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

*December 96 Vol.7 No.3*

# Statistical COMPUTING & GRAPHICS

***A WORD FROM OUR CHAIRS***

## Statistical Computing



*This is Sallie Keller-McNulty's last column as 1996 Chair of the Statistical Computing Section. The editors would like to thank her for her contributions to the newsletter.*

This is my final column as your Section Chair. I am pleased to say that I have now placed the leadership of the Statistical Computing Section in the able hands of Daryl Pregibon. Some of our other new Section Officers that will be performing important functions for the Section this coming year are James Marron, our new Publication Liaison; Russell Wolfinger, the 1998 Joint Statistical Meetings (JSM) Section Program Chair; Karen Kafadar our Section Chair-Elect; and our new Council of Sections Representatives, Naomi Altman and Terry Therneau.

CONTINUED ON PAGE 2

## Statistical Graphics



*William DuMouchel is the outgoing 1996 Chair of the Statistical Graphics Secion. The editors would like to thank him for his help in providing this column during the past year.*

As my year as Chair of the Statistical Graphics Section ends, I would like to thank all the officers of the Section for working so hard during the year to make our activities successful. Many of them deserve special mention: Stephen Eick organized and shepherded a great program for last August's meetings in Chicago. Dianne Cook, our incoming Program Chair, has lined up another great program for next summer. Robert Newcomb, our Secretary/Treasurer, has, in addition to his official duties,

CONTINUED ON PAGE 3

***SPECIAL FEATURE ARTICLE***

## Computational Anatomy: An Emerging Discipline

By Ulf Grenander and Michael I. Miller

The last several decades have witnessed a revolutionary change in medical imaging. New imaging instrumentation has transformed the field from one that was dominated by the time honored flat Xrays to a quickly expanding technology with powerful tools like MRI (Magnetic Resonance Imaging), PET (Positron Emission Tomography ) and SPECT (Single Photon Emission Computed Tomography ) to mention but a few. This has radically changed the diagnosticians' ability to acquire patient data with more or less non-invasive procedures.

Some sensors, for example ultra sound, produce very noisy data; others like MR have a better signal-to-noise ratio. In both cases the presence of randomness necessitates statistical treatment, but it should be emphasized that the main difficulty is not due to noise, whose probabilistic properties can be derived from the physics of the sensor with some degree of accuracy. Instead the overwhelming challenge facing the analyst of data from medical imaging is *understanding the randomness that represents biological variability.* This is more *terra incognita* than noise analysis, although pioneering work was done as early as the 1970's by Bookstein (1978) and others studying shape change. Today an increasing number of mathematicians, statisticians, anatomists and radiologists are exploring the new discipline *computational anatomy.*

Statisticians are wont to complain about the scarcity of data, but computational anatomy should be the statisticians El Dorado, since data come in huge quantities, especially in 3D imaging where $10^8 - 10^{10}$ data points are not unusual. We believe that with the emergence of fundamentally new technologies in CRYOSECTIONING

CONTINUED ON PAGE 4

*EDITORIAL*

# One Year Later...

We have been editing the newsletter for a year, now, and proudly notice that it hasn't folded yet! On the contrary, we have been able to bring you stimulating articles on a wide variety of topics and to keep you abreast of recent developments concerning our sections and our areas of interest. Of course, in this era of "electronic information," it is difficult, if not impossible, for the printed medium to remain fully competitive as far as timeliness is concerned. We remind you, however, that the electronic version of the newsletter becomes available ahead of the printed version and can be easily accessed by following the links to either of our sections from the ASA web page (`http://www.amstat.org/`). In our sections' web pages you will also find an abundance of additional professional information and resources.

This issue contains the last columns as section chairs by Sallie Keller-McNulty and Bill DuMouchel. Looking back at what happened over the past year, they gladly recognize the numerous accomplishments of our sections and share with us their thoughts for the future. We are certain to convey the sentiment of all of our readers by extending to both of them our thanks for their managerial and administrative efforts.

The special feature article by Ulf Grenander and Michael Miller that begins on page 1 is an excellent introduction to a fascinating emerging discipline: computational anatomy. The authors argue that, to take full advantage of the huge amount of diagnostic data generated by novel medical imaging techniques, a clear understanding of biological variability is needed. In the article, they describe their approach to examining the geometric properties of the brain substructures based on the construction of statistical measures of anatomical variation. They also point out the challenges that lie ahead in this area of research.

After a year-long hiatus, we are happy to bring you back the "Bits from the Pits" column. Al Liebetrau, who used to contribute regular pieces featuring statistical computing and statistical graphics activities in science and industry, has kindly agreed to become the column editor. For this issue, he has secured an interesting contribution from two of his colleagues at the Columbus headquarters of Battelle Memorial Institute. Beginning on page 9, Stephen Wall and John Orban describe a computer simulation program that they developed to illustrate the use of different statistical sampling methodologies for collecting highway traffic data.

In the "Unix Computing" column starting on page 14, Phil Spector picks up from where he left off in the August issue and discusses more advanced applications of the Perl programming language. In another of our regular columns, "Topics in Information Visualization" (page 16), Daniel Carr and Suzanne Pierson address the issue of redesigning choropleth maps using micromaps. Their article points out visual and representational problems encountered with typical choropleth maps and describes a new template that overcomes these problems by linking micromaps and row-labeled plots.

This issue also contains a (belated) summary of the highly successful scientific programs that our sections sponsored at the Joint Statistical Meetings in Chicago and a report on the annual joint business meeting and mixer. A clear sign of the vitality of our scientific community is given by the many conferences on topics related to statistical computing and graphics that will take place in the next several months. Please take a minute to check out the notices printed in the last few pages of the newsletter and decide if you would like to attend. As usual, if you have any comments or if you wish to send us a contribution, we would be happy to hear from you.

Thank you and enjoy your reading.

Mark Hansen
*Editor, Statistical Computing Section*
*Bell Laboratories*
`cocteau@bell-labs.com`

Mario Peruggia
*Editor, Statistical Graphics Section*
*The Ohio State University*
`peruggia@stat.ohio-state.edu`

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

# Statistical Computing

CONTINUED FROM PAGE 1

The Section Executive Committee has made an important addition to the committee. We have added an appointed position of Electronic Communications Liaison to the committee. Our first Electronic Communications Liaison is Tom Devlin. Tom will serve a three year term. Even before this appointment, Tom has been active in this capacity for the Section. He is the creator of our homepage (`http://www-stat.montclair.edu/asascs/`).

Tom (devlin@mozart.montclair.edu) welcomes suggestions about the homepage and any other thoughts you have about activities with which he should become involved.

Section dues have been used on a variety of exciting activities this past year. We again sponsored a student paper competition session for the JSM this past August. The session was a huge success and the Journal of Computational and Graphical Statistics has requested that the students submit their papers for publication. We sponsored three short courses at the JSM and one at the Interface Symposium in Australia. We help support an undergraduate data analysis competition. The winners of the competition (Theresa Crofts, Victoria Field, AAron Holt, Jeffrey Edwardy and Amy Stai from Winona State; Matthew Haubrich and Matthew Schwab from Iowa State; Aidan Palmer from Carnegie Mellon; Ellie Nagel and Todd Nelson from Brigham Young) gave presentations at the JSM last August, as well as joining in the festivities at our Joint Mixer.

In the upcoming year, the Section will sponsor another student paper competition for the 1997 JSM this August in Anaheim. We will also be supporting travel for Peter Huber to do a special presentation on massive datasets at the 1997 JSM. We are helping to support the Third North American Conference of New Researchers in Statistics that is being organized by Snehalta Huzurbazar and Aparna Huzurbazar (see page 25 for information). Finally, the Section is co-sponsoring a symposium this spring at New York University on Recent Developments in Smoothing Methods, (see page 26 for information).

The year promises to be an exciting one for our community. The 1997 JSM will focus on "Shaping Statistics for Success in the 21st Century." With this theme we will see many sessions dealing with the topic of massive datasets. There will also be much discussion at the JSM about the impact of computing and computational sciences on statistics in the 21st century. In February, the International Association for Statistical Computing will host the Second World Conference in Pasadena California. The conference theme is Computational Statistics &Data Analysis...on the Eve of the 21st Century (http://www.stat.unipg.it/iasc.html). We should all be sure not to miss this year's Interface Symposium being chaired by David Scott. The symposium theme is "Mining and Modeling of Massive Data Sets In Science, Engineering, and Business," (http://www.stat.rice.edu/). Another meeting of interest is being sponsored by the American Association for Artificial Intelligence. It is the Third International Conference on Knowledge Discovery and Data Mining. It will be held immediately after the JSM in Newport Beach, California (see page 8 for information).

So, as you can see, there are many events of interest that the Section is involved in this year! I wish you all a productive new year and I look forward to seeing many of you at the various conferences and symposia.

Sallie Keller-McNulty
Kansas State University
sallie@cecil.stat.ksu.edu



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

# Statistical Graphics

CONTINUED FROM PAGE 1

continued to develop and maintain our Web page. Mario Peruggia, as the Graphics Section co-editor of this newsletter, has, with Mark Hansen, continued to maintain its wonderfully high quality. And Sally Morton, the incoming Chair, assisted with many administrative tasks throughout her year as Chair-Elect. We can all look forward to a great year for the Statistical Graphics Section with Sally as Chair!

I hope all of you have had reason to access the ASA's Web site [www.amstat.org] by now. Primarily because of the support of our two Sections, it has become very useful and professional-looking – we should all be proud. I find the member index really useful. And of course the link to our own Graphics Section web site is especially valuable! I would like to direct your attention in particular to two parts of our site. First, the list of JSM-96 Exhibitors who contributed prizes to the raffle conducted at our mixer in Chicago. Second, the description of the 1997 Data Exposition.

The data for this year's Exposition were compiled by Colin Goodall, who is also a Council of Sections Representative for our Section. The challenge for participants in the 1997 Exposition is especially relevant and topical – Can statistical and graphical methods help improve the quality of our health care system? Let's assume the prerequisite for improving the quality of any system is deciding how to measure its quality. Our data consist of over a hundred variables describing costs, patient outcomes and sample sizes for subpopulations of patients in each of over 1000 hospitals. Is it possible to prepare report cards that fairly summarize hospital performance and separate the good hospitals from the not-so-good?

Health care in the U.S. represents almost a trillion dollar annual expenditure. Shouldn't the tools for improving process quality that our profession is famous for having developed be able to increase the efficiency of this process by at least a few percent? If statisticians don't participate in the solution of this problem, we will have only ourselves to blame if whatever methodology gets eventually adopted ends up supplying endless examples for "how to lie with statistics".

So please check out our web site for more discussion and description (and the data, of course) and enter the contest!

Best wishes to you all for the new year.

Bill DuMouchel
AT&T Labs - Research
dumouchel@research.att.com

## SPECIAL FEATURE ARTICLE

CONTINUED FROM PAGE 1

and high field strength MR there is a need for examination of statistical measures of anatomical variation in the context of the geometric properties of the brain substructures. This has been the principal focus of our work. It is now possible to study the cortical surfaces, folds and subvolumes of the brain, simply because the new imaging technologies are defining upwards of 10 times the resolution of previous 1-2mm MRI supporting the geometric properties of these brain structures.

To illustrate examples of brain structures which can now be imaged, in the left panel of Figure 1, we present a cryosection through a macaque brain, the whole 3-D brain consisting of $640 \times 480 \times 200$ voxel volume elements at approximately $(0.1mm)^3$ resolution. Notice the exquisite cortical folds of white and gray matter. The right panel shows a whole brain MRI-MPRAGE image volume with a section through the whole brain showing a clear delineation of the surface of the hippocampus.

