Using Spatial Data Visualization to Motivate Undergraduate Social Science Students

Richard LeGates, San Francisco State University, San Francisco, CA

Many undergraduate students who are beginning majors in sociology, political science, anthropology, urban studies, and other social science disciplines fear math and feel inadequate to undertake rigorous empirical social science research. This is quite different from students drawn to natural sciences, mathematics, and engineering because they enjoy math and are confident in their quantitative reasoning ability. Most undergraduate social science majors take courses in research methods and data analysis only because they are dreaded requirements for the major.

Research methods and data analysis courses are usually required as soon as social science students declare their major. Subjecting idealistic students who hate math to rigorous methods courses at the start of their education poses a serious motivation and retention problem. Some students are bored by or feel incompetent in methods courses and drop out of social science majors. Other students are turned off to the majors, and the balance of their course work is tainted by their negative experience with methods courses. Still other students grudgingly complete social science methods courses but never use the methodology they learn in other course work or in their careers.

A major reason beginning social science majors are drawn to these fields is because they feel courses such as "The American City," "Race, Class, and the Environment," or "Social Problems" are substantively interesting and relate to their own life experience. They are often idealistic and see mastering a social science discipline as a vehicle to make the world better.

How can social science professors draw upon students' interests and channel their enthusiasm into rigorous application of scientific methods to address issues that passionately concern them? How can they turn methods courses from dreaded requirements into courses students enjoy? What will motivate students to enthusiastically apply what they learn in their methods training to the balance of their course work and in their careers?

Spatial analysis and data visualization using GIS software can contribute to all of these goals. This is the premise of the NSF "Space, Culture, and Urban Policy" CCLI Educational Materials Development (EMD) project at San Francisco State University [1]. The pedagogical theory we have developed and operationalized in instructional modules for use in upper-division undergraduate social science research methods and data analysis courses helps students "see" social phenomena and the consequences of different policy choices.

This chapter begins by describing how GIS software can help students see social relations and social policy options. It argues that visualization should be an integral tool in exploratory data analysis; not just a final step in presenting the results of analysis. Indeed, a major thesis of the paper is that visualization should be one of the first topics taught in undergraduate social science research methods courses. The balance of the article describes a specific model of how to teach spatial analysis and data visualization in the social sciences that the author and colleagues are developing with NSF support.

Seeing Social Relations and Social Policy Options Through GIS

GIS software links computerized maps to attribute tables that contain data about map features. In addition to physical features such as rivers and roads, a GIS can display information about social and policy phenomena such as race, income, housing affordability, voting behavior, historical events, crime, and immigration.

Until recently, GIS software was expensive and technically demanding. GIS software is becoming more user-friendly and is now widely available on university campuses. Modern GIS software built around point-and-click graphical user interfaces provides intuitive and powerful digital tools for analysis accessible to beginning college students.

Beginning students can open an existing GIS file showing world cities (such as Figure 1) in the same way they open a Microsoft Word document or Excel spreadsheet. Students can then use digital tools to examine and analyze spatial and attribute data. A few dozen intuitive commands will permit beginners to do some powerful analysis. Students can answer questions such as, "What is where?", "How are features distributed in geographical space?", "Are patterns discernible?", and "What features are spatially coincident?" Beginning students can produce cartographically respectable analyses and maps that reflect their own interests. As they develop conceptual and operational skills, they can ask deeper questions and get more subtle answers. As they learn to build more complex models of reality and symbolize map output more elegantly, they can produce ever more revealing and more cartographically excellent maps.

Unlike a static paper map, digital GIS maps link spatial representations of geographical map features to data about attributes of the features. The results of operations on map features are reflected in data tables. Operations on the data underlying a map are immediately visible on the digital map. Students can immediately see the results of analysis in a form that is much easier to grasp than correlation coefficients or regression formulas. They can learn to formulate queries that produce answers to questions the students themselves tailor. A beginning student who knows only enough GIS to ask a GIS software program to select all the cities whose 2015 population is expected to be greater than 10,000,000 can immediately produce a new map showing the answer visually. Another query with Boolean operators asking for a display of cities whose populations are expected to be over 10,000,000 *and* whose median income is expected to be under $10,000 U.S. will instantly produce a new map. How many of these cities are there? Where? What are they like? Once students start asking spatial questions and getting visual spatial answers, they begin to ask fundamental questions about the nature of data and the logic of social scientific inquiry.

