Uplink

Sept. • Oct. 2005
Volume 17, Number 5

Published bimonthly by the National Institute for Computer-Assisted Reporting

www.nicar.org

WORKPLACE

Immigrants' deadly rides

By Mike Schneider
The Associated Press

The early morning fog and darkness were still two hours from lifting, but already the parking lot of the La Mexicana grocery store in the farming community of Immokalee, Fla., was packed with 200 farmworkers waiting to be picked up for a day of harvesting in southwest Florida's orange and vegetable groves.

Creaky buses that looked two decades old pulled up. Some farmworkers jumped into the back of U-Haul trucks that were then pulled closed, shutting out fresh air. Others jumped into vans that had their seats pulled out to make room for more people.

I had come to Immokalee to write about a farmworkers' campaign to pressure Taco Bell, one of the largest purchasers of Florida tomatoes, into demanding that Florida growers improve farmworker wages.

But I was struck by the decrepit vehicles the farmworkers were forced to ride in to work. Some had

continued on page 14

PHILIP MEYER JOURNALISM AWARD

Contest honors stats pioneer

By Amanda Buck, *IRE and NICAR*

The first time Ted Mellnik met Philip Meyer was at a seminar in the late 1980s. Meyer, already considered the father of precision journalism, was teaching journalists like Mellnik how to use social science research methods in their work.

Mellnik, database editor at *The Charlotte Observer*, still remembers that seminar. A few years ago, as he considered the impact of Meyer's use of social science research methods on modern journalism, Mellnik decided

continued on page 3

SPOTLIGHT: BUSINESS

Borrowers buried under debt, fees

By Carrie Teegardin, *The Atlanta Journal-Constitution*

James Oglethorpe founded the colony of Georgia as a haven for people who had been imprisoned for their debts. But today, Georgia has become perhaps the least forgiving state in the nation for people who owe money.

A team of reporters at the *Journal-Constitution* – Ann Hardie, Alan Judd and myself – went to work last year to understand how Georgia was transformed from a debtor's colony to a creditor's paradise. We mined lots of data to understand the lending landscape in Georgia. The series, which kicked off with four parts in January and is continuing, was filled with findings unearthed by our analysis of various databases. To illustrate these findings for our readers, we relied heavily on the personal stories of people in the lending business and their customers.

Here's a look at some of the major findings and the data we used:

continued on page 16

SPOTLIGHT:

For more about tracking business see:

- Detecting rise in bankruptcies, p. 8
- Deflating job program claims, p. 10
- Taking stock of top companies, p. 12
- Resources from IRE, p. 13
- The September-October 2005 *IRE Journal*

Bits & Bytes

Visiting journalist

Yong-Jin Kim, a veteran reporter for KBS, South Korea's largest television network, has joined IRE and NICAR as a visiting scholar. At KBS, Kim served as editor and senior reporter for the Investigative and Current Affairs teams. He also is the former director of the National Union of Media Workers.

Kim will spend the school year learning about U.S. investigative reporting, including computer-assisted reporting, and working on a reporting project. He received his master's degree in journalism from Dong-A University in Busan, Korea, and is the recipient of numerous awards from the Journalists Association of Korea.

The Korea Press Foundation is sponsoring his visit with IRE and NICAR.

Global gathering

IRE and NICAR participated at the third Global Investigative Journalism Conference Sept. 29-Oct. 2 in Amsterdam. Top international journalists led debates, workshops, and discussions of case studies on topics affecting global investigative journalism. David Boardman, president of IRE's board of directors, and Brant Houston, executive director of IRE and NICAR, as well as other IRE members, participated.

The VVOJ, the Dutch-Flemish Organization for Investigative Journalists, organized the event. More than 25 computer-assisted sessions were on the schedule, with topics ranging from "Life Beyond Google" to

continued on page 4

Change of seasons in Database Library

By David Herzog, *NICAR and Missouri School of Journalism*

It's the start of a new academic year here at the Missouri School of Journalism – home to IRE and NICAR – so it's time to introduce the graduate research assistants in the Database Library who also serve as contributing editors to *Uplink*.

We have a talented crew, as usual. You may have already talked to the staff if you have called to order data or ask a technical question.

Brian Hamman is a second-year master's student who is returning to the Database Library. He has worked as a reporter, graphics reporter and online assistant city editor for the *Columbia Missourian*. He is also an editor for MyMissourian, a citizen journalism publication based at the Missouri School of Journalism.

Catherine Rentz Pernot is a second-year master's student specializing in business reporting and is returning to the Database Library. Over the summer she had the Pulliam Fellowship for business reporting at *The Indianapolis Star*. She also has reported for the *Columbia Missourian*.

Ben Welsh is a second-year master's student in newspaper journalism and he is new to the Database Library. Before coming to Missouri, where he

continued on page 23

About our contributors ...

Michelle Breidenbach covers state government for The *Post-Standard* in Syracuse, N.Y.

Lee Davidson is a special projects reporter for *The Deseret Morning News* in Salt Lake City. He is a 2005 alumnus of the IRE and NICAR Computer-Assisted Reporting Boot Camp.

Helen Huntley is personal finance editor at the *St. Petersburg Times*. Her column, "On the Money," runs in the Sunday business section.

Patrick Lakamp is an investigative reporter for *The Buffalo News*. He is a 2005 alumnus of the Mapping Data for New Stories mini-boot camp and a graduate of the Missouri School of Journalism.

Mike McAndrew is a general assignment reporter for *The Post-Standard* in Syracuse, N.Y.

Scott North specializes in investigative reporting and special projects at *The* (Everett, Wash.) *Herald*. He is a 2004 alumnus of the mapping mini-boot camp.

