September 2000

Uplink

MAPPING

Census and GIS

By Steve Doig
Arizona State University

Still trying to persuade the newsroom bean counters that they should pay for a good geographical information system mapping program? Tell them the U.S. Constitution requires it.

Well, okay, the Constitution doesn't actually mention computer mapping. But it does mandate a national census every 10 years. And a newsroom that tries to do census stories without a GIS will be about as effective as one that covers breaking news without cell phones and digital cameras. It can be done, but it's a lot more difficult.

The census, in fact, is a perfect story for mapping. The reason is that every scrap of census information is connected to some particular piece of geography. Moreover, a key part of most census stories is explaining how the demographics of various places compare to each other. Which do you think readers, viewers and editors would rather see: A page full of numbers, or a nicely shaded map?

Of course, journalists were printing and broadcasting maps long before GIS programs came along. But even relatively simple maps can take a graphic artist hours to complete using a drawing program. And census maps are likely to be complex. For instance, a census tract map of a typical metro area will contain hundreds of polygons to be drawn and shaded, while creating and shading a map of the nation's 3,156 counties could mean days of work. A GIS, however, can make such maps in seconds.

What to buy?

Okay, you're convinced. So start shopping for a good GIS program. There are a variety of them available, including MapInfo and Maptitude. But I and many other CAR specialists would recommend ESRI's ArcView 3.2.

ArcView is pricey – expect to pay about $1,200 for a copy. But ArcView's advantages are many. It works well on both PC and Macintosh computers. ArcView maps are easily exported into files that can be opened and enhanced by your graphic artists. The widespread use in the CAR community means you can get quick help from journalism mapping experts via the NICAR-L and JAGIS-L (Journalism and GIS) e-mail lists. You can get excellent training at NICAR's mapping boot camp scheduled for Oct. 20-

Continued on page five

CRIME

Rampage killers

By Ford Fessenden
The New York Times

In March, *The New York Times* published "Rampage Killers," a four-part series that revealed the results of a database project we undertook of 100 cases of explosive public murder over the past 50 years. These killings are not new, we said, although they may have increased in the 90s. We found more evidence of serious mental illness than of fascination with violent video games, and pointed out some important ways these criminals differed from regular murderers.

The series showed what you can do with the techniques of precision journalism. It showed how you can speak with authority when you let computer tools help you manage and distill large amounts of information.

The seeds were sowed at *The New York Times* in 1999, when Thom Shanker, an editor in the Washington bureau, suggested a few days after the Kosovo pogrom

Continued on page fourteen

Inside Uplink

This month's issue of *Uplink* focuses on one of the hottest CAR topics: mapping. With the Census 2000 data coming out soon, mapping is a skill that will be in high demand. Jeff Thomas of the *Colorado Springs Gazette* shares his experience getting started with mapping this past year (see page two). Ron Nixon of *The Roanoke Times* used mapping to reveal faults in plans to relieve flooding (see page four). Jennifer LaFleur, of *The St. Louis Post-Dispatch*, writes about how journalists in Europe are using mapping (see page six).

Access Reports

The report tool in Access is an often overlooked part of the database manager, but Mike Himowitz of *The* (Baltimore) *Sun* explains how he found it particularly useful for a project investigating Maryland's sentencing system for crimes with handguns.

SEE PAGE TWELVE

No more confusion

In the midst of a large analysis project, it's likely you've saved what seems like a zillion queries. Richard Dalton, of *Newsday*, offers some naming conventions that promise to simplify this process and make it easier to retrace your steps later.

SEE PAGE TEN

Uplink

September 2000

Volume 12, Number 7
A newsletter of the National Institute for Computer-Assisted Reporting

EDITOR
Brant Houston

MANAGING EDITOR
Mary Jo Sylwester

ASSOCIATE EDITOR
Mike Sherry

ART DIRECTOR
Kerrie Kirtland

NICAR is a joint effort of Investigative Reporters and Editors and the University of Missouri School of Journalism.

NICAR services include hands-on newsroom training in computer-assisted reporting, special academic and advanced training in data analysis.

DIRECTOR OF PUBLICATIONS
Len Bruzzese

SUBSCRIPTION ADMINISTRATOR
John Green

Uplink is published every month by the National Institute for Computer-Assisted Reporting.
138 Neff Hall Annex
Columbia, MO 65211.
(573) 882-0684. Subscriptions are $40 for IRE members, $60 for nonmembers.

Postmaster: Please send address changes to NICAR.
Send e-mail to jgreen@nicar.org

MAPPING

A beginner's lessons

By Jeff Thomas
Colorado Springs Gazette

The first moral of this story is: New CAR tools don't always make life easier. They can help you tell a smarter story – as our first attempt at using mapping software did for us in one case – but you're still likely to break a sweat getting up the learning curve.

Shortly after I returned from NICAR's three-day mapping seminar in Columbia, Mo., last October, voters in our largest school district rejected a bond issue. Three years earlier, they had approved an even larger bond. Naturally, we asked: Why the turnaround? Finding the neighborhoods where the drop was most dramatic would steer us to the kind of people we needed to talk to. Mapping software could help us do this. It also could help us see if the swing precincts had older or younger people, more or fewer kids, higher or lower income, etc.

Mapping software gave us the ability to automate the colored-pencil part, giving us much more lead time to report the story.

We had used spreadsheets in the past to unearth voting trends in the precincts, but the chore usually wasn't complete until we had broken out the colored pencils and a large precinct map. Mapping software gave us the ability to automate the colored-pencil part, giving us much more lead time to report the story. It also would show us the underlying demographics of the precincts where the vote swing was most pronounced.

Getting started

I dusted off a copy of ESRI's ArcExplorer CD, which I had picked up at NICAR's Boston conference several months earlier. This is a stripped-down version of ArcView, the standard in most GIS-equipped newsrooms. As I found out, it did the job, barely. For example, ArcExplorer lacks the extensions necessary to export a map in a format the art department can work with. Publishing a map would have been nice, but more-informed reporting was our first goal, and the tight turnaround on the story precluded building a map from scratch on the Macs, so we went without the map.

We poured the precincts' "yes" and "no" votes on the bond issue into two columns of a spreadsheet, and poured the votes from the 1996 election into two other columns. For each precinct in each year, we calculated a support ratio (yes votes/no votes). Then, for each precinct, we calculated the percentage change in the ratio from 1996 to 1999. Precincts with the most-negative change numbers were those where support had dropped most dramatically. We added a column indicating voter turnout in each precinct. This took just a few minutes.

The bigger job was to assemble the needed digitized maps, which ESRI calls shapefiles: voting precincts (to plot the change in voting behavior), Census tracts (to plot demographics), and streets (to help the eye navigate the map).

I asked the county's GIS department for a digital map of precinct boundaries. No problem, they said: That will be $500. The GIS manager, however, was on my side and sneaked me onto the next day's county commission agenda, where I made my plea for a fee waiver, which was granted. I left the meeting with the CD.

