

Combining Usability Techniques to Design Geovisualization Tools for Epidemiology

**Anthony C. Robinson, Jin Chen, Eugene J. Lengerich,
Hans G. Meyer, Alan M. MacEachren**

GeoVISTA Center
Department of Geography
302 Walker Building
The Pennsylvania State University
University Park, PA 16802, USA

Phone (814-865-3433) **Fax** (814-863-7943)
Email [arobinson, jxc93, elengerich, hmeyer, maceachren]@psu.edu

Abstract

Designing usable geovisualization tools is an emerging problem in GIScience software development. We are often satisfied that a new method provides an innovative window on our data, but functionality alone is insufficient assurance that a tool is applicable to a problem *in situ*. As extensions of the static methods they evolved from, geovisualization tools are bound to enable new knowledge creation. We have yet to learn is how to adapt techniques from interaction designers and usability experts toward our tools in order to maximize this ability. This is especially challenging because there is limited existing guidance for the design of usable geovisualization tools. Their design requires knowledge about the context of work within which they will be used, and should involve user input at all stages, as is the practice in any human-centered design effort. Toward that goal, we have employed a wide range of techniques in the design of ESTAT, an exploratory geovisualization toolkit for epidemiology. These techniques include; verbal protocol analysis, card-sorting, focus groups, and an in-depth case study. This paper reports the design process and evaluation results from our experience with the ESTAT toolkit.

1. Introduction

The task of understanding the causes, controls, and distribution of cancer presents an array of diverse and intricate challenges to those who wish to develop geovisualization tools to support this work. Within the discipline of cancer epidemiology there is still relatively little insight into the complex, interconnected set of factors that influence outcomes (mortality and incidence) and their primary spatial components. An important new focus in cancer epidemiology research encourages exploration of geographic health data to generate new hypotheses (Cockings et al. 2004; Khan and Skinner 2003). Cartography has a tradition of enabling the development of new and enhanced hypotheses with respect to health data, from the seminal spatial epidemiology of John Snow (1855) to contemporary atlases (Pickle et al. 1999; Wennberg et al. 1999). Therefore, we expect that new methods of geovisualization will foster this kind of knowledge creation as well. Indeed, there is a potential that the interactivity and dynamism inherent in geovisualizations will facilitate exploration in a more effective and dynamic manner than prior cartographic techniques (MacEachren et al. 1998).

There are a handful of recent software packages designed to facilitate spatial epidemiology. Examples include SaTScan (Kulldorff and Information Management Services 2004) and ClusterSEER (Jacquez and Estberg 2003), both of which were developed through projects with the National Cancer Institute (NCI). These tools were designed to aid in the detection and analysis of disease clusters. To date there has been no prior example of software for epidemiology that supports visual exploration of multivariate geographic data.

In this paper, we introduce one attempt at facilitating this kind of exploration – ESTAT (Exploratory Spatio-Temporal Analysis Toolkit), a geovisualization toolkit developed for the Statistical Research & Applications Branch of the National Cancer Institute by the Geographic Visualization Science, Technology, and Applications (GeoVISTA) Center at The Pennsylvania State University. ESTAT features four primary visual analysis elements. We have combined a scatter plot, bivariate map, time series plot, and parallel coordinate plot (PCP) in a dynamically-

linked interface (Figure 1). In each analytic tool, brushing, selection, classification, and color scheme choice is linked in real-time; this supports quick exploration of both multivariate relationships and the geographic aspects of those relationships. ESTAT is available as a standalone product, or as a design document for use in its parent application, *GeoVISTA Studio* (Takatsuka and Gahegan 2002).

