Please help to continue the tradition by saying yes when John Schollenberger, Jerry's successor, asks for assistance. Robin Lock will be handling the Data Exposition for the Orlando Meeting. He would welcome suggestions for data sets to be used in the Exposition.

Thanks also to Abbe Hertzig who is stepping down as head of continuing education for the Section. Her replacement, Karen Kafadar, has been busy and has at least one session ready for Orlando entitled "Using tcl/tk in building interfaces to statistical software and robust visualization" which will be given by Joseph McKean and John Kapenga from Western Michigan University.

Mike Meyer deserves everyone's praise (even those of us with deadlines) for co-editing (with Jim Rosenberger from the Statistical Computing Section) the Newsletter you are reading. Mike is finishing his term as editor and he and the Section are looking for suggestions for a new editor.

Our strong budget position has allowed us to take several new initiatives. We will be funding more outside speakers at our meetings and providing them with first class equipment to give their presentations. To provide an incentive for authors to send their best work on graphics to the Journal of Statistical Computing and Graphics (JCGS), I have appointed a committee consisting of Rick Becker (Chair), Howard Wainer, and Jerry Moreno charged with setting up a process that will lead to the awarding of a prize or prizes for the best papers published in JCGS. This committee may also be asked to consider an evaluation process and prizes for the best graphics presentation at the annual meetings.

We also will be spending more to promote Section activities, to subsidize continuing education when necessary, to provide incentives for Section members (reduced fees and subscription charges), and to encourage student participation at all levels. This year the Section supported a Data Analysis Contest for undergraduate mathematics and statistics majors run by David Robinson at St. Cloud University. We also want to help the ASA and all Sections get involved the Internet, Mosaic, and the World Wide Web since these tools can now be used to provide graphical communication. This may require some funds as well.

Rick Becker as Chair of the Nominations Committee has proposed the following candidates for the 1995 Statistical Graphics Election:

Chair-Elect: Sally Morton and Linda Clark

Program Chair-Elect: Robert Wesley and Di Cook

Council of Sections Representative (96-98): Lorraine Denby and Jerry Moreno

Final notes:

Don't forget the Interface '95 Symposium in Pittsburgh, PA June 21-24, 1995.

Please give some time and energy to your Section. That is what makes things happen.

My apologies to David Scott for not returning his black transparency pen that I borrowed to make some slides for the Section business meeting/mixer. David mentioned this at his talk when he had only slides made with a red pen. At least I used to prepare my talks on the plane coming to the meeting, not at the meeting!

It is now a pleasure (and relief) to introduce David Scott to this column. Good luck, David!

> Roy Welsch
> *Chair, Statistical Graphics Section*
> `rwelsch@sloan.mit.edu`
>
> ⟨⟩

Call for Ideas

The Statistical Graphics Section officers are interested in hearing from members about effective and appropriate ideas for budgeting and expending section funds. General or specific ideas should be sent to David Scott, Chair-elect, scottdw@rice.edu.

> David Scott
> *Chair-Elect, Statistical Graphics Section*
> Rice University
> `scottdw@rice.edu`
>
> ⟨⟩

Statistical Computing

CONTINUED FROM PAGE 1

Our Continuing Education representative, Tom Devlin, proposed to enhance the scholarships given by our section for students to attend short courses co-sponsored by our section. These courses are given at both the Interface meetings and the ASA. The executive approved his proposal, and the scholarships will cover both the course and conference registration fees for up to 20 students per year. Details on how to apply will be provided in the next Newsletter. We also voted to augment the support of presenters of short course sponsored by our section at the ASA meetings. Currently they receive an honorarium of $500 and expenses of $1000 towards travel and lodging, from the ASA, to be shared by all the presenters. Our section will augment this by $250 per presenter, for up to three presenters.

The deadline for the student paper competition is approaching (January 10, 1995). This is a wonderful opportunity for students to attend the ASA meetings fully funded, give a paper in a special contributed paper session, and have their paper summarized in this newsletter! Details appeared in the last newsletter, August 94, page 3, and is reprinted here on page 5.

Leo Breiman of U.C Berkeley is giving a tutorial on statistical modeling at the 1994 Neural Information Processing Systems (NIPS) conference in Denver this December. This is the flagship conference for Neural Networks, and our section is co-sponsoring his tutorial. Other section members are participating at this meeting as well: Rob Tibshirani, Jerome Friedman and I are all giving papers.

Statistical Computing Issues

Now to an issue that has been on my mind lately: **STATISTICAL COMPUTING!** Statistics has certainly been changing over the years, and by far the biggest impact has been the effect of computing speed, data transmission and storage capacity. Think of some of the hot topics in the field right now: the bootstrap and resampling schemes for inference; Gibbs sampling in Bayesian posterior computations; maximum-likelihood estimation of complicated and highly connected genetic models; likewise for graphical Bayesian belief models; wavelet modeling of high dimensional signals and images; statistical models where observations are im-

ages; neural network and other non-parametric models for classification and regression with huge training sets; high dimensional inverse problems (PET); pattern recognition in high-resolution satellite images; real time (interactive) statistical graphics. In some cases (e.g. bootstrap, Gibbs) these techniques are supplanting some of our traditional statistical and probability models, created for mathematical and computational convenience but relying on unjustified assumptions. In other cases (e.g. images, interactive graphics) we can dabble in areas that were simply not available to us in the past. These are all computer intensive areas, and require significant resources and computing skills. One reflection of these changes has been the recent burst of new statistical computing journals—it has become legitimate to publish computing type articles (and get credit from our peers for doing so!)

What about our computing needs, in statistics departments at universities and other research centers? We are beyond the age of the "package" that can do all things statistical. We need well integrated and planned hardware and software environments, and facilities and abilities for programming at a variety of levels. Some of the areas require really specialized equipment and software. Most important, these are not one time needs that a few committee meetings and a lump of money can solve.

Here comes a change that statistics departments have the most difficulty dealing with. With all these computing demands, a Computer Resource Administrator is needed. This is a specialist who can manage the planning of the environment; who manages software and hardware acquisition and maintenance; who can take notice of and try out innovative software and freeware; who can teach introductory and not so introductory computing classes to the faculty and graduate students; who can manage data bases, create WWW resources and manage ftp sites. The list goes on.

Why is it so hard to justify such a person? Mainly because it sounds like a faculty position will have to be sacrificed, and they are like hen's teeth around universities. In many cases the people making these kinds of decisions have not been swept along by the tides of change, and don't feel the need as strongly.

I recently conducted a mini-survey of departments who had managed in some form or another to create such a position. Create is the word: in a large number of cases the University and department together imaginatively conjured up a way to fund the positions from grants, department funds and campus wide resources.

Unfortunately for us (and fortunately for the future gen-

erations) we are the pioneers in this endeavor. Soon these positions will be an obvious requirement for any modern statistics department, given the nature of the field. But in the meantime, don't give up the battle!

With that fiery flame I conclude my final newsletter article as chairman of the statistical computing section — it has been a pleasure serving you.

Trevor Hastie
Chair, Statistical Computing Section
`trevor@playfair.stanford.edu`

Student Paper Competition

The Statistical Computing Section of the ASA will sponsor a Student Paper Session next year at the Orlando Joint Statistical Meetings in 1995. The topic of the session will (naturally) be *Statistical Computing*. Three to four students will be selected to participate in this session, which will include a discussant nominated by the selection committee. Fees associated with registration, accommodation and travel to the conference will be awarded to the participants in this Session.

Call for Papers

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 1994. 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 authorship, the mentor should indicate what fraction of the contribution is attributable to the applicant.

All application materials MUST BE RECEIVED by January 10, 1995. 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 23, 1995. 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 the following:

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

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

All electronic submissions of papers should be in postscript.

Statistical Computing Section

Minutes of the Monday, August 15, 1994 Section Executive Committee Meeting, Toronto, CANADA

Attendance

Trevor Hastie, Chair
Mary Ellen Bock, Chair-Elect
James L. Rosenberger, Newsletter Editor
Tom Devlin, Continuing Education Coordinator
Deborah Swayne, Secretary-Treasurer
Karen Kafadar, Publications Liaison Officer
John Rice, Program Chair-Elect
Ron Thisted, Section Representative 1994-96
Russell Lenth, Council of Sections Representative

Rick Becker, Graphics Section
Roy Welsch, Graphics Section
Mike Meyer, Graphics Section and Newsletter Editor
Terry Therneau, 1995 Secretary-Treasurer

Elected to new offices for 1995

Chair Elect: Sallie Keller-McNulty
Program Chair-Elect: Robert Tibshirani
Secretary/Treasurer: Terry Therneau
Council of Sections Reps: MaryAnn Hill (1995-97), Michael Meyer (1995-96)

Assuming new offices in 1995

Chair: Mary Ellen Bock
Program Chair: John Rice

Previous minutes The minutes of the June, 1994 meeting, presented by Deborah Swayne, were approved.

Membership as of 8/4/94

Stat Computing Section alone:	761
Stat Graphics Section alone:	171
Joint Computing and Graphics:	2362
Total Computing:	3123

Treasury as of 6/30/94

Section Cash	$36,450.31
Joint Projects Cash	$8,253.29

Student Paper Competition Trevor handed out brochures announcing the student paper competition.

Council of Sections Meeting Ron Thisted and Russell Lenth reported on the Council of Sections meeting that had just taken place.

ASA plans to hire a consultant on organizations to suggest changes in the structure of ASA, with the particular goals in mind of improving outreach and member services.

There was also a lot of discussion about certification. The membership was recently polled on certification issues. Approximately 5000 members responded, and they were roughly equally divided on the subject—but each section is not equally divided. The section on quality and productivity is the most strongly in favor of certification, and it is possible that the ASQC may go ahead and certify statisticians whatever the ASA does. A resolution was passed and then tabled at the board meeting, saying that the ASA will have discussions with other societies about certification.