Mike Schneider is the Orlando, Fla., correspondent for the Associated Press.

Carrie Teegardin is an investigative reporter for the *The Atlanta Journal-Constitution*. She is a 1999 alumna of the CAR boot camp.

Rosemary Winters is a business reporter for *The Salt Lake Tribune*.

Scott A. Zamost is an investigative producer for WTVJ-Miami. His article about flaws in the Broward County, Fla., 911 emergency response system ran in the May-June 2005 *Uplink*. His biographical information was omitted in error.

Contest

continued from page 1

he should do something to help honor Meyer's contributions.

Furthermore, although Mellnik knew from personal experience that journalists across the country use computer-assisted reporting and social science methods to produce "spectacular" stories, their work often was not recognized in prestigious contests.

That, Mellnik thought, should change.

Three years later, it has. In September, the National Institute for Computer-Assisted Reporting (NICAR), IRE and the Knight Chair in Journalism at Arizona State University's Walter Cronkite School of Journalism and Mass Communication introduced an award designed to recognize the best journalism done using social science research methods. NICAR is a joint program of IRE and the Missouri School of Journalism.

Named in honor of Meyer, who holds the Knight Chair in Journalism at the University of North Carolina at Chapel Hill, the Philip Meyer Journalism Award is planned as an annual contest that will make real the kind of recognition Mellnik imagined.

"I felt we had a lot of people who did spectacular work, and that they needed to be recognized," Mellnik says. "In awards, we tend to honor stories that get results – that put people in jail or result in reform legislation or something like that. But there's also a benefit to stories that establish a methodology and give reporters across the nation a tool that enables them to do good work for their readers."

Mellnik approached others with his idea, including IRE Board Member Steve Doig, the Knight Chair in Journalism at Arizona State University. Formerly associate editor for research at *The Miami Herald*, Doig has years of experience using social science research methods. For the past several years, he also has served as a member and chair of the IRE Awards committee.

Doig's work with the committee convinced him that some award-worthy work was being overlooked.

"The problem with investigative awards (in this context) is that one of the real criteria is that the story should help uncover information that others were trying to keep secret," he says. "Yet there's a great deal of good CAR, precision journalism kinds of things, for instance dealing with the census, that isn't strictly investigative. You're not uncovering wrongdoing; you're uncovering how society works, for instance."

Meyer is known for doing that. He pioneered in using survey research as a reporter for Knight Ridder newspapers to explore the causes of race riots in the 1960s.

"He inspired the profession to be more sophisticated and precise in its understanding of and reporting on social issues," says IRE and NICAR Executive Director Brant Houston. "He has been a mentor to many, many journalists – including myself – and he encouraged us to learn and use methods that made us and our profession much more credible."

Doig called Meyer the "prime mover" in raising awareness among journalists to the need for using social science methods.

"It started with his 1973 book, 'Precision Journalism.' He wrote that really 10 years before it started becoming easier to do the things he was talking about," Doig says.

Now in its fourth edition, "Precision Journalism," is widely used as a textbook in journalism schools.

Social science research methods that would make stories eligible for the award draw on tools used across a variety of disciplines, from sociology and demographics to geography and political science.

The award "will help recognize stories that have used social science methods to reveal patterns, trends and sometimes abuses in areas including social or economic or criminal systems," Houston says.

Because IRE has been the leading organization in the area of using social science methods in investigative reporting and training people to do that, it made sense for the organization to cosponsor the awards, Houston says.

Nominations for the award must be postmarked by Oct. 31. Appropriate work published or aired between Oct. 1, 2004, and Sept. 30, 2005, will be considered.

A panel of judges including journalism educators and social scientists will choose first, second and third place entries. Judging will be complete by late December, and the awards will be presented during the 2006 CAR Conference scheduled for March 9-12 in Newark, N.J.

To see a list of judges, download an entry form and read further details about the contest, visit *www.ire.org/meyeraward*. Additional questions should be directed to the IRE Resource Center at 573-882-6668 or rescntr@ire.org.

All entries will be placed in IRE's Resource Center story library.

Contact Amanda Buck by e-mail at amandab@ire.org.

Bits & Bytes

continued from page 2

"Social Network Analysis" to three sessions covering "Mapping the News."

For more information about the conference, and other IRE and NICAR training events, visit *www.ire.org/training*.

Pollution data

The IRE and NICAR Database Library recently updated its Toxics Release Inventory (TRI) data from the U.S. Environmental Protection Agency for 2003. The data is available starting with 1987.

The EPA now requires reporting of about 650 toxic chemicals and toxic chemical categories. The TRI consists of information about on- and off-site releases of these chemicals and other waste management activities reported annually by industries and federal facilities. The data shows what happened to the chemical after it left the facility, such as whether it was recycled, treated or released into the environment, and where the chemicals were sent. One-time accidental releases are also included in the database.

Though this data is available for free through a download from the EPA Web site, IRE and NICAR's Database Library analysts have cleaned the data significantly, making it easier for journalists to use immediately. The Database Library provides additional fields containing standard date formats and latitudes and longitudes that can be mapped using geographic information system (GIS) software. For a sample of the table, samples of other tables and more information about TRI see *www.ire.org/datalibrary/databases/toxic*.

FIRST VENTURE

Looking anew at dams, taxes, bridges

By Lee Davidson, *Deseret Morning News (Salt Lake City)*

My children say I am ancient. They may be right, because I now see how much older than computer-assisted reporting I am.

The realization came as I made some first ventures into CAR by essentially updating stories that I wrote 20 years ago – then without the help of computers – about the conditions of local dams and bridges, and about those who fail to pay property taxes.