As well, mapping tools now exist which support the geometry of these brain structures (see, for example, Christensen, Rabbitt, and Miller 1996). Thus far we have principally worked in the homogeneous anatomy setting, in which the space of anatomical imagery is assumed topologically equivalent. Space limitations do not allow us to give more than a rough sketch of the pattern theoretic methodology. To fix ideas, say that we are interested in a nucleus or brain structure such as the hippocampus $H$ in the brain (right panel of Figure 1) and for the moment only consider its shape as a volume $H \subset X \subset \mathbf{R}^3$ and its texture given by a function $I : X \rightarrow \mathbf{R}$ inside and outside the nucleus located in a background space $X$, and where $X$ will usually be a rectangular set in 3-D. To represent the shape variability of $H$, we introduce a *similarity group* $S$ with element mapping

$$s : X \rightarrow X$$

homeomorphically so that we represent the actual shape as the *deformed template*

$$I(x) = I_{temp}(sx); s \in S; x \in X .$$

Then the anatomical ensemble (space of anatomical realizations) is the orbit (deformable template) under $S$ so that any two images $I_1$, $I_2 \in \mathcal{I}$ are topologically

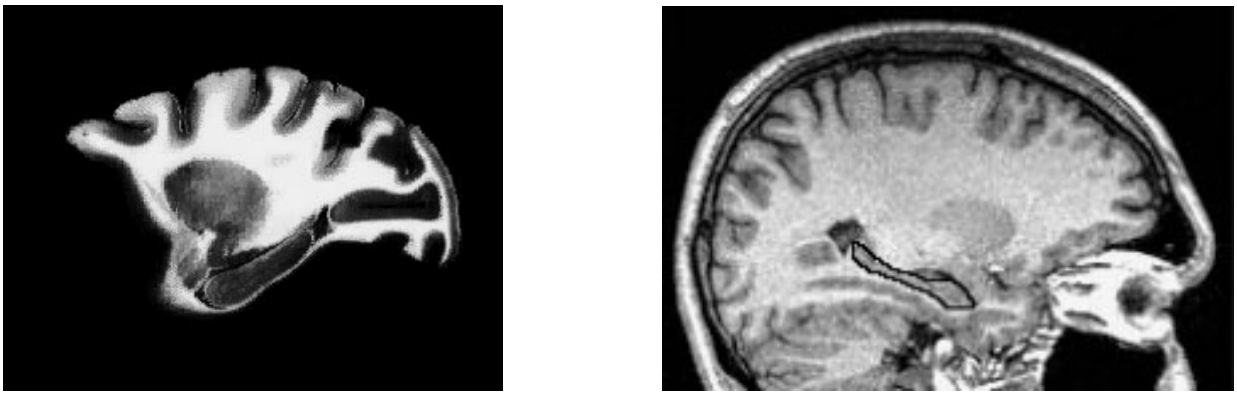


Figure 1: The left panel shows a whole brain cryosection through a macaque brain illustrating white and gray matter folds. Data taken from David Van Essen of the Department of Anatomy and Neurobiology at Washington University. The right panel shows a section through the whole brain MRI scan delineating the hippocampus. Data taken from Dr. John Csernansky of Department of Psychiatry at Washington University.

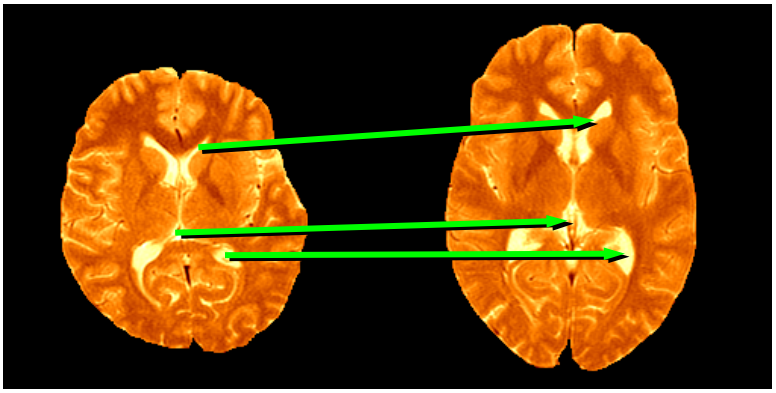


Figure 2: Showing transverse sections through MRI images with three correspondence points (arrows) depicted between the two images. These are three elements of the vector field transformation determining the similarity $s \in S$ mapping one anatomical coordinate system to another.

equivalent in the sense that

$$\exists s \in S \ni I_1(x) = I_2(sx), x \in X \ . \qquad (1)$$

After choosing one image, call it the template $I_{temp}$, all anatomies can be generated from it. Characterizing normal anatomy becomes an empirical procedure of constructing probability laws on the transformations $S$ from the family of transformations observed in actual anatomies. In this setting, disease or abnormality then corresponds to a transformation which is a *large deviation* from the identity in the group of transformations as reflected by the perhaps normal and disease measures. Building anatomical representations becomes an empirical procedure of constructing probabilities on the transformations from the various populations of anatomies.

As human anatomy is exquisitely complex, we have been studying similarities which are constructed at their highest resolution from vector fields, products of the basic translation groups. They correspond to transformations of dimension equivalent to the number of voxels in the imagery, on the order of 100,000,000 parameters. Figure 2 shows two sections from volume imagery associated with MRI image data, depicting three points in a similarity corresponding to the vector fields defining the mapping of one anatomical coordinate system to another, as well as vectors associated with three points forming the vector fields.

The major challenges facing computational anatomists in the future is that even in this "homogeneous anatomy setting," representing variability of shape and other entities in anatomy requires new mathematical tools. One of them is pattern theory as presented for example in Grenander (1993) in which probability measures are introduced on groups and other families of transformations in order to create knowledge representations of anatomical variability.

Here we have started from one or several templates $I_{temp}(\cdot)$ that describe the intensity field in and around a *typical* shape $H_{temp}$ in $X$. How to determine the template from data is a problem in statistical estimation of unconventional type that cannot be discussed here. Anyway, once $I_{temp}$ has been chosen, we represent biological variability by a prior probability measure $\Pi$ on the group $S$. Call the probability density $\pi(\cdot)$ with respect to an invariant measure $m$ on $S$

$$\pi(s) \ = \ \Pi(ds)/m(ds) \, .$$

In passing we point out that we insist on *modelling the anatomy in the continuum* $\mathbf{R}^3$ *in which it lives*, not on any discrete lattice as has been customary in most pattern recognition. Of course, the computing will have to be done discretely since we rely on digital machines, so that discretization has to be done sooner or later. Our advice is: do it later!

Let us denote the observed image by $I^{\mathcal{D}}(y), y \in Y$, which is an array indexed by $y$. For an MRI, for example, $Y$ could be $256 \times 256 \times 256$ so the image would be an enormous 3D array, say of dimension $d$. Now we specify the likelihood function $L(I^{\mathcal{D}}|s)$, i.e. the conditional probability density of $I^{\mathcal{D}}$ in $\mathbf{R}^d$ so that we can apply Bayes' theorem and get the conditional probability density of the unknown group element $s$ given the deformed image $I^{\mathcal{D}}$ through the proportionality

$$p(s|I^{\mathcal{D}}) \ \propto \ \pi(s)L(I^{\mathcal{D}}|s) \, .$$

We have avoided the question of how to perform the actual modelling of images through configuration spaces, connectors and other pattern theoretic concepts. The reader is referred to Grenander and Miller (1993) for a

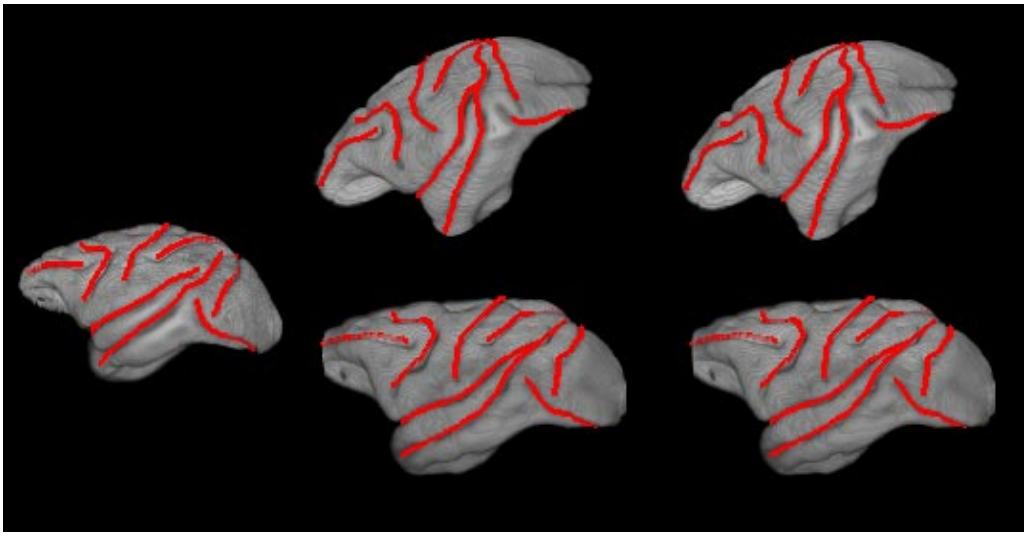


Figure 3: The left panel shows the template $I_{temp} = 87A$. The middle panels shows the transformed templates $I_{temp}(s_1 x), I_{temp}(s_2 x)$. The right panel shows the two targets $I_1 = 93G$ and $I_2 = 90C$.

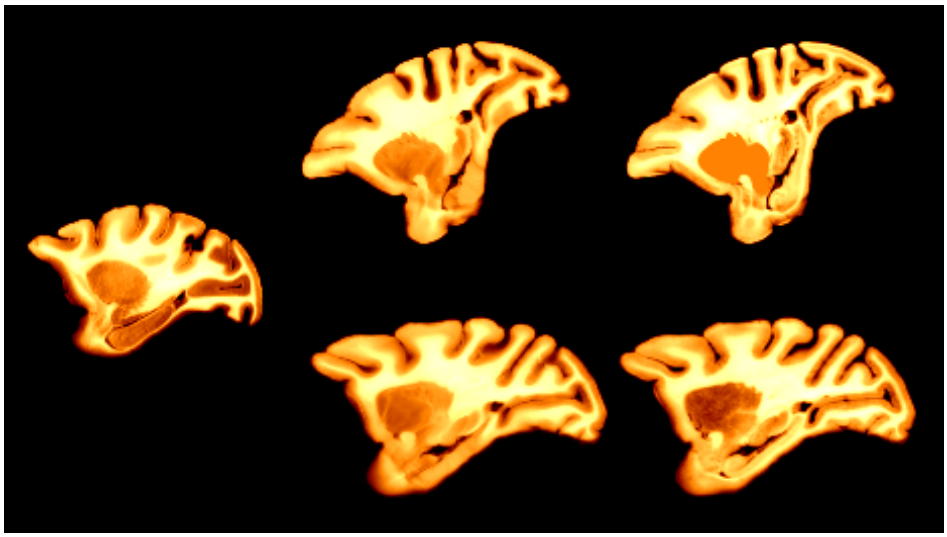


Figure 4: The left panel shows a section through template $I_{temp} = 87A$. The middle panel shows sections through the transformed templates $I_{temp}(s_1 x), I_{temp}(s_2 x)$. The right panel shows two target sections $I_1 = 93G$ and $I_2 = 90C$.

detailed discussion of these topics in the context of cytological micrographs as well as to a more concise treatment of brain anatomies in Miller, Christenson, Amit, and Grenander (1993).

Note that the posterior $p$ contains all the information available for analyzing the data given only the MRI: both medical knowledge from an "anatomical textbook", and empirical knowledge residing in the observed image. Any inference should therefore be derivable from $p$, so that the main question has now been reduced to handling $p$ analytically and/or computationally. Once we have an estimate $s^*$ of the group element $s$ we obtain an analysis of the brain data by computing $I_{temp}(s^* x); x \in X$.

Automated tools for generation of whole brain maps corresponding to these similarities $s \in S$ from these fabulously high resolution scans are only now starting to be generated in laboratories throughout the country. To illustrate, Figures 3, 4 taken from Joshi, Grenander, and Miller (1996) demonstrate whole brain maps in the macaque monkey. The left panel of Figure 3 shows locations of the deep sulcul folds, called the fundi of the sulci. These have been labeled in David Van Essen's laboratory in the Department of Anatomy and

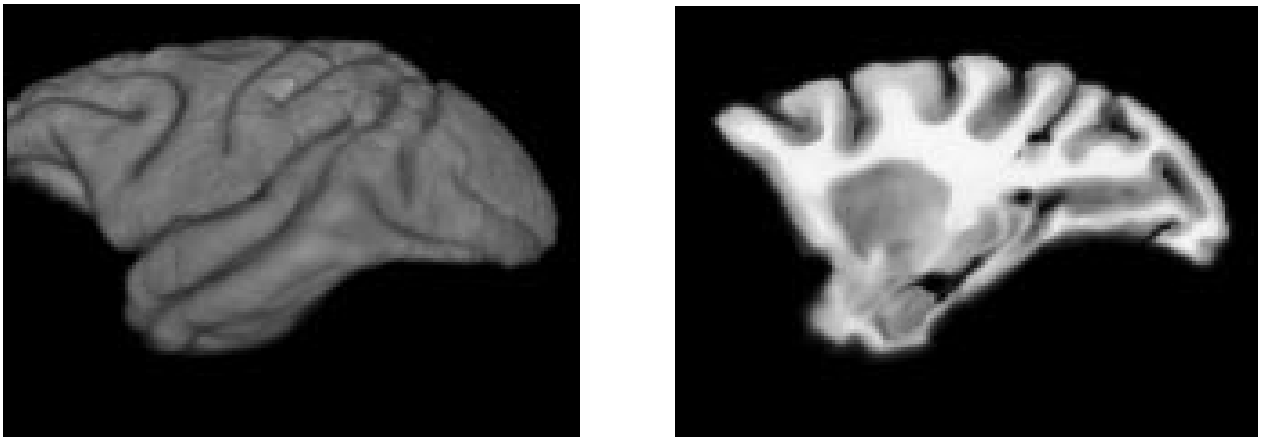


Figure 5: Showing the sample mean brain with a section through it. Generated by mapping three whole brains to one common coordinate system. Whole brain volume reconstructions taken from the laboratory of David Van Essen of the Department of Anatomy and Neurobiology at Washington University.