Undergraduate students can answer many spatial questions without statistical training[2]. University of California, Santa Barbara, geography professor Keith Clarke, the author of a leading undergraduate GIS textbook (Clarke, 2002)[3] and the author of the Environmental Science Research Institute's (ESRI) widely read book on GIS analysis (Mitchell, 1999)[4], presents lists of spatial questions that can be answered using GIS without statistical training.

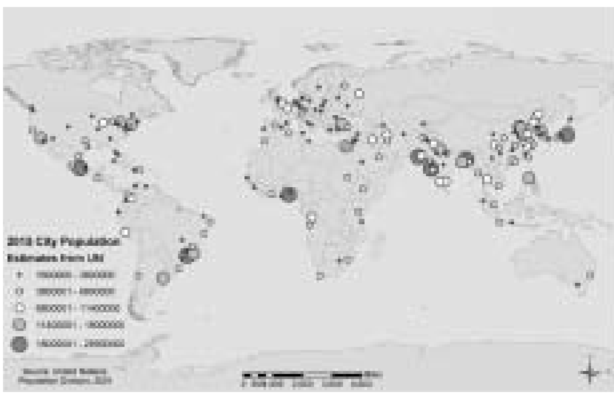

Figure 1. Cities of the future: world population estimates for 2015.

Inverting the Role of Exploration and Visualization in Social Research

Historically, most social scientists research followed a rigid linear model: formulating hypotheses, collecting data, and then testing the hypotheses using the data. Yale political science professor John Tukey (Tukey, 1977) inverted this way of approaching social research. Rather than requiring his students to formulate all their hypotheses first and then test them, Tukey advocated exploratory data analysis. He encouraged his students to formulate and refine hypotheses continuously as they explored data. Tukey also invented ways to visually depict the intermediate results of exploratory data analysis such as "box and whisker" charts and "stem-and-leaf" diagrams so that student would see patterns in the data that could guide their iterative exploratory data analysis.

Visual representation in social science research was historically tacked on as part of the final stage in the rigid hypothesis-testing model. Only after analysis had been completed and research findings were in their nearly final form should researchers create graphs and charts. The purpose of data graphics at this point was to visually embellish written research reports. Contrary to Tukey's method, this approach grants data visualization no role in exploratory data analysis. It views creation of data graphics as a small, incidental part of the social science research enterprise that occurs only after the analysis is complete. Regrettably, even today, most undergraduate social science research methods and data analysis courses treat data visualization as an afterthought.

French semiologist Jacques Bertin launched a frontal assault on the marginalization of visual analysis in research that was as revolutionary as Tukey's theories about exploratory data analysis (Bertin, 1967). Bertin developed theory about what he called the semiology of graphics—the art and science of creating graphics that would communicate meaning to the viewer. Bertin developed a body of theory about the way in which humans perceive "retinal variables" such as color and texture. He argued that the conventional model of using visual representation only to summarize research findings after analysis was complete was entirely wrong. In contrast, Bertin advocated "using vision to think," creating visual representations of data continuously throughout the research process as a vehicle for understanding the data and refining the research. Like Tukey, Bertin argued that research should be exploratory and iterative, with initial findings informing further research directions.

There is now a substantial body of scholarship around scientific data visualization. In addition to Tukey and Bertin, Yale political science professor, Edward Tufte, has written provocatively on the visual representation of quantitative data (Tufte, 1983, 1990, 1997). William Cleveland, a scientist at Bell Labs, has systematized much of what is known about excellence in scientific data graphics (Cleveland, 1994). Stuart Card, Jock D. Mackinlay, and Ben Shneiderman have summarized material on scientific information visualization (Card et al., 1999). San Jose State University urban and regional planning professor Earl Bossard has developed theory for envisioning neighborhoods (Bossard, 2005).

Our pedagogical model draws upon this body of theory. We emphasize exploratory data analysis using data visualization. We go even further, arguing that data visualization should be the first topic taught in social science research methods courses.

Visualization First

An instructor could tack two weeks devoted to spatial analysis onto the end of a 15-week beginning social science methods course that covers such conventional quantitative and qualitative social science research methods as field research, interviewing, focus groups, survey research, experiments, and secondary data analysis. There is a better way.

Our pedagogical model turns this approach on its head. In our own courses, we have become convinced that data visualization motivates students. Because many social science students begin research methods and data analysis classes with little motivation and fearful of the course material, engaging students interest early on is critical to the success of these courses. Accordingly, we advocate starting these courses with spatial analysis and then moving to other methods than can illuminate issues the spatial analysis suggests. Once students literally "see" the power of analysis to illuminate issues they care about, they are motivated to tackle additional less visual methods that can deepen their understanding, including statistics.