It is much easier to dig into these stories with computers, even for a CAR rookie like me who must ask my teenage and twenty-something kids how to fix computer crashes and burps. I feel like the caveman who discovered that hunting with a spear beats hunting with rocks. (I also envy the true CAR experts who hunt with the equivalent of rifles and bazookas).

After 17 years as the Washington, D.C., correspondent for the *Deseret Morning News*, I recently returned to Utah as the newspaper's special projects reporter. Because I had limited knowledge of CAR, my editors sent me to an IRE and NICAR Boot Camp in January.

One of the boot camp training exercises used data from the National Inventory of Dams compiled by the U.S. Army Corps of Engineers. Using a Microsoft Excel spreadsheet, I saw how quickly I could sort and count dams by inspection date, river, nearby cities – and whether they are "high hazard," meaning people would likely die if they fail.

It brought back memories of 20 years ago when I sat in the state engineer's office for days poring over paper records for a time-consuming but good project about the most dangerous dams in Utah.

I decided to update my old story by testing my new CAR skills. At that moment, the National Inventory of Dams was a couple of years out of date. (The IRE and NICAR Database Library has updated data that it provides to journalists for a fee; see *www.nicar.org/data/dams/*).

Because I wanted to see, in part, whether Utah was regularly inspecting high-hazard dams, I surfed the Web to find a more current and local version of the database. A Google search for Utah dams led me to an online database at *http://des.utah.gov/nathaz/damsafety.htm*. It had the same information as the national inventory and is updated monthly. Plus, it has links to online inspection files.

I copied and pasted data from the state Web site into my own Microsoft Excel spreadsheet for sorting and counting.

I called the state dam safety engineer to tell him I was updating my old story, and wanted to interview him (as I did his predecessor long ago) about which dams he felt were the most dangerous. He told me his office maintains another database – unavailable online – that gives a numerical safety rating for each dam. He e-mailed it to me as an Excel file, and I could quickly see which dams the state considered "unsafe" because they did not meet current guidelines.

Sorting allowed me to see how many unsafe dams were upstream from major Utah cities and to make lists of the worst dams. I discovered that the engineer's office also had a database about dam failures in the state and what caused them. I obtained a copy,

continued on page 20

PARKING

Uneven hand in enforcement

By Patrick Lakamp, *The Buffalo News*

Our idea to map Buffalo's parking tickets originated in a coffee shop where employees shout warnings to customers when they spot a parking enforcement officer. The ritual can render them practically hoarse.

Having coffee with a source for an unrelated story, we heard four warnings in an hour. Meanwhile, the source complained about her eight tickets in two years and $300 in fines.

After that meeting, staff reporter Jay Rey and I – under the supervision of Susan Schulman, head of *The News'* investigative reporting team – decided to study the hundreds of thousands of parking tickets issued in the city using computer-assisted reporting. (For more details about reporting the story, see the September-October issue of *The IRE Journal*.)

We filed a state Freedom of Information Law request for an electronic database of all tickets written January 2003 through October 2004. The city provided, at no cost, a Microsoft Access database on CD containing 366,500 parking tickets.

We ran hundreds of queries in Access and analyzed the data using ESRI ArcView 3.2 geographic information system (GIS) software. News researcher Andrew Bailey helped with the data work.

The data was fairly clean. Parking officers issue tickets by using a hand-held device capable of scanning a vehicle registration sticker and automatically recording the time and date. The officers enter the location and violation type.

We found inconsistencies in the recorded street addresses. Some officers, for example, entered "Bailey Ave." while others entered "Bailey Avenue." We needed consistency with addresses, because we planned to use ArcView's geocoding function to turn those addresses into points on a map. We wanted one map showing the tickets by neighborhood.

Bailey created a new table from a GroupBy and Count query of street name to identify streets with the most parking tickets.

Mapping patterns

Bailey had a good handle on the geocoder's preference for how street names should be identified (Northrup West, rather than W. Northrup), so that's how he standardized street names.

We found inconsistencies in the recorded street addresses.

Sometimes he used Access' Find and Replace function to standardize addresses, but more often he relied on update queries. When Bailey standardizes addresses, he usually creates a table that has one column containing the variations of the street name and then creates a column that contains the correct, standardized, street name. He then uses this table in the update query.

Bailey geocoded the cleaned-up addresses and we ended up with 290,000 matches out of our original 366,500 records. Complete success is elusive when geocoding. Some ticket addresses couldn't be mapped because officers sometimes failed to record the street number in the violation location field.

We could overcome that for violations on short streets contained in one neighborhood, but not for those on streets that cut through several neighborhoods. Ashton Place, for example, is a one-block street in Buffalo's South Park neighborhood. Tickets written on Ashton Place, but lacking a precise address, could therefore be attributed to South Park.

Bailey calculated the number of tickets for each neighborhood by using ArcView's spatial join. Spatial joins allow you to assign information from one map layer to another. Bailey overlayed the polygon neighborhoods map from City Hall and joined it to a map layer of parking ticket points. Then, after the join, he calculated the number of tickets in each neighborhood. That number was then displayed for each neighborhood as the percentage of all tickets mapped.

We could see downtown had the greatest percentage of all parking tickets. But several Upper West Side neighborhoods, each having 5 percent to 10 percent of the tickets, stood out, too. This indicated the Upper West Side was the most ticketed residential area of the city.