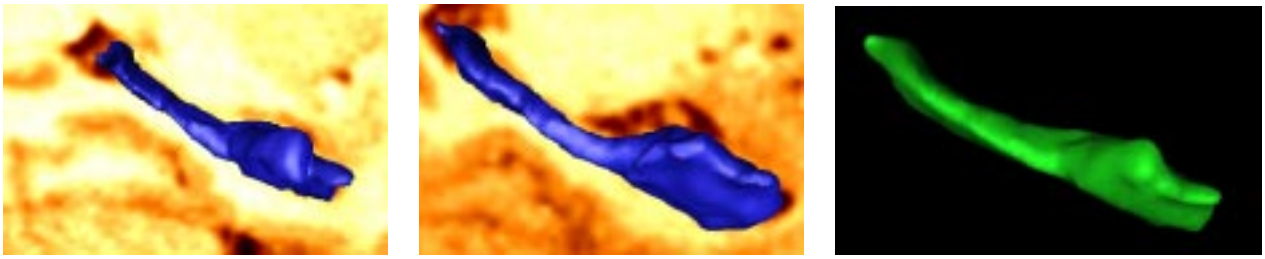


Figure 6: The left and middle panels show rendered surfaces from two of the target volumes resulting from mapping the template hippocampus onto two targets. The right panel shows the mean hippocampus generated from a population of 10 anatomies. Data taken from Dr. John Csernansky of the Department of Psychiatry at Washington University. Figures taken from Joshi, Grenander, and Miller 1996.

Neurobiology at Washington University, and are depicted via the superimposed lines. The brains were then mapped one to another; the template brain 87A was mapped to two target brains 90C and 93. Figure 3 shows the results from the whole brain mapping procedure. Figure 4 shows sections through the whole brains. The left panels show the template $I_{temp}$. The middle panels show $I_{temp}(s_1 x), I_{temp}(s_2 x)$ resulting from mapping the template to the two targets. The right panels show the targets $I_1, I_2$. The similarity transformations consisted of vector fields the dimension of which is equivalent to the number of voxels. In this case, the mapping consisted of about 10,000,000 parameters. Notice the fantastic correspondence and the detail that is accomplished by the whole volume transformation.

Statistical properties of such mapping experiments are beginning to emerge. Groups have already begun characterizing large populations; (see Evans, Collins, and Holmes, 1996, and Mazziotta, Toga, Evans, Fox, and Lancaster 1995, for example). Average brains can now be generated, with variations around the average studied via probability measures on the similarities. Such an example is depicted in Figure 5 showing the mean brain generated from 3 whole macaque brains.

The mean brain was generated by mapping a whole macaque provisory template brain onto two whole macaque brains, then averaging the coordinate system transformations and applying the average transformation to the template. This average transformation applied to the provisory template is then the closest man-made brain to the population.

Shown in Figure 6 are average surfaces associated with the hippocampus. Panels 1 and 2 show two rendered surfaces embedded in two of the target volumes $\mathcal{M}(s_i x), i = 1, 2$. These surfaces were generated by transforming the provisory template hippocampusvolume through the volume transformations carrying the template onto the target,and then composing the map

with the 2-dimensional surface manifold $\mathcal{M}$ representing the provisory template hippocampus. The right panel shows the average hippocampus generated by combining ten anatomical maps from a population of ten anatomies.

This is an emerging discipline in its earliest stages of development, with tremendous opportunity for the development of new statistical methods for characterizing human variation. We call upon the statistical community to engage itself in research in the field of computational anatomy. Not only can the results be of value to medical science and ultimately to patients; the intellectual challenges are overwhelming, and while much progress has been made during the last five years we can expect major practical breakthroughs as well as theoretical advances in the years to come.

### References


Bookstein, F.L., (1978), *The Measurement of Biological Shape and Shape Change*, New York: Springer-Verlag.

Christensen, G.E., Rabbitt, R.D., and Miller, M.I. (1996), "Deformable Templates Using Large Deformation Kinematics," *IEEE Transactions on Image Processing*, 5, 1435–1447.

Evans, A.C., Collins, D.L., and Holmes, C.J., (1996), "Computational Approaches to Quantifying Human Neuroanatomical Variability," in *Brain Mapping: The Methods*, 343–361, eds. Toga, A.W. , and Mazziotta, J.C., San Diego: Academic Press.

Grenander, U., (1993), *General Pattern Theory*, Oxford, UK: Oxford University Press.

Grenander, U., and Miller, M.I., (1994), "Representations of Knowledge in Complex Systems," *Journal of the Royal Statistical Society, Ser. B*, 549–603.

Joshi, S., Grenander, U., and Miller, M.I., (1996), "On the Geometry of Brain Sub-Manifolds," *International Journal of Pattern Recognition and Artificial Intelligence, Special Issue on Processing of MR Images of the Human Brain*, to appear.

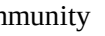
Mazziotta, J.C., Toga, A.W., Evans, A., Fox, P., and Lancaster, J. (1995), "Probabilistic Atlas of the Human Brain: Theory and Rationale for Its Development," *Neuroimage*, 2, 89–101.

Miller, M.I. Christensen, G.E., Amit, Y., and Grenander, U. (1993), "Mathematical Textbook of Deformable Neuroanatomies," *Proceedings of the National Academy of Sciences*, 11944–11948.


Also visit the Web Sites:
`http://www.dam.brown.edu/pattern/`
`http://www.cis.wustl.edu`


Ulf Grenander
Brown University
`ulf@brownvm.brown.edu`

Michael I. Miller
Washington University
`mim@cis.wustl.edu`

*CALL TO ACTION*

# Data Mining

By Daryl Pregibon

While it has been considered a dirty word in Statistics, "data mining" is an emerging area that is projected to be a multi-billion dollar industry by the year 2000. The modern usage of the term data mining connotes the goal of "extracting information from data", that is, refining crude and abundant raw data into high grade information for decision making. The field of data mining emerged from the applications side of the Machine Learning community, a largely theoretical subgroup of Artificial Intelligence (AI). As such, data mining is a blend of statistics, AI, and database research. There is much overlap between statistics and data mining and therefore tremendous opportunity for statisticians who are looking for challenging new areas in which to apply their skills and experience in data analysis and inference.

Statisticians interested in learning about the area will have several opportunities to do so at the Joint Statistical Meetings in Anaheim this summer. Special sessions are planned of both a tutorial and research nature. Immediately following the Joint Meetings in nearby Newport Beach CA, the American Association of Artificial Intelligence, in cooperation with the ASA, will be hosting the 3rd International Conference on Knowledge Discovery and Data Mining (KDD-97). This is the preeminent conference in data mining and a good place to see first hand what all the hype is about. For more details point your browser to `http://www-aig.jpl.nasa.gov/kdd97/`.


Daryl Pregibon
AT&T Labs Research
`daryl@research.att.com`

# A Simulation Tool For Evaluating Design And Analysis Options For Monitoring Highway Traffic Characteristics

By Stephen M. Wall and John E. Orban

## *1. Introduction*

Transportation planning agencies, faced with a need to improve highway system performance and safety, are making greater efforts to monitor travel trends and the changing characteristics of the vehicles used on the nation's roadways. For example, the Federal Highway Administration (FHWA) established the Highway Performance Monitoring System (HPMS) to monitor traffic volumes, annual vehicle distance traveled (AVDT), annual average daily traffic (AADT), vehicle classification, and truck weight (Traffic Monitoring Guide 1995). Truck safety compliance rates and vehicle occupancy rates are also quantities of interest to transportation planners. Each of these monitoring programs requires the development of sampling plans and estimation procedures that are statistically valid and practical to implement. Kinateder et al. (1997) discuss many of these issues from a statistical perspective. This article discusses how a computer simulation program was developed to evaluate and illustrate the statistical properties of different sampling strategies and analysis techniques. We begin by discussing some of the statistical sampling and analysis options that might be under consideration in a given situation; then, we present an overview of how the simulation program is used to evaluate these options.

## *2. Options for Sampling Designs and Analysis Procedures*

The choices of sampling designs and analysis procedures to be employed on a specific monitoring program depend on several factors. The first consideration is the level of stratification required. The monitoring program might involve several regions (states, counties, etc.) or functional classes of roadway (interstate, rural, etc.). Furthermore, the planners might be interested in estimating traffic characteristics at various times of the day or different days of the week. The second factor to consider is the choice of performance measures. Often this defines constraints on the sampling approach. For example, when estimating traffic volumes or average vehicle distance traveled, it is possible, in most instances, to count all vehicles that pass a sampling location. On the other hand, determining if commercial vehicles are complying with safety regulations imposes constraints on the number of vehicles that can be inspected. Other considerations include whether or not prior information is available to "optimize" the sampling allocation, limits on the number of sampling locations or level of sampling at each location, and regulatory issues such as the requirements to include certain locations or to inspect certain types of vehicles.

Just as there are different options for designing the monitoring program, there are also options on how the results should be reported. From a statistical perspective, one would like to have estimators that are unbiased with minimum variance. However, depending on the complexity of the sampling design, the "optimal" approach may not be obvious. Furthermore, regulators and administrators may not be comfortable with complex analysis routines; especially when a "simple" approach produces essentially the same answer.

Thus, in considering the various options for designing the monitoring program and analyzing the results, a tool is needed to evaluate the statistical implication of the decisions. Such a tool is the topic of our article.

## *3. Simulation Modules*

In designing the simulation program, we identified the following four modules that would be incorporated into the first version of the software: Stratum Module, Sampling Design Module, Simulation Module, and Results Module. Because each of these modules represents a distinct component of the sampling problem, our goal was to make them as independent as possible. Each module would be developed with well-defined inputs and outputs that could be used by the other modules, but all data manipulation would be localized within the modules. This approach would provide flexibility for modifying individual modules without affecting the operation of other modules. This approach also minimizes the initial development and testing effort, and the amount of integration testing required when individual modules are modified.

### *3.1 Stratum Module*

The Stratum Module is used to define the stratification strategy. In its current form (see Figure 1), one can select different locations, functional classes, and sampling times. However, as statistical methodology is developed to handle complex stratification schemes, this module

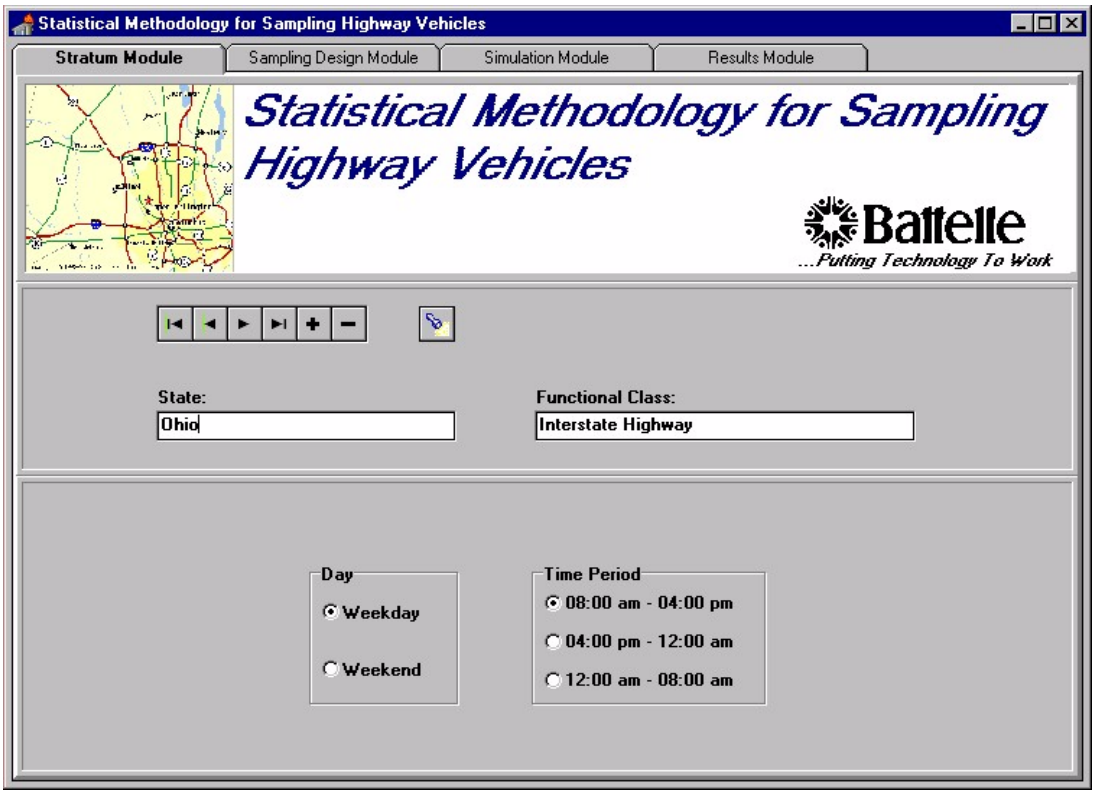



Figure 1: Stratum Module.