Most social and public policy phenomena have spatial dimensions that can be analyzed and mapped. There is a large and rapidly expanding body of spatial data related to issues that interest undergraduate students. Our instructional modules use real data on the urbanization of the human population; endangered animal species habitats; immigrant clusters in American cities; the relationship between poverty, race, and the location of toxic sites; economic and racial segregation in school districts; and the location of Head Start centers in relation to preschool-aged children eligible to attend them.

Even in the first class sessions of our GIS modules, beginning students learn to open GIS data files, move around in digital space, zoom to examine geographic features of particular interest to them (often areas they know personally), classify data, and construct queries like the one described above. Students feel empowered because they immediately produce interesting, visual answers to questions that they themselves pose. It is only a short step from asking and answering these kinds of "what is where," "how many," "near what" questions to more analytic questions looking at the way in which data values are distributed, spatial coincidence, and unknown values interpolated from known ones. Once students see the power of key GIS commands, they are motivated to learn more complex operations and tackle the conceptual and operational challenges of good spatial analysis—indeed all scientific inquiry.

In addition to its motivational value, introducing undergraduate social science students to spatial analysis addresses an important gap in social science research methods teaching. Space matters. Geographers know this, but even sophisticated social scientists in other disciplines often neglect geographical space in their teaching and their own research. A political scientist describing voting behavior without considering where votes were cast misses an important dimension of voting behavior. An economist analyzing an economic recovery will undoubtedly find the recovery happening in different ways in different places. A sociologist studying issues of race will find that people of different races are not uniformly distributed in metropolitan areas today. Historians know that the French revolution did not unfold uniformly in different regions of France. All of these social scientists can make good use of spatial analysis. Verbal, tabular, and statistical representations of where phenomena occur are rarely as effective as maps.

How to Teach Spatial Analysis and Data Visualization in the Social Sciences

In addition to the broader pedagogical approach we have developed (using spatial analysis and data visualization in social science research, inverting the order of visualization in research, and teaching visualization first), our experience has resulted in a specific model of how to teach spatial analysis and data visualization in undergraduate social science research methods and data analysis courses.

Our project is developing two instructional modules. Ayse Pamuk, project co–principal investigator, is developing a module titled GIS Methods in Urban Analysis. The author is developing an instructional module titled *Thinking Globally/Acting Regionally.* By "module," we mean 1) a short, visual, paperback book with material appropriate for four sessions of an upper-division undergraduate social science research methods or data analysis class; 2) GIS data sets to accompany the modules; and 3) additional resources on a website accompanying the modules.

Thinking Globally/Acting Regionally will be a soft-back textbook with a CD-ROM containing datasets for analysis. The book is divided into three parts. Each of the three sections of the book follows the same pattern: 1) a chapter introducing a substantive theme undergraduates find interesting, including maps, but without discussing spatial analysis concepts or operations, 2) a chapter on spatial analysis concepts that uses examples related to the theme, and 3) a chapter that describes GIS operations at a general level. These chapters do not describe the details of using specific GIS software, but accompanying exercises take students through one widely used GIS-ArcGIS.

Our modules are appropriate for beginning students in geography or environmental studies courses who may go on to study GIS in depth, but they are designed for students in all of the social sciences. This fills a notable gap. Currently, introductory texts on spatial analysis are either intended to teach GIS or geographical information science (GIScience), a broader field of inquiry pioneered by Michael Goodchild at University of California, Santa Barbara. There are a number of good GIS books aimed at semester-long undergraduate GIS courses (Bolstad, 2002; Clarke, 2002; DeMers, 2000; Heywood and Cornelius, 2002; Theobald, 2003). There are also books aimed at GIScience (Longley et al., 1999). Suggested model curricula to teach GIS and GIScience have been developed by nationwide consortia—the University Consortium of GIS and the National Center for Geographical Information and Analysis Core Curriculum in GIScience (University Consortium for Geographical Information Science, 2004; National Center for Geographical Information and Analysis, 2000). There is no existing book designed to introduce GIS and spatial analysis to students in college-level social science courses.

GIS software is evolving very rapidly, and most GIS texts are out of date almost as soon as they are published. The conceptual chapters in our modules have been written to emphasize fundamental spatial analysis concepts that will not go out of date with the next version of existing GIS software. They are "software independent."

The module exercises emphasize "learning by doing." "Step-by-step" exercises are included to teach software applications preceded by a precise specification of learning objectives at the beginning of each exercise and a section titled "Your Turn" at the end that asks students to repeat each of the operations they have learned using different data. The culminating project for the module calls for students to apply all they have learned to real data without rigid guidance.