Zooming in

Another map created using ArcView showed the Upper West Side close up. Here, Bailey took the county property parcel map and zeroed in on the residential neighborhood. He grouped individual parcel polygons into street blocks – grouping roughly 3,500 parcels into the roughly 160 blocks in our area and assigning a name for each block. Bailey did this by manually selecting the polygon parcels and then using the Dissolve Feature tool in the Geoprocessing Extension

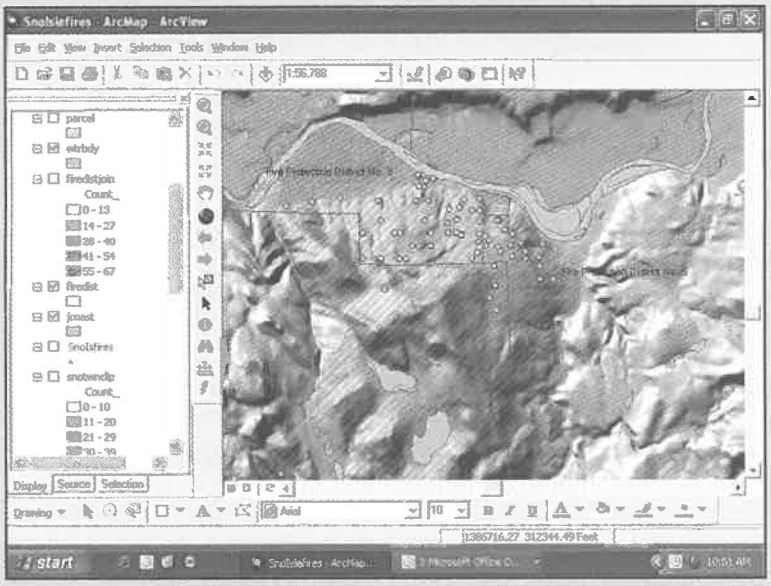
continued on page 21

MAPPING IT OUT

The latest uses of mapping in news reporting.

Homes, drought a fiery mix

By Scott North
The (Everett, Wash.) Herald



It was a moment of potentially dangerous convergence.

My editors looked at their story budget and saw a hole on the front page a few days hence. At that moment, I didn't have anything pressing on my schedule.

Tag. I was it.

The assignment – find out how an unusual winter drought had affected this normally soggy place – could have been just another weather yarn. Instead, some quick data analysis using geographic information system (GIS) software highlighted a developing hazard that had been taking shape for years in the forestland outside area cities and towns.

The *Herald* serves a community of more than 630,000 people in rapidly growing Snohomish County. The county, roughly a half-hour drive north of Seattle, is about the size of Connecticut. Most of the population is concentrated not far from the tideland shores of Puget Sound. To the east are rugged foothills, evergreen forests and thousands of acres of Cascade Mountain Range wilderness, including a slumbering volcano.

Lots of rain and snow usually keep this place green. Not this year. The driest winter in 110 years led the governor to declare a drought emergency. Officials forecast what could be the worst wildfire season in years.

If the fires got going, where would they likely appear, and who would be at greatest risk?

Using mostly free data arrayed in a GIS, we were able to report that 5,500 people in our county appear to be living in the areas most at risk for wildfires. We found planners have been approving about 100 new homes in those fire-prone areas each year, many outside fire-protection districts.

My partner on the story, Diana Hefley, knew the state Department of Natural Resources had identified portions of the county as particularly at risk for wildfires. The determinations were based on a number of factors: fire history, presence of brushy fuel, terrain, ease of access, a low level of organized fire protection.

A quick check showed the department had created mapping shapefiles of these fire zones and had posted them on the Web. The state also had 30 years' of wildfire incidents, their locations mapped using X,Y coordinates.

After consulting the metadata for these records, we opened the files in ESRI ArcView 9. It is at moments like this that I love Washington's strong public records laws. In addition to the free data on wildfire threats, past and future, my GIS data cache already contained shapefiles of property parcel lines, a decade of building and construction permits already geocoded into points for use in another project, fire district and city boundaries, major roads and rivers, population estimates at the Census block level, plus a fat database containing property tax assessment records.

After making sure all the GIS data was in the same projection, I set ArcView to the task of clipping information that would form the core of our analysis.

Clipping has become one of my most-used GIS tools. The process is simple. First, decide what data you

GIS software discount for IRE members only

Members of Investigative Reporters and Editors, Inc., qualify for discounts on geographic information system (GIS) software from ESRI, the publisher of ArcView.

ESRI is offering ArcView GIS single-use licenses at no charge to IRE members who agree to attend a GIS training event conducted by IRE and NICAR or ESRI. Purchasers must sign a three-year maintenance agreement with ESRI at a cost of $400 a year, with the first year's fee waived. ArcView, the GIS program most widely used by journalists, lists for $1,500. During the maintenance agreement period, purchasers will receive software upgrades and technical support.