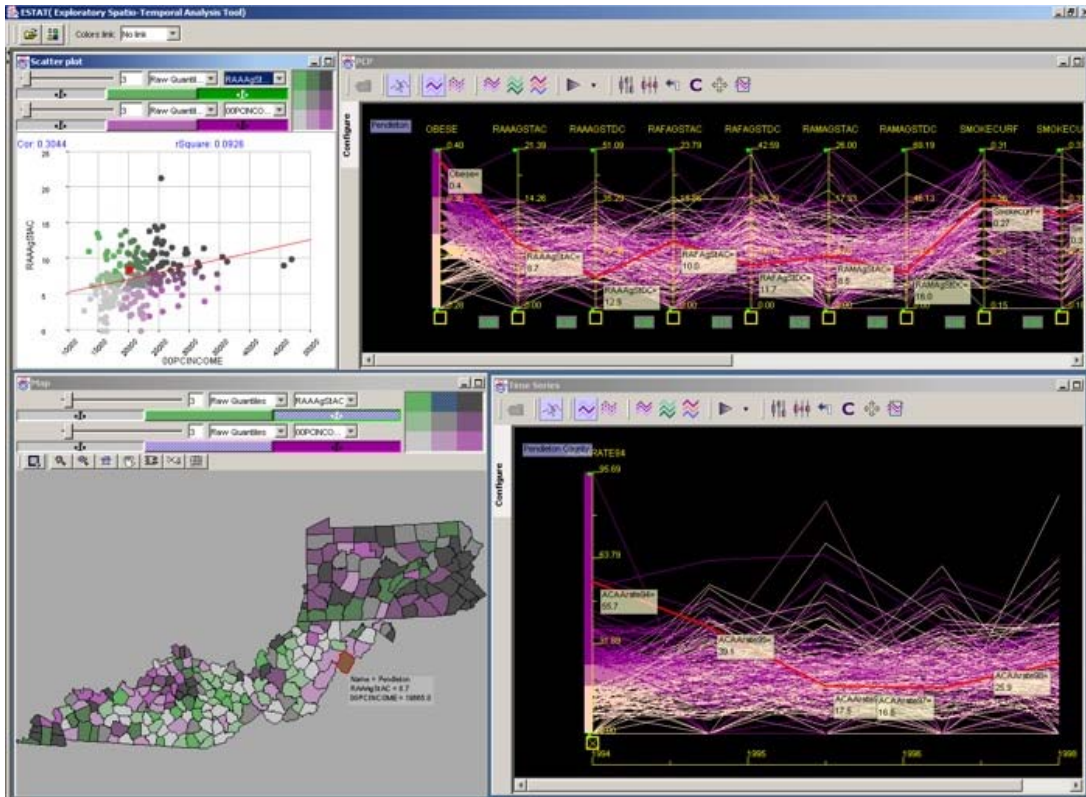


Figure 1: The ESTAT Geovisualization Toolkit

This paper presents the development history of the ESTAT application, specifically reporting on our efforts to make geovisualization tools usable for epidemiology. ESTAT is the result of a user-centered design process, and the series of assessments we have employed are described and discussed. In particular, we focus attention on a case study application using ESTAT for a specific problem in cancer epidemiology. Additionally, we detail the process of usability research that led up to the decision to embark in a long-term case study and collaboration. The results of our case study are discussed, but our primary focus is on the methods we have employed to assess the utility and enhance the interfaces of our geovisualization tools.

2. Background

User issues and interface design are common themes in current geovisualization research. The work reported here attempts to address interface and cognitive/usability outlined by MacEachren and Kraak (2001), particularly their call, "...to develop a comprehensive user-centered design approach to geovisualization usability."

There are a number of recent examples of usability studies with geovisualization tools (Andrienko et al. 2002; Edsall 2003; Haklay and Tobon 2003; Montello et al. 2003; Slocum et al. 2003; Suchan 2002), and many of the methods described herein are inspired by aspects of this body of work. Of these examples, we draw the most from Slocum et al. (2003), who describe a six-stage design process for the creation of a user-centered tool to visualize issues related to water-balance. In the research they report, however, end-user input is incorporated only after key functionality and interface design issues have been decided; the authors ultimately lament the lack of early and repeated user input in their discussion of results. Their recommendation is that user participation should happen from start to finish, rather than after key elements have been decided by developers. This recommendation matches common practice in user-centered design outside of GIScience (Gabbard et al. 1999; Nielsen 1993; Norman 2002).

As a result, we have adopted a design process (Figure 2) that incorporates end-users throughout. The ways in which user input and information are incorporated into this process differ across each stage, and these differences are detailed below along with brief descriptions of each procedural element.

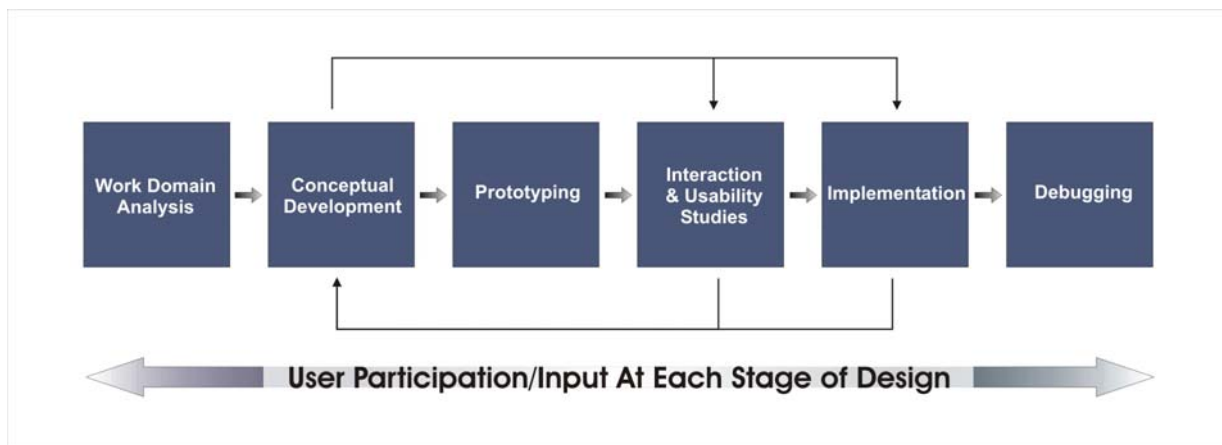


Figure 2: Our adaptation of the user-centered design process.

The first stage, *work domain analysis*, represents the initial communication of ideas and requirements between the client (in this case, NCI) and developers (GeoVISTA) as well as our focused research into the tasks and traditions of epidemiology. The latter aspect of this stage is reported on in another paper we are preparing. As input to the broader method and tool development project that ESTAT builds upon, we interviewed domain members and studied their published work in order to develop a detailed picture of epidemiologists and the work they pursue.

Conceptual development refers to the outline of desired features that comes from understanding the work domain. During this stage, the layout, tools, and architecture are discussed and the application is drawn as a graphical concept prototype. This stage iterates through multiple designs, and each iteration benefits from stakeholder feedback. During the ESTAT development process, designs are discussed through regular meetings and via informal email communication.

After conceptual development defines the core of the application, *prototyping* begins. In this stage, working models of the application are created. During the development of ESTAT, this stage has been essentially concurrent with the stage that follows it in our process diagram – interaction/usability assessment. In this paper, we use the term ‘assessment’ when we refer to formal and informal evaluation of both the overall usability of our tools as well as the interactions that they foster. We are especially interested in the latter of these two aspects.