The Council of Sections representatives were asked to bring feedback from section meetings to a second meeting to take place toward the end of the joint meetings. We agreed to send back a statement saying that we en-

courage sections to work with other societies on *their* certification issues, taking in effect a consulting role.

Continuing Education Tom Devlin reported that the section's continuing education courses were very successful at the Interface conference this year, and he's interested in some suggestions: At what other meetings might we sponsor courses? What new courses could we offer? What new avenues could we use to find more speakers and topics?

John Rice suggested a topic: non-linear random effects models are of great interest in the pharmaceutical industry. Two avenues were mentioned as well: Could we somehow make use of the world wide web? Shall we solicit suggestions and volunteers in Amstat News?

CE: Cancelling courses Ron Thisted said that there had just been a discussion about continuing education at the Council of Sections meeting. Many short courses were cancelled by ASA last year because enrollment was low (below 25). This caused much unhappiness, and a decision was just made that if enrollment falls below some threshold, the sponsoring section can choose whether to cancel the course or to split any financial losses with ASA.

CE: Student scholarship program Tom continued with a proposal to extend the section's support of students who attend our continuing education courses at either the Interface or at the joint meetings. We approved a proposal to pay for the conference registration, as well as the course registration, for up to twenty students per year.

CE: Additional support for course presenters We next discussed proposals for expanding financial support for presenters of continuing education courses at the joint meetings. Currently, speakers receive from the ASA a $500 honorarium and up to $1000 to cover travel and lodging costs. (This money is shared among all presenters.) We approved a proposal to give some additional money to presenters, to help them defray necessary expenses: $250 for each speaker up to three speakers. Despite describing the money as expense money and not an honorarium, we will not ask for receipts. (Note that this funding applies only to the joint statistical meetings and not, as yet, to the Interface conference.)

Both of these plans—the student scholarship program and the additional support for course presenters—will be announced both in the newsletter and in Amstat News.

Newsletter editors Two new newsletter editors should be identified by the end of the year. It was also agreed

that some arrangement should be made so that the terms begin to be staggered instead of beginning and ending in unison; for example, one of the new editors could have a two-year term and the other a three-year term.

Term of Secretary-Treasurer We clarified the previous decision to extend the term of the secretary-treasurer to two years in this way: A new secretary-treasurer will still be elected every year, and each one will stick around for two years, acting as Past Secretary-Treasurer during the second year. (Note: check by-laws)

Internet connection for the ASA Mike Meyer commented that electronic communication with ASA officers is quite difficult, and volunteered to investigate getting them a better internet connection. He suggested that we consider paying for the better connection for a year in order to encourage them to proceed. We agreed to consider it once he had investegated the issue.

> Deborah Swayne
> *Sec/Treas, Statistical Computing Section*
> Bell Communications Research
> `dfs@bellcore.com`

⟨O⟩

BUSINESS MEETING MINUTES

Statistical Graphics Section

Minutes of the Monday, August 15, 1994 Section Executive Committee Meeting, Toronto, CANADA

Attendance

Rick Becker, Linda Clark, Bill DuMouchel, Jane Gentleman, Colin Goodall, Dave Hoaglin, Sallie Keller-McNulty Robin Lock, Mike Meyer, Barbara Milton, Sally Morton, Bob Newcomb, David Scott, Roy Welsch

Election Results

Roy reported the election results for 1995:
Chair elect: Bill DuMouchel
Program chair elect: Steve Eick
Secretary/treasurer: Bob Newcomb
Publications Officer: Deborah Donnell
COS Representative: Colin Goodall

Committee Reports. Bill Tucker, Vice Chair of the Council of Sections gave a brief report on the function of the COS, and urged members to interact with/give feedback to our COS representatives.

Data Analysis Exposition. Robin Lock will be taking this over for next year. As of yet there is no dataset for next year's competition. Since it is already late in the year, we need to start publicizing this asking for suggestions. Notes should go in AMSTAT News and the computing/graphics newsletter. David Coleman has done this in the past.

Poster Competition. Jerry Moreno attended the meeting to give a report on the activities of this committee. He gave an oral report and also had a written version, which is available for further detail. Some issues that we discussed were:

Roy thanked Jerry on behalf of the section for all of his efforts on this project. It has been, and continues to be, a great success.

Last year the executive committee approved John Schollenberger to replace Jerry, who wishes to step down.

Mike Meyer will include Jerry's report in the next newsletter, along with pictures of this year's winners.

After some discussion of the $3 entry fee, and whether it was prohibitive to some, the executive committee agreed to leave the fee as it is and generally give the committee freedom to do things on their own.

Committee to Nominate Fellows. Lorraine Denby becomes the third member of this committee along with Larry Cox and Dan Carr. Andreas Buja, nominated by this section, was made a Fellow this year.

Dynamic Graphics Video. Sally Morton reported on the activity of the video lending library. There were three new videos submitted this year. She is currently arranging for an Australian university to copy the videos into PAL format. Some issues raised were:

She is looking for suggestions on making a history video.

She's also looking for suggestions on how to advertise the video library.

Sally would like to step down from this committee in about a year, so we need to find a replacement.

Continuing Education. Karen Kafadar has agreed to replace Abbe Hertzig on this committee. We are looking for suggestions of courses the Section could sponsor. One idea mentioned was to do something at SIG-GRAPH.

Charter. Dave Hoaglin reported on the proposed changes to the charter of the Section, and had copies available. The changes are modest and will be voted on by Section members next Spring.

Newsletter. Mike Meyer reported on the status of the Computing/Graphics Newsletter. A replacement for Mike will need to be found by the end of next year. They have made the last two years' editions of the newsletter available on Statlib and the World Wide Web.

Treasurer's Report. Linda Clark reported the following amounts in the Section accounts:

Graphics Section account $33,881.30
Joint computing/graphics account $ 8,253.29

There was some discussion on whether we should continue to solicit funds for the mixer from software vendors. Also, the questions was raised as to how to decide the amount of money to budget for the mixer. Roy will look into this.

Program reports. Sallie McNulty reported on two tentative invited sessions for 1995. She also said that we are at risk of losing sessions because of low attendance at graphics section talks, and also because the number of contributed papers has decreased. Sallie also commented on the need to have better equipment available for speakers at our sessions. Good graphics talks often need more than just an overhead projector, and we should look into the availability and cost of additional equipment.

Report of the council of sections representatives. Sally Morton and Jane Gentleman reported on the COS meeting held earlier in the day.

Here are some highlights:

Barbara Bailar has resigned effective Jan 15, 1995.

The idea of a second senior staff position being added has been postponed until further study by a consultant. The ASA board wants member feedback to Mitch Gail by early November and is circulating questions to guide input.

Results of the survey on certification show opinions are very split. The board put forth a motion(tabled until Dec) to discuss the issue with other societies (ASQC, IEEE).

Continuing education: course cancelation policy will depend on 1) minimum enrollment or 2) the section taking on the loss if enrollment is too low. Currently the section makes $2,000 if there are 35 or more enrolled. We could subsidize enrollment by giving scholarships to encourage participation or buying slots to keep a course form being canceled.

JCGS graphics paper prize committee. Roy will ask some people to serve on this committee.

Expenditures for outside speakers. There was some discussion about what the policy should be on paying speakers. Bill DuMouchel suggested we give the program chair money to 1) find outside speakers (whether they need to be ASA members would be up to the discretion of the chair) 2) promote section activities. The program chair will be informed that they can request funds for these purposes. The executive committee did vote to approve funds for this purpose.

New section activities. Some suggestions on how to spend section funds:

Mike Meyer - 1) offer a free one year membership to the section 2) connection to the Internet for the ASA (possibly jointly with computing). Bill DuMouchel - 1) money to help support the poster competition 2) help on the poster competition brochures

Linda Clark
Sec/Treas, Statistical Graphics Section
AT&T Bell Laboratories
`lac@research.att.com`

ⓒ

MIXER NOTES

Mixer at the Joint Meetings

The joint business meeting and mixer at the joint statistical meetings of the ASA in Toronto, August 1994, followed the traditional pattern. A short reporting session from the section chairs, Trevor Hastie for Statistical Computing and Roy Welsch for Statistical Graphics, brought to the members in attendance news from the respective executive meetings held earlier in the day. Following the reporting and announcements, the mixer began with a random drawing for door prizes, with the winners receiving various books and materials donated by the vendors at the meeting.

Following is a list of publishers and software companies who donated door prizes for the mixer:

- Academic Press, Inc.
- Addison-Wesley Publishing Co
- Chapman & Hall
- Duxbury Press
- Irwin Publishing
- John Wiley & Sons
- Marcel Dekker, Inc
- Oxford University Press

- Prentice-Hall and MacMillan
- Springer-Verlag, Inc.
- W.H. Freeman & Co
- SAS Institute, Inc
- BMDP Statistical Software, Inc

Software companies who made cash donations for the mixer are following:

- BMDP Statistical Software, Inc.
- SAS Institute
- SPSS, Inc.
- Stat. Graphics Corp
- Systat, Inc.
- Cytel Software Corp
- Resampling Stats

We extend our sincere thanks to these companies for helping to make the mixer a success!

Debby and Linda
Sec/Treas, Computing and Graphics

ⓒ

Statistical Computing

by Thomas F. Devlin
Statistical Computing CE Chair

Short Course Proposals Sought

Proposals for short courses on topics in statistical computing or statistical graphics are invited for presentation at the Interface '95 in Pittsburgh, on June 21. This year's themes include graphics, manufacturing and the environment. If you are interested in presenting a 2 hour or 4 hour course, please send a proposal immediately, including topic and course outline, to the address below (email preferred).

You may also be interested in presenting a course at another professional society meeting. Courses may be either one-day (6 hours of instruction) or half-day (3 hours) in length. They may be based on a new text, however 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.