I downloaded a street shapefile from the Census Bureau's TIGER service (*http://tiger.census.gov*). A Census tract shapefile was downloaded from CEISIN (*http://sedac.ciesin.org*).

Files in hand, I opened ArcExplorer, launched a new project, and used the "add theme" function to bring the three files in. The first beads of sweat began to form when I tried to view all three themes at once. When viewing the street and Census-tract themes, I could not add the precinct theme. Or, when viewing the precinct theme, I could not add streets and tracts to the view.

Glitches

Which brings me to the second moral of the story: Always check the projections of

Continued on page three

Continued from page two: **Lessons**

your digitized maps. After wailing to the NICAR listserv, I was instructed to check the metadata of each map to make sure the projections matched. Sure enough, the precinct map was in a different projection than the street and tract maps. This is a potential pitfall that Andy Lehren spent a fair amount of time hammering into our heads at the NICAR seminar, but I guess it took a real-life example to drive the point home.

In any case, ArcView can convert projections, but ArcExplorer cannot (and won't let you add the necessary conversion extensions), so the county converted the precinct map projection and shipped me the new shapefile. All three themes now could be seen at once: precinct boundaries, census tracts, and beneath them both, streets.

The next step was to pour the demographic into the Census-tract map. This required joining a table of demographics to the table underlying the tract shapefile in ArcExplorer. I bummed a MapInfo CD of 1996 Census-tract demographics from our marketing department and moved the selected records to an Excel sheet. To make the join, I had to modify the short tract ID numbers (e.g., 2.01) in Excel to match the longer ID contained in ArcExplorer tract shapefile (e.g., 080410002.01).

Finding the neighborhoods where the drop was most dramatic would steer us to the kind of people we needed to talk to.

I saved the table as a .dbf file, closed it and used the "add table" function in ArcExplorer to open it. More problems. For reasons still unknown, the lengthened tract ID numbers on the Excel sheet did not survive the transition to a table within ArcExplorer. Again I ran to the NICAR list, where I was advised to move the Excel data to Access, export the table from Access as a .dbf file, then pick up the table with ArcExplorer. This worked. I used the same routine to move the precinct voting numbers into ArcExplorer.

In ArcExplorer, I joined the voting-numbers table to the table underlying the precinct shapefile, and the demographic table to the table underlying the Census-tract shapefile. This was the point where ArcExplorer could start doing its job.

Results

Now that the demographic numbers were a part of each census tract's record in the shapefile, we could identify the tracts with the greatest concentration of, say, school-age kids. The ones with the highest concentrations were indicated by, say, vertical lines.

And now that the vote-swing numbers were part of each precinct's record in the shapefile, the precincts with the biggest swings against the bond issue could be indicated with a color. Precincts where the swing was less dramatic appeared in successively lighter shades of the color. This is quick, point-and-click stuff.

Looking at both themes – demographics (lines) on top of vote swing by precinct (color), – we were able to produce this story nut: "A Gazette computer-assisted analysis shows that [the] steep drop in support was spread across the district. Perhaps surprisingly, the reversal was evident in precincts on the northeast and northwest sides, where simple demographics would seem to indicate strong support for schools. Those neighborhoods are filled with young families. Median household income is higher – by several thousand dollars – than it is countywide. Voter turnout tends to be high."

We never published a map, but the reporter was able to zero in on the critical precincts, ask better questions, and force the school district to confront the reality that it has serious credibility issues to deal with in neighborhoods it had counted on.

Which brings me to the final moral of this story: The effort is worth it.

Jeff Thomas can be reached by e-mail at jeff@gazette.com

Mapping Bootcamp:

This three-day bootcamp will be held in Columbia, Mo., Oct. 20-22.

Andy Lehren of *Dateline NBC* and Jennifer LaFleur of the *St. Louis Post-Dispatch* will teach the basics of mapping, geocoding and spatial analysis using ArcView GIS.

More information, including registration and hotel details, is available at *www.ire.org/training/oct20*.

Flood dangers

By Ron Nixon
The Roanoke Times

Flooding is a way of life in the city of Roanoke. The city has had major floods in 1940, 1972, 1977 and 1985. The last flood was a hundred-year flood causing millions of dollars in damage and costing several lives. In the late 1990's the city decided to enlist the U.S. Army Corps of Engineers to relieve the flooding problems.

The Corps proposed channeling the Roanoke River, which runs through the city and several surrounding localities. The Corps and the city said the project would reduce the amount of flooding by 50 percent. But other localities pulled out of the project saying it would provide little benefit in reducing flooding.

They were right.

Using mapping, flooding data and interviews with flooding experts we were able to show that the city was actually helping to create the very problems that the flood control was trying to fix. Our analysis showed that there were over a thousand properties in the flood plains in the city.

It's a recipe for disaster that we were only able to show using GIS.

And despite the numerous floods that have occurred in the last 20 years, more structures continued to be built. Even a new multi-million dollar bio-medical center is slated to be built in the flood plains.

It's a recipe for disaster that we were only able to show using GIS. Even the city had no idea of just how many structures were in the flood plain.

The stories prompted dozens of letters to the paper, and local, state and federal officials are beginning to take a closer look at building in the flood plains in Roanoke. Many residents are pushing the city to remove more structures from the flood plains. Officials from FEMA are also looking at the flood plain development in the city.

This project showed me the power of GIS, but it also showed me that you have to be careful when using the powerful tool.

I began looking into the flood dangers when three nearby localities pulled out of the flood control project and started moving homes and business out of the flood plains. I asked the city why it wasn't doing the same and how many structures were in the flood plains. They said they did keep a database of buildings in the flood plains, so we began to try and find out for ourselves.

For a couple hundred dollars, I purchased the building permit database of structures built since the last great flood, and bought digital maps from Federal Emergency Management Agency.

With help form Andy Lehren at Dateline NBC, I culled the flood plains from the digital maps. I then overlaid the locations of buildings with maps of the city and the local river.

I did several interviews and thought I had my story done until I noticed something in the building permit database. It was a structure owned by *The Roanoke Times*, my employer, but no one had ever heard of it. My analysis showed that it was in the flood plain and had been built fairly recently. I called former employees to see if they knew about it. No one did. Finally, with maps in hand, I drove to the address. I found an empty lot.

Panicked, I called the city back to find out what was wrong with the data. It took several days, but they finally got back to me and said the address showing a *Roanoke Times'* owned building in the flood plain was actually a temporary structure that had been put up for a local festival. There were dozens of other address in the building permit database like that. The city had no way of showing permanent structures in it database; using it, you couldn't tell a sign from a house.

There goes the story I thought. Then I remembered that a few weeks earlier, I had gotten the city's real estate database. The database had all permanent structures in it, the year built, type of building, everything I needed to complete the project. The impact of this project is still on-going.

Ron Nixon can be reached by e-mail at ronn@roanoke.com

Tipsheets:

"An introduction to mapping as a CAR tool," tipsheet #1159 from the 2000 IRE national conference.

This tipsheet explains the uses and importance of mapping, and it includes some suggestions for data to map.