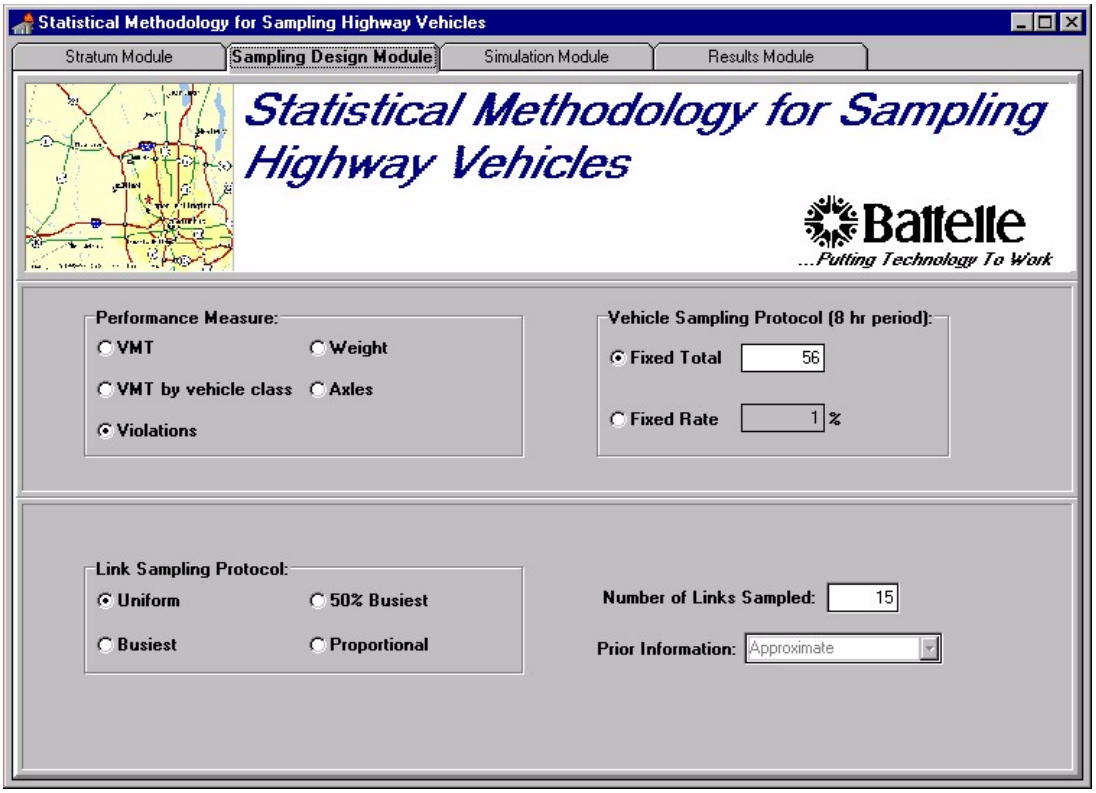
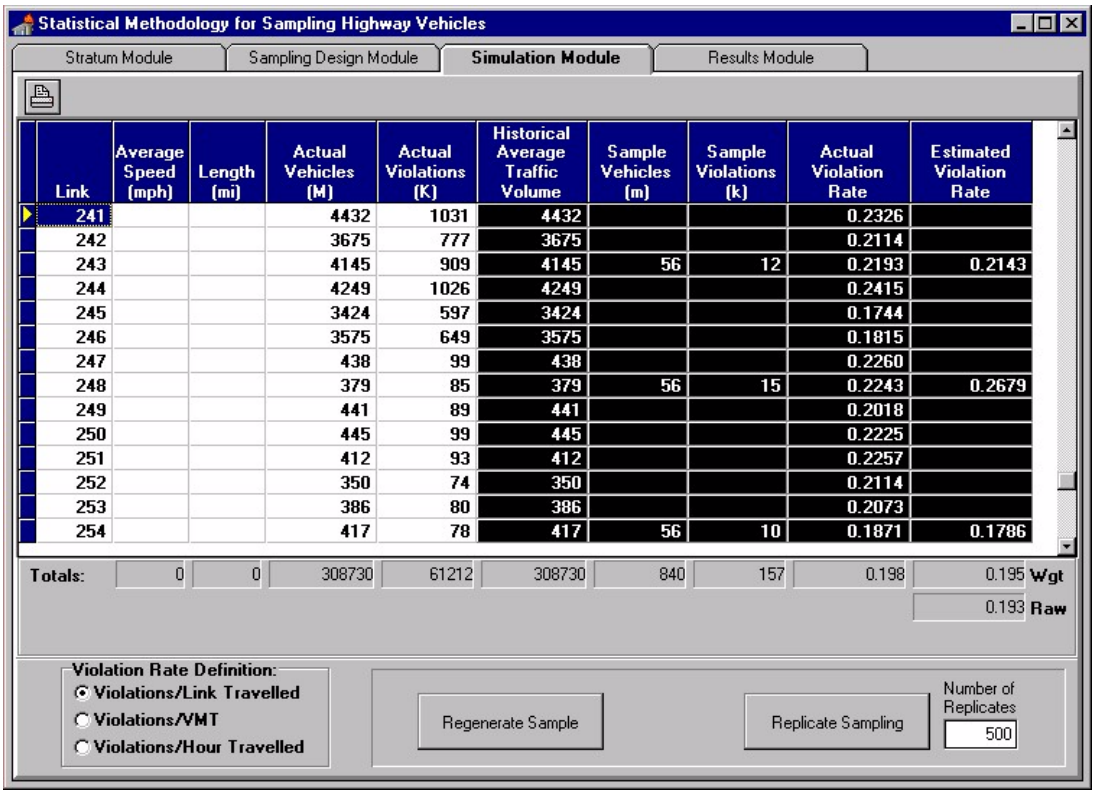



Figure 2: Sampling Design Module.




| Link | Average Speed (mph) | Length (mi) | Actual Vehicles (M) | Actual Violations (K) | Historical Average Traffic Volume | Sample Vehicles (m) | Sample Violations (k) | Actual Violation Rate | Estimated Violation Rate |
|---|---|---|---|---|---|---|---|---|---|
| 241 | | | 4432 | 1031 | 4432 | | | 0.2326 | |
| 242 | | | 3675 | 777 | 3675 | | | 0.2114 | |
| 243 | | | 4145 | 909 | 4145 | 56 | 12 | 0.2193 | 0.2143 |
| 244 | | | 4249 | 1026 | 4249 | | | 0.2415 | |
| 245 | | | 3424 | 597 | 3424 | | | 0.1744 | |
| 246 | | | 3575 | 649 | 3575 | | | 0.1815 | |
| 247 | | | 438 | 99 | 438 | | | 0.2260 | |
| 248 | | | 379 | 85 | 379 | 56 | 15 | 0.2243 | 0.2679 |
| 249 | | | 441 | 89 | 441 | | | 0.2018 | |
| 250 | | | 445 | 99 | 445 | | | 0.2225 | |
| 251 | | | 412 | 93 | 412 | | | 0.2257 | |
| 252 | | | 350 | 74 | 350 | | | 0.2114 | |
| 253 | | | 386 | 80 | 386 | | | 0.2073 | |
| 254 | | | 417 | 78 | 417 | 56 | 10 | 0.1871 | 0.1786 |
| Totals: | 0 | 0 | 308730 | 61212 | 308730 | 840 | 157 | 0.198 | 0.195 Wgt |
| | | | | | | | | | 0.193 Raw |

Figure 3: Simulation Module.

will be modified to allow the user to classify each stratum and determine the sampling times for that stratum. This information is used by the Simulation Module to predict the number of vehicles that will be traveling on the stratum and the number of violations that will occur during sampling.

### *3.2 Sampling Design Module*

The Sampling Design Module is used to define the sampling strategy. The program screen representing the Sampling Design Module is shown in Figure 2. The data input to this module define how the Simulation Module will select the highway links and the number of vehicles to be sampled on each link. Also, the user can choose among several different performance measures and specify whether the prior information (e.g., traffic volumes) is known exactly or approximately. Each performance measure dictates how the simulation algorithm is executed and what data items are displayed as part of the Simulation Module.

### *3.3 Simulation Module*

In setting up the Simulation Module (Figure 3), the user enters the data that characterize the links in each stratum. The data that are displayed depend on which performance measure is selected in the Sampling Design Module. In Figure 3, the module displays a simulation of safety violations. Eventually, the program will be set up to display different simulations depending on the performance measure selected in the Sampling Design Module.

The Simulation Module is also where the user executes the simulation algorithm, views the simulated sampling protocol, and performs replicate sampling. The random sampling process is illustrated by "clicking" on the "Regenerate Sample" button. The new sample immediately appears on the screen. The "Replicate Sampling" function generates multiple simulated samples in order to evaluate the sampling istributions of estimators. The results are displayed in the Results Module. The user specifies the number of replicate samples to generate. Eventually, we foresee that there may be several different simulation algorithms, or at least "flavors" of the original algorithm, incorporating different performance measures. Currently, we only have one algorithm in this version of the software.

### *3.4 Results Module*

The Results Module presents a statistical summary of the simulated estimators. The screen representing the Results Module is shown in Figure 4. Currently, the module is set up to compare two methods for estimating violation rates – one based on simple averages and the other using statistical sampling weights. As we

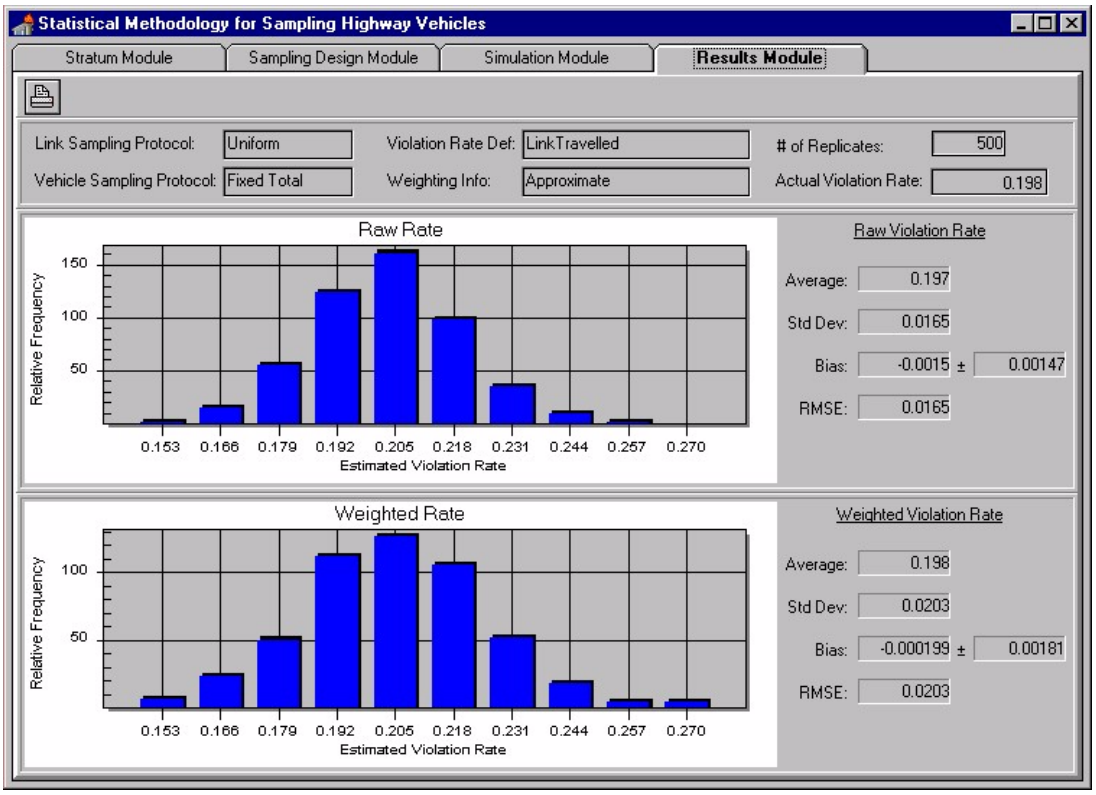



Figure 4: Results Module.