Proposals, inquiries and suggestions should be sent to:

Thomas F. Devlin
CE Chair, Statistical Computing Section
Mathematics & Computer Science Dept.
Montclair State University
U. Montclair, NJ 07043
Voice: 201-655-7244
`devlin@mozart.montclair.edu`

ⓒ

Graphics Video library

In the August 1993 issue of this newsletter (Vol 4, No. 2, pp. 7–8) we published an article about the Statistical Graphics video library, and listed the titles available. Here we present some updated information about the library and the new titles. For the full list of titles we refer you to the earlier article.

Information about the library is available via e-mail from `Sally_Morton@rand.org`.

New Submissions

We have had three new submissions to the library since August 1993:

Edge Information at Landmarks in Medical Images Fred L. Bookstein and William D. K. Green, University of Michigan Medical Center (1993). 26:15 minutes run time.

Grand Tour and Projection Pursuit Dianne Cook, Andreas Buja, Javier Cabrera, and Deborah F. Swayne, Bell Communications Research (1993). 19:00 minutes run time.

Exploring Time Series Using Interactive Graphics Andrew McDougall and Dianne Cook, Rutgers University (1994). 16:00 minutes run time.

The Dynamic Graphics Video Theater at the 1994 ASA Meetings ran from 8:30am to 6:34pm, and included all library videos in chronological order. The ten minute 'Tokyo Data Map' was shown continuously from 6:04pm to 6:34pm to finish the Video Theater.

Borrowers and PAL Satellite Library Branch

We have approximately ten borrowers per year, mostly from the U.S. with some interest from Korea. At present, we are arranging for an Australian University to copy all the videos, or perhaps just those post-1987, into PAL format. The University is willing to act as our satellite library branch, lending out the PAL videos to other PAL borrowers from Australia or New Zealand for example.

New Directions

The Dynamics Graphics Video Committee welcomes comments from Computing and Graphics section members regarding the video library. If you have suggestions on how to improve the library, please send them to Sally_Morton@rand.org.

Rules and Contents for the Library

The Video Lending Library of the ASA Statistical Graphics Section.

The library is currently composed of the material that has appeared in the ASA Dynamic Graphics Video Theater, but over time will pick up new contributions. The library also serves as an archive where the material will be preserved in a central and accessible place. The library contains 1/2 inch VHS video tapes and, for selected tapes, an abstract, related publications and technical reports—even comments from previous viewers.

The video tapes are great for educational institutions: several people have used them as the basis for a "stat day" production, for statistical computing classes, or even for a video lunch complete with popcorn.

Lending policies for the library are:

- Loans up to 3 weeks are permitted
- A limit of 3 videos at one time is permitted (with exceptions for statistics day presentations where more videos will be available for a shorter period)
- The borrower will be responsible for including postage payment along with the request for videos.
- Borrowers who are not members of the Graphics Section will also be asked to pay a lending fee of $10; borrowing from the video library is one of the many services that members receive in return for their section dues.

Authors of the video tapes that are distributed as part of the lending library have given permission for the library to lend the tapes but have NOT authorized extra copying. After viewing a tape, should you wish to acquire a personal copy, you should contact the author.

To find out more about the video library, or if you have your own video tape that you think should be part of the video library, please contact Sally.

Sally Morton
Video Librarian, Statistical Graphics
310-393-0411 x7360
The Rand Corporation
Sally_Morton@rand.org

$\textcircled{0}$

FEATURE ARTICLE (Cont.)...

The WWW is a mechanism over the Internet by which multimedia information can be disseminated from site to site. Through the use of client programs such as the National Center for Supercomputing Application's (NCSA) Mosaic and Netscape, information, including documents, sounds, images, and video clips, can be transferred between machines world-wide. Information is requested automatically from thousands of server machines on the network through the use of the client program Mosaic. The server machine informs the client what type of information is requested, i.e. whether it is text, pictures, video, hyper-text, or audio, and the client displays or presents the information in the appropriate form using the appropriate viewer program, such as xdvi for dvi files, ghostview for postscript files, xv for gif, tiff, and jpeg graphics files, netaudio for audio files, etc.

Statistical data can also be viewed as a type of information, and therefore sent through the Internet and displayed with a "viewer". For example, one can retrieve data using FTP or Mosaic from sites such as Statlib, the Census bureau, or the NIH. The natural way to "view" this information is with a statistical package! The package provides the means of summarizing and displaying the data for the user to understand.

Using the concept of duality, statistical procedures such as linear regression or robust estimation can also be viewed as information—instead of the server supplying the data, imagine instead the server supplying the procedures. Statistical procedures can thus be downloaded and illustrated on data, as opposed to data being downloaded and viewed on the procedures. Interactive statistical packages therefore become the tool which allows the data and procedures to interact and therefore

can be considered "viewer" programs for this information.

Teaching Paradigm

Interactive statistical packages and interesting datasets provide a rich combination for teaching and learning statistics. Just as one interacts with a postscript or dvi viewer to change pages, resize the view, and act on the visualization of the document, statistical packages can be thought of as a means of acting on the visualization ("view") of the data and the elucidation of statistical procedures.

Hypertext documents are designed for user interaction. In the context of learning statistics, the student can view demos, cut and paste sample commands into the program, try out suggestions, quickly find the definitions for new words, and follow up on related concepts. For example, in HTML, help for concepts and keywords can be linked inside the document, so that unlike a textbook, simply clicking on the word allows the user to easily find definitions for unfamiliar concepts. In addition, just as one might follow a link from the document to a definition, a picture, movie, or sound-byte, one can link data for an assignment as a similar piece of information. Instead of viewing a picture, the data is "viewed" or elucidated, using statistical procedures such as EDA, estimation, or hypothesis testing. The student can load the data and procedures as needed and will save or store information only if it is useful. Changes to data sets and updates to procedures and macros can be distributed this way. An example of this is the WWW interface to Statlib.

This facility gives students online help for statistics programs and related packages. Existing documentation is easily converted into HTML, the hyper-text markup language. This is relatively trivial in the case of plain ASCII text and LaTeX documents for which an automatic translation program exists. In addition, there is a wealth of documentation which has already been converted.

One motivation for this work stems from a need to supply students with reference information about statistics and statistical computer packages. The three applications which the first author is currently developing are instructions and aids for statistical packages (Xlisp-Stat and Minitab), supplemental materials such as course handouts or additional notes, and finally additional exercises which are optional and might assist students in performing the required exercises. These materials have been designed on the assumption that the course has an accessible instructor, though these materials are especially suited to distance learning. An entire course could be taught over the Internet—one model is the *Introduction to Object Oriented Programming using C++*, available through the WWW , (connect to these materials using `http://uu-gna.mit.edu:8001/uu-gna/text/cc/index.html`).

Mean versus Median Example

Consider a tutorial on describing differences between the mean and the median to describe the center of a distribution. A tutorial document discusses the mean and median for continuous data, which is not different than standard class notes or handouts. However, we can embed hypertext links to separate documents which focus specifically on the mean and the median to help remind the students of the definitions and properties if necessary. Exhibits can also be used to illustrate the point, just like a textbook. However, unlike a textbook, a demonstration program is linked in, which the student can use to explore how changes in a single data value influences the sample mean and median. This demonstration can use either text-based or graphical input for interactively changing the values of data points one at a time.

Example on Line

In fact, this tutorial or lab session is available online. To try it out, use a WWW client like Mosaic to connect to `http://www.stat.psu.edu/statlearn/` and proceed to the *center of distributions* document. The section below provides more details for the new user.

Detailed Demonstration Instructions

To try out the demonstration, three main steps must be taken.

1. Retrieve and Install an HTML viewer such as NCSA Mosaic.
2. Retrieve and Install Xlisp-Stat on your machine.
3. Retrieve a copy of the modified mailcap and Xlisp-Stat viewer script.

The first step has been documented in a number of other places and so will not be repeated here (See e.g. Meyer, M.M. "Mosaic and WWW", *Statistical Computing and Statistical Graphics Newsletter*, 4(3):19-20, 1993) After the first step has been taken, connect to the document `http://www.stat.psu.edu/statlearn/` for further instructions for steps 2 and 3. The second step is also documented well in Jan de Leeuw's article on page 16 of this Newsletter. Pointers for retrieving Xlisp-Stat source, binaries, the modified mailcap file, and the Xlisp-Stat viewer script are provided. In addition, links to connect the tutorial session on Mean versus the Median are given.

Statistical Information as Objects for Viewing

Implementing procedures for displaying statistical information can take two approaches, referred to as server-side and client-side, depending on where the statistical package resides.

For the *server-side* approach, the server computer does all of the computations, and the client only needs to be able to view the resulting text, graphics, movies and sound that are generated from the client's request. This could require a complex user interface in order to handle the general problem. Of course packages are available which aid in the design of the HTML forms used for this interface. The basic idea is that the forms used as the interface for generating the statistical macro will produce the desired results from the statistical package. The server machine handles the statistical computation, formats it as a document, and returns the document to the user. This resulting document could include links to other text, graphics, and perhaps audio or video clip results which are pertinent replies to the statistical query.

The second approach, referred to as the *client-side* approach, is to have the client computer run the statistical program locally, and have the serving machine send any special data and statistical procedures required to complete the exercise. This has the advantage of minimizing the load on the server. However, it requires the client to **trust** the server, since the server will have some access to the client's machine; this is because most statistical packages allow one to escape to the shell (or MS-DOS, in the case of IBMs and compatibles), and there are potential security problems. For example, one could download a virus or a malicious shell script, and care must be taken to guard against this!

The problems of the first method include a large load placed at a central site and the difficulties of constructing a general interface. Software licensing might also be a big problem! However, the second method is not without its faults, as it both produces potential security problems and needs technical details to be worked out so that the statistical package does not have to be restarted for each piece of information viewed. The latter is a current project.

Implementation

We choose to use Xlisp-Stat as the basis for the client-side, and SAS for the server-side implementation. Xlisp-Stat is free and is available for most common platforms on the Internet such as Unix workstations, MS-Windows, and the Macintosh. In addition, there are extensive help files which are available for converting to HTML, and can be used to learn Xlisp-Stat !