Interaction/usability assessment activities are crucial to understanding the pieces of an application that work well, as well as those that need further re-design. Formal assessment efforts may take place in a usability laboratory where audio and video can be captured while users attempt to work with an application. They may also include interviews and focus groups that discuss the application in question. Informal assessment occurs as end-users are asked to try out prototypes and forward comments, questions, and ideas. Additionally, there are a wide range of informal assessment activities that occur on the development side as the application is critiqued internally. Each of these methods has been employed during the design of ESTAT, in addition to a hybrid of these techniques that emerged through long-term case study collaboration.

Implementation follows assessment activities, and typically spawns its own fresh set of design issues. It is difficult to simulate ‘real’ work well enough during the assessment stage in order to ensure that you have no significant problems during implementation. Therefore we include feedback loops between both assessment and implementation stages as they work backward into conceptual development and initial design. Our experiences have shown that the results of both stages have often caused us to return to the proverbial drawing board and try other options. Similarly, we have been inspired to explore new designs for different domains/tasks via things we learned during these activities. Furthermore, if one can imagine concept development as a stage occurring across multiple projects, the lessons learned from usability assessment on one project are extremely valuable inputs to the beginning steps of other similar projects.

The final stage in this process is *debugging*. In this portion of the design process the application is adjusted to enhance stability, compatibility, and make the most out of the computing infrastructure in which it has been implemented. Mechanisms for user feedback at this stage include web based issue trackers like JIRA (<http://www.atlassian.com/software/jira>), links to email support in help documentation, and follow-up phone calls to individual users.

The remainder of this paper focuses particular attention on stage four, that of interaction and usability assessments. The following sections outline the array of techniques we have applied and the things we have learned from them about our tools, our users, and strategies for usability assessment of complex, exploratory geovisualization environments.

3. Assessing ESTAT

We have opted to combine several methods of usability assessment in order to examine the ESTAT toolkit. The following sections briefly describe the techniques of card sorting, verbal protocol analysis, focus groups, and ethnographic case studies.

Card sorting (Nielsen 1993) is a simple and fast method of assessing the structure of an interface. Users are given 3 x 5 note cards labeled with individual functions, and asked to arrange them in categories and an order that they think makes the most sense. This method has found particular success in website usability research, as designers seek end-user guidance to help them arrange web pages.

Verbal protocol analysis (VPA) gathers user experiences in real-time as they ‘think aloud’ (Ericsson and Simon 1993). It is especially valuable for understanding both the critical needs of an application as well as its expected behavior from the perspective of the end-user. Typically, users are given a task to achieve using the application in question, and they are instructed to verbalize their thought processes as they work. VPA has the inherent benefit (or detriment) of generating a massive amount of data very quickly.

Focus groups (Morgan et al. 1998) solicit ideas and feedback through a group discussion. They are moderated by a discussion leader who asks questions and prompts for elaboration as described in advance by those who are sponsoring the session. Focus groups are even faster than VPA and generate a similarly large amount of data. In general, focus groups allow users to share their experiences with an application and develop hypothetical situations/ideas in a quick and efficient manner. These aspects make focus groups an applicable method across multiple stages of the software design process. We have chosen to combine VPA with follow-up focus groups to capture both the details inherent in the epidemiological workflow as well as reflections our users have about how our tools could be situated more appropriately for their daily usage.

Ethnographic case studies combine a real-world application of methods with participant-observation practices (Yin 1994). Case studies are often undertaken as proof-of-concept exercises to demonstrate the utility of a particular tool or method. Ethnographic case studies are different in the sense that they are undertaken not only to evaluate tool utility, but also so that researchers may observe how work takes place around these tools. Data from such studies are usually collected ad hoc, in notes, through informal and formal communication with subjects, and from direct observation. Ethnographic case studies require a significant time commitment in order to complete, and the results are subject to interpretation. They do, however, yield deep knowledge of the situated work experience.

There is a clear emphasis in our work with ESTAT on methods that elicit qualitative results. While we recognize the value of quantitative measures to answer some categories of questions, at the stage of work presented here, we are interested in *how* people are working (or not working) with exploratory geovisualization tools; thus our focus is on formative rather than summative evaluation. Formative evaluations are carried out early in the design process (focusing on assessing the needs of users and the extent to which the overall conceptual approach fits those needs) while summative evaluations occur after a design has been completed (and their aim is to directly compare the new design to other applications designed to accomplish the same tasks)(Gabbard et al. 1999). In general, formative studies lend themselves to qualitative methods, while summative evaluations rely more often on quantitative measures.

In many instances, usability seeks to reduce the time it takes to perform a routine task or limit the number of errors that might occur when solving specific problems. Recently, Saraiya et al.

(2004) presented an intriguing model for summative studies that quantifies the number of ‘insights’ generated during exploration. Insights were judged for validity by outside experts, in their case from microbiology. In contrast to Saraiya et al., the tasks we are examining are exploratory, and ideally result in a hypothesis. In our case it is not clear that we could rely on measuring validity, as our users are in fact the experts we might solicit to judge this – though their insight-based study is a useful model for future work we may pursue. In the future we will shift toward summative evaluations in order to critique the utility of ESTAT against existing methods of spatial data exploration in epidemiology.