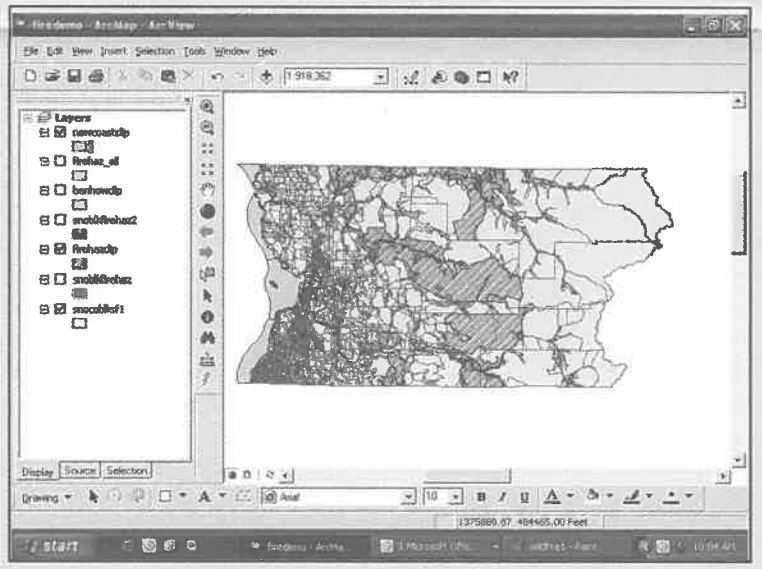
IRE members must attend a qualifying training session within one year of entering the agreement with ESRI, which is based in Redlands Calif., and has been a regular exhibitor at the annual IRE and CAR conferences. Qualifying sessions are IRE and NICAR's Mapping Data for News Stories mini-boot camp, offered two times a year with the next scheduled for Jan. 6-8, 2006; an online ESRI Virtual Campus course, and ESRI classroom training.

For more information about IRE and NICAR training see: *www.ire.org/training/advbootcamps.html.*
For more information about ESRI training see:
www.esri.com/training_events.html

IRE members also can purchase discounted extension programs, which expand the analytical capabilities of ArcView. The single-license cost for Spatial Analyst, 3D Analyst and Geostatistical Analyst is $1,500 each. That is a 40 percent discount off the list price of $2,500 each.

To obtain an order form, please contact John Green, membership services coordinator for IRE, at jgreen@ire.org or 573-882-2772.

want to clip, and place that into an ArcView map. Next, overlay the shapefile that contains the area you want to use as a cookie cutter to punch out the relevant records. If you have your GIS ducks in a row, clipping works equally well to extract a section of points (say wildfire locations, or building permits) or to snip out relevant polygons, such as census blocks, or property parcels.



We clipped data using the state's shapefiles of fire hazard zones. ArcView automatically creates a new shapefile of the data it clips, complete with its own underlying database table in dBASE format.

ArcView has SQL query capability, but I still like to export data tables to Microsoft Access database manager for analysis. In short order, I knew how many people lived in the fire zones (by analyzing clipped Census block data), how many new homes on average had received construction permits each year (by analyzing clipped building permits) and the parcel numbers of each piece of property within the hazard areas.

Parcel numbers are the key to making joins to property data. Using Access, I joined the data table of clipped parcels to property assessment records. Because assessors record the year of construction for each home they record, I was able to track the development history in the clipped area, year by year.

It was then time to help the data start talking.

After tabulating the home construction data, I dropped the results into a Microsoft Excel spreadsheet and prepared a chart. The resulting trend line, and some quick math, showed that more than three-fourths of the homes in the fire-prone areas had been built

continued on page 22

SPOTLIGHT: BUSINESS

Database answers bankruptcy questions

By Rosemary Winters, *The Salt Lake Tribune*

Since 2002, Utah has held the dubious distinction as the personal bankruptcy capital of the nation, with one of every 37 households filing for bankruptcy each year.

Yet the reasons for the state's high rate were largely unknown before *The Salt Lake Tribune* launched a computer-assisted project last year.

The Tribune's business desk first studied Utah's rising bankruptcy rate in 1999 as the state was soaring to record levels, but before it hit No. 1. This time we decided to take a closer look at the cases and record more information from each filing.

Business editor Lisa Carricaburu led discussions about what kind of data needed to be collected for the project and what questions we wanted to answer. Beat reporters had theories they wanted to test. For example, real estate reporter Lesley Mitchell wondered whether Utah's high rate of home ownership and high mortgage foreclosure rates play into bankruptcy filings. Health care reporter Linda Fantin wanted to test a common perception that aggressive collection tactics by a Salt Lake City-based health care provider led to many bankruptcy filings.

Because more than 20,000 personal bankruptcy cases are filed in Utah each year, we decided to study a representative sample. We consulted statisticians at the University of Utah and Brigham Young University to determine a reliable sample size. Selecting every 20th case, we obtained a random sample of 1,053 cases filed between July 1, 2003, and June 30, 2004. At a 95 percent confidence level, the sample had a margin of error of plus or minus 3 percentage points.

Electronic copies of bankruptcy case documents are kept as Adobe Portable Document Files by the U.S. Bankruptcy Court for Utah in Salt Lake City. Using PACER, Utah bankruptcy cases can be accessed through the court's Web site at *www.utb.uscourts.gov*, but they cost 8 cents per page and are about 30 pages or longer. Downloading the 1,053 cases we studied would have cost more than $2,000.

Instead, we obtained permission to use public computer terminals at the bankruptcy clerk's office to access the records for free. *Tribune* reporters Julie DeHerrera and I compiled the database, using laptops and the court's Internet access to open case files on the court's computer and then transcribe information into a Web form, which submitted the data into a *Tribune* database. The Web form and underlying Microsoft Access database were created by the *Tribune*'s new media editor, Tony Semerad.

We used sequential case numbers to open and record every 20th case, assuring the randomness of our sample. We recorded up to 170 items from each case, including income, property, number of dependents and types of debt. We spent about 15 minutes on each case.

The data entry took nearly four months to complete and was finished in August 2004.

Once the database was completed, Semerad chose Microsoft Excel to analyze the data because of its easy sorting and statistical functions. He primarily used means, medians and ratios.

DeHerrera and I also flagged potential outlier cases based on data that looked unusual when it was being entered. Semerad distilled general trends from the data, including credit card use, housing expenses, average income, family size, charitable giving, medical debt and whether incomes increased or decreased in the years prior to filing. DeHerrera, I and business reporters Steven Oberbeck, Linda Fantin and Bob Mims helped flesh out the findings. Data from the story was compared to state and naitonal averages.