Client Side

The HTML specifications allow one to specify file types identified by suffixes, using the MIME specification for multi-media mail. This is done at two points; the server must acknowledge and inform the client of the type of document being transfered (plaintext, HTML , postscript, etc) and the client must know what should be done with the document. The server side description is done by specifying a new type in the configuration file (different for different http servers) and the client-side configuration is done by specifying a viewer for "statistical information" files through the system or user's *.mailcap* file (on a Unix system). The final necessary tool is a script which runs Xlisp-Stat inside an xterm window (a terminal window in X11 which contains an interactive shell, like a dumb terminal), to be the *Statistical Viewer*.

The user starts up the client and opens the server which contains the course notes or other information to be browsed. At this point, the procedure is similar to reading a book, except that the experience is easily made non-linear.

Server Side

For the server-side mechanism, scripts must be written to take the input from the forms interface and produce a batch file suitable for the statistical package. Then the script must reformat the output from the statistical package. The demonstration uses SAS, with an interface developed for the script language TCL. This is based on work done at the Department of Medicine, Section for Artificial Organs, at the Hershey Medicine College at Penn State. These scripts were originally designed to construct an easy interface to SAS for residents who were interested in graphically examining a body of medical experiment data. For the server-side construct, contact the first author for sample scripts written in TCL for the SAS system.

Discussion

In conclusion, the Internet and the World Wide Web can make a fine supplement to teaching a course. There is much useful information out there, and a little time spent on collecting it can save instruction time.

Many people are starting to produce documents; the one advantage is that in general, others can use the material for their courses, so that people do not have to continually re-invent the whole wheel. There is work on this currently being done at UCLA and U Minnesota, as well as both in the Penn State Statistics and Biostatistics departments, and probably others are developing materials that we are not familiar with.

The paradigm of being able to treat statistical procedures and data as viewable information is a useful one. One extension which is immediately possible based on this work would be in the possibilities of multi-media e-mail. One could conceivably discuss a proposed analysis through e-mail, with the relevant information passed back and forth.

There are further mechanisms, not examined here, which can be extremely useful for distance learning and the spread of knowledge. For example, there is a mechanism, MBONE, similar to usenet news and WWW , for transmitting audio and video (movies) over the Internet. For the course on C++ programming which was mentioned, students communicated with the teaching assistants through email, "talk", and the Internet Relay Chat (the latter two are means for real-time communication by typing). Other mechanisms include MUDs and MOOs (Multi-User Dungeons and MUDs, Object Oriented), which are similar to the old text adventure games, except that multiple players are involved, and objects can be left by different users for each other. For visual communication, there is the *CuSeeMe* package, which transmits audio and visual between two workstations, almost like a visual phone call. In this manner, a virtual classroom can be constructed.

Appendix: Technical Details for UNIX systems

In order to use the server software developed at Penn State, the client must be able to translate the document. For this purpose, a new information type must be defined for Mosaic, application/x-stat-xlisp, and the viewer needs to be specified. The viewer is called *xlispstat.wrap* to denote that it is a wrapper around Xlisp-Stat , i.e. a program which makes sure xlispstat starts up properly. With this installed, the following information should be put into a global *mailcap* file or into a user's personal *mailcap* file (i.e. *.mailcap* in the user's home directory).

```
# Statistical information
application/x-stat-xlisp; xlispstat.wrap %s
```

A sample *mailcap* file and the script for the wrapper are available via anonymous ftp from `ftp.stat.psu.edu` in the file `/pub/statlearn/xlispclient.shar`.

The other side of the story is that the server must know that files ending in `.lsp` are xlispstat files. This is usually declared in the configuration file of the http server (the program that serves hypertext documents to clients). This requires that the configuration files be modified. An example of the code to be added for the NCSA HTTPD server is as follows:

```
# Added for statistical information,
# AJR (091694).  EXPERIMENTAL!
AddType application/x-stat-xlisp  lsp
```

These lines must by placed into the `srm.conf` configuration file. This would vary for different HTTPD servers.

One could do a similar thing for other statistical packages; while this transfers easily to Splus and Minitab, it also could be set up for SAS or other generally batch systems. For Splus or Minitab, it would suffice to choose a local suffix—perhaps `.q` for Splus and `.mtb` for Minitab. Then one could construct a universally accepted MIME type such as application/x-stat-splus and application/x-stat-mtb and write the respective interfaces.

Anthony J. Rossini
Center for Biostatistics and Epidemiology
Hershey Medical Center at Penn State
`arossini@biostats.hmc.psu.edu`

James L. Rosenberger
Department of Statistics
Penn State University
`JLR@stat.psu.edu`

⬭

SECOND FEATURE

The Lisp-Stat Statistical Environment

by Jan de Leeuw

A Question and an Answer

Name a statistical package which

- is infinitely extensible, open, and comes with an elegant object-oriented extension language;
- is fast, use garbage collecting, and has a byte-compiler for additional speed;
- is interactive, dynamical, and graphical, and allows you to build to own graphical interfaces;
- is portable, and uses the native window system on Mac, MSW, X11;
- is free.

After this blatant commercial beginning, everybody is waiting for an answer. Here it is. **Lisp-Stat !!**

Some History

Lisp-Stat is written in a combination of C and Lisp, just like GNU Emacs. The basic language used to write the interpreter and the build-in functions is C, and Lisp is used as the extension language of the system.

We shall not discuss the complicated history of the Lisp language. If you want to know more, you can read

`~/common/docs/Evolution-of-Lisp.ps`

Note: In the sequel we shall refer to various documents on our ftp server. The root for paths we give are `ftp://ftp.stat.ucla.edu/pub/lisp`.

For our purposes it is sufficient to know that in the early eighties Lisp systems were huge and expensive. Therefore David M. Betz decided in 1985 to write a small and portable interpreter for a subset of the Lisp language. PC and Mac users could run their own Lisp systems by using his Xlisp, and they could experiment with object-oriented programming as well. The interpreter was written in C. Xlisp 2.0, released by Betz in 1988, has many additional functions and data types build in, and the newer releases which are based on the work of Tom Almy, Niels Mayer, and Luke Tierney are getting closer and closer to Common Lisp. Documentation is in `/xlisp/docs`. There is a reference manual (for Xlisp 2.1) by Tim Mikkelsen, and introduction to Object-Oriented Programming (OOP) by Tim Mikkelsen, and a little manual on how to extend Xlisp at compile time by adding additional C functions by Jeff Prothero.

In 1988 Luke Tierney published the first version of Lisp-Stat, which was documented in the technical report `/xlisp/xlisp-stat/docs/xlispstatdoc.ps` Tierney took Xlisp 2.0 as his starting point, and added functions for linear algebra, optimization, statistics, and graphics programming. Some of these functions are added at the C level (i.e. at compile time) and some at the Lisp level (i.e. at run time). Then in 1990 he published the book *Lisp-Stat. An Object-Oriented Environment for Statistical Computing and Dynamic Graphics*. Unfortunately Xlisp 2.1g, which is used in the newest versions of XLisp-Stat, is so much larger than Xlisp 2.0 that the account of the system given by Tierney in his book is quite incomplete. An update to the book is in `/xlisp/xlisp-stat/docs/changes.ps`.

The various names illustrate, by the way, the relationship between Lisp-Stat and Xlisp-Stat. Any Lisp system with the statistics and graphics added is a Lisp-Stat system, but adding the Stat component to Xlisp gives Xlisp-Stat. There is an (incomplete) version of Common Lisp Stat, and we can easily image Elisp-Stat or Scheme-Stat.

Some Comparisons

We have called Lisp-Stat a statistical package. This is somewhat misleading, because there are so many different types of packages, and Lisp-Stat is different from almost all of them. It is not similar to the dinosaurs SAS or SPSS, which apply relatively routine statistics to huge data sets, with gigantic amounts of overhead. It is also not similar to the many PC/Mac packages that do more or less the same thing as SAS and SPSS, only on a somewhat smaller scale, and with a little bit more graphics. I think the only useful comparisons are with Matlab, Mathematica, DataDesk, and in particular with S-plus.

A fairly extensive comparison between Xlisp-Stat and S is contained in a collective review of List-Stat in the November 1991 issue of Statistical Science. The issue has reviews by Baxter and Cameron, Weihs, Young, and Lubinsky. Especially Lubinsky makes fairly detailed comparisons with S. While S, especially in its S-plus implementation, is more complete and glossy, Xlisp-Stat is generally faster and has (much) better memory management. The speed advantage is more pronounced now than in 1991, because Xlisp-Stat has a byte-compiler in its newer versions, and because of its portability, it has been rapidly set up on newer and faster systems. S has more statistics, Xlisp-Stat has more graphics. S is perhaps more appropriate for doing statistics, Xlisp-Stat is more appropriate for producing statistics software, because it comes with a complete GUI builder. S is better for those who write papers in which statistics are applied, Xlisp-Stat may be better for those developing statistics.

Another important comparison is the language. Xlisp-Stat uses Lisp, a general purpose language with many applications. S uses the S-language, a special purpose language geared to statistics. I happen to think that if you work in a UNIX environment, then learning Lisp is a good investment (think of Emacs, Maxima, and so on). In fact, some free Common Lisp systems (such as cmucl) are now available that can even do fast floating point, and code for these systems is often very close to Xlisp code.

What it comes with

Xlisp-Stat comes with everything in Xlisp-2.1g, including many of the functions in the various Lisp standards. In fact I would guess (but I am not sure) that the system contains well over 500 Lisp functions not related to statistics. This provides an extraordinary powerful basis for program development. Many of the functions are completely undocumented in the Tierney book, and even in the Mikkelsen reference manual. For instance, Xlisp now contains the system package from Common Lisp, with the many functions that come with it. Xlisp now can also save workspaces, similar to APL

workspaces, which makes it easy to customize and save work between sessions.