4. The Assessment Process

In accordance with work pursued to understand the work domain and devise initial conceptual designs, we have been conducting usability and interaction assessments for ESTAT since October of 2003. Initially, we focused our efforts toward building prototypes based on contract specifications and informal communication with NCI staff that could be quickly evaluated by GIScience graduate students who had expertise with spatial data analysis. This was followed by formal user testing with health researchers at NCI, the results of which were later augmented by a long-term case study collaboration with an epidemiologist. Each of these stages is described below.

4.1 Rapid Prototype Assessment

As noted above, the first prototype of ESTAT featured the core features and functionality that had been requested by our colleagues at NCI and augmented by our research into their work domain. Our initial evaluation efforts focused on a single component of ESTAT, the parallel coordinate plot tool, and used the card-sorting method and verbal protocol analysis with GIScience graduate student participants. Our focus here was to quickly determine whether or not the tool was understandable enough to accomplish the tasks we were trying to facilitate. These evaluations were easy for us to execute, but in general they provided more questions than answers. Few of our initial testers understood how to use a Parallel Coordinate Plot, and more importantly, most were unable to simulate the tasks of epidemiology. When we asked users to explore epidemiological outcomes along with population information and predictor variables none were able to develop the kind of hypothesis that ESTAT was supposed to yield. These initial assessments did, however, provide some useful information. Analysis of the protocol transcripts identified a number of instances in which testers expressed frustration with the basic layout of the tools, the data loading process, and the lack of consistency in our interfaces.

We applied the card-sorting method to try and reorganize our interface. Again, our participants were GIScience graduate students. Our card-sorting results showed that there appeared to be two general interface groupings that we could implement. While we were able to uncover these possible reconfigurations of the interface controls, we still had no sense of how epidemiologists might choose to organize the ESTAT interface differently.

4.2 Assessment with Domain Experts

At this stage, we turned to our collaborators at NCI for additional input. A formal usability assessment of the alpha stage ESTAT prototype occurred in February of 2004 with a group of testers identified by our primary NCI contacts as likely ESTAT end-users. Each user worked through a tutorial and brief set of epidemiological tasks before participating in a focus group discussion to verbally assess the tools.

During the tutorial and task session, we captured audio and video of the session and encouraged participants to both ask questions (of the two moderators) and discuss what they were attempting to accomplish. Although it was essentially a cognate of a protocol analysis session, we did not use a ‘keep talking’ prompt or otherwise force our participants to vocalize their interactions. Our method was driven by the fact that we had a limited time to work with a relatively large number of users (17) in two short sessions over a period of two days. During these sessions the moderators took notes, and subsequent analysis of the video and audio helped to augment these notes with further detail.

Immediately following the tutorial and task session, a focus group was held to discuss various aspects of the ESTAT toolkit. During the focus group, the same two moderators led the discussion, with occasional input and questions from two of the NCI project leaders who collaborate with the GeoVISTA Center. In hindsight, it would have been better to not have this additional input, as we were not aware before the session of the questions they wanted to ask or the issues that they would be most interested to explore. In general, the NCI project leaders were interested in hearing about what new features were still required, while we were focused on determining whether or not ESTAT functioned effectively as an exploratory visualization toolkit. As a result, our focus group discussion was somewhat more discontinuous than it may have been had we had the full time and scope within our control.

Both the modified VPA and focus group approaches to tool assessment by domain experts were extremely valuable to our development effort. From the tutorial/task sessions we were able to determine that our data loading mechanism needed to be completely redesigned in order to be reasonably efficient. While we had our own informal debates regarding this part of ESTAT prior to testing at NCI, having documentation of end-users trudging through the interface had a greater impact on our developers. The tutorial/task sessions revealed that most users never got very far into actual epidemiological analysis because of the clumsiness of the interface and their lack of familiarity with the visualization methods being applied. This latter point was an aspect of our work that we had not anticipated – we found that the visualizations that NCI had requested we build in ESTAT were not widely understood by the users NCI had in mind. In particular, most users needed tutoring to understand the Parallel Coordinate Plot and Time Series graphs. We had included descriptions about how these work in our tutorial. However, for users who were experiencing this kind of analysis tool for the first time, more focused training was clearly needed.

From our focus group discussions we were provided with insights into the modifications and additions needed for ESTAT in order to make it usable for epidemiological work. In general, our users were excited about the potential that geovisualization tools hold for their discipline. The

version of ESTAT that was tested, however, lacked a number of the essential ingredients needed to make the software practical for use by typical public health researchers.

Specifically, our users repeatedly mentioned a desire to see descriptive statistics of the data displayed, in order to help them assess the character of the data patterns and relationships in question. One user went so far as to ask, “Why is it called ESTAT if there are no stats?” When we pressed our users for more details regarding the specific statistics they would require, we were met with a wide range of possibilities, including; regression values/lines, correlation coefficients, means/modes/medians, and significance values. A small number of users had specific suggestions for more advanced statistics, including; Poisson regression, Bayesian models, and the ability to use spatial analysis methods. For these more advanced functions we were advised that it would be worthwhile for our tools to be able to ‘talk to’ a statistical software package so that users could create and execute complex and customized routines on data they have viewed, initially, in ESTAT.