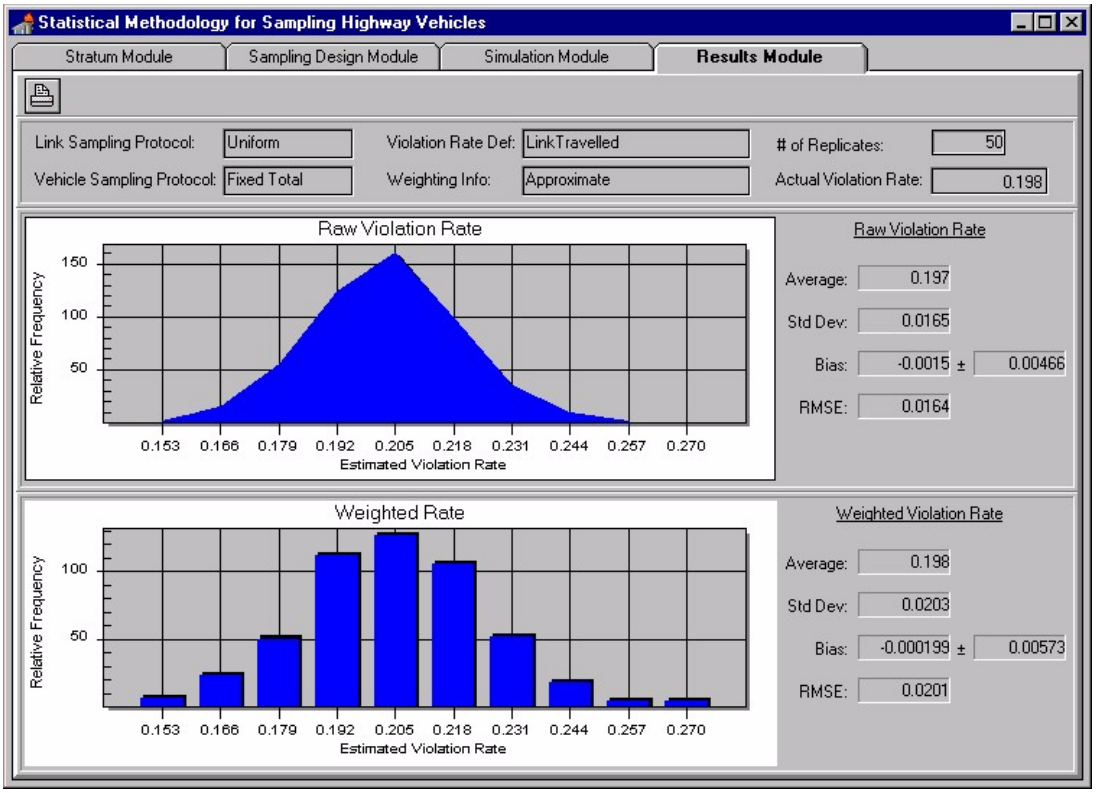



Figure 5: Results Module with New Title and Plot Style.

implement new performance measures and simulation algorithms, the Results Module will be expanded to include other comparison methods.

These plots show the sampling distributions of the two estimators. Specific properties, such as bias, standard deviation, and root mean squared error (RMSE) are also presented for each estimator. We developed the Results Module in such a manner to allow the user to customize these plots. For example, the user can change titles, fonts, or type of plotting routine in much the same way one makes these changes in a modern spreadsheet package. Figure 5 shows an alternative way to present the results of the simulation.

## 4. Software Development Environment

Because our goal was to develop an easy, intuitive user interface for the simulation program, we selected Microsoft Windows 3.x as our development platform. Throughout the development process, we made every effort to make the program feel like a typical Windows application. Most users would be familiar with these types of programs.

As our software development tool, we chose Delphi—a rapid application development (RAD) product from Borland, Inc. Delphi provides a highly optimized compiler and a visual programming environment that is excellent for prototyping applications. With the prototype approach, we were able to design and evaluate the user interface early in the development process. Changes to the interface could be made before developing the simulation and analysis routines. The prototype approach also allowed us to develop and unit test the program on a module by module basis.

Another reason for choosing Delphi is its drag &drop, component-based architecture. This was important to us because we could build our application using components that were developed and tested by Borland and other software vendors. Delphi includes such components as tabbed notebooks, database controls, grids, and edit boxes. A good example of a software vendor component is embedded in the Results Module. To provide the level of customization required, we included a plotting component called GigaSoft ProEssentials from Gigasoft, Inc. By incorporating these pre-tested components in our application we were able to greatly increase productivity while decreasing development costs. We chose Paradox from Borland, Inc. as our database because it provided fast data retrieval, a relational data model, and seamless integration with Delphi.

## 5. Future Plans

When we began the development of our simulation program, we recognized the need for a customized tool that could evolve as we explore different statistical design and analysis approaches. For example, we may want to look at different sampling designs, add the ability to include stratification, or evaluate different performance measures. Or, if we decide there are better ways to compare the simulation results, we can add them to the Results Module. With the software selected, we have complete flexibility to change nearly every aspect of the program.

We started the development process before Microsoft, Inc. released Windows 95. Because we believe that a 32 bit operating system will provide greater stability and enhanced program execution, we plan to port the software to operate with Windows 95 and Windows NT. Since these are the predominate platforms for today's desktop computing, we don't expect to develop the software to operate with any other operating system.

Other enhancements to the program include: incorporating the remaining performance measures, adding additional stratum classification parameters, and adding the ability to simulate a sampling protocol that includes multiple strata.

## References

*Traffic Monitoring Guide, 3rd ed.* (1995), U.S. Department of Transportation, Federal Highways Administration, Office of Highway Information Management.

Kinateder, J.G., McMillan, N.J., Orban, J.E., Skarpness, B.O., and Wells, D. (1997), "Sampling Designs and Estimators for Monitoring Vehicle Characteristics When Inspection Capacity is Limited," *Transportation Research Board Publication No. 971372*, to appear.

Stephen M. Wall
Battelle Memorial Institute
walls@battelle.org

John E. Orban
Battelle Memorial Institute
orbanj@battelle.org

# Advanced Perl Applications

By Phil Spector

In the previous article in this series, I gave a brief introduction to some of the features of perl that make it so attractive for rapidly developing applications like data formatting and report generating. In this article, I'd like to expand the range of possible applications by introducing the idea of packages. Basically, packages are collections of perl functions, either distributed with perl, or contributed by other authors. You include packages in your program with the `require` directive; perl then searches the directories stored in the `@INC` array to find the necessary programs, and then includes them. After that, you can refer to the functions in the packages in your program, and greatly expand the range of what you can do with your perl programs.

To illustrate these points, I'll provide simple programs to perform two interesting tasks; the first will use the system-provided chat2 package to communicate with the University of Michigan weather server, and the second will use Steven Brenner's cgi-lib package (available at `http://www.bio.cam.ac.uk/cgi-lib`) to produce and process an HTML form suitable for a web browser.

## *Communication with other Computers*

The University of Michigan operates a computer which provides weather reports for the entire United States. In normal use, you telnet to the machine `madlab.sprl.umich.edu` through port 3000, and make a variety of menu choices to find the information you want. After displaying a banner, the following prompt appears:

```
Press Return for menu, or enter 3
letter forecast city code:
```

At this point, I would select the code "`sfo`" for my local forecast, which would be displayed on the screen. After the forecast, it displays a menu entitled "`CITY FORECAST MENU`" and prompts for a menu selection with the phrase "`Selection:`". I thought it would be convenient to have a program that I could execute which would supply the city code, print the forecast, and then exit from the menu, without any intervention on my part.

```
#!/usr/local/bin/perl
require 'chat2.pl';

&chat'open_port("madlab.sprl.umich.edu",3000) || die "Couldn't open port";
*S = *chat'S;

# Change input separator to read up until the first prompt
$/ = "3 letter forecast city code:";
$str = <S>;

# The plus sign suppresses paging on this particular server
print S "sfo+\n";

# Now we can read the input until it says "CITY FORECAST MENU", so we'll
# make that the record separator
$/ = "CITY FORECAST MENU";

# read up to the prompt, and eliminate the final line
($str = <S>) =~ s#\n.*$/##;

print $str;   # print the weather report

$/ = "Selection:";
$str = <S>;
# Send an "x" to exit, and we're done!
print S "x\n";
```

To do this task in perl, one feature I repeatedly used is the ability to change the definition of the input record separator. This variable, which is known as `$/` in perl, defaults to newline, so that each record read by the <> input operator will normally read one line of text. Since the menu prompts from the weather server don't have newlines, I redefine the separator to read up until the next prompt, so at each stage, I can read everything I want with one execution of the input operator. Since the chat2 package opens a socket, I can use the same perl file handle to both read from and write to the remote computer.

To use the chat2 package, you first open the required port using the function `&chat'open_port`. The apostrophe in the name indicates that the `open_port` function is part of the chat package. This should return a non-null value, unless the program was unable to make a connection. Next, you associate a local file handle with the one defined within the chat2 package; from that point on, reading from the filehandle is equivalent to reading the output from the remote machine and printing to the filehandle is equivalent to typing responses to be interpreted by the remote machine. The perl program is listed on page 14. For communicating with standard protocols (like NNTP, HTTP, etc.), port numbers can be determined by using the system call `getservbyname` (provided in perl) or by looking at the file `/etc/services`.

## Processing HTML Forms

The general topics of HTML, forms and configuration of HTTP servers is beyond the scope of this article. An excellent introduction to the general topic of HTML can be found at `http://www.ncsa.uiuc.edu/General/Internet/WWW/HTMLPrimer.html`; an "instantaneous" introduction to forms and CGI scripts can be found at url http://kuhttp.cc.ukans.edu/info/forms/forms-intro.html.

The basic idea behind the CGI interface is that if you point your web browser at an executable script, and if the HTTP server on the machine where the script is located permits that script to execute, the script will be run with its printed output going back to the web browser which originally called it. HTTP servers are generally configured to only allow scripts in certain directories to be executed, so check with your server's administrator to find out where to install your scripts. The cgi-lib package provides a routine called `&ReadParse` which gets the input from a form and creates an associative array, indexed by the name of the input fields in the form, which contains the values which were entered in that particular field. Since the script's output is displayed in the browser which called it, most CGI programmers write what are known as "comboforms"; that is a script which both displays the form on the browser, and processes the results.

Before this starts sounding too complicated, we should look at an example, because it really is remarkable how simple it is to produce a comboform. In the example below (which is continued on the bottom of page 18), the perl subroutine `doform` is written to display the blank form. The `PrintHeader` subroutine of the cgi-lib

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

require 'cgi-lib.pl';

# The ReadParse routine from cgi-lib.pl creates an associative array
# containing all the variables entered from the html form

&ReadParse(*input);
if($input{"submit"}){
        print &PrintHeader;
        print("<html><title>Results from Sample Form</title><body>\n");
        print("<h1>Thanks for using the Sample Form!</h1><br>\n");
        foreach $k (keys(%input)){
                printf("%s:<BR>%s<P>",$k,$input{$k});
                }
        print("</body></html>");
        }
else{ &doform; }
```

package prints the necessary "Content-type" line to inform the world that it is producing an HTML document; all other fields need to be filled in by the script. The program starts by calling the `&ReadParse` routine. If the `submit` field of the form has a value, that means that the program is being called from a form; otherwise, it's being called directly, and it simply needs to print the necessary HTML to display the form. To keep this example simple, we'll just have one fill-in-the-blank field, and a set of three checkboxes, but any of the more complex form elements can be easily accommodated with this scheme. Similarly, the results are just printed back to the browser, but the full range of perl's capabilities could be used to do something more interesting with this information.

### Further Resources

This article doesn't begin to show you the wide variety of packages and scripts that are available to extend perl's capabilities. As always, the first place to start is the CPAN archives (`http://www.perl.com/perl/CPAN/CPAN.html`). Another source of scripts is the Metronet archive at `http://www.metronet.com/perlinfo/scripts`.

Phil Spector
UC at Berkeley
spector@stat.berkeley.edu



*TOPICS IN INFORMATION VISUALIZATION*

# Emphasizing Statistical Summaries and Showing Spatial Context with Micromaps

By Daniel B. Carr and Suzanne M. Pierson

### 1. Introduction

This article concerns redesigning a choropleth map. During my (Dan) fellowship at the Bureau of Labor Statistics (BLS), the staff showed me press releases with maps similar to the map in Figure 1. They did not like the map and asked me to develop new and innovative methods for displaying the data. My thoughts turned quickly to micromaps. In response to Tony Olsen's guiding query, I had previously proposed micromaps for linking row-labeled plots to ecoregion maps (Olsen, Carr, Courbois, and Pierson 1996). Sue Pierson's first implementations and Pip Courbois' variations demonstrated to the team that micromaps do more than provide links. Micromap sequences directly reveal spatial patterns. Below is our rationale for evolving from a traditional choropleth map to a new and powerful template that links micromaps and row-labeled plots.

CONTINUED ON THE NEXT PAGE

```
sub doform{
# The action field in the form below is the location of this script.