Another important addition to current versions of Xlisp-Stat is the byte compiler. The lisp code can be byte-compiled, and byte-compiled code generally loads and executes much faster than the original lisp. For code with lots of loops and lots of consing, using the byte compiler can mean speed ups of 500% or more.

Of course all the basic arithmetic and transcendental functions are there, including the beta and the gamma. Descriptive statistics such a mean, median, standard-deviation, quartile, max, min, and fivenum are available. The numerical linear algebra functions are the usual, with inverse, determinant, solve, eigen, sweep, qr-decomposition, cholesky-decomposition. Arrays with unlimited dimensionality are native in Lisp, and there are various primitives to handle indexing (this has been extended greatly with add-on Lisp code, see below). There are two general purpose optimizers, one based on Newton's method and one based on the Nelder-Mead simplex method, and there is a somewhat idiosyncratic but very useful (they tell me) module to approximate the posterior mode using Laplace expansions.

A nice generalized linear modeling package is added at the Lisp run-time level. It is similar to the nonlinear regression package because it inherits much of the functionality of the regression model prototype. Regression is one of the areas in which Lisp-Stat really shines. It forcefully demonstrates the elegance of an object-oriented organization, and combination of the regression and graphical facilities has lead to many great tools. (Note: The OOP system in Xlisp-Stat is prototype-based, while the OOP in Xlisp is class-based. This means that actually two systems are available, but Xlisp-Stat users will routinely use the prototype-based one.) The recent *Introduction to Regression Graphics* by Cook and Weisberg is an impressive example of the power of these modules. It is a clear illustration that Xlisp-Stat is ideally suited to develop modules for statistics teaching, modules that can be used by students who do not have to learn a single parenthesis of Lisp.

The uniqueness of Xlisp-Stat are its graphical interface building facilities. It will take some climbing of a relatively steep learning curve to use this effectively, but once you are at the top (for somebody who has many other things to do this will take at least a year, for others it will take a couple of months and the usual amounts of coffee) an amazing tool chest is available.

There are also some weak points, which are partly unavoidable and partly easy to repair. Interpreters running their own shell are relatively closed systems, and consequently communication with the outside world (input, output) is somewhat of a problem. Reading data into the system is slow and tedious. This could be improved a great deal by moving some of the input facilities from the Lisp to the C level. There is no spreadsheet data editor, although it would be possible to add one. The help system is rather primitive. And there are no facilities to make publication quality plots. Some of this has been taken care of in the Lisp add-ons discussed below, but the basic facilities should really be part of the core system (which means part of the C code).

For many users it may be a nuisance that generalized additive models, cluster analysis, and many other advanced techniques which are in S-plus or SAS are not in Xlisp-Stat. Again, this merely reflects the fact that the systems have different objectives. SAS is for production, S-plus is for research and production, Xlisp-Stat is for research and teaching. It also is a self-fulfilling prophecy. As the sections below show, extensions can be easily added to Xlisp-Stat at various levels. C and FORTRAN subroutines can be loaded at runtime, and linked dynamically into the system. This makes complete libraries of techniques available with a minimal effort. But somebody has to make the effort. Tierney is maintaining and upgrading the core system, which is a lot of work. Extensions have to come from the users. Once again I will mention GNU Emacs (or tcl/tk or X11) to illustrate that if a free and open tool is adopted by a user community, while the core is maintained by experts, we can create a tool which commercial companies cannot possibly compete with.

Why I love it

In the first place because of the GUI building tools. Second, because of the Lisp language (but that's just taste). Third, because it is free. This may look trivial if you have money, but it means that we can give it to our students – not only to the graduate students, but also to the undergraduates. They can use it not only on our $ 10K elitist workstations, but also on their lowly 386 or Performa machines at home. And fourth, it is open. We can see exactly what it does, because we have the source. We can extend it at compile-time, at link-time, and at run-time level. If we get the latest greatest UNIX box, or a Pentium, or a PowerMac, we recompile to use the power of the new machine. It is easy, the system is eminently portable. Because of these four reasons, I personally don't even think of S or S-plus as an alternative.

How to get it

The complete Xlisp-Stat code distribution can be found in

`~/xlisp/xlisp-stat/dist/`

and binaries for various types of machines are in

`~/xlisp/xlisp-stat/systems/`

There is also quite a bit of Xlisp-Stat code on `statlib` in the directory `xlispstat`. (Note: By the way, `ftp.stat.ucla.edu` is no competitor of `statlib`, it is more specialized and may be more complete in this area.)

We try to keep a collection of everything connected with Xlisp-Stat in the ftp directories. There is always a version of the source code, and there are binaries for selected systems. We actually mirror the `pub/xlispstat` directory from `umnstat.stat.umn.edu`. As the path to the directory shows, there are other lisp systems besides xlisp, and other xlisp extensions besides xlisp-stat. Our server can also be reached via `gopher.stat.ucla.edu` or `www.stat.ucla.edu`.

Code

The most interesting sub-directories on our server are

`~/xlisp/xlisp-stat/code/contrib/`

and

`~/xlisp/xlisp-stat/code/homegrown/`.

In `homegrown` we have Xlisp-Stat code written at UCLA. This includes:

- the SIR package by Ker-Chau Li and his students, for Slicing Inverse Regression;
- the Terrace package by James Hilden-Minton for Hierarchical Linear Modeling with diagnostics;
- the Gifi-gardens package by James Hilden-Minton for MVA with optimal scaling;
- many utilities by Jan de Leeuw for monotone regression, b-splines, the basic APL array operators, multidimensional scaling, nonlinear MVA, FORTRAN-type formatted input, log-linear modeling;
- stem-and-leaf-plots by David A. Betz;
- bootstrap/jackknife algorithms by Jan de Leeuw. This includes bootstrap-t, percentile, abc, bca confidence intervals, bootstrap and jackknife standard error, bootstrap and jackknife bias estimation, bootstrap prediction error estimation.

The bootstrap algorithms for regression are written as methods for regression-model-proto. There is also a graphical bootstrap demo. And there is code for permutation and bootstrap hypothesis testing. There are also versions of Randy Sitter's algorithms for bootstrapping in survey situations with clustered and stratified random samples (without replacement).

- Didactic Modules by Jan de Leeuw and Jason Bond. Moving the regression line, moving points in regression scatterplots, the central limit theorem.
- UCLAData. Evolving package for Introductory Data Analysis teaching by Jan de Leeuw.
- Help. Evolving graphical help system for Xlisp programmers and Users by Jan de Leeuw and Jason Bond.

In `contrib` we have xlisp-stat code written by others. This includes everything on statlib, and then some. We mention

- axis: Introductory statistics package by Robert Stine (Wharton).
- bpois: Bayesian Poisson Regression using the Gibbs Sampler by Hani Doss (OSU) and B. Narasimhan (PSU).
- cw: Dynamic Graphics for Regression, by Dennis Cook and Sandy Weisberg (Minn). Called dyndiag on statlib. This has been superseded by the much-improved code that comes with their recent book.
- eltoy: Introductory Statistics and Bayesian Elicitation, by Russell Almond (U Wash).
- expsurv: Exploratory Survival Analysis by K. Neely Atkinson (U Texas).
- hasse: F tests and ANOVA tables for factorial designs by Philip Iversen (Iowa State).
- jflisp: Three packages for Model Selection in Regression Analysis by Julian Faraway (Univ. Michigan).
- naras: Code contributed by B. Narasimhan (see also bpois). This includes utility code for sliders and boxplots, as well as packages to demonstrate Markov chains, and to fit kernel density estimates. There is also a package for programmers who want to dynamically link arbitrary C functions at runtime.
- plottools: Don't know who wrote it. It provides nice graphical windows to selects symbols and colors.

- puranen: Some demos by Juha Puranen (Univ Helsinki). Demoes Box-Cox and some of the common densities.
- simulation. General purpose simulation tool by Robert Stine (Wharton).
- students: Projects by students in Nancy Reid's course at the U Toronto. Code for repeated measures with missing data, teaching demos, correspondence analysis, Weibull MLE, the general linear hypothesis.
- surv: Code for survival analysis by E. Neely Atkinson and Meg Gelder.
- times: Digital filtering code by Bill Hatch (Coleman Research Corp).
- udina: gnuplot interface to produce beautiful plots from Xlisp-Stat by Jan de Leeuw and Frederic Udina. kde kernel density estimation system by Frederic Udina.
- xlisp-to-s: Interface from Xlisp-Stat to something called S or S-plus or something. By Steve McKinney (U Wash)
- young. Vista: Visual Statistics System by Forrest Young (UNC).

Some of these packages are just snippets, some are major research efforts. Some have been abandoned in mid-stream, others are still being developed. I know of quite a few projects that are going on that we can hopefully add in the near future.

If you have code to contribute, no matter how incomplete, imperfect, tentative, please send it to `deleeuw@stat.ucla.edu`.

Conclusion

Any knowledgeable instructor can make Xlisp-Stat into the ideal tool for teaching statistics at the graduate level. With a considerable investment from many persons, we can make it into the ideal toolkit for developing introductory statistics teaching modules. With another such investment, we can make it the ideal tool for statistics research (and small-scale production). It is imperative that more statisticians start using Xlisp-Stat in the next few years, because otherwise it will either wither and die or go commercial and freeze. This would be a major loss for the statistics landscape.

> Jan de Leeuw
> Statistics Program
> UCLA
> `deleeuw@stat.ucla.edu`

⚭

Establishing a Server

by Michael Conlon

Should your department have a "server" — a computer that performs some collection of tasks for other computers in the department? What will be gained? How difficult is a server to maintain? If you already have a server are you getting the most out of it? Do you offer the right mix of services to your "clients" — the computers that receive services?