Reporters added interviews with people who filed for bankruptcy, attorneys and financial experts. Using names from the database, they were able to speak with several people who had gone through the bankruptcy filing process. This was perhaps the most challenging piece of the project, since most people didn't want to have their financial troubles publicized.

We found that Utah's economy, with wages below the national average but living costs at or above average, combines with a culture that emphasizes large families and material success to push people to the brink of financial ruin more often than in other states. Many Utahns live paycheck to paycheck, juggling mortgage payments and other debt with living expenses, so a catastrophe such as a medical problem or job loss makes them more vulnerable to bankruptcy.

And, belying the myth that bankruptcy filers are irresponsible spenders, we found evidence that filers chose bankruptcy as a last resort, trying alternatives such as debt counseling, borrowing money from relatives or even selling blood plasma to pay off their debts.

Although the Federal Housing Administration recommends that no more than 30 percent of net monthly

income pay for housing, we found that bankruptcy filers, on average, spend 47 percent. We found that 85 percent of people who declare bankruptcy are gainfully employed and nearly 40 percent of filers experienced a significant decrease in their annual incomes during the three years prior to filing. Average take-home pay was $1,577 a month, but average monthly living expense was $2,126.

The cost of health care was also significant factor. Few filers reported health insurance as a monthly expense but 60 percent reported having unpaid medical bills, a percentage double the national average. Intermountain Health Care's collection tactics were not deemed a significant factor in the bankruptcy rate.

Religious culture also appeared to be a factor. In a state where about 70 percent of the population are members of The Church of Jesus Christ of Latter-day Saints, social norms often dictate that residents marry young and have large families. Many Utahns, in turn, get married, buy homes and have children at a time when they are at the lowest point of their earning power. Often, bankruptcy was filed in an effort to avoid a home foreclosure.

Almost two-thirds (63.6 percent) of people who filed for bankruptcy had one or more dependent children, making them twice as likely to be supporting children than the average household nationally. Three cases even had seven, eight or 10 kids.

In addition, members of the LDS Church typically donate 10 percent of their income to the church. Tithing, which was reported by 12 percent of filers, may play an ancillary role in some cases by reducing disposable incomes and allowing people to obtain more credit than they can afford when loans are based on their pre-tithe income.

Through our research, we were able to show that Utah is No. 1 for bankruptcy for a web of related reasons. The state's socioeconomic forces place more people on the edge, making it easier for people to slip into financial disaster when pushed by unexpected debt.

Contact Rosemary Winters by e-mail at rwinters@sltrib.com.

EXTRA! EXTRA!

YOUR GUIDE TO THE LATEST INVESTIGATIVE WORK

Looking for ideas of business stories you can do with computer-assisted reporting? Extra! Extra! on the IRE Web site (*www.ire.org/extraextra*) carries links to and summaries of stories that used data analysis. Here are some recent examples from the business beat:

The Charlotte Observer analyzed 2.2 million mortgage applications from the top 25 U.S. lenders and found that blacks were four times as likely as whites to get high interest rates for loans. The disparities existed even when blacks had much higher incomes.

The Dallas Morning News used Census and state economic numbers to show that the wealthy residents of Collin County, Texas, carry larger debts than residents in other high-income counties across the country.

Census data helped the *San Jose Mercury News* tell the story about California's decline in young farmers. The number of farmers younger than 35 fell 44 percent in California and 18 percent nationwide from 1997 to 2002.

The (Allentown, Pa.) *Morning Call* used restaurant inspection data to scrutinize food safety in the Lehigh Valley and the rest of Pennsylvania. School cafeterias faired well but the report concluded that the "patchwork of food safety laws and public health agencies often fails to provide even the minimal monitoring of restaurants and food retailers."

The Charleston Gazette compared color overlays of old mine maps to newer digital maps in their analysis of a coal company building silos outside its permit area and within 300 feet of a school.

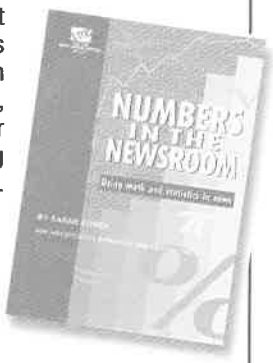
The (Staunton, Va.) *Daily News Leader* analyzed five years of state Department of Labor and Industry safety inspections and found that many Virginia businesses are not inspected each year. They also reported "the labor department performed too few safety inspections between 1999 and 2003 to reach even a quarter of the construction businesses in more than a third of the state."

The *San Diego Union-Tribune* analyzed county grant receipts and discovered "records for 54 grants totaling nearly $1 million are missing."

REQUIRED READING For Your Newsroom

Numbers In The Newsroom
Using Math and Statistics in News, by Sarah Cohen

Pulitzer Prize-winning journalist Sarah Cohen guides reporters through fractions, rates, percents and per capitas. Making inflation adjustments. Understanding averages. Doing the budget story. Questioning surveys and polls.



ORDER TODAY!
Call 573-882-3364
with your VISA, MasterCard or American Express
-OR-
Visit our Web site at **www.ire.org/store**
for online ordering or order form downloads

SPOTLIGHT: BUSINESS

Detecting bloat in job claims

By Mike McAndrew and Michelle Breidenbach, *The (Syracuse, N.Y.) Post-Standard*

Politicians sometimes sweep into town to announce they will create or save jobs by giving an economic development grant to a local business.

It's usually a surefire way for elected officials to get good publicity.