print &PrintHeader;
print <<EOF;
<html>
<title> Sample Form </title>
<body>
<h3 align=center><b> Welcome to the Sample Form! </b></h3>
<form method="post" action="http://wherever/cgi-bin/sampleform.pl">
Fill in the blank:<br>
My name is <input name="name" size=40>
<p>Choose your favorite color:<br>
  <input name="check" type=checkbox value="Blue"> Blue<br>
  <input name="check" type=checkbox value="Green"> Green<br>
  <input name="check" type=checkbox value="Red"> Red<br>
<p>
<input type=submit name="submit" value="submit">
</form>
</body> </html>
EOF
}
```

## 2. Visual and Representational Problems

The choropleth map in Figure 1 has several visual and representational problems. Visually, the map is reminiscent of pen plotter era. Representing values by line density and crosshatched patterns is predicated upon notions of reproduction ease and cost, not upon notions of aesthetic communication. That is, lines copy better than half-toned gray and less expensively than color. Questions about state grouping reveal the groups to be administrative divisions that have little bearing on analysis. The state grouping can be dropped. However, the problems extend beyond unaesthetic appearance and irrelevant grouping.

When evaluating Figure 1, consider the story to be presented. The story focuses on sample-based estimates of unemployment rates with associated uncertainty estimates, counts of unemployed, and spatial indices. Directly to the point, The Power of Maps by Wood (1992) contains an intriguing chapter entitled "Every Map Shows This ... But Not That." If we look though statistical eyes at what is and is not represented in Figure 1, we see cartographic bias and representational problems.

The cartographic bias in Figure 1 reflects representational choice and relative emphasis. The cartographic literature (for example, see Bertin 1983, MacEachren 1994) provides systematic treatment of ways to represent variables on maps. Cartographic choices implicitly assume that the best representation, position along a common scale, is devoted to the two spatial coordinates. This leaves second best choices to show data values and other information. The systematic treatment fails to show scatterplot alternatives favoring statistics, e.g., melanoma rates on the y axis, cloud-free days on the x axis, latitude encoded as circle area and longitude encoded as circle color. Carr, Littlefield, Nicholson, and Littlefield (1987) provide an early example of a balanced representation. They use position along a common scale for data values, spatial coordinates, and time while linking subsets across panels with color. Our redesign of Figure 1 also uses position along a common scale for both data values and spatial coordinates, improving representation of the statistical information.

The traditional cartographic choice plays out in terms of emphasis. Figure 1 emphasizes state boundaries. The often-used Albers projection preserves the relative areas of the continental U.S. states. A large number of vertices is devoted to boundary representation and substantial graphic space is dedicated to representing state area. In contrast, detail associated with unemployment rates is limited to a few class boundaries. The unemployment rates appear as a caricature because the conversion to class intervals adds noise.

Class interval options are caricature options for statistical distributions. Common choices include equal size intervals and gap-based intervals covering the range. Even for percentage point options, the default is often based on the number of regions. Carr and Olsen (1995a and 1995b) argue for class intervals based on percentage points of a cumulative distribution function chosen for interpretation purposes, such as the percent of people involved. Carr and Olsen also propose a visual summary of the distribution that can appear in a small legend. In the current case, only fifty-one estimates are to be represented. We take the radical approach of directly showing all estimates.

The representational problem is more than caricaturizing the unemployment rates. The map does not show estimate uncertainties. While MacEachren (1994) describes methods for representing uncertainty on maps, they are rarely used. (A notable exception is Pickle, Mungiole, Jones, and White 1996). Beyond failing to provide details about estimate precision, the omission of estimate uncertainties is serious on two counts.

First, the presence of uncertainties suggests that sound statistical methodology produces the estimates. The world is awash in convenience-based guesstimates. The public needs clues to decide if data is statistically sound. Omission of clues helps politics and sales compete on an equal footing with science. Frederick Mosteller (*) says, "It is easy to lie with statistics, but easier to lie without them." A corollary is: it is easy to lie with confidence intervals, but it is much easier to lie without them.

A second consideration is that the public needs to be educated about uncertainty. People may not like the probabilities of weather forecasters, but over time the reporting convention has become familiar. The failure to show confidence intervals for estimates is a missed chance to educate the public.

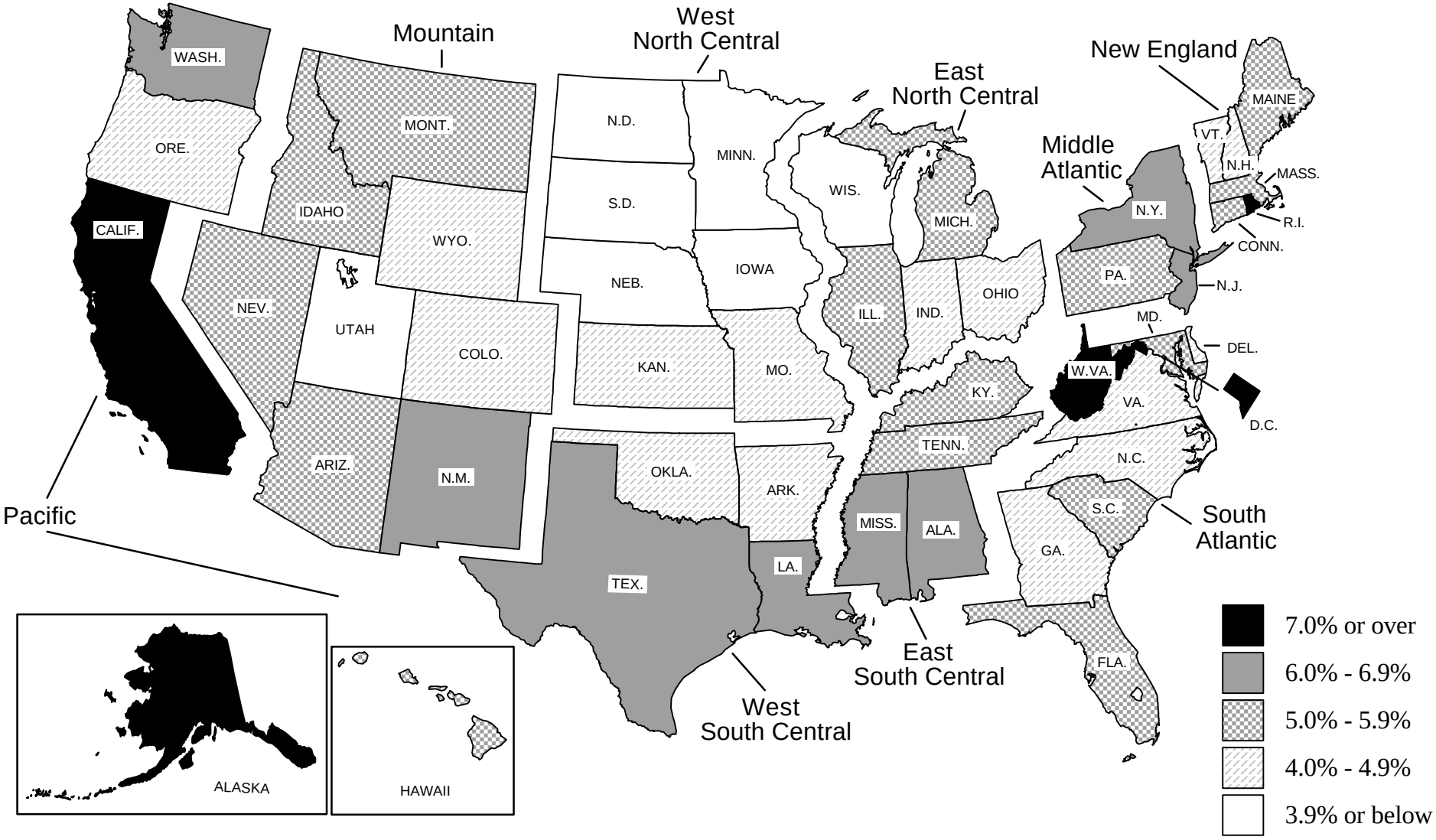
The most important design task is to represent the statistical summary. The spatial component of the summary is important, but secondary. The new design should reflect this priority.

## 3. Dot Plots, Visual Simplicity and Grouping Considerations

The variable of interest is the state unemployment rate. Labeled dot plots (Cleveland 1985, 1993) provide a good way to show such estimates. Unfortunately, dot plots have been slow to appear in government publications. The efforts of Carr, Valliant, and Rope (1996) are intended to help remedy the situation by providing

# Unemployment rates by state, 1995 annual averages

**(U.S. rate = 5.6 percent)**



JAVA-based network tools and examples using government data. Our effort extends the scope of the examples. The design goals for the current example include 1) adding information to ease interpretation and 2) simplifying visual appearance to facilitate communication with the public. This section addresses the second task, simplifying visual appearance.

For improved visual interpretation, Cleveland (1985) advocates presenting dots in sorted order. Carr (1994) and Carr and Olsen (1996) echo this advice and note that sorting improves plot simplicity by reducing the visual path between dots.

One can further simplify plot appearance. Important visual simplification techniques include grouping and layering. For example, see Kosslyn (1994) for a discussion about grouping and Tufte (1983 and 1990) for discussions of small multiples and layering. Graphics software has not successfully automated thoughtful grouping and layering, so some thought about the current example is instructive.

A list with fifty-one lines can be visually intimidating (see Carr and Olsen 1996). By analogy, a fifty-one line paragraph may visually intimidate many people. One can lose one's place. Breaking a long paragraph into short paragraphs helps the reader with visual tracking. Similarly, breaking a list of names into groups helps the reader in visual tracking. A further benefit of creating smaller perceptual groups is that readers can easily spot names at group edges. Spotting an interesting name draws the reader into the graphic. In other words, small perceptual groups increase the number of interest-based entry points.

How do we divide fifty-one states into visually effective groups? (Here we include the District of Columbia as a state, but it is often preferable to treat D.C. as a city.) One choice would be to partition the states using large jumps in the sorted rates. This has merit, but can get awkward when many states have similar values. Our approach starts with regular partitioning into groups of five. (For many applications Kosslyn (1994) recommends groups of four or less.) For vertical grouping we find that groups of five facilitate counting and still allow quick label and value matching by relative position. For example, one can easily match the third label in a group of five labels to the third dot in a group of five dots. This works even when the labels and dots are separated by nearly a page width. Consequently, grouping obviates the need for the horizontal dots that Cleveland (1985 and 1993) used to assist in visual tracking. The partitioning produces ten groups of five and one group of one.

The first layer of grouping produces eleven groups. Eleven groups are too many to put into a single simple-appearing perceptual unit. The groups need to be grouped. As shown in Figure 2, we create an additional information layer using three larger groupings with a 5-1-5 pattern. This creates symmetry and calls attention to the median.

The basic layout for Figure 2 is an eleven row (5-1-5 pattern) by four column matrix. The state names appear in the second column and unemployment rates with 95 percent confidence intervals appear in the third column. (The confidence intervals are model-based and subject to refinement. The BLS has not extended the confidence interval calculations to the seasonally adjusted estimates it often shows.) The confidence intervals detract a bit from the goal of visual simplicity, but the above discussion motivates their inclusion.

Representation of confidence intervals is a design challenge. Typical error bars draw visual attention to the least precise estimates (Carr 1994). Further, error bars detract from the visual flow in following the estimates. In Figure 2, the interval endpoints appear as small gray dots. Connecting adjacent endpoints with a thin black line reinforces desirable vertical flow and brackets the group of estimates. The pinch points call attention to the most precise estimates. In other examples, an estimate dot can overplot the confidence interval dots. In isolated cases, the slope of confidence interval lines from above or below can suggest the size of the hidden interval. In the new design, even those new to confidence intervals may surmise a connection between the California pinch point and the large number unemployed. Perhaps better approaches will emerge, but the approach in Figure 2 has considerable merit.

The next steps are to show the remaining secondary information. The information includes spatial positions and estimated numbers of people unemployed. The final step is to link everything together with the state names.

## *4. Micromap Design*

Associated with each unemployment rate is a spatial position, the state location. In the current example, as in many studies, the exact spatial position and precise boundaries are not important. All that is needed is a map caricature showing the general position and neighborhood relationships. As demonstrated in the first column of Figure 2, a micromap for each group of five can show the state locations. A full page map is not required.

The design of small maps requires attention because distinguishing hues in small regions can be difficult. The map caricature needs to enlarge small regions while retaining enough features to provide region recognition.

The task is not as simple as it might seem. After several independent attempts, we chose to modify a state visibility map developed over a decade ago by Monmonier and illustrated in *Mapping it Out* (Monmonier 1993).

Figure 2 shows ten micromaps on a standard page in portrait orientation. Making the micromaps much smaller will complicate color perception for relatively small states like Rhode Island. The example is pretty close to the minimum size limit. Observe that Illinois has the median rate and appears in black in the two middle maps. This avoids the need for an eleventh micromap.

## 5. Related Data

Carr, Valliant and Rope (1996) argue that graphics should provide metadata to facilitate proper interpretation. The current example is static, so we cannot exploit web-based access to the Bureau of Labor Statistics Handbook of Methods or to other information. However, the design readily accommodates one or two additional columns. Figure 2 shows the number of unemployed in the fourth column. While unemployment rates are useful for comparison purposes, the number of unemployed shows the importance of the rates in terms of human lives.

The confidence intervals for the counts were not available for this article. As an approximation, the rate intervals could be scaled by the population size. We chose to focus attention on the rates and to omit confidence intervals for the secondary information. The elegance of the confidence interval representation depends on sorting. Adding confidence intervals to the count estimates competes with the goal of visual simplicity.

## 6. Region Labels and Linking

In Figure 2, the state labels link the information. The relative vertical position of a label and a dot within a group of five is an adequate positional link. The colored dot beside the state name may not increase the matching speed over the positional link, but may remove doubt about a correct match. The colored dot in the label is an important link to a particular state. Some may assume that everyone knows state positions and question the need for a color link. However, many informed U.S. citizens may hesitate when labeling all the states on a map. Further, altered boundaries in the map caricature may slow recognition for some states. We conjecture that color links to the map increase matching speed and sometimes provide an educational tool. The color links go both ways. Some people use maps rather than names to find their state and the corresponding estimate.