Benefits of a server

As networked desktop computers become more common, the role of the server continues to grow. A server (regardless of platform—Mac, PC or UNIX) can be expected to offer one or more of the following services:

Print sharing

In some configurations, certain computers act as "print servers." Other computers rely on these machines for printing across the network. The benefit is reduced investment in printers, and a potential to support multiple types of printing — laser printers, high volume printers, large-format printers, color printers. To get access to a variety of print devices, the client computers are connected to a network and one or more print server computers connect to print devices. It is also possible for many types of printers to be put directly on the network without a print server computer.

Backup

A server typically is maintained by a system administrator. That person's job functions should include backing up the disk space on the server. The server may also remotely backup file space on client machines across the network. This is a tremendous improvement over the "everyone for themselves" model of backup in which some people do a better job than others. Intelligent backup systems pool hardware resources by only having tape drives on the servers and only backing up files that have changed (incremental backups).

File sharing

Disk space may be available on the server for individuals to use. This space may be accessible by other users, using logon and permission systems to enable users to control who they share files with. Such arrangements are a great improvement over file sharing by file copying

— "can I get a copy of x?" and then when x changes, you don't get the updated copy. In file sharing systems, you are getting access to the actual file. As it changes, you get the newest material. File sharing is naturally important in collaborative consulting environments where more than one person may be involved in a data analysis.

Software maintenance

Many servers simply provide software to client computers. This is a variation on the file sharing idea above, but happens at the system level. By installing software on the server, the software is centrally maintained. Only one copy needs to be updated. All client machines receive the software at run-time — no copying – so that each client is always running the current version and all clients are running the same version. This scheme also makes accounting for software much easier than distributing the software out to individual hard disks. And potentially large amounts of disk space can be saved. The software collection at the Department of Statistics at the University of Florida currently occupies approximately 2 gigabytes of disk space. We have 75 client machines. Even if it were feasible to put all that software on each machine (it isn't), we save 150 gigabytes of space by managing it centrally.

Internet services

Servers are increasingly important to provide basic network services such as email, access to Usenet newsgroups, and for serving web pages. While some of these services may be available from corporate or University mainframes, the services available from these other providers tend to be inferior to what can be provided locally. Few mainframe email systems support MIME — the Internet standard for multi-media mail, while many department mail systems do support MIME. Similarly, establishing web services on a department server is straightforward, whereas interacting with a mainframe computer for web services may be quite difficult.

Server models

Regardless of platform, the organization of client computers with respect to servers can generally be classified into three "models."

No server

Here every client operates independently, as computers did before networks. A local printer, a local tape drive, no file sharing, no large software library, Internet services either provided on the desktop or not at all. Many people continue to operate happily this way. When considering the establishment of a server, one needs to think clearly about the impact on such people and whether they will benefit from access to a server. In my college, many of the PCs operate this way. They are on the net with telnet and ftp and may have mail service from a mail server, but all other services are provided locally.

Weak server

The server provides some services but not all. The key feature of this model is that the desktop computers continue to operate normally if the server is not available. Software is installed locally rather than accessed from the server. If the server is unavailable, clients may not get email, but they will still be able to access data (no file sharing) and perform analyses (with local software. Of course some services may be available (access data) and some may not (printing) depending on the mix of services offered by the server. This model makes sense in places where the network is unstable — if the server is unavailable because the network is down, people continue to work. Software must be installed locally, so the variety of software is limited. Many of the Macintosh computers in my college operate in this model.

Strong server

Most services come from the server. If the network or server is down, the client machines can not be used. Sounds risky, doesn't it? Actually, our department servers (three) have a better uptime record than the campus mainframe (which is considered rock-solid). Network outage is a more serious problem. We have designed our system so that we have a server in each major location, so that if the interconnecting network is unavailable, each work group continues to function. Almost all of the UNIX computers in my college operate in this model. PCs on Novell Netware often operate in this model, receiving necessary application software from their server.

The downside

Print and file sharing, Internet services, backup and access to a large centrally maintained software collection are attractive reasons to have a server. There are two major problems.

The first is cost. A capital outlay for server software and hardware and network connections to the clients may be be required. A continuing outlay for maintenance contracts, software renewals will be needed. The most important cost is the time and effort of a competent system administrator to oversee the server and its parts. This is, by far, the largest part of the investment and the largest deterrent to the second risk.

The second risk is server failure. If a strong server model is established so that client systems will not function without the server, and the server is not available, department members will sit around unable to use the system — just as we did twenty years ago when the mainframe was down. Many of us have memories of those days and no desire to return to them. The risk of server downtime is substantial. A good system administrator is the best protection against server downtime, since most server problems are related to software and its configuration —- not to hardware failure. One can buy a hardware maintenance contract and/or have spare parts available to recover from server hardware failure. Recovering from poorly configured server software is far more difficult.

Multi-platform issues

In any consideration of services being offered centrally from a department facility, thought should go into the variety of machines in the department. Will everyone benefit from a department mail server? If the server is a POP (Post Office Protocol) server, client software can be obtained for Mac, PC and UNIX machines, enabling all to take advantage of the new department service. Proprietary vendor email systems rarely cover all platforms.

On balance

The benefits of server installation outweigh the risks where qualified system administration can be dedicated to the task of maintaining the server. Strong server models are common because of the productivity advantages they deliver to their departments. Where system administration is weak or non-existent, the benefits of department servers are rarely realized.

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

◑

A Colorful Variation On Box Plots

by Daniel B. Carr

Introduction and Background

The box plot provides a useful caricature for a batch of data. The word caricature is appropriate since the box plot hides some distributional features, such as gaps, while calling attention to others, such as the median and quartiles. Since box plots receive such widespread use, it is natural for statistical graphics designers to see if they can make even minor improvements on the graphical representation. This article describes my efforts to improve on the box plot. The article builds on the last newsletter article about grids (Carr 1994a). Both articles come from a technical report on converting tables into row-labeled plots. Interested readers can find more details and coverage of other topics such as representation of confidence bounds in Carr (1994b).

In terms of a brief history, Tukey (1977) introduced the box plot. McGill, Tukey, and Larsen (1978) added the idea of notches for comparing medians. Frigge, Hoaglin, and Iglewicz (1989) described different choices for the five- number summary portion of the box plot. Tufte (1983) provided alternative representations that reduced the amount of redundant ink in the box plot. His example suggested that others also search for a better graphical design. Tukey (1989) took up the challenge to develop a better representation. His design strategy used more easily distinguished graphical elements for different components of the box plot. In one variation an area symbol, the filled dot, represented the median. This area symbol stood out from other graphics elements and helped when comparing medians. Tukey (1993) provided more variations on letter displays and box plots and more thoughts on graphical comparison of several linked symbols. As discussants of Tukey's article, Becker and Cleveland (1993) reinforced the use of filled dots to represent medians and brought out further design considerations, such as sorting box plots by their medians. A simple look at the Becker and Cleveland example shows that filled dots and sorting provide a better basis for comparison. This history on black and white methods, the availability of color and variations that appear in software suggest that perhaps there is still room for improvement.

Design Objectives For The Box Plot

A useful starting point in graphical design is to review the design objectives. The box plot symbol is a composite symbol that represents five summary numbers (median, quartiles and adjacent values) plus outliers. The design of the box plot involves selecting a composite symbol to represent these five numbers. One design objective is for viewers to perceive the box plot as a single perceptual unit, as one symbol rather than five. A second objective is to ease comparison of elements within the box plot. For example, are the quartiles symmetric about the median? Other objectives involve comparisons among box plots: comparison of medians with other medians, upper quartiles with upper quartiles and so on. The multiple objectives complicate settling on an optimal design. Design efforts here focus on promoting comparisons among box plots.

Linkages that tightly bind elemental symbols into perceptual units make the elemental symbols difficult to disaggregate for comparison across perceptual units. Kosslyn (1994) discusses several principles of perceptual organization of which two, proximity and good continuation, are particularly relevant to composite symbol binding. For example, the box in the box plot binds the quartiles together. The visual flow along the box length and along the lines from the box to the adjacent values helps people perceive the outliers as part of the same distribution. A related principle for perceptual grouping is visual enclosure. Enclosing the median symbol in a box helps make the symbol become an integral part of the composite symbol. Weakening the tight binding provides a means of encouraging comparisons among box plots.

One design objective is for viewers to perceive the box plot as a single perceptual unit, as one symbol rather than five.

The filled dot suggested by Tukey (1989, 1993) is an area symbol that creates an alternative perceptual grouping:filled area symbols versus line symbols. The alternative grouping counteracts the tight binding in early versions of the box plot. This approach suggests the next step, using color to create alternative perceptual groups. When color is available, using an area symbol for the median becomes less advantageous. The box plot design shown in Figure 1 uses black vertical lines to call out medians. The design represents the ranges between the quartiles and adjacent values using filled boxes. Different colors distinguish the median, values below the median and values above the median. A simple white line shows the median comparison interval. The availability of color provides such options.

Several considerations motivate the particular design choices. Comparing medians across box plots is one of the most important comparison tasks. The current design gives the medians a different color and shape from the other box plot elements. The distinctive symbol enables people to rapidly find the medians using preattentive vision (see Julesz 1986). One can focus attention on all the black line segments just as one could focus attention on all filled dots. In contrast to the early box plot design, the line for the median comprises the vertical extremes of the composite symbol. This avoids total entrapment of the median symbol within the composite symbol. The vertical line segment orientation encourages the vertical visual flow needed to make comparisons. Note that by itself the horizontal orientation of the composite symbol impedes vertical visual flow. When comparing medians from adjacent rows the separation of the nearest endpoints is nearly horizontal. This helps in judging the correct distance. In the dot design it is difficult to judge the distance between dot centers. When adjacent medians are nearly identical in value, they can be resolved with vernier resolution. Thus the symbol for the median facilitates comparison.