This project proposes a new approach to teaching social science research methods and describes a first step in implementing it. Data visualization and spatial analysis are not substitutes for the important and pedagogically demanding teaching of scientific method, quantitative reasoning, and statistics that should remain the background of a solid social science methods course. It leaves more advanced visualization (Cleveland, 1994; Card et al., 1999) to more advanced courses. Once students are engaged in the analysis process and produce good basic visual output, discussion of more advanced techniques and the complexity of understanding social reality are very much in order.

Conclusion

Social science research and data analysis should be a joyful exploratory experience in which researchers uncover understanding about issues they care about and continuously formulate and test additional hypotheses. Beginning students should be excited about doing research. Data visualization, beginning right away, motivates students because they can "see" the results of data analysis. Using visualization early and often is more than a pedagogical gimmick. Space matters and social scientists have too often neglected the spatial dimensions of social phenomena. Creative theorists such as John Tukey and Jacques Bertin have advocated a style of visual, exploratory data analysis that encourages recursive research. As research reveals truth, particularly as researchers can see what is true, they can deepen their inquiry. This NSF CCLI-EMD grant will develop materials that can help a new way of teaching social science research methods and data analysis.

BIBLIOGRAPHY

Bertin, J. (translated by William J. Berg). 1967. Reprint 1983. *Semiology of Graphics: Diagrams, Networks, Maps.* Madison, WI: University of Wisconsin Press.

Bolstad, P. 2002. *GIS Fundamentals.* White Bear Lake, MN: Eider Press.

Bossard, E. G. 2005. *Envisioning Neighborhoods.* Redlands, CA: ESRI Press, forthcoming.

Card, S. K., J. D. Mackinlay, and B. Shneiderman. 1999. *Readings in Information Visualization: Using Vision To Think.* San Francisco, CA: Morgan Kaufmann.

Clarke, K. 2002. *Getting Started with Geographic Information Systems.* 4th ed. New York: Prentice-Hall.

Cleveland, W. 1994. *The Elements of Graphing Data.* Revised edition. Murray Hill, NJ: AT&T Bell Labs.

DeMers, M. N. 2000. *Fundamentals of Geographic Information Systems.* 2nd ed. New York: Wiley.

Heywood, I., S. Cornelius, and S. Carver. 2002. *An Introduction to Geographical Information Systems.* 2nd ed. New York: Prentice-Hall.

Longley, P. A., M. F. Goodchild, D. J. Maguire, and D. W. Rhind. 1999. *Geographic Information Systems: Principles, Techniques, Management and Applications.* New York: Wiley.

Mitchell, A. 1999. *The ESRI Guide to GIS Analysis, Volume 1: Geographic Patterns and Relationships.* Redlands, CA: ESRI Press.

National Center for Geographic Information and Analysis. 2000. *Core Curriculum in GIScience Santa Barbara.* Santa Barbara, CA: National Center for Geographic Information and Analysis.

Nyerges, T., and R. Golledge. 1997. *Asking Geographic Questions.* Santa Barbara, CA: National Center for Geographic Information and Analysis.

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

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

Tufte, E. R. 1997. *Visual Explanations: Images and Quantities, Evidence and Narrative.* Cheshire, CN: Graphics Press.

Theobald, D. M. 2003. *GIS Concepts and ArcGIS Methods.* Boulder, CO: Natural Resources Ecology Laboratory, Colorado State University.

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

University Consortium for Geographic Information Science. 2004. *The Straw Report: Model Curriculum.*

Endnotes

1. CCLI-EMD 0228878.
2. Nyerges and Golledge (1997) suggest that geographical analysis can seek to answer the following: Where is it? Where does it occur? What is there? Why is it not elsewhere? How much is there at that location? How far does it extend already? Is there regularity in its distribution? What is the nature of that regularity? Where is it in relation to others of the same kind? What else is there spatially associated with that phenomenon? Do these things usually occur together in the same place? How has it changed spatially (through time)? Why has it spread or diffused in this particular way?
3. Clarke suggests geographical questions that undergraduate students can explore without statistics such as size, distribution, pattern, contiguity, neighborhood, scale, and orientation. He notes that beginning students can explore whether distributions of things in space are sparse, uneven, random, regular, uniform, scattered, or clustered.
4. Mitchell's widely used introduction to GIS analysis describes how to map the most and the least, density, what's inside, what's nearby, and change. While the precise methods Mitchell describes are subtle and statistical training provides more powerful ways to do this, the basic concepts are easily grasped by undergraduate students without statistical training.