The tipsheet can be downloaded from the IRE Web site at *www.ire.org/resourcecenter*.

Search the tipsheet database, then follow the directions below the abstract for downloading a PDF file.

Continued from page one: # Census

22. And ESRI's "Virtual Campus" has online courses in all sorts of GIS topics.

Even better, ArcView and its big brother, ArcInfo, are widely used by federal, state and local government agencies. This means a big step forward in meeting your second major need: Acquiring map files.

A GIS program like ArcView works by merging data tables, such as you might see in an Excel spreadsheet, with map files which contain the information necessary to draw and label political boundaries, streets, bodies of water and other geographic features. Thus, you'll want a collection of map files (which ArcView calls "shapefiles") to cover all the kinds of maps you're likely to need for stories: Census tracts, streets, ZIP codes, election precincts, police patrol zones, etc.

Even better, ArcView and its big brother, ArcInfo, are widely used by federal, state and local government agencies. This means a big step forward in meeting your second major need: Acquiring map files.

Shapefiles

Chances are good, though, that most such shapefiles can be had from your local government GIS specialist via a public records request. Shapefiles useful for working with older census data, such as 1990 tract maps, also can be downloaded for free from the Census Bureau at *www.census.gov/geo/www/cob* or from ESRI's ArcData Online site at *www.esri.com/data/online*.

However, the shapefiles for the 2000 census geography are not yet available. Worse, the Census Bureau indicates it may not be until next summer before those shapefiles can be downloaded from the Census site. That's too late for anyone wanting to work with the first wave of data, which will hit in March.

Happily, there is another alternative - The Census Bureau's TIGER/Line files, the national map database used by the bureau to create all its maps. The updated 2000 version of TIGER, the one that will be used to create the official boundaries for all the 2000 census geographic units, is expected to be ready just before that first wave of data is released.

TIGER doesn't come in the shapefile format that is used by ArcView. But ArcView does have a so-called extension – an add-on script that gives the program additional capabilities – that will translate TIGER files into shapefiles. If you have ArcView, you can practice with the "TIGER Reader" extension now by downloading the 1998 TIGER files for your state from *www.census.gov/geo/tigerline/tl_1998.html*.

Different geographies

One other 2000 Census mapping problem also awaits resolution. It's the question of how different your 2000 census geography is from the 1990 version, particularly if you want to look at demographic change during the decade.

State and county boundaries should be the same, but some cities will have annexed territory during the '90s. Even more likely are changes to census tracts. While most of them will be unchanged, tracts that gained or lost a lot of population may be split or combined. The Census Bureau promises to release "Relationship Files" comparing 1990 and 2000 blocks and tracts – but probably not until next summer.

Rather than wait, you'll want to make your own comparison table for your area once the 2000 geography is set. And that's another good task for a GIS.

The next IRE and NICAR Census workshop will be held in College Park, Md., Dec. 9-10. More information, including registration forms and hotel details, is available at *www.ire.org/training/censusworkshops.html*

Steve Doig can be reached by e-mail at steve.doig@asu.edu

"Reporting Census 2000": *http://cronkite.pp.asu.edu/census*

Shapefiles can be downloaded from the Census Bureau at *www.census.gov/geo/www/cob* or from ESRI's ArcData Online site at *www.esri.com/data/online*.

You can practice with the "TIGER Reader" extension now by downloading the 1998 TIGER files for your state from *www.census.gov/geo/tigerline/tl_1998.html*.

No boundaries

By Jennifer LaFleur
St. Louis Post-Dispatch

As mapping skills spread, journalists are plotting data beyond U.S. boundary files. Reporters in Denmark, the Netherlands, Sweden and other European countries use GIS (geographic information systems) software for showing crime incidents, business data or demographic trends.

At *Morgenavisen Jyllands-Posten*, Denmark's largest newspaper, journalists Nils Mulvad and Flemming Svith do mapping regularly using a GIS program called MAPS by Denmark-based ESS. Currently, according to Svith, about five people are using mapping in the newsroom.

Recently, Svith has mapped the wages of municipal employees compared to costs and wages in the private market. His story showed that about 1.4 billion Kroner (about $188 million) were being put in the wrong pockets and that about 100,000 employees earned too little. "I could only have done this story using a map," Svith said.

GIS also speeds up production, Svith says. "Before, graphics did the maps and you couldn't do any analysis and it took a whole day to create. Now, I can do a map and send it to graphics. We have a lot more maps in the paper now."

Netherlands

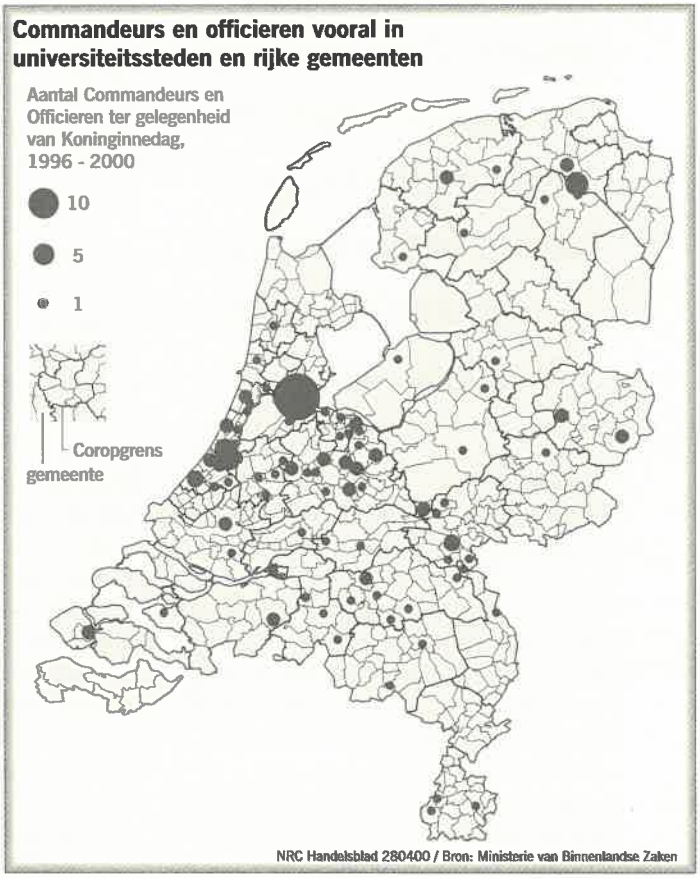
Dick van Eijk a reporter and editor with *NRC Handelsblad* in the Netherlands also uses mapping in reporting. Some are quicker mapping projects. "A growing number of maps is not the result of extensive data analysis, but not a copy of a map provided by others either. In these cases we receive a data file, e.g. from a government agency, from Eurostat, from the OECD or another organization. This file contains data for every municipality, region of country in a particular area. From these data we generate our own thematic maps using mapping software. Some simple calculations may precede the creation of the map, e.g. dividing by number of inhabitants of the area. Sometimes this is done in a spreadsheet (Microsoft Excel), sometimes in the mapping program itself (MapInfo, both on the Mac and on PC)."