4.3 Case Study Collaboration

As a result, we were faced with the challenge of deciding which statistical methods to incorporate into our tools and which we should leave to other software. Additionally, the suggestion to enable our tools to ‘talk to’ a statistical package is not easy to implement because most statistical software is designed to be self-contained and coordination with our Java-based software involves a substantial software engineering effort. While we gathered many useful ideas from these tests at NCI, it was clear that we could pursue a more elegant and effective solution by embarking on an in-depth collaboration over a significant period of time with an epidemiologist. Doing so would provide deeper insight into common methods of epidemiological analysis, help determine the kinds of statistical/mathematical methods that are used most frequently to assess the quality and general characteristics of data, and allow us to understand in greater detail the kinds of interfaces that epidemiologists encounter regularly and therefore might be most comfortable adopting. Furthermore, we could augment the results gathered from our other assessment efforts and begin to triangulate areas of common agreement.

The process we followed draws upon the practice of interaction design (Preece et al. 2002). Specifically, we adopted a participant observation approach to observing and cataloguing the actions and ideas of our collaborator, with goals similar to those we had in mind during the work domain analysis stage of our design process. Over a period of roughly four months, Anthony Robinson worked together with Dr. Eugene Lengerich, an epidemiologist from the Penn State University College of Medicine. Lengerich has been a colleague in several medical geography projects we host at the GeoVISTA center, and therefore was familiar with our staff, software, and approach. Our goal from the beginning was to identify a problem in epidemiology that we could use ESTAT to explore, and by doing so together we would attempt to uncover a wide array of issues regarding the functionality and usefulness of our geovisualization toolkit. The case study work took place in several semi-formal sessions where we worked together using a laptop, sometimes augmented by a projector so that other colleagues could participate as well. Outside of these sessions, we often collaborated via email and phone.

Initially our meetings were largely informational, as we began to know each other and determine how we might work together on a specific epidemiological research question. These first steps allowed us to build an understanding of the kinds of problems that were interesting to our colleague, and conversely provided Lengerich with ample opportunity to see a wide array of what we were working on and what kinds of tools were at his disposal. This is the kind of mutual understanding that is difficult to achieve when an outside expert is brought in for a limited amount of time, as is typically the case in many structured evaluations of software tools. Participant observation provided us with the ability to effectively develop this understanding into real synergy between our group of experts and a researcher focused on epidemiology.

Over the course of several meetings, Lengerich decided he would like to try using these tools to augment a more traditional epidemiological study. Specifically, he wanted to be certain that ESTAT could echo the results he would obtain from a structured mathematical analysis. The example analysis was a study of colon cancer incidence in the Appalachian counties of Pennsylvania, Kentucky, and West Virginia. This area is a focus of the Appalachian Cancer Network, for which our colleague is a research director. He had a working hypothesis that there are differences between the spatial patterns exhibited for incidence of colon cancer depending on whether or not that cancer occurs in the ascending or descending colon. This idea stems from current research in epidemiology that is examining potential etiologic differences in colon cancer that occurs in the ascending colon versus the descending colon (Iacopetta 2002). Moreover, Lengerich was interested in exploring how ascending/descending malignancies might differ according to variables representing prevalence of health screening and access to healthcare facilities. Our initial work with ESTAT on this problem centered on issues related to the data we wanted to analyze, and a great deal of work became necessary in order to create the right multivariate dataset that covered enough detail to explore a rather broad hypothesis.

The final sessions of our case study collaboration focused on using ESTAT to explore the colon cancer hypothesis (Figure 3). Lengerich verbally confirmed his findings as we used ESTAT to examine the relationships he had examined statistically. He was able to visually explore the same data he had previously analyzed and identify corroborating evidence to support his conclusions. Our experiences with NCI research staff indicated that most cancer researchers will be unsatisfied with purely visual analysis, that they want a range of descriptive statistics as well. For this case study, Lengerich had carried out statistical analysis of the data prior to using ESTAT, thus he had less need for integrated statistical tools. Perhaps more importantly, ESTAT presented the geographic picture of Lengerich's analysis that he had not seen before. Robinson and Lengerich spent time iterating through each state in the three-state study region to explore the geographic pattern in greater detail. In this stage of data exploration it would have been valuable to have access to spatial statistics in ESTAT to examine the geography more systematically.

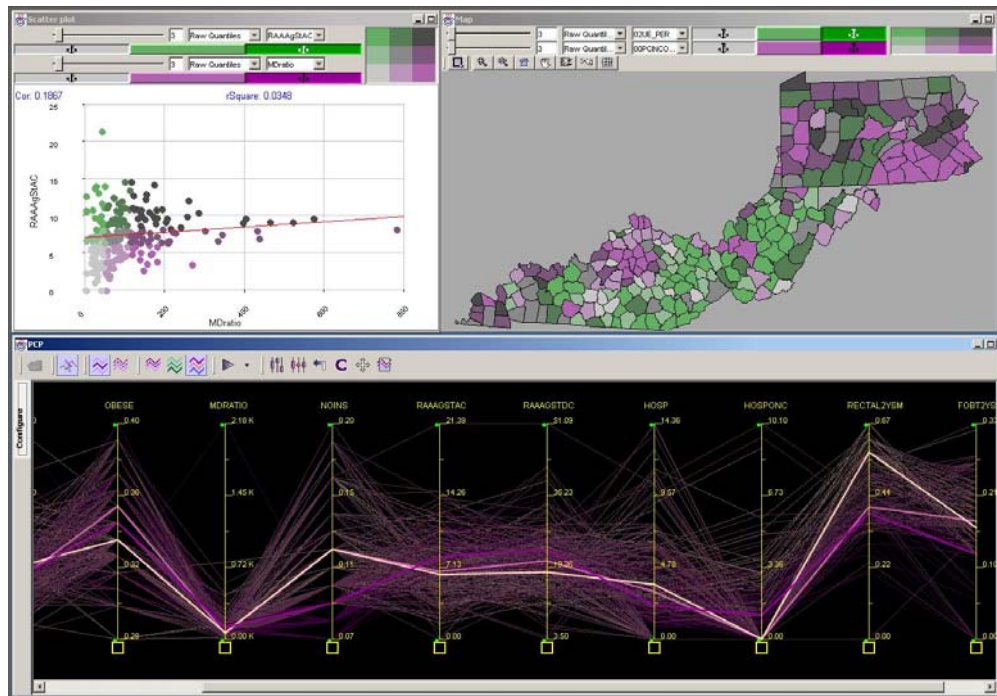


Figure 3: Exploring colon cancer incidence in Appalachia.