That's because journalists often report those rosy predictions without going back to see if they come true.

In New York, 47 percent of the companies awarded $195 million in state grants or loans by a state public authority failed to employ as many workers in 2004 as they promised they would, *The Post-Standard* discovered.

Companies had to promise to create or retain specific numbers of jobs to get the taxpayer-funded grants.

Gov. George Pataki's top economic development arm, the Empire State Development Corp. (ESD), can impose penalties on grant recipients that miss their job targets. But more often than not, ESD waived penalties. Sometimes the authority offered even more aid to businesses that had not delivered the required jobs.

To get this story, we used a Microsoft Excel spreadsheet to bolster some old-fashioned digging.

As a public authority, ESD operates under a governor-appointed board of directors in New York City, outside the normal scrutiny of the state legislature and comptroller.

We heard that several years ago a state senator had filed a Freedom of Information Law request, just as an ordinary citizen would, for a copy of the authority's annual report. She discovered that, unlike many other state authorities, ESD does not produce one. Instead, ESD provided the senator with a jobs report – one that showed in clear terms that many companies receiving economic development grants were not creating jobs.

We filed our own Freedom of Information request for the ESD's 2004 jobs report in electronic format.

We wrote formulas in Excel...to show businesses failing to live up to promises.

The Excel file the ESD sent by e-mail identified nearly 500 economic development grants or loans it made across the state. It included the date the grant was approved, each recipient's starting employment total, the number of workers each company promised to hire or retain, and the businesses' total jobs in the spring of 2003 and 2004.

An accompanying summary prepared by the ESD staff for the authority's directors noted that ESD had required some companies that missed job targets to repay some of the grants. But the summary didn't identify which companies were penalized or which got waivers. The database also included no addresses for the businesses.

So we filed a second FOI request with ESD, asking for fields that provided the location of each grant recipient and details about penalties imposed on the nearly 200 companies that had failed to comply.

We received an Excel spreadsheet with addresses quickly. But when it came to providing details about penalties and waivers, ESD dragged its feet. It took months of telephone calls and a letter hand-delivered to ESD's chairman before we received that information.

Once we had it, we asked our Excel spreadsheet some questions. Which companies missed their job target by the largest number? What businesses had fewer workers in 2004 than they did when they got their grant? Who eliminated the most jobs, but escaped without a penalty? What companies got ESD to reduce their jobs obligation for 2005?

We mined the list for potential anecdotes. For instance, why was an aerospace museum in Niagara Falls given $750,000 to create only two jobs?

We wrote formulas in Excel subtracting current employment from employment at the time of the grant to show layoffs, and subtracting current employment from target employment to show businesses failing to live up to promises. Then we sorted the data.

Computer-assisted reporting enabled us to identify faster the grant recipients who were likely to become the best examples to shine a spotlight on in our story.

The same calculations pointed out success stories, too. Then we read published reports, corporate Web sites and grant documents, and conducted interviews with officials at

companies, to check the accuracy of some job totals that ESD provided. We discovered the state was overstating some of its successes in its report.

For instance, ESD claimed General Motors had hired 2,000 new workers since it received a $5 million grant. In reality, we learned that GM's Buffalo area engine plant was employing about 1,000 fewer workers than it did before it got the state aid. When we asked ESD about that, the authority admitted it made an error.

We found the state also was taking credit in its jobs report for helping to retain jobs that had been around for more than a century at one firm. The owner of another business admitted he would have expanded and hired more workers with or without the state's money.

ESD declined to allow us to interview anyone but its spokesman. But, in minutes of meetings of the ESD board of directors – a document we had obtained in electronic format via a FOI request – we discovered one nugget. A director had questioned in 2002 whether ESD would be criticized by the press because 27 percent of the economic development grant recipients had missed their job targets that year. Two years later, ESD could only dream of the days when it had a 27 percent failure rate.

By using CAR tools to take a deeper look at the numbers, *The Post-Standard* discovered the results of ESD's economic development program were sometimes gloomy – certainly in sharp contrast to the happy talk of politicians announcing they are going to save jobs by giving away a grant.

Contact Mike McAndrew by e-mail at mmcandrew@syracuse.com.

Contact Michelle Breidenbach by e-mail at mbreidenbach@syracuse.com.

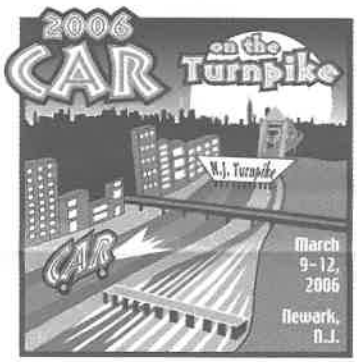
IRE

Annual Conferences

Be sure to put these dates on your calendar today!

Using computer-assisted reporting techniques, doing investigations in the Latino community and the Americas, digging into border and immigration issues, cultivating sources, producing better broadcast investigations, and holding government and business accountable. All this and much more will be offered at the next conferences. Be sure to put these three on your calendar and make your hotel reservations today!

2006 CAR Conference
March 9-12 in Newark, N.J.
Hilton Newark Gateway Hotel



2006 IRE Conference
June 15-18 in Dallas/Fort Worth
Renaissance Worthington Hotel

2007 IRE Conference
June 7-10 in Phoenix
Arizona Biltmore
Resort & Spa

Visit www.ire.org/training/conferences.php for more information!

SPOTLIGHT: BUSINESS

Dissecting biggest public companies

By Helen Huntley, *St. Petersburg Times*

Does your area have a billionaires' club? That's what the *St. Petersburg Times* calls its list of the 10 largest public companies in the Tampa Bay area. You can tally up the wealth of public companies in your area fairly easily by using the Internet to access U.S. Securities and Exchange Commission filings and a spreadsheet to crunch the numbers.