With the holiday season as motivation, we seriously considered using colored names in the label column and dropping the dots. This drops one visual element and simplifies plot appearance. However, Monmonier (1993) recommends against colored labels because they need to be large enough to carry color and are difficult to read. Switching to a bold font in Figure 2 works fine for carrying color. We concur that changing colors makes reading a little harder. Further, the colorful names draw visual attention away from other columns. Still, we think the problems are minor and that colored names might be used on occasion to add variety.

## 7. Interpretation and Comparison

Figure 2 provides much more statistical information than Figure 1. Consider four questions.

1. What is the unemployment rate for California?
2. Is the unemployment rate higher for California or Alaska?
3. What are the confidence bounds for the California estimate?
4. On the average how many were unemployed in California?

For the first question, Figure 2 provides a more precise determination of the estimated rate. For perceptual accuracy of extraction (Cleveland and McGill 1984), dot plots with grid lines are a hard graphic to beat. For question two, Figure 2 provides a complete ranking while Figure 1 only provides a ranking of equivalence classes. Figure 1 does not provide answers for questions three and four.

The traditional choropleth map does not do as well as the linked micromap row-labeled plot template in regard to the above questions. This motivates the search for tasks in which the large choropleth map has performance advantages. Two tasks seem evident, finding the value of a position-known state and locating the values of neighboring states. The micromap template requires a scan of small maps or a list of fifty-one names before linking to a value. The scan is a slow process. An additional alphabetic-label position-link column (not illustrated here) may speed locating a given state by name. The position link following a name can be little black dots in a 5-1-5 pattern with the location dot highlighted by color (or shape). For Hawaii, the third dot would be red and link to red in the third vertical group. The real memory and search intensive task in using micromaps is to find values for a state's neighbors. With micromaps one can quickly observe if neighboring states have similar rankings, but that is not the same as finding all the values. Traditional choropleth maps have a few advantages.

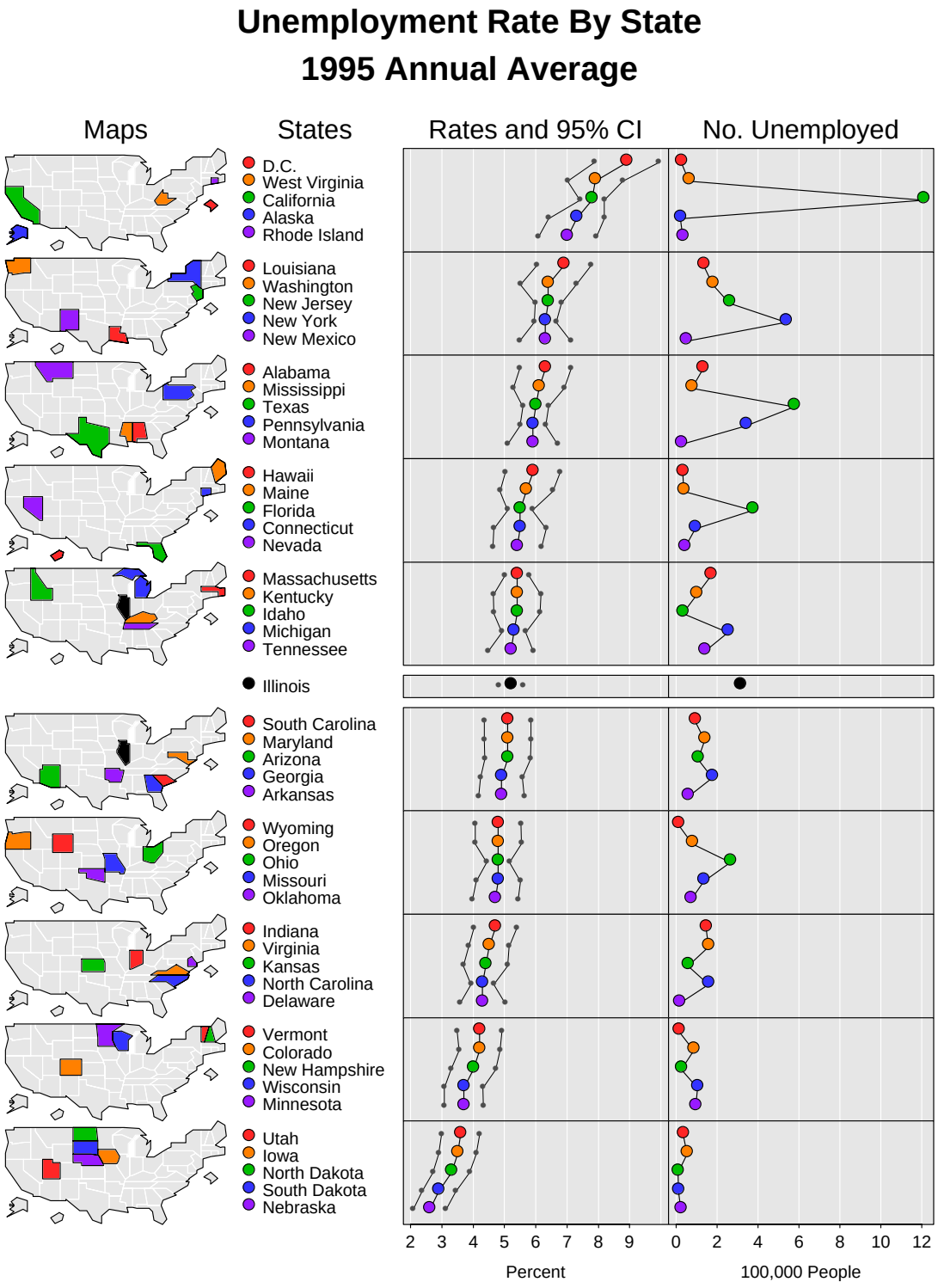



Figure 2: Improved display with micromaps.

Cartographers may also suggest that a single choropleth map provides better global spatial pattern perception than a sequence of micromaps. Since integrating information while scanning across all the micromaps is nontrivial, the claim is likely true. However, many important tasks involve local pattern perception.

For tasks involving local pattern perception, micromap sequences may be very competitive to a single choropleth map. While the colors in Figure 2 may be distracting, ten micromaps provide the rough equivalent of ten class intervals. For sound perceptual reasons, typical choropleth maps show six or fewer class intervals. This complicates direct comparison in cognitive testing. Different class intervals bring out different patterns, so there may be no clear winner.

The micromap patterns can be quite suggestive. Figure 2 shows many small groups and raises questions about economic similarities. As two of several examples, Vermont and New Hampshire form a pair while Virginia and North Carolina form another. The two bottom micromaps show a larger group of states in the upper Midwest and Northern Plains. It is not hard to integrate patterns across two small juxtaposed maps. While beyond the domain of local pattern perception, the micromap sequence wins hands down when it comes to ranking. The cyclic colors in a rough spectral order break the ties within the ten micromaps. The combination of multiple maps and color provides a complete ranking without reference to the other columns.

Variations on micromaps may strengthen global spatial pattern perception or at least bring out additional patterns. A darker shade of gray can distinguish all states above median and provide another layer of information in the top five micromaps. The darker gray region appears the same in all five maps, except for the overriding hue-linked states that provide compositional detail for the high unemployment region. New patterns emerge. For example Appalachian states have above average rates. A similar approach accentuates the low unemployment region in the bottom five micromaps.

## *8. Closing Remarks*

Viewing graphics as puzzles to be solved is often instructive. How do the pieces fit together and what do they mean? We conjecture that the template illustrated by Figure 2 is a puzzle accessible to many. Learning to read a dot plot is easy. Linking by color and position is simple. Determining that the visual islands are Alaska, Hawaii, and D.C. should be manageable by those not familiar with the U. S. A deep understanding of confidence intervals goes beyond the graphic, but readers new to statistics may surmise that the big dots for rate estimates are not exactly THE TRUTH. For those with a little background, the figure is pretty close to being self-explanatory.

Figure 2 has educational merit beyond showing unemployment rates. One can learn about the positions of the states. The figure can prepare people to answer the question, to what side does the median belong? More importantly, the figure can prepare people to learn about uncertainty. To the thoughtful, the figure suggests that rank orderings are a bit arbitrary in the presence of uncertainty. The figure also suggests that rate magnitude and rate importance are distinct concepts. The template of linked micromaps and row-labeled plots extends to many other spatial contexts. For example, one can show county data within a state. When fifty or fewer counties will be displayed, a layout similar to Figure 2 will likely suffice. The main challenge would be to develop a county within state visibility map. Some states are not easy.

Many variations of the template are possible besides those mentioned above. Some may prefer a different set of hues. For example, one can pick a set designed for the color blind. A good reference concerning color choice and mapping is Brewer (1994). Other variations may slightly improve perceptual accuracy of extraction for dot plots. For example translucent dots in the right two panels will help keep the grid lines visible. A small, similar hue dot inside each big dot may help locate the dot center precisely without being too distracting. The possibilities are numerous.

The data and Splus source code used to generate Figure 2 are available through anonymous ftp (`galaxy.gmu.edu`). The newsletter software directory seems to change periodically. The current path for this article is `pub/dcarr/newsletter/micromap`. For Splus users, the matrix layout tools should be of interest by themselves and a technical report with documentation details should be available in the same time frame as this article.

I (Dan) continue to seek design challenges and opportunities for collaboration. Also, I appreciate gentle comments about potential improvements. Please contact me at the address below.

## Acknowledgments

Conceptual work behind this paper was supported by the EPA under cooperative agreement No. CR8280820-01-0. Specific work on this application was supported by the BLS. The article has not been subject to review by either the EPA or the BLS, does not necessarily reflect the view of the agencies, and no official endorsement should be inferred. Special thanks go to Tony Olsen for the design challenges, Pip Courbois for experiments in micromap variations, and to Rick Valliant and Dan Rope for coordinating the BLS data and graphics access.

## References

Bertin, J.B. (1983), *Semiology of Graphics Diagrams Networks Maps*, Translated by Berg, W.J., London, UK: The University of Wisconsin Press.

Brewer, C.A. (1994), "Color Use Guidelines for Mapping and Visualization," in *Visualization in Modern Cartography*, 123–147, eds. MacEachren, A.M., and Taylor, D.R.F., Oxford, UK: Pergamon/Elsevier Science.

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

Carr, D.B., Littlefield, R.J., Nicholson, W.L., and Littlefield, J.S. (1987), "Scatterplot Matrix Techniques For Large N," *Journal of the American Statistical Association*, 82, 424–436.

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

Carr, D.B., and Olsen, A.R. (1995b), "Parallel Coordinate Variants of CDF and Quantile Plots," *Statistical Computing and Statistical Graphics Newsletter*, Vol. 6, No. 1 13–18.

Carr, D.B., and Olsen, A.R. (1996), "Simplifying Visual Appearance By Sorting: An Example Using 159 AVHRR Classes," *Statistical Computing and Statistical Graphics Newsletter*, Vol. 7, No. 1, 10–16.

Carr, D.B., Valliant, R., and Rope, D. (1996), "Plot Interpretation and Information Webs: A Time-Series Example From the Bureau of Labor Statistics," *Statistical Computing and Statistical Graphics Newsletter*, Vol. 7, No. 2 19–26.

Cleveland, W.S., and McGill, R. (1984), "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods," *Journal of he American Statistical Association*, 79, 531–554.

Cleveland, W.S. (1985), *The Elements of Graphing Data*, Summit, NJ: Hobart Press.

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

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

MacEachren, A.M. (1994), *Some Truth with Maps: A Primer on Symbolization &Design*, Washington D. C.: Association of American Geographers

Monmonier, M. (1993), *Mapping It Out*, Chicago, IL: The University of Chicago Press.

Olsen, A. R., Carr, D.B., Courbois, J.P., and Pierson S.M. (1996), "Presentation of Data in Linked Attribute and Geographic Space," Poster presentation, ASA Annual Meeting, Chicago, Il.

Mosteller, F. (*), Personal communication—The quote is not from a publication but likely originated in a speech.

Pickle, L.W., Mungiole, M., Jones, G.K., and White A.A. (1996), *Atlas of United States Mortality*, Hyattsville, MD: Public Health Service Pub. No. 97-1015.

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

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

Woods, D. (1992), *The Power of Maps*, New York, NY; The Guilford Press.

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

Susanne M. Pierson
Anteon Corporation
spierson@heart.cor.epa.gov



# Scientific Program

by James Rosenberger and Stephen Eick

## Statistical Computing

At the Joint Statistical Meetings in Chicago, in August 1996, the Statistical Computing Section sponsored 5 invited sessions and 17 contributed paper sessions. In addition there were 16 posters and 5 roundtable luncheons sponsored by the section.

The invited sessions were organized by Rob Tibshirani, Program Chair, and included the following stimulating themes. A session on Monday focused on "Bayesian Inference for High-Dimensional Problems," chaired and organized by Radford Neal, it covered topics on A Multiresolution Time Series Model; Bayesian Regression using Gaussian Process Priors; and Spatial Subordinates and Density Estimation. Art Owen organized a session on Tuesday titled "Statistics and Numerical Integration" which included talks on: Monte Carlo and Quasi-Monte Carlo; Number-Theoretic Methods in Multivariate Statistical Tests; and Lattices and Dual Lattices in Experimental Design for Fourier Models. Also on Tuesday a session on "Wavelets and Time-Frequency Analysis," organized by Jonathan Buckheit, included talks on: Curve Estimation with the Stationary Wavelet Transform; Improved Local Discriminant Bases Using Empirical Probability Estimation; and Alternatives to Karhunen-Loeve Expansions Using Libraries of Orthogonal Bases. A Wednesday session organized by Martin Schumacher on "Resampling and Cross-Validation in Model Building" included talks on Cross-Validation in Survival Analysis, The Bootstrap and Modulation Estimators, and Reduction of Bias Caused by Model Building. The final invited session, "Algebraic Algorithms" organized by James Stafford, included talks on: Symbolic Ito Calculus: A Further Application to the Statistical Theory of Shape; Stochastic Differential Equations; A Computer Algebra for Sample Survey Theory; and An Operator for the Symbolic Calculation of Properties of Bootstrap Estimates.