The differences between what our formal testing at NCI and our case study suggested in terms of statistics were likely influenced by the fact that our users at NCI had a short amount of time to make themselves familiar with ESTAT, and their alpha version was without even simple measures of correlation or regression, which *were* implemented by the time Lengerich was using ESTAT. Furthermore, for the case study we situated ESTAT as a tool that would augment and confirm a ‘typical’ epidemiological analysis – a departure from the ‘explore and hypothesize’ approach we had encouraged at NCI. This change in focus happened primarily because Lengerich wanted to make sure ESTAT would echo his traditional analysis before he would begin to rely on it for exploratory tasks. His conservative approach was mirrored in many instances by users at NCI who vocalized their skepticism about visualization techniques and how they may be misrepresenting various aspects of the data.

5. Discussion

As a result of our efforts to systematically evaluate ESTAT in terms of its design and usability, we have been able to better shape our toolkit for the tasks of epidemiology.

We were especially fortunate to have a window into a discipline through our case study collaboration. We knew this would allow a limited viewpoint because we were dealing with a single user alone. In practice it is sometimes difficult to separate individual idiosyncrasies from the more valuable common themes in order to refine and enhance a design. However, our one-on-one collaboration enabled a deeper understanding of one expert’s data analysis strategies, providing us with a basis for strategic decisions regarding which suggestions to follow (from this expert and others) and which we should reserve for further input from future evaluative efforts.

Insights derived from our case study are complemented by those from our group evaluation sessions at NCI as well as by our early, internal usability testing. The synergy among multiple usability methods with multiple participants (having different backgrounds) provided an intricate and comprehensive input into our software design process. Virtually all aspects of the interface as well as the functionality of ESTAT were addressed by at least one component of our assessment strategy. Furthermore, focusing on user input and participation throughout the design process has shifted emphasis from developing new methods in isolation to developing pieces of *toolkits* that have an immediate application.

The case study portion of our usability work with ESTAT was particularly beneficial to our general understanding of the problem domain, and our colleague's specific, as well as implicit, suggestions led us to focus our attention on issues that we had previously neglected – loading/sorting data and interface icons/controls. It cannot be stressed enough that without attention to the *actual* use of our geovisualization tools, it would have been difficult for us to accurately predict these deficiencies, though it may have been possible had we taken a proactive stance and modeled user interactions with our design. Moreover, our experiences in the design process presented here (and in other related work) support Buxton's (1999) argument that converging methods are essential to assessing complex, geospatial interfaces. While it is quite possible to learn a great deal from a single exercise, it is not a way in which we can effectively tailor designs to a particular context of use. That evaluative effort must mirror the development effort, in that it should be constant throughout the progress of the project. We advocate incorporation of use and usability assessment from the start of any software development project. Furthermore, developers should not be expected to 'take care' of usability issues on their own. Internally, our development process has undergone major changes since we have dedicated resources to research assistants who do not develop software and are explicitly focused on user issues.

A series of modifications were performed on our software in response to usability issues brought up by our case-study collaboration and other interaction/usability assessments. Our initial in-house focus on the PCP component led to a total redesign of its interface and behaviors. The subsequent evaluation activities have led to a similar redesign of the entire ESTAT environment. The following sections describe the major issue areas that emerged as a result of our assessment work with ESTAT.

6. Issues Emerging as a Result of Assessments

6.1 Data

ESTAT has gone through an extensive reworking of the way it loads and handles datasets for epidemiological research. Our initial testing at NCI revealed that our data loading tool was far too complicated as a single panel item (Figure 4). Armed with audio/video detail showing our end-users struggling with its interface, we were able to demonstrate the need for a redesign to our developers. As a group, we decided to create a data loading wizard that would guide users through each step more efficiently. Following an initial prototype, we changed each prompt so that it used natural language in place of the technical 'programmer-speak' that had been implemented. At the same time, icons were reworked to create visual cues to the kinds of

functions they represent. Both of these changes, as well as others we have implemented, are inspired by interface design guidelines discussed by Schneiderman (2004).