At *NRC Handelsblad* there are three people familiar enough with MapInfo and Excel to do mapping – two reporters and one person in the graphics department.

Van Eijk says they also do bigger CAR projects using mapping. In April, Van Eijk did a project on royal decorations awarded on Queen's Day.

"Typically some 2,000 people a year receive such an award on Queen's Day, and some 2,000 more on other occasions. It turned out that the geographical distribution of these decorations was very uneven. That was remarkable. I had a hunch that this was a structural phenomenon, but I could not

Continued on page seven



For more information:

MAPS: *www.mapping.dk/*

ArcView: *www.esri.com*

Mapinfo: *www.mapinfo.com*

Danish Institute for Computer-assisted Reporting (DICAR)
www.dicar.org

The map to the right, created by Dick van Eijk, shows the absolute number of higher medals per municipality that he found for a story on royal decorations.

Continued from page six:

No boundaries

prove that using one year of data. So this year I analyzed all the royal decorations awarded on Queen's Day since the new law on these decorations in 1996." The new law requires uniform rules in every part of the country for determining who gets awards.

"Using five years of data I could show not only that the geographical distribution of royal decorations was very uneven, but I could also prove that this uneven distribution was consistent over the years. In some municipalities the number of decorations per head was more than twenty times as high as in others."

New journalists are also learning mapping. Peter Verwey, director of the new media department at the School of Journalism in Utrecht, the Netherlands, incorporates ArcView training into digital journalism coursework. Verwey's students used maps to study the relationship between the crime rate in a police district and the money spent on the police force in that district. They also have used maps to look at restaurant inspections in Utrecht.

As CAR has become a more common tool in newsrooms in Europe so have geographical information systems. Will records access issues slow European journalists' ability to do mapping?

Not in Denmark, says Mulvad of *Morgenavisen Jyllands-Posten* and the chair of the Danish Institute for Computer-Assisted Reporting.

"The country is divided geographically in lots of ways – postal codes, local authorities, and so on. It's very easy to divide political thinking, income, police solving crimes, prostitutes, foreigners and so on geographically and put it in a map in the paper," he said. "Actually, because of that and because of the easy way to use maps, we think mapping perhaps will be used before using database managers for a lot of journalists."

Jennifer LaFleur can be reached by e-mail at jlafleur@postnet.com

MAPPING

School test scores

By Seth Hemmelgarn
San Jose Mercury News

In California, school funding is tied to schools' performance on standardized tests. Starting with the 1999 scores, each school is given a 1 to 10 ranking. Schools with a 1 are in the bottom 10 percent of the state, and schools with a 10 are in the top 10 percent.

The state released the data in January. We wanted to map where schools with the highest and lowest ranks were in our county. The only problem was, we had the score for each school, but we didn't have the address. So we got the addresses from the state, and created a dbaseIV file with all of the information we needed.

After I geocoded the file in ArcView, I used the query wizard to pinpoint the schools with ones and tens. Not surprisingly, most of the highest-ranked schools formed a blob running along the county's western foothills, where millionaires are commonplace. The lowest-ranked schools showed up in a patch on the city's eastside, where incomes are generally lower.

This was my first official mapping assignment after I attended NICAR's incredibly worthwhile mapping bootcamp, so I wanted to have somebody else review my work. That's where Carl Neiburger, mapping guru and *Mercury News* cohort, came in.

I'd used two different queries to show the highest and lowest ranking schools, which resulted in two different shape (shp) files. Carl suggested selecting all of the ranks – from 1 to 10 – into one shp file. Then, I just eliminated the ranks we didn't need. This meant one less shp file to keep track of.

Carl also enlisted the help of Nadine Selden, from the paper's marketing research department. Nadine pulled the median income for each of our county's Census tracts from Claritas, and Carl included this in the map. The result was a map that not only showed where the high and low-scoring schools were, but what the income ranges in those areas were.

Seth Hemmelgarn can be reached by e-mail at SHemmelgarn@sjmercury.com

Mapping Listservs:

NICAR-L is a busy list (expect 10-20 messages a day) hosted by the National Institute for Computer-Assisted Reporting.

It's frequented by hundreds of working journalists whose CAR skills range from expert to beginner. To join the list, send this one-line message: SUB NICAR-L firstname lastname to *listproc@lists.missouri.edu*.

JAGIS-L: Tom Johnson's JAGIS homepage is at *http://online.sfsu.edu/~jagis/*. There's a sign-up link there.

CENSUS-L was started early this year as an offshoot of NICAR-L. Discussion is primarily about the Census, but there may be talk about mapping as well.

To join, send this one-line message: SUB CENSUS-L to *majordomo@nicar.org*.

Forming a users group

By Doug Haddix
The Columbus Dispatch

It all started with the self-proclaimed "BAG lady."

Media analyst Gwen Brown maps data extensively to promote advertising through direct mail and in "The BAG," a collection of circulars distributed weekly to 700,000 households in central Ohio.

Last summer, Gwen suggested the creation of a mapping users group open to all of the local companies under the umbrella of Wolfe Enterprises Inc. In addition to On Target Marketing, the direct marketing arm of the Dispatch Printing Co., the family-owned company's holdings include *The Columbus Dispatch* and WBNS-TV, the local CBS affiliate.

The idea for a GIS users group emerged during a three-day course on using ArcView mapping software. The class included 14 employees from the newsroom, the newspaper's marketing department and the direct marking group.

Several of the members have used mapping software for years, so they bring a wealth of experience and technical support to those of us in the newsroom who ventured into mapping just a year ago.

The class diehards formed the nucleus of the GIS users group, whose mission is to share data resources, technical expertise, mapping-related news and ideas about creative uses. Members meet quarterly, rotating among the company sites, and correspond regularly between sessions.

Several of the members have used mapping software for years, so they bring a wealth of experience and technical support to those of us in the newsroom who ventured into mapping just a year ago.

Many newsrooms could benefit from becoming part of a GIS group, whether it involves marketing and advertising or other departments at their media company, a more far-flung corporate group such as ours, or a community group of users in local government and business.

Already, we've seen numerous benefits:

- Substantial discounts on software and licenses bought at the corporate level with multiple users.
- Professional bonds with mapping experts outside the newsroom who can help trouble-shoot technical problems and assist on enterprise work. Gwen spent several hours in the newsroom looking over data for a project on the Columbus School District and offering ideas on how to map it effectively. Cherie Smith of the marketing department took the same data and ran a regression analysis on it to find the most influential demographic factors in student proficiency test scores.
- The sharing of data that already has been cleaned up and standardized. For example, our marketing department provided a database table of all drug stores in central Ohio with latitudes and longitudes. That enabled us to produce a quick map to run with a business story about an escalating retail war in the Columbus area among the big drug chains.
- Creative brainstorming on tackling upcoming news stories. Our group includes several people who have worked extensively in analyzing census data; their expertise will be pivotal in helping *The Dispatch* prepare for the mapping mother lode of demographic data from the 2000 census.
- The sharing of local shape files.
- Connections with other mapping users. Newsroom representatives have attended the quarterly meetings of a local government GIS users group and reported key developments to the Wolfe Enterprises group.
- The sharing of training tips and handouts from conferences, such as the mapping boot camp last fall by NICAR and the annual IRE convention.