The contributed paper sessions included many excellent presentations, nicely organized around coherent themes, due to the large number of submissions. Excellent sessions were organized on Linear Models and Regression, Analysis of Variance and Mixed Models, Tree-Based and other Models, EM Algorithm and Simulation, Bayesian and Hierarchical Bayes Methods, Multivariate Analysis, Markov Chain Monte Carlo Implementation, Clustering, Classification, Time Series and Nonparametrics. Noteworthy sessions also included the Special Contributed Session with the winners of the Student Paper Competition sponsored by the Statistical Computing Section. And a special co-sponsored session highlighted presentations from the Undergraduate Data Analysis Contest.

As program chair for the 1997 meeting, I look forward to your many contributed paper and poster submissions to complement the invited program, and am putting together an exciting program for the Anaheim meeting. Expect to see more electronic presentations, and special poster sessions highlighted this year. Feedback and suggestions for improving the meetings are always welcome. A preliminary look at the program will be provided in the next Newsletter.

James L. Rosenberger
*Statistical Computing*
*1996 Program Chair-Elect;*
jlr@stat.psu.edu

## Statistical Graphics

The 1996 summer meetings featured four invited Statistical Graphics sessions, Information Visualization: The Next Wave In Statistical Graphics, organized by Steve Eick, Innovations in Graphics, again organized by Steve Eick, Statistical Graphics and Multimedia Education, organized by David Scott, and Transactional Data Analysis, organized by Daryl Pregibon. As past graphics program chair, I now appreciate how difficult it is to organize new and innovative invited sessions. Please, those of you with session ideas, become involved. The chair needs your help! In my opinion, as you might expect, all of the this year's sessions were outstanding.

Visualization, of which Statistical Graphics is part, is an exploding research area. A new IEEE journal recently started (March 1995), IEEE Transactions on Visualization and Computer Graphics, and has over 6 thousand subscribers. A new symposium, Information Visualization, is now in its third year. Much of the interest in this area has been generated by the explosive growth of the World-Wide-Web, widespread availability of networked databases, and recent surge of interest in Data Mining. Computer Scientists, particularly those involved with Human Computer Interaction, have been building many novel and dynamic interfaces for displaying data.

One of the themes for this years sessions was to invite some of the well-known computer scientists to participate in our meetings. I occasionally see powerful and effective statistical graphical displays that lack the engaging user interfaces that characterize much of the work in

computer science. Conversely, the computer scientists' tools often lack the data analysis sophistication that we take for granted. By bridging the two communities, I hoped to encourage progress in both disciplines and to broaden our respective perspectives. Several of the sessions were followed by intense and stimulating discussions, suggesting progress. I welcome your thoughts, feedback, and suggestions.

Stephen G. Eick
*Statistical Graphics, 1996 Program Chair*
eick@bell-labs.com



# Business Meeting and Mixer at JSM 96

By Nandini Raghavan

The Statistical Graphics and Computing Sections held their joint business meeting and mixer on the evening of Monday, Aug 5-th. With snacks and wine aplenty, the buzz of conversations filling the air and the pleasant anticipation of prizes to be won later, it made for a very enjoyable evening. Sallie Keller-McNulty, Chair, Statistical Computing Section and William DuMouchel, Chair, Statistical Graphics Section informed the attendees about the happenings in their sections. And then, it was time for the raffle drawings! The door prizes were generously donated by the following exhibitors. Their email addresses and URLs can be found through the section home page at: http://orion.oac.uci.edu/~rnewcomb/statistics/graphics/graphics.html

- Academic Press Inc.
- Arnold Publishers
- Chapman &Hall
- Conceptual Software, Inc.
- Current Index to Statistics
- Data Description, Inc.
- The Interface Foundation
- Mathsoft, Inc.
- McGraw-Hill
- NCSS Statistical Software
- Resampling Stats
- Sage Publications, Inc.
- SAS Institute, Inc.
- Scientific Computing Associates
- SIAM: Contact Vickie Kearn, kearn@siam.org
- Springer-Verlag New York, Inc
- SPSS, Inc
- Statistical Graphics Corp.
- Texas Instruments
- John Wiley &Sons Inc, Publishers

Nandini Raghavan
The Ohio State University
raghavan@stat.ohio-state.edu



## *CONFERENCE NOTICES*

# The Third North American Conference of New Researchers

July 23-26, 1997

Laramie, Wyoming.

The purpose of this meeting is to provide a venue for recent Ph.D. recipients in Statistics and Probability to meet and share their research ideas. All participants will give a short expository talk or poster on their research work. In addition, three senior speakers will present overview talks. Anyone who has received a Ph.D. after 1992 or expects to receive one by 1998 is eligible. The meeting is to be held immediately prior to the IMS Annual Meeting in Part City, Utah (July 28–31, 1997), and participants are encouraged to attend both meetings. Abstracts for papers and posters presented in Laramie will appear in the IMS Bulletin.

The New Researchers' Meeting will be held on the campus of the University of Wyoming in Laramie, and housing will be provided in the dormitories. Transportation to Park City will be available via a charter bus. Partial support to defray travel and housing costs is available for IMS members who will also be attending the Park City meetings, and for members of sponsoring sections of the ASA. The Section on Statistical Computing is one of the many sponsors of this meeting and applicants who are members of this section should indicate this on their application. Additional information on the conference and registration is available at the website: http://www.math.unm.edu/NR97.html. Or contact Prof. Snehalata Huzurbazar, Department of Statistics, University of Wyoming, Laramie, WY 82071-3332, USA; e-mail: lata@uwyo.edu; fax: 307-766-3927.

This meeting is sponsored in part by the Institute of Mathematical Statistics; the National Science Foundation, Statistics and Probability Program; and the ASA Sections on Bayesian Statistical Sciences, Statistical Computing, and Quality and Productivity.

# New York University Symposium on Recent Developments in Smoothing Methods

Friday, May 30th, 1997

• Sponsored by the Department of Statistics and Operations Research, Leonard N. Stern School of Business
• Cosponsored by the Section on Statistical Computing, American Statistical Association

The existence of high speed, inexpensive computing has made it easy to look at data in ways that were once impossible. These computational advances have led to great interest in what might be termed flexible models, where strict parametric forms are replaced with smooth representations of underlying patterns and relationships.

The symposium will bring together leading researchers in the theory and practice of smoothing methods. The objective is to create a forum for the discussion of recent developments and long standing issues in the field. All interested researchers and practitioners are welcome.

Speakers and tentative titles of talks:

• *Assessing lack of fit for parametric regression models via nonparametric regression techniques*
R. L. Eubank (Texas A &M University)

• *A versatile approach to local modeling*
J. Fan (University of North Carolina)

• *Extended linear modeling and an application to speech recognition*
C. L. Kooperberg (University of Washington)

• *Bayesian wavelet shrinkage*
J. S. Marron (University of North Carolina)

• *Remarks on making smoothing methods work better in several dimensions*
D. W. Scott (Rice University)

For additional information contact Jeffrey S. Simonoff:

| | |
|---|---|
| phone: | (212) 998-0452 |
| Fax: | (212) 995-4003 |
| E-mail: | `symposium97@stern.nyu.edu` |
| WWW: | `http://www.stern.nyu.edu/SOR` (Click on "Events") |

# INTERFACE '97




## *29th Symposium on the Interface: Computing Science and Statistics*

May 14-17, 1997

Holiday Inn - Houston Medical Center
Houston, Texas

The theme of Interface '97 is "Mining and Modeling of Massive Data Sets in Science, Engineering, and Business," with subthemes in Environmental Statistics and Graphics. The keynote speaker is Jerry Friedman, Stanford University. Over 25 invited paper sessions have been organized, and contributed papers are sought. Partial funding for young investigators and students may be available, subject to final grant funding.

The meeting is sponsored by the Interface Foundation of North America. Cooperating Institutions are the American Statistical Association, the Institute for Mathematical Statistics, the International Association for Statistical Computing, the Society for Industrial and Applied Mathematics, the Institute for Operations Research and the Management Sciences, and the Biometrics Society (ENAR and WNAR).

Organizer and program chair is David W. Scott, and host institutions include Rice University and M.D. Anderson Cancer Institute. Full information on hotel, program, exhibitors, and activities can be obtained accessing the World Wide Web site `http://www.stat.rice.edu/`, or sending e-mail to `interface97@stat.rice.edu`, or writing to Interface '97, c/o David W. Scott, Department of Statistics, MS-138, Rice University, 6100 Main Street, Houston, TX 77005-1892, USA., 713-527-6037.

## SECTION OFFICERS

### Statistical Graphics Section - 1996

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

**Michael M. Meyer,** Chair–Elect
412-268-3108
Carnegie Mellon University
MikeM@stat.cmu.edu

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

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

**Edward J. Wegman,** Program Chair–Elect
George Mason University
703-993-1680
ewegman@gmu.edu

**Mario Peruggia,** Newsletter Editor (96-97)
614-292-0963
Ohio State University
peruggia@stat.mps.ohio-state.edu

**Robert L. Newcomb,** Secretary/Treasurer (97-98)
714-824-5366
University of California, Irvine
rnewcomb@uci.edu

**Michael C. Minnotte,** Publications Officer
801-797-1844
Utah State University
minnotte@math.usu.edu

**Lorraine Denby,** Rep.(96-98) to Council of Sections
908-582-3292
Bell Laboratories
ld@research.att.com

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

**Roy E. Welsch,** Rep.(97-99) to Council of Sections
MIT, Sloan School of Management
617-253-6601



### Statistical Computing Section - 1996

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

**Karen Kafadar,** Chair–Elect
303-556-2547
University of Colorado-Denver
kk@tiger.cudenver.edu

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

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

**Russel D. Wolfinger,** Program Chair–Elect
SAS 919-677-8000
sasrdw@sas.com

**Mark Hansen,** Newsletter Editor (96-98)
908-582-3869
Bell Laboratories
cocteau@research.att.com

**Evelyn M. Crowley,** Secretary–Treasurer (97-98)
317-494-6030
Purdue University
crowley@purdue.edu

**James S. Marron,** Publications Liaison Officer
919-962-5604
University of North Carolina, Chapel Hill
marron@stat.unc.edu

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

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

**Terry M. Therneau,** Rep.(97-99) Council of Sections
507-284-1817
Mayo Clinic
therneau@mayo.edu

**Naomi S. Altman,** Rep.(97-99) Council of Sections
607-255-1638
Cornell University
naomi_altman@cornell.edu



## *INSIDE*


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

SPECIAL FEATURE ARTICLE
- Computational Anatomy: An Emerging Discipline 1

EDITORIAL . . . 2

CALL TO ACTION
- Data Mining . . . 8

BITS FROM THE PITS
- A Simulation Tool For Evaluating Design And Analysis Options For Monitoring Highway Traffic Characteristics . . . 9

UNIX COMPUTING
- Advanced Perl Applications . . . 14

TOPICS IN INFORMATION VISUALIZATION
- Emphasizing Statistical Summaries and Showing Spatial Context with Micromaps . . . 16

REPORTS FROM JSM 96
- Scientific Program . . . 24
- Business Meeting and Mixer at JSM 96 . . . 25

CONFERENCE NOTICES
- The Third North American Conference of New Researchers . . . 25
- New York University Symposium on Recent Developments in Smoothing Methods . . . 26
- Interface '97 . . . 26

SECTION OFFICERS
- Statistical Graphics Section – 1997 . . . 27
- Statistical Computing Section – 1997 . . . 27



# Statistical COMPUTING & GRAPHICS

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



Mark Hansen
*Editor, Statistical Computing Section*
Statistics Research
Bell Laboratories
Murray Hill, NJ 07974
(908) 582-3869 • FAX: 582-3340
cocteau@bell-labs.com
http://cm.bell-labs.com/who/cocteau/

Mario Peruggia
*Editor, Statistical Graphics Section*
Department of Statistics
The Ohio State University
Columbus, OH 43210-1247
(614) 292-0963 • FAX: 292-2096
peruggia@stat.ohio-state.edu
http://stat.ohio-state.edu/~peruggia/



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

American Statistical Association
1429 Duke Street
Alexandria, VA 22314-3402 USA
(703) 684-1221 • FAX (703) 684-2036
asainfo@amstat.org

Nonprofit Organization
U. S. POSTAGE
**P A I D**
Permit No. 50
Summit, NJ 07901

American Statistical Association
1429 Duke Street
Alexandria, VA 22314-3402
USA

This publication is available in alternative media on request.