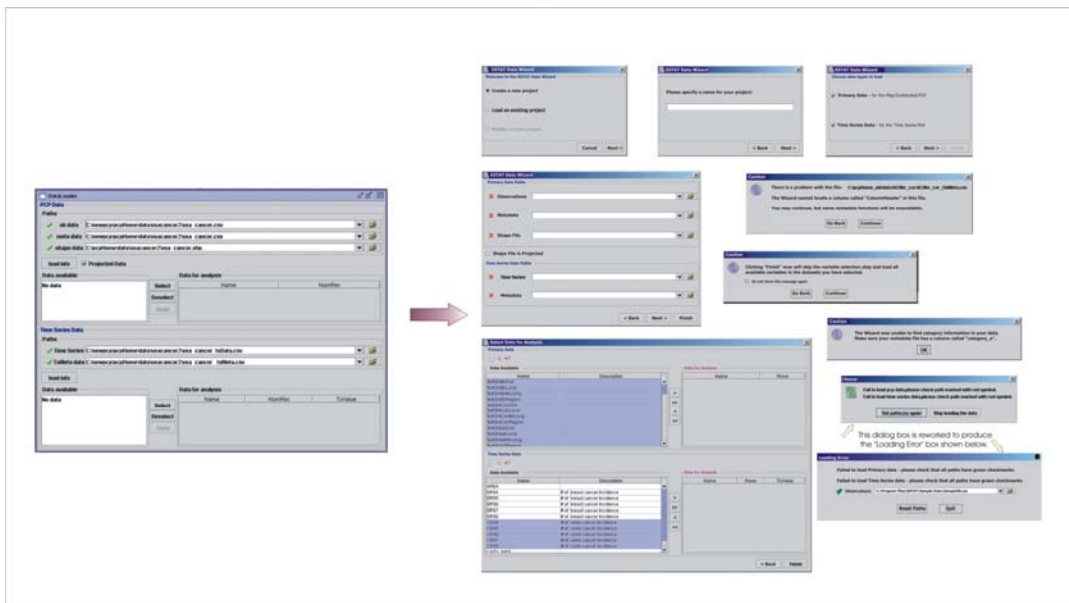


Figure 4: From a single panel data loader to a comprehensive wizard.

The case study collaboration was particularly valuable toward our understanding of data handling in epidemiology. We had placed less importance on this stage of analysis than on the visual exploration tools, as the visual exploration stage is the focus of our GeoVISTA research center and of the primary research grant supporting this particular work. What we found was that our epidemiologists were in need of a mechanism for visualizing and making sense of the complex datasets before they could select variables and explore them using our exploratory geovisualization tools. While we work on a visual method to support variable selection, we have implemented categorizing and sorting tools into our data loading wizard. This has provided a structure for data selection that complements the analysis strategy we were able to observe from our case study work with Lengerich. We were able to confirm the utility of these added features in our case study meetings as we explored the colon cancer hypothesis.

We have also created a simple metadata file to accompany each project that provides detailed variable descriptions to help alleviate problems of comprehension that emerge when a large number of truncated variable names are on screen at once. Rollovers in our tools now display the full description of each variable.

6.2 Supplementary Statistics

A crucial issue we came across a number of times during our evaluations was a strong desire for basic descriptive statistics to characterize data that were visualized. The initial ESTAT prototype did not incorporate such statistics, and our users were reluctant to explore the data visually without some sense of the mathematical structure of the data up front. Since this sentiment had

been declared previously by our tester group at NCI, we implemented basic descriptive statistics prior to embarking on our case study collaboration.

Correlation coefficients now are available between each pair of PCP and Time Series axes as well as for the distribution displayed in the scatter plot. In addition, the scatter plot can show a regression line and r-squared value. The scatter plot calculates these statistics on the fly, and selecting a subset on the plot will cause values to change to reflect that grouping (Figure 5).

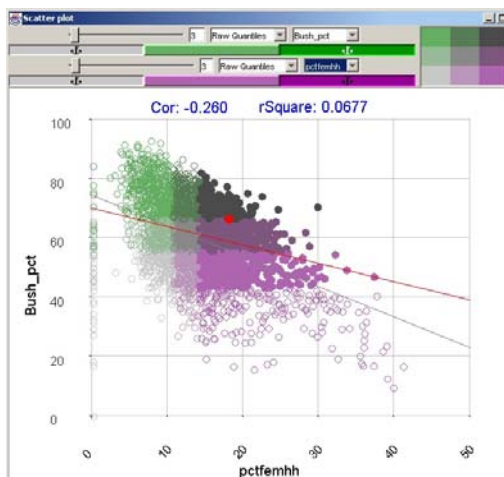


Figure 5: Subset correlation as displayed on the ESTAT scatterplot.

6.3 Performance/Computing Infrastructure Issues

The disparity between the computers we have access to in our research center and those that are seeing daily use by our colleagues in Epidemiology (typically the latter have less memory, lower resolution screens, and slower processors) has caused us to focus attention on increasing performance of our software tools for use on machines that are not on the cutting edge of computing technology.

A more vexing issue is that our visual tools require a lot of screen space, and many of our developers are building them on dual-panel machines. Not surprisingly, few of our end-users have or will soon have this capability at their disposal. While it is reasonable to expect that higher-resolution monitors will become more affordable and commonplace over time, we have begun to encourage scrutiny of interfaces in terms of how much space they consume. Innovative window management tools, such as the methods available in Macintosh OSX may also alleviate some of these issues.

7. Future Directions

At present we are working on further methods of evaluating and redesigning ESTAT for epidemiology. In the near future an in-depth individual user assessment is scheduled to take place at NCI with a small number of participants. This activity will focus again on providing domain experts with the ability to perform quick, visual exploration of multivariate spatial data in order to formulate new and enhanced hypotheses. The verbal protocol analysis technique will

be employed in its more traditional form, and the users will discuss their experiences with the latest version of ESTAT after their individual sessions in a focus group. It is hoped that through this evaluation effort we will gain a greater understanding of which aspects of ESTAT are useful for exploration and how epidemiologists from a wide range of backgrounds situate visualization within their current research.

In the near future, we hope to incorporate a suite of spatial statistics in ESTAT and its parent application GeoVISTA *Studio* as a result of our software development collaboration with the creators of the GeoDA spatial analysis toolkit at the Spatial Analysis Laboratory of the University of Illinois at Urbana-Champaign.