The *Times* has been publishing an annual report about the top public companies in our area for 10 years. Initially, we hired an outside vendor to collect and analyze the statistics for us, but we quickly figured out we could do it ourselves. Over the years we have made numerous changes in how the report is prepared. We now do a streamlined report, focused on the 10 largest companies (down from 50), called the *Times* 10.

Our starting point is the running list we keep of all public companies in our area traded on the New York or American stock exchanges or the NASDAQ Stock Market (bulletin board stocks are excluded). We publish closing prices on these stocks daily. I narrow the list to the ones that have large enough revenues to have a chance of making the top 10 cutoff before I get down to serious data collection. (After you've done this one year, you'll have a good idea of the contenders the following year.)

We make our comparisons based on the most recently completed fiscal year. Since most companies have a Dec. 31 year-end, I start compiling the report in April and we publish it in May. An alternative approach is to use the four most recently completed quarters. However, that becomes more complicated, so I do it only for one company that is always late getting its results out.

The primary source for my information is the Form 10-K annual report that companies file with the SEC at the end of the fiscal year. You can find these reports on the SEC Web site (*www.sec.gov*) or via links on the company's Web site. I use the commercial 10k Wizard (*www.10kwizard.com*), which provides easy searches of the filings.

We now do a streamlined report, focused on the 10 largest companies

Although financials can be downloaded, I prefer to write down the information I want and print the pages to check my numbers against later. I collect these numbers for the most recent year and the previous year: Revenues, net income, net income per share and shareholder equity. I average the two years of equity to get the figure used in calculating return on equity.

I use a Bloomberg terminal to get the closing stock price on my chosen date (typically the last Friday in April) and the date two years earlier, to calculate the total return for that period and to get the dividend and market capitalization. If you don't have access to a Bloomberg terminal, you can use Yahoo Finance (*http://finance.yahoo.com*) to get closing stock prices. If you multiply shares outstanding (from the 10-K) by the share price, you have market cap. I type all this information into a Microsoft Excel spreadsheet.

The top 10 rankings are based solely on total revenues. Once I have the top 10, I use the spreadsheet to calculate rankings for those companies that made the list. First I type in the formulas to get the information I want, such as the year-over-year change in revenue (2004 revenue - 2003 revenue / 2003 revenue) and the return on equity (2004 net income / average shareholder equity for 2003 and 2004). Then I use the Sort function to rank the companies by fastest growing, most profitable, most valuable (based on market cap) and best and worst investments.

We publish a short profile about each company in the *Times* 10, including basic information such as address, telephone number, type of business, top executives and number of employees. Normally all this can be found in the 10-K. If you don't follow the company closely, call to verify that the executives haven't changed since the 10-K was published.

One caution: when companies lose money, it will do funky things to the calculations of year-to-year changes. When I calculate the fastest-growing profits list, I deal with this issue by only considering companies that were profitable in both years. For the individual profiles, I calculate a change in total profits and use positive or negative values based on what makes sense even if it is not mathematically correct. I show a positive number for a company that went from a loss to a profit and a negative number for a company that went from a profit to a loss. If a company lost money both years, I show a positive number if the loss got smaller and a negative number if it got larger. I note this in a footnote and have never had anybody complain about this approach.

Contact Helen Huntley by e-mail at huntley@sptimes.com.

Resources for covering business

IRE and NICAR offer resources for journalists interested in using computer-assisted reporting to cover business topics.

Journalists can download the "CAR Can Drive the Business Beat" presentation developed by IRE from the Donald W. Reynolds National Center for Business Journalism at the American Press Institute Web site. (*www.businessjournalism.org/content/6864.cfm*). The presentation lists useful Web sites for the business beat, and shows how journalists can integrate simple data analysis into their coverage.

Labor: Find out which local companies have bucked minimum wage, family leave and other federal labor laws with the Wage and Hour Enforcement database. It includes employee and employer information, number of violations and the amount of penalties. In 1998, *USA Today* did a computer-assisted analysis that showed working children are at risk and that the federal government does little to find or punish those employers who break child labor laws. (Story No. 15123)

Business loans: Which local businesses have had loans guaranteed by the federal government? The SBA7a Business Loans database includes information about loans guaranteed by the U.S. Small Business Administration. It shows which businesses have repaid their loans and under which program the businesses received the loans. A computer-assisted investigation by *The Kansas City Star* revealed that the SBA gave most of its government-guaranteed loans to firms located in middle- and high-income areas. The paper discovered that minority-owned businesses received fewer than 15 percent of the SBA loans, despite efforts by the administration to target minority owners. (Story No. 16986)

Federal contracts: The Federal Procurement Data System was created by the Government Services Agency. Contracts included in the 2004 data show services being performed or items being produced in all U.S. states, as well as U.S. territories and some foreign countries. The list of services or products being contracted includes telecommunications, maintenance, office furniture, food products, nursing home care contracts, consulting services, military equipment, computer equipment and software, janitorial services, removal and cleanup of hazardous materials, hotel/motel lodging, construction of troop housing, textile fabrics and fuel products. The *Corpus Christi* (Texas) *Caller-Times* did a series of computer-assisted stories in 2002 on the House Armed Services Committee third-ranking Democrat, Rep. Solomon Ortiz, and his influence on and personal gain from defense federal contracts in his district. (Story No. 19602)

For more details and a comprehensive list of what the Database Library has to offer for reporting on business go to *www.ire.org/datalibrary/databases