As for the thicker box between the quartiles than between the adjacent values, Tukey (1993) points out that symbols near the edges of the plot have higher visual emphasis. Stressing symbols in the center of the distribution provides balance. Turning the lines from quartiles to adjacent values into thin boxes is a small change that accommodates the drawing of median comparison lines as described below.

Distinguishing the left and right sides of the box plot with color makes it easier to make comparisons among the respective halves of the distributions. For tasks like comparing third quartiles, one restricts attention to light red symbols. The almost instantaneous sorting by color narrows the field for comparisons. Evaluation of the interquartile range for a single box plot remains straight forward since the vertical line for the median and the color change does not seriously interrupt the visual flow along the thick box between the quartiles. A few considerations should govern the color selection. The colors should be easy to distinguish from each other and the median color. Colors for low and upper values should be brightness matched (see Carr 1994c) and not fully saturated (see Tufte 1990). When gray is the only option, the most important objective is make the median a distinctive color. This suggests dropping the color distinction between upper and lower values in gray level plots.

Cancer Death Rates

Low-Rate Sites By Sex For Two Decades

Box Plots Based On State Rates
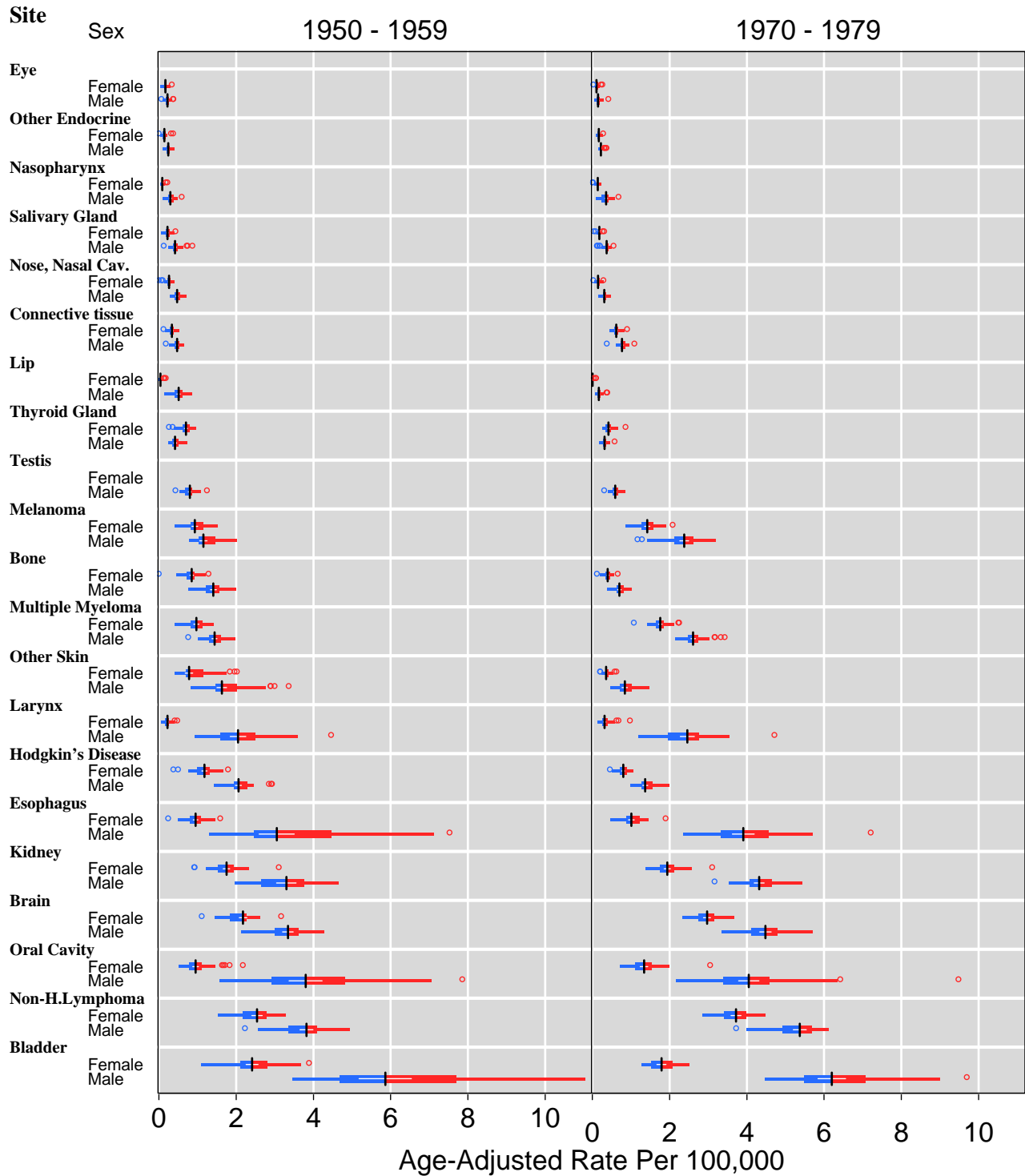
White Line - 95% Test For Equal Medians

Figure 1: A Three-Factor Plot Showing Box Plots

The treatment of outliers in Figure 1 is standard. An alternative shows outliers as vertical line segments with the same height as the adjacent value box. This reduces overplotting, encourages vertical visual flow and plots symbols at their location centers rather than around them. Open circles help increase visibility when the segments are very short.

The proposed box plot design needs to be tested in cognitive experiments to see if it helps people to compare box plots more quickly and accurately.

Figure 1 uses a white horizontal line at the vertical center of each box to represent the median comparison interval. This line representation works when a confidence bound goes past a quartile. Notches appear strange when they extend past quartiles. The internal line representation has merits in addition to handling asymmetric distributions. Since the confidence limits are symmetric about the median, the confidence line allows more accurate comparison of symmetry for the first and third quartiles. To assess quartile symmetry, one looks at the difference between the ends of the white line and the ends of the quartile box. According to Weber's law, people can compare these small differences more accurately (on an absolute scale) than they can compare the larger distances to the median. The symmetry of the white line about the median encourages mentally folding the box plot about the median to judge symmetry. Since the white line is trapped inside the composite symbol, it detracts little when comparing symbol outline elements such as third quartiles. With median comparison lines inside the boxes, comparison of lines may not be as precise or easy as with other possible designs. This is a reasonable compromise because approximate hypothesis tests are not the primary box plot feature. In other words the comparison lines are easy to ignore but serve several comparison purposes when needed.

The interpretation of median comparison lines is the same as for notched box plots. The medians differ at the given significance level when the lines do not overlap and adjustments should be made in multiple comparison situations.

The Data Set, Plot Construction and Interpretation

The data set includes age-adjusted cancer rates for four factors: states (50), body sites (34), sexes, and decades (3). Figure 1 represents states using box plots, body sites and sex using indented rows and decades using panels. The 68 body site by sex combinations stretch the limits of the legible labeling and symbol plotting.

Consequently this plot shows values for the 21 body sites with the lowest rates. Sorting by the male median values for the 1950's determines the body site order. The range of death rates remains large for these sites and a log scale would be reasonable for the x-axis. Figure 1 does not use the log scale in order to show the symmetry of the confidence lines and the performance of some box plot symbols under restricted resolution conditions. Figure 1 shows just two decades to provide more resolution for some symbols. As it stands the one page plot represents over forty percent of the 10,200 values in the data set.

The interpretation is straight forward even in the overplotting cases. For cancer of the eye, the median lines completely cover the quartiles and the median comparison lines. For males and females one can see the horizontal difference between the median lines. This indicates that the hidden comparison lines do not overlap. The conclusion is that rates for males and females are significantly different. For the twenty comparable body sites in the plot, male rates are higher with one exception, the thyroid gland. The plot design makes it easy to compare the rates for sexes. Comparing body site rates for a given decade and sex is not much harder since all values have the same scale. Comparisons across panels is a little harder but here the grid lines come into play (see Cleveland 1993 and Carr 1994a). Comparing box plot elements against the grid lines makes the rate increases from the fifties to seventies quite apparent.

Figure 1 provides a tremendous amount of information but is not a stopping point. To go further, one should study the national rates directly. The box plot outliers represent atypical state rates, so are of some interest. However careful comparison adjusts for estimate variability that includes the influence of differing population sizes.

Closing Comments

The plot that cannot be improved is rare. At the same time attempted improvement do not always prove successful. The proposed box plot design needs to be tested in cognitive experiments to see if it helps people to compare box plots more quickly and accurately. In the mean time, readers who like the design should give it a try. S-Plus users can modify the template for Figure 1. The cancer data, script file, rowplot function and box plot function are available by anonymous ftp to galaxy.gmu.edu. The directory `/pub/submissions/rowplot` contains this and many other row-labeled plot examples.

References

Becker, R.A. and W.S. Cleveland (1993), "Discussion of Graphic Comparison of Several Linked Aspects: Alternative and Suggested Principles," Journal of Computational and Graphical Statistics, Vol. 2, No. 1, 41-48.

Carr, D.B. (1994a), "Using Gray in Plots", Statistical Computing and Statistical Graphics Newsletter, Vol 5, No. 2, pp 11-14.

Carr, D. B. (1994b), "Converting Tables into Plots." Technical Report No. 101, Center for Computational Statistics, George Mason University, Fairfax, VA.

Carr, D. B. (1994c), "Color Perception, the Importance of Gray and Residuals, on a Choropleth Map," Statistical Computing and Statistical Graphics Newsletter, Vol 5, No. 1, pp 16-20.

Frigge, M., Hoaglin, D. C. and Iglewicz, B. (1989), "Some Implementations of the Box Plot," The American Statistician, 43, 50-54.

Julesz, B., (1986), "Texton Gradients: The Texton Theory Revisited. Biological Cybernetics, No. 54, 245-251.

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

McGill, R., Tukey, J. W., and Larsen, W. A. (1978), "Variations of Box Plots," The American Statistician, 32, 12-16.

Tukey, J. W. (1977), Exploratory Data Analysis, Reading MA: Addison- Wesley.