Finally, we hope to address the challenge of modifying common summative evaluation techniques for our exploratory tools. We wish to answer the call for more definitive evidence that visualization, and geovisualization in particular, provides clear advantages over other methods of exploratory analysis. Our case study results in particular provide some evidence that we are within reach of this goal.

ESTAT, along with sample datasets, is available for download on the web at:
<http://www.geovista.psu.edu/grants/nci-esda/tutorials.html>

Acknowledgements

The research reported here has been supported in part by a contract from the National Cancer Institute (to construct the initial ESTAT application) and by grant CA95949 from the National Cancer Institute (that supported development of our approach to user-centered design and the study of ESTAT use).

References

- Andrienko, G. L., N. V. Andrienko, H. Voss, F. Bernardo, J. Hipolito, and U. Kretchmer. 2002. Testing the usability of interactive maps in CommonGIS. *Cartography and Geographic Information Science* 29:325-42.
- Buttenfield, B. 1999. Usability evaluation of digital libraries. *Science & Technology Libraries* 17:39-59.
- Cockings, S., C. E. Dunn, R. S. Bhopal, and D. R. Walker. 2004. Users' perspectives on epidemiological, GIS and point pattern approaches to analysing environment and health data. *Health & Place* 10:169-82.
- Edsall, R. M. 2003. Design and usability of an enhanced geographic information system for exploration of multivariate health statistics. *Professional Geographer* 55:605-19.
- Ericsson, K. A., and H. A. Simon. 1993. *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.
- Gabbard, J. L., D. Hix, and J. E. I. Swan. 1999. User-centered design and evaluation of virtual environments. *IEEE Computer Graphics and Applications* 19:51-9.
- Haklay, M., and C. Tobon. 2003. Usability evaluation and PPGIS: Towards a user-centered design approach. *International Journal of Geographical Information Science* 17:577-92.

- Iacopetta, B. 2002. Are there two sides to colorectal cancer? *International Journal of Cancer* 101:403-08.
- Jacquez, G. M., and L. Estberg. Clusterseer 2.0: Software for the detection and analysis of spatial, temporal, and spatio-temporal patterns. BioMedware, Inc. <http://www.terraser.com/products/clusterseer.html>.
- Khan, O. A., and R. Skinner, eds. 2003. *Geographic information systems and health applications*. Hershey: Idea Group Publishers.
- Kulldorff, M., and I. Information Management Services. Satscan v5.1: Software for the spatial and space-time scan statistics <http://www.satscan.org/>.
- MacEachren, A. M., F. Boscoe, D. Haug, and L. W. Pickle. 1998. Geographic visualization: Designing manipulable maps for exploring temporally varying georeferenced statistics. Proceedings of IEEE Information Visualization Symposium, Research Triangle Park, North Carolina, October 19-20.
- MacEachren, A. M., and M.-J. Kraak. 2001. Research challenges in geovisualization. *Cartography and Geographic Information Science* 28:3-12.
- Montello, D. R., S. I. Fabrikant, M. Ruocco, and R. S. Middleton. 2003. Testing the first law of cognitive geography on point-display spatializations. Proceedings of Proceedings, Conference on Spatial Information Theory (COSIT '03), Lecture Notes in Computer Science 2825, Ittingen, Switzerland, September 24-28.
- Morgan, D. L., R. A. Krueger, and J. A. King. 1998. *The focus group kit*. Thousand Oaks, CA: Sage Publications.
- Nielsen, J. 1993. *Usability engineering*. Boston, Massachusetts: Academic Press, Inc.
- Norman, D. 2002. *The design of everyday things*. New York: Basic Books.
- Pickle, L. W., M. Mungiole, G. K. Jones, and A. A. White. 1999. Exploring spatial patterns of mortality: The new atlas of united states mortality. *Statistics in Medicine* 18:3211-20.
- Preece, J., Y. Rogers, and H. Sharp. 2002. *Interaction design: Beyond human-computer interaction*. New York: John Wiley & Sons.
- Saraiya, P., C. North, and K. Duca. 2004. An evaluation of microarray visualization tools for biological insight. Proceedings of IEEE Symposium on Information Visualization 2004, Austin, TX, October 10-12.
- Slocum, T., D. Cliburn, J. Feddema, and J. Miller. 2003. Evaluating the usability of a tool for visualizing the uncertainty of the future global water balance. *Cartography and Geographic Information Science* 30:299-317.
- Snow, J. 1855. *On the mode of communication of cholera*. New York: The Commonwealth Fund.
- Suchan, T. A. 2002. Usability studies of geovisualization software in the workplace. Proceedings of National Conference for Digital Government Research, Los Angeles, CA, May 19-22.
- Takatsuka, M., and M. Gahegan. 2002. GeoVISTA *Studio*: A codeless visual programming environment for geoscientific data analysis and visualization. *Computers and Geosciences* 28:1131-44.
- Wennberg, J. E., M. M. Cooper, J. D. Birkmeyer, K. K. Bronner, T. A. Bubolz, D. E. Campbell, E. F. Fisher, G. T. O'Connor, J. F. Poage, S. M. Sharp, J. Skinner, T. A. Stukel, and D. E. Wennberg. 1999. *The Dartmouth atlas of health care 1999*. Chicago, IL: American Hospital Publishing, Inc.
- Yin, R. K. 1994. *Case study research: Design and methods*. 2nd ed. Thousand Oaks, CA: Sage Publications.