Involvement in the GIS users group has kept the newsroom on a steady course as we

Continued on page nine

Past projects and investigations of *The Columbus Dispatch* are available online at *www.dispatch.com/news/special/*

The stories include:

"Dividing Lines": *The Dispatch* investigated the division of Columbus public elementary schools along lines of race, poverty, student achievement, teacher experience, building conditions and educational resources. Published June 25-28, 2000.

"Appalachia: Hollow Promises": In this five-day series, *The Dispatch* investigated the Appalachian Regional Commission's spending to find out where the money from the commission was going and why it often bypassed some of the neediest counties in the region. Published Sept. 26 - 30, 1999.

Continued from page eight:

Users group

continue to explore the practical uses of mapping software. Think of the group as a local version of the IRE and NICAR listserves – a source of inspiration, story ideas, trouble-shooting and emotional support.

Reporting uses

Over the past year, *The Dispatch* has used mapping software on projects as well as daily stories.

For a series on Appalachia, we determined how much the federal Appalachian Regional Commission had spent in more than 400 eligible counties across 13 states since the agency's creation in 1965, adjusted the figures for inflation, then linked the county totals to a color-coded map. It instantly became clear that areas in the poorest parts of Appalachia – southeast Ohio, eastern Kentucky and West Virginia – had received a lot less money than many urban areas on the outskirts of the region in places such as Spartanburg, S.C.; Scranton/Wilkes-Barre, Pa.; Birmingham, Ala.; and Pittsburgh.

Our newsroom's foray into data mapping at times has felt a bit foreign or obscure, but it has been well worth the investment of time.

On a smaller scale, when the U.S. Census Bureau released initial mail response rates for the 2000 census, we downloaded the results by census tract and mapped them on a shape file provided by the agency. Rather quickly, we spotted the census tracts in Columbus with the highest and lowest response rates. With printouts of the map in hand, reporters went to those neighborhoods to find out why residents had been so cooperative or so unresponsive.

Data mapping also enabled reporter Bill Bush to produce a compelling two-day look at property assessments in Franklin County. He analyzed 114,000 sales of residential parcels from 1990-98, calculated the percentage by which properties were over-assessed or under-assessed based on sale price, and plotted all of those dots on a map. For his base, he downloaded an up-to-date parcel map from the county government's Web site. Bill produced maps of poor neighborhoods that had been over-assessed and wealthier areas that had been under-assessed to support the project's theme of unfair valuations.

The newspaper's four-day series on disparities in Columbus public elementary schools showed us the real strength of ArcView as a reporting tool. Without mapping software, it would have been tedious if not impossible to spot geographic trends and differences among the city's 88 elementary schools.

After creating a master Excel spreadsheet file with all of the key data, we saved it as a .dbf table and moved it into the mapping software. Suspicions about differences in inner-city schools and those on the district's more affluent outskirts proved to be true in several cases. For example, we mapped the median teacher salary by school and found that the more-experienced, higher-paid teachers generally worked in the middle-class neighborhoods while rookie educators found themselves assigned to the poorest schools in the central city.

As I edited the stories for the series, I found myself using ArcView in an unexpected way: as a spot check on school locations. With 88 buildings, it was hard to keep track of which ones are on which side of the city. The electronic map helped me quickly find particular schools and scan the underlying data for particular buildings to see whether further context or information was needed in a story.

As survivors of the first mapping boot camp at NICAR and nearly a year of using ArcView in the newsroom, Bill and I can attest to the power of mapping software to produce stories that once seemed impossible.

Our newsroom's foray into data mapping at times has felt a bit foreign or obscure, but it has been well worth the investment of time.

Doug Haddix may be reached by e-mail at dhaddix@dispatch.com

Tipsheets:

"MORE Maps for You: Multiuser GIS in Newsroom," tipsheet# 1250 from 2000 IRE national conference. This handout provides a list of reasons and benefits for using mapping software in the newsroom.

"Mapping as an Investigative Tool: "Do's and don'ts," tipsheet #1223 from 2000 IRE national conference. This handout lists several do's and don'ts for using mapping as an investigative reporting tool.

The tipsheets can be downloaded from the IRE Web site at *www.ire.org/resourcecenter.*

Search the tipsheet database, then follow the directions below the abstract for downloading a PDF file.

Below are some recent projects using mapping. They're available from the IRE and NICAR Resource Center at *www.ire.org/resourcecenter/*

"Crime in the City," #15637. The Sacramento (Calif.) Bee. This 1998 special report examined thousands of records with extensive mapping to pinpoint the most crime-plagued areas of the city.

"Sticker Shocked?," #16510. Seattle Times. In this eight-part series, which ran March 5 to 12, 2000, the Times reported on real estate in the Puget Sound area and examined "the gap in value between neighborhoods that have appreciated rapidly and those that haven't and why this is happening." The series examined prices for different neighborhoods, explained the prices, and looked ahead to the future for real estate in that area.

Naming queries

By Richard Dalton
Newsday

Sometimes it seems I need a Dewey decimal system to manage my queries in Microsoft Access.

A colleague and I have created more than 260 queries in our current project. So it's essential to develop naming conventions to find the right query.

It's also important to carefully choose the names of tables, forms, reports and other Access objects. Here are some tips from myself, reporters and a Microsoft document at *http://msdn.microsoft.com/LIBRARY/BACKGRND/HTML/MSDN_20NAMING.HTM.* (I've adapted some of the tips.)

Consider using the following format to name queries: [prefixes]tag[Basename][Qualifier]. (To avoid enclosing your query name in brackets when you refer to it, don't use spaces between words in query names.)

Prefixes:

• "zz" – Instead of deleting queries (tables, etc.), prefix them with "zz" to drop them to the bottom of the list.
• "zq" – Use the prefix "zq" for temporary queries to avoid cluttering your query window.
• "__" – Prefix a query with two underscores ("__") when you're working on it. It will appear at the top of your list.

Some CAR specialists, including Dan Keating of *The Washington Post* and Geoff Dougherty of *The Miami Herald*, use a prefix to identify the order in which queries were created. Use an underscore before the number because you can't begin queries with digits. Here's a sample from Keating:

_01 merge and make combined table
_02 summary by district
_03 summary by district w_pcts.

"Having them sequenced makes it a lot easier to come back later and figure out what you were doing and why," Keating wrote via e-mail.

Query Properties:

Alternatively, you could also use the query properties (or description) field instead of the query name field to indicate the order. (see below). Using this method, "NassauTaxes" and "NassauAssessments" would appear close together in the query window, even if one was the first query and the other was the 50th.

Creation Date:

You can also sort the queries by the creation date by clicking on the top of the "Created" column in the query window.

Tag:

After the prefix, you may use a tag to identify whether you're dealing with a table "tbl", or query "qry." More advanced users might identify the query type, ("qsel" for select, "qapp" for append, "qxtb" for crosstab, etc.).

This naming convention might seem superfluous at first. But in some cases, such as selecting a table or query in report Wizard, Access merges queries and tables into one long list. So using the prefixes makes it easier to distinguish tables from queries.

Basename:

This identifies the contents of the query: "accidents," "taxes," etc.

David Herzog of the *Providence Journal* further breaks down the basename into Geography, Content, Time frame, e.g. RIGunDealersWithCriminalRecords1995-1999.

Qualifier:

Qualifiers distinguish basenames that are similar. You might use it to identify the source of the data if you're using similar data from two sources, e.g. "TaxesHuntington1999FromAssessor" and "Taxes Huntington1999FromTaxReceiver." I would also tack on "detail" or "summary." You also can indicate whether you've counted or summed up the records.

If you're focusing on taxes, regardless of location, list "taxes" first in the basename. If you're focusing on given locations for, say, both taxes and property assessments, put the location first. Then you can sort by the location, and all the queries about taxes and assessments for a given location will appear together.

Continued on page eleven

Continued from page ten:

Queries

Here's an example:

"qselNYPlaneAcc1990CountByCounty" would list airplane accidents in New York in 1990, with a count by county.

Using the properties field, accessible through "view / properties," you can further elaborate on the contents of query in plain English. You can even list the contact phone numbers of the source of the data, any quirks about the query or table and the initials of the reporter who created the query.

In the properties box, you also can check the "hidden" box to hide the query you no longer use (instead of prefixing it with "zz"). To view hidden objects, choose "tools / options," click on the "view" tab, and check "Hidden objects" under the "Show" bar.

Richard Dalton can be reached by e-mail at rdalton@newsday.com

BEST OF NICAR-L

Averaging averages

Here's a question thrown out by Griff Palmer, of *The San Jose Mercury News* recently:

Let's say we've decided to compare various investment brokerages, based on their associates' scores on a standardized test. At each level of experience, associates are given a different standardized test.

For each brokerage, we've got two years' worth of average national percentile rankings for brokers in each experience-level group. We have a reporter who wants to see if she can measure how well each brokerage has improved from 1999 to 2000. She proposes averaging the average scores of all experience groupings for each brokerage; then ranking each brokerage by the averaged average scores.

An editor who oversees this area, though, expresses discomfort with the idea of averaging average scores to come up with a single indicator for each brokerage. It is easy to demonstrate the problems inherent in averaging averages. However, these are the data we have to work with.

Anybody want to offer an opinion on proceeding in this fashion? We know the number of associates tested in each experience group, so we could at least weight the averages for each experience group before calculating an overall average.

Can anybody suggest a better way of making year-to-year comparisons of brokerages?

Answer from Sarah Cohen, *The Washington Post:*

Others may disagree with me, but I see averaging averages as legitimate (and preferable) in some situations.

The idea is to identify what question you want to ask, then decide whether you can get to the best answer. If you can't, then admit you're getting to the second-best answer.

I think it depends on the target unit of analysis. In this case, do you care about comparing the brokers within companies, or companies to other companies.

Here's an example where averaging averages seems right to me: You want to compare your city (in crime, for example) or schools (in test scores) or company (in this case) to others, not to the national average or all-student average.

Weighting the averages (the same thing as having every broker's score) would tell you how, on average, brokers in one company did against brokers in others. But leaving them unweighted tells you how a brokerage did against other brokerages.

This comes up a lot in comparing crime rates – do you compare your city to the nation (weighted) or to other cities/states (unweighted)? They're two different analyses – the first is whether residents of your city are more or less likely to be robbed or something than in the nation as a whole. The second tells you whether your city is doing any better/worse than other cities. If you weighted it, you'd end up just comparing your city to Los Angeles and New York because they're so big.

Extreme values can affect either one, depending on whether they occur in smaller places or larger places. So I'd consider checking them against medians or geometric means as well.

Below are some recent projects using mapping. They're available from the IRE and NICAR Resource Center at *www.ire.org/resourcecenter/*

The Air We Breathe, #15491. The Record (Hackensack, N.J.). An analysis of EPA data showed that New Jersey had a high level of toxic chemicals in the air. The 1999 report included maps of hazardous chemical concentrations and gave small tips for maintaining air quality.

College Community Crime Risk, #16337. APB Online. This 1999 report analyzed the crime risk in U.S. college communities, complete with risk maps and ratings. The authors used GAO reports and Department of Education investigative reports.

Other messages on the topic of "Averaging averages" can be found by searching the NICAR-L archives at *www.ire.org/datalibrary/nicarl.html*

Creating a 'rap sheet'

By Mike Himowitz
The Sun

The findings that reporter Caitlin Francke detailed in a January series were sobering: less than 25 percent of those charged with gun crimes received the five-year sentence. A third of the handgun cases were dropped before trial, often because witnesses disappeared and prosecutors made little effort to find them. Worse yet, only a handful of the guilty wound up with handgun crimes on their records.

A copy of the story will soon be available in the IRE Resource Center.

Search the database at *www.ire.org/resourcecenter*

With 300 homicides a year and many more shootings, armed robberies, rapes and assaults, Baltimore has one of the highest rates of handgun crime in the nation.

Many of its criminals are repeat offenders with long records; yet Maryland has one of the toughest handgun laws in the country. Pull a pistol during a crime and you're supposed to get five years in prison without parole – no ifs, ands or buts. Repeat offenders are supposed to get even longer sentences.

Obviously, the system wasn't working. To find out why, we obtained 27 months of criminal records from Baltimore City Circuit Court and began one of the most complex computer-assisted projects we've attempted.

Sorting through the muddle

Given the variables that go into crime and punishment, sentencing projects are fraught with statistical peril. We spent months riffling through the data, trying to make sure we were getting the right picture. Eventually we got Microsoft Access to sort through the muddle and generate reports that enabled Caitlin Francke to document the cases of hundreds of violent criminals who escaped the serious punishment the law envisioned.

While most of us are familiar with Access tables and queries, we often overlook or ignore the program's superb reporting tools. To the uninitiated, they smack of those "Sales by month and department" tables that pop up in corporate earnings statements. But reports provide a critical, third-level approach to analyzing complex data and presenting it to reporters, editors and ultimately, our readers.

I'm not going to turn this into a how-to manual for would-be report makers. If you're tolerably good at constructing queries in Access, you can learn to construct simple reports in a day or two. Instead, this piece is designed to convince you that it's worth your while to give it a try.

Reports

What makes a report different than running a query and hitting the PRINT button?

It's all in how you look at it. A report is a presentation based on information in a query, a table or even a form. Reports are inherently visual – they're designed for printed output and can be simple or fancy. Unlike queries, reports give you complete control over data placement, typefaces and even graphics. If you design a report properly, the information can jump off the page.

At the quick-and-dirty level, you can use reports to distribute information to the staff in a hurry. For example, when Maryland released school test scores in December, we imported the data into Access and turned out detailed reports for each county and the entire metro area, with lists sorted by school name, reading scores, math scores and overall performance. We posted the reports on our network as PDF files, which saved a few trees and gave reporters in competitive situations an early start on their stories.

Anne Arundel County 3rd Grade Reading Scores Page 2 of 33

This table shows the percentage of students in each school who scored satisfactory in Maryland MSPAP reading tests since 1994. The last column shows improvement or decline from 1994 to 1999.

	1999	1998	1997	1996	1995	1994	Change
Maryland	41.2	41.6	36.8	35.3	34.0	30.6	10.6
Countywide	47.8	46.7	43.9	44.3	40.4	37.8	10.0
Annapolis	19.6	35.7	12.8	44.4	40.0	25.6	-6.0
Arnold	51.6	61.5	59.3	69.4	56.7	51.7	-0.1
Belle Grove	38.5	13.2	37.9	48.0	25.0	21.9	16.6
Belvedere	64.5	60.2	51.0	45.6	57.4	48.1	16.4
Benfield	82.9	78.3	64.2	68.3	71.2	63.3	19.6
Bodkin	75.7	55.8	52.1	52.5	47.4	41.9	33.8
Broadneck	57.0	72.7	55.1	50.5	52.0	56.8	0.2
Brock Bridge	35.8	42.5	41.8	36.0	17.2	47.8	-12.0

But a report is more than a pretty face on a query. Unlike queries, which can display lists of detailed data or summary tables, a report can display both simultaneously. It can also organize data on multiple levels. For example, a single report can summarize a candidate's political contributions by ZIP code on one line, display individual contri-

Continued on page thirteen

Continued from page twelve:

'Rap sheet'

butions by donor and date below that, and then total each contributor's payments.

Handgun project

For our handgun project, we started with the court system's proprietary database, a monolithic text file that contained data from 30 different tables, some of which had more than 800,000 records. The court's administrators were cooperative, but they told us flatly that there was no way we could turn the data into a relational database.

We did, of course, using old-fashioned BASIC programming to slice the data into chunks that Access could import as tables. From these we learned that prosecutions were organized around the "case:" a particular crime such as a liquor store holdup, for which a defendant has been charged.

There were more than 40,000 cases, each consisting of multiple counts. The counts are individual charges, each numbered and titled according to the criminal code, such as robbery, attempted murder, assault or use of a handgun in the commission of a crime.

Each count, it turned out, was linked to a particular disposition, including the plea, verdict and sentence. Sentences were in turn divided into time to be served, time suspended and probation. Some of these sentences were concurrent, others consecutive. Worse yet, sentences were recorded in the "xxDxxMxxY" format, so that a sentence of 5 years, 3 months and 15 days would appear as "05Y03M15D." Not exactly arithmetic-friendly.

To complicate matters further, a single crime with multiple victims can generate several cases. During plea bargaining, those cases are usually consolidated, so that a defendant pleads guilty to only one or two counts from a single case. The rest are dropped.

Prosecutors and defense lawyers also consolidate unrelated cases against the same defendant at a single hearing because it's more convenient for them.

Developing a rap sheet

So how do you make sense of it all?

It was easy to search for handgun counts and their dispositions, and that's what we did first. What we found confirmed what prosecutors and defense lawyers had told the reporter anecdotally. In all but a tiny fraction of the cases, the handgun charge was dropped in plea bargaining.

But that didn't tell the whole story. In fact, to courthouse veterans, this behavior is perfectly normal.

So we had to switch gears and analyze what happened to each defendant who was charged with a handgun count as part of any case. That meant calculating the total sentence for all charges in all cases (minus suspended time). From that, we developed a list of defendants who got less than the minimum time, or no time at all. Finally, it was time to see what happened to each of them.

This is where Access' report functions came in. Based on a query that linked four tables (defendants, cases, charges and dispositions), we developed a "rap sheet" report for each defendant that showed exactly what happened to each count in each case, organized in a fashion that made it possible to identify plea bargains and sort out consolidated cases.

Now, armed with statistical summaries and examples from hundreds of individual cases, Francke talked to prosecutors, defense attorneys, judges and victims. What she found was a court system overwhelmed by sheer numbers.

With too many bad guys and too little support, prosecutors were dismissing any questionable case and settling for bargain-basement pleas in many more.

As a result, even those who got more than the minimum five-year sentence rarely wound up with convictions for use of a handgun on their records. The next time, they appear in court (and history shows that many will be back), the system won't mark them as a repeat handgun offender, and their lawyers will be free to bargain once again.

As part of the reporting process, Francke gave copies of our reports to prosecutors and judges. Most had never seen that kind of analysis before. And though our series was highly critical of the way they conducted their offices, none of them challenged our statistical conclusions.

Mike Himowitz can be reached by e-mail at mike.himowitz@baltsun.com

Continued from page one: **Killers**

More about the reporting of the "Rampage Killers" story is featured in this month's issue of the *IRE Journal.*

"Rampage Killers" is available in the IRE Resource Center, story #16521. To order a copy, call (573) 882-3364.

began that we try to get a better handle on the atrocities that were just beginning to be reported by refugees at the province's borders.

The *Times'* database team – at the time comprised of Josh Barbanel and me – suggested we take the reports, which the research department was pulling from the wires and Web sites of human rights groups and governments, and parse them into a few simple but useful fields. Josh designed a Lotus Notes application that would allow a clerk to input the important fields.

It showed how you can speak with authority when you let computer tools help you manage and distill large amounts of information.

Within a few days, we had 200 records in that database, and could analyze it based on the source of report, the location of the atrocity, and the details. We were able to show, when Kosovo was still being purged, that there was a geographical pattern to the Serbs' actions.

Selling CAR

We were in the midst of a "campaign" to sell computer-assisted reporting to the newsroom that Terry Schwadron, the editor of research and technology, had launched last year. Josh's Kosovo database became a paradigm for Times editors to think about how to apply computer assistance to reporting projects. In late summer, Dean Baquet, the national editor, came to me with the idea of establishing our own database of explosive public murder, and use it to unravel the apparent epidemic of school and workplace shootings in America over the last few years.

I set up a Lotus Notes application, with consulting help from George Landau of NewsEngin. I've been doing computer-assisted reporting for 15 years, and I've managed databases on PCs with, in turn, KnowledgeMan, Q&A, Alpha4, FoxPro and Access. This is my formula for migrating to Lotus Notes: Take everything you have learned and throw it out the window.

Using Notes 4.6 to create a Web application only made the learning process steeper and less intuitive, because, George explained, it wasn't optimized for that task. That's why they created version 5.0. But the people who run computer systems for the business side of *The Times* supported 4.6. And aside from Josh's experiments, news technology, where I work as a reporter, had no experience with Notes. (At one point, in frustration, I stormed in to Terry's office and demanded that he assign someone with technical skills to learn Notes so we could get this now slow-moving project going, and his answer, essentially, was: "OK. You.")

Nevertheless, Notes was the right way to go, for a database with limited records that amounted to a kind of hybrid between a text database and a formal parsed database. With Notes, I set up a form that the reporters used to record information into 90 fields. Fox Butterfield, one of the reporters on the project, liked to be able to paste huge sections of text into the form, and Notes allowed me to include text fields of nearly unlimited length. Notes' ability to handle multiple answers in a single field also was important. Over about three months, we entered information about the victims, the guns, the warning signs, the mental health histories, as well as demographics.

One of my jobs at *The Times* is to spread computer skills. On this project, I was always torn between developing applications for the Web and trying to get someone to learn, say, Excel, so they could creatively pose and test their own hypotheses with the data. Ultimately, I developed six or seven canned Web-page views so the reporters could sort various fields on the Web. I didn't spread many CAR skills that way, but I did spread enthusiasm.

Analysis

My analysis was pretty straightforward, and accomplished simply by sorting the views. A view in Notes allows you to specify which fields you want to display across a page, and gives you some ability to sort the records,

Continued on page fifteen

Continued from page fourteen:

Killers

just like a spreadsheet. To know how many cases in which an offender had previously been institutionalized, I'd call up the view that included the mental health variables, click to sort the "PrevHosp" column, and count the "Y's". With only 100 records, that was perfectly satisfactory and manageable.

It became clear in the research department that it would be impossible to use newspaper databases to develop a full picture of these crimes. Too many would slip through the cracks, especially in early years. We would not be able to develop prevalence, so we concentrated on putting together as many as possible that fit the description of multiple victims, in a public place, not connected to a robbery or domestic strife. We would acknowledge that this wasn't a sample or an entire population, and go with 100 cases.

We were able to show that our group was vastly different from the one that criminal justice institutions were used to handling.

We've been asked many times, by criminologists, psychiatrists and people who participated in the Web forums, to publish the entire database. But at this point, largely for proprietary reasons (we might write some more stories using the data), we've decided not to do so.

To get at the issues of prevalence and trend that went unanswered by our 100-killer database, I tried an analysis of FBI homicide reports, using some more usual database tools on that piece of quantifying. I downloaded the data files – 15,000 to 22,000 murders a year going back to 1976 – from a University of Michigan site (*www.icpsr.umich.edu/NACJD/ucr.html*). NICAR sells them as well, and it might be a good idea to get them all from one place. I had to buy the most recent year from NICAR, and the formats were different, which was an additional headache.

It took a few weeks of sorting through the data, and reading the criminology papers that had used them, to figure out if we could make the data work for trend analysis of rampage murder. The data are not collected in a way that makes this easy, but perhaps that's not surprising. They are rare crimes, and only recently had events suggested the need for a census of them.

I used SPSS to isolate murders with three or more victims, and classified them into three groups: Armed robbery or gang killings; familicide, in which the principal victim was a family member; and non-robbery, non-family murder. I ran crosstabs by year, and copied the results to an Excel spreadsheet to create charts. There were only about two to three dozen events in each category each year. They varied from year to year, as rare events often do.

I used yearly population estimates to calculate rates, and it was immediately clear that one of the groups had jumped in the 1990s - rampage killings. Was it significant? I tested them for a linear trend, and there was no significant one for the entire period, 1976 to 1997.

Josh suggested an independent-samples t-test to compare the years before and after 1990, and the increase, about 30 percent, was significant at the $p<.01$ level. The test tells you if the means of two groups are significantly different, taking into account the variability and number of cases.

To illustrate how these killers were different from regular murderers, I also grabbed a couple of great datasets from the Michigan site: A survey of inmate populations that allowed me to calculate education, employment, and military background for regular murders; and the National Crime Victimization Survey, from which I could get the time of day most regular crimes happen.

We were able to show that our group was vastly different from the one that criminal justice institutions were used to handling. They're whiter, older, better educated but less fully employed. They're much more likely to have a military background, kill in the morning and make no plan of escape.

Ford Fessenden can be reached by e-mail at fordfess@nytimes.com

Uplink story ideas

Have you or one of your colleagues recently published a story using CAR that has not been done before or involved particularly difficult data work?

Do you know of a technical problem (or its solution) that others may like to hear about?

Or is there some issue or beat that we haven't covered?

If you have a story idea, we'd like to hear from you. Please contact managing editor MaryJo Sylwester either by e-mail at maryjo@nicar.org or by phone at (573) 884-7711.

Bits, Bytes and Barks

FECinfo offer

IRE members can now receive a very discounted subscription rate to FECinfo's enhanced campaign finance data site. The subscription to FECInfo Pro enables you to research and extract highly specialized reports and information on the sources of federal political contributions.

This rate is going to news organizations for $2,500 a year. But as an IRE member, you can receive the same yearlong subscription for $150.

This offer is good though Dec. 31, 2000. More information is available at *www.ire.org/fecinfo*

NICAR data updates

The NICAR Database Library has updated the INS Legal Residency Data for fiscal years 1997 and 1998, and the Federal Procurement Database for fiscal year 1999 – the most current data available for both datasets.

The Federal Procurement Database, which is from the U.S. General Services Administration, includes all federal contracts worth more than $25,000.

NICAR has data for each fiscal year, 1992-1999. Cost for one fiscal year, entire United States, is $150, $175, $200 based on news organization size. A single state slice of one fiscal year costs $50, $75 or $100. Data for all years, entire United States, cost $325, $375, $425.

More information, including sample data, is available at *www.ire.org/datalibrary/databases/fedcontacts/*

The Legal Residency data, from the Immigration and Naturalization Service, includes characteristics of aliens granted permanent residence in the United States. Names and exact addresses are redacted under the Privacy Act.

More information, including sample data, is available at *www.nicar.org/data/ins/*

Cost for each fiscal year is $40 for circulation below 50,000 or 50-200 market; $50 for circulation 50,000 to 100,000 or 25-50 market; $60 for circulation over 100,000 or top 25 market. Prices for all years are $125, $150 and $175, based on size of news organization.

To order, call the database library at (573) 884-7711.

Upcoming bootcamps

Here's the schedule for IRE and NICAR basic bootcamps during the upcoming year. These bootcamps offer hands-on training in Microsoft Excel and Access, as well as tips on searching for information and importing data from the Web.

More information, including registration and hotel details, is available at *www.ire.org/training/bootcamps.html*

November 12-17, 2000 — St. Petersburg, Fla., at The Poynter Institute
January 7-12, 2001 — Columbia, Mo.
March 25-30, 2001 — Columbia, Mo.
May 20-26, 2001 — Columbia, Mo.
July 15-20, 2001 — Columbia, Mo.
August 5-10, 2001 — Columbia, Mo.

An advanced bootcamp on statistics will also be held at the University of North Carolina in Chapel Hill, May 13-17, 2001. This bootcamp features hands-on training in statistics and SPSS.

Investigative Reporters and Editors, Inc.
138 Neff Annex
University of Missouri
School of Journalism
Columbia, MO 65211

NON-PROFIT ORG.
U.S. POSTAGE
PAID
PERMIT NO.286
COLUMBIA, MO. 65211