Tukey, J. W. (1989), "Data Based Graphics: Visual Display in the Decades to Come," Proceedings of the American Statistical Association Sesquicentennial Invited Paper Session, Alexandria, VA: The American Statistical Association, 366-381.

Tukey, J. W. (1993), "Graphic Comparisons of Several Linked Aspects: Alternative and Suggested Principles," Journal of Computational and Graphical Statistics, Vol. 2, No. 1, 1-33.

Tufte, E. R. (1983), The Visual Display of Quantitative Data, Cheshire, CT: Graphics Press.

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

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

⊙

WWW viewers—Is there one for you?

by Mike Meyer

The Netscape Browser

In the last issue of the Newsletter I hinted at a software offering from Mosaic Communications. Well, the infobahn is racing ahead at full speed. The company changed its name to Netscape Communications (but kept the `mcom.com` Internet domain), released at least 3 beta versions of its software and won the praise or wrath of just about every denizen of the Internet. Even Time Magazine, available online at `http://www.timeinc.com`, in its December 5 1994 edition, has jumped into the fray, naming one of the Netscape developers, Marc Andreessen, to its list of 50 people with the "requisite ambition, vision and community spirit to help guide us in the new millennium". At least the electronic version of Time, from which I cut and pasted the above quote, mentions Andreessen. I haven't checked the printed version.

So what is all the noise about? Clearly not just a particular piece of software—but I will limit my comments to that.

Netscape is currently available at `ftp:ftp.mcom.com` and several mirror sites. There are versions for most popular Unix platforms (using X windows) and Macintosh and Windows versions. The user interface is very slick and remarkably similar across all three platforms, winning praise from many people. I regularly use the Mac and Unix versions and I have no difficulty switching between the two. The original Mosaic software, which by most measures would still be considered in its infancy, looks old and weary in comparison. Many of the people who praise the user interface also decry the memory requirements of the software. The Mac version requires 2Mb of memory and works best with 4Mb. Lots of people are spending lots of money on memory. Of course the people who complain about Netscape's memory requirements probably have not tried to install the latest commercial software from companies like Microsoft.

Netscape tries to do it all, and it largely succeeds.

Why should Netscape require so much memory? One reason is that Netscape tries to do it all, and it largely succeeds. Where other WWW viewers rely on external helper applications, especially to show non-inlined gifs

and other images, Netscape includes its own integrated viewer. One effect is faster overall operation, and the memory requirements of Netscape are probably on a par with another viewer plus helper applications.

Another area where Netscape receives both praise and damnation is the issue of licensing. The current version is (more or less) freely available to anyone. That seems like a very generous offer. Furthermore the Netscape developers have repeatedly stated on Internet bulletin boards that they intend to provide the browser for free to not for profit organizations, like most universities. Even the quoted prices for Netscape Communications' http daemon look very reasonable. So, what is the problem. Many people, including me, can not (or at least could not) find this promise as part of the Netscape license or elsewhere in the Netscape documentation, and that causes us to pause. Should we start relying on this software for critical applications? An even more bizarre problems arises for certain foreign nationals. Presumably because of US export restrictions on encryption software, Netscape is not allowed to be exported to certain countries and even certain foreign nationals living the US are not allowed to use Netscape. At least that is the how one of my Iranian colleagues, who is a professional online information developer, reads the license. That is making it increasingly more difficult for him to do his job.

Where other WWW viewers rely on external helper applications, especially to show non-inlined gifs and other images, Netscape includes its own integrated viewer.

An area of Netscape that seems to have won largely praise, is its performance over lower speed communications lines. One innovation is the handling of "interlaced" images where one can quickly get a fuzzy picture that progressively becomes clearer as more of it is downloaded. Even StatLib (`http://lib.stat.cmu.edu/`), now uses interlaced images. Quite fun. Another useful feature is the ability to click on links even before all the document is loaded. Savvy network providers are now realizing that it is important to provide the important information early in the document, so that power Web cruisers can continue exploring.

Much more emotion is vented at the HTML (the native authoring language of the WWW) extensions that Netscape supports, and thereby promulgates. Some people like, and are actively using, the ability to have more control over font sizes, and a number of page layout capabilities like centering text. Only Netscape recognizes this new formatting—all other HTML-compliant readers are supposed to ignore it. In practice it seems that the HTML extensions are at worst harmless to users who do not use Netscape, and quite useful to those who do. So, again, what is the fuss about? These HTML extensions are not the same as those proposed by other groups, and some people fear that the momentum behind the WWW will begin to dissipate if more and more incompatible HTML extensions proliferate. There is some merit to that argument. However, if enough people find the Netscape extensions to be useful I feel confident the extensions will be incorporated into the appropriate standards.

An area of Netscape that seems to have won largely praise, is its performance over lower speed communications lines.

Netscape is also the first client (and server) on the block to offer secure communications—in this case by incorporating encryption technology from RSA. This opens the possibility of relatively secure banking and electronic funds transfer. It may not be too long before you can feel secure about entering your credit card into, say, Netscape and having the appropriate server contact the credit card company, get approval for the transaction, bill you, and then ship you the information or product that you ordered. The full text of the Encyclopedia Britannica (`http://www.eb.com`) is already available, *provided* your educational institution has paid the appropriate fee. With Netscape it becomes feasible to charge individuals for each transaction.

The final area where Netscape is either a winner, or a loser, is its ability to have multiple connections active at the same time. Many WWW browsers can have multiple browse windows open, but so far only Netscape allows each of those windows to be retrieving different information at the same time. This can be very useful. While I am waiting for an image from the Louvre (`http://sunsite.unc.edu/louvre` or `http://mistral.enst.fr/louvre`) I can also ask Lycos (`http://lycos.cs.cmu.edu/`) to search through its database of almost 1 million unique URLs. Some people predict the demise of all (or at least all popular) servers, because of the increased load. I suspect otherwise.

In summary, I believe the good of Netscape, far outweighs any potential bad, and I heartily welcome another important WWW player.

Other Browsers

Many users either find Netscape more than they need, or are deterred by licensing concerns. Many

of them with Macintoshes or Windows machines are turning to MacWeb and WinWeb from EINet (`http://galaxy.einet.com/`). These browsers are only in alpha release, and do have some problems with HTML forms, but are otherwise very useful tools. There are several other commercial vendors of "enhance" Mosaic, but I have not sampled these wares.

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

ⓒⓞ

CONFERENCE NOTICES

Interface '95

27th Symposium on the Interface
Computing Science and Statistics
June 21–24, 1995
(note change of date)
Convention Center and Vista Hotel—Pittsburgh PA.

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

First Announcement and Call for Papers

The Symposium is being organized around the theme of "Statistics and Manufacturing," with a sub-theme of green manufacturing, the environment, and quantitative environmental science.

The keynote speaker will be Raj Reddy, Dean of the School of Computer Science at Carnegie Mellon and the most recent recipient of the A.M. Turing award.

Call for Papers

Sessions will include invited and contributed papers. Authors who wish to give a contributed paper should submit two copies of an abstract by February 28, 1995. The abstract should not exceed one-half page (text=6.5 inches wide by 4 inches tall) with centered title, author(s), and address. Invited and contributed papers are to be submitted for the proceedings by July 15, 1995. Proceedings submissions must be camera ready.

General Information

Registration

The registration fee is $155 for members of the cooperating societies, ASA, IMS, SIAM, ORSA, the Biometrics Society (ENAR and WNAR) and for persons affiliated with Penn State University and Carnegie-Mellon University. For others the fee is $175. The registration fee covers the reception, coffee breaks, banquet, and proceedings.

Schedule

Conference registration begins on Wednesday afternoon (June 21). The first official conference event is a Wednesday evening mixer. Technical sessions will be held on Thursday and Friday from 8:15 a.m.–5:15 p.m. and from 8:15 a.m. to noon on Saturday. Breaks are scheduled between the sessions and also for lunch. All conference meetings will be held in the David L. Lawrence Convention Center, which is linked to the hotel by an enclosed walkway.

Entertainment

Baseball game. If you are a baseball fan, the Pittsburgh Pirates will be hosting the San Francisco Giants on Wednesday, June 21st at 3:00 pm.

Thursday night banquet. This year's banquet will take place on board one of the boats in the Gateway Clipper Fleet. The three hour cruise will include sightseeing along Pittsburgh's three rivers, musical entertainment, and the Captain's Buffet Dinner. The cost is included as part of your registration fee.

Registration

For further registration and hotel information, contact:

Email: interface95@stat.cmu.edu
Phone: (412) 268-3108 Facsimile: (412) 268-7828
Mail: Interface95, Department of Statistics
Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213, USA.

Further Information

For current information, use the World Wide Web to access the URL:
`http://www.stat.cmu.edu/interface95/`

Mike Meyer
Carnegie Mellon University
`mikem@stat.cmu.edu`
James L. Rosenberger
Penn State University
`JLR@stat.psu.edu`

SECTION OFFICERS

Statistical Graphics Section - 1994

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

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

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

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

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

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

Linda A. Clark, Secretary/Treasurer (93-94)
   908-582-4807
   AT&T Bell Laboratories
   lac@research.att.com

Howard Wainer, Publications Officer (93-94)
   609-734-5623
   Educational Testing Service
   Princeton, NJ 08541
   hwainer@rosedale.org

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

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

<center>◎</center>

Statistical Computing Section - 1994

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

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

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

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

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

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

Deborah F. Swayne, Secretary-Treasurer
   908-829-4263
   Bell Communications Research
   Morristown, N.J.
   dfs@bellcore.com

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

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

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

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

<center>◎</center>

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 (93-96)
    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

James L. Rosenberger, Newsletter Editor (93-96)
    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
    301-496-8556
    National Cancer Institute
    kk@helix.nih.gov

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

ⓘ

Statistical
COMPUTING & GRAPHICS

PENN STATE

Department of Statistics
University Park, PA 16802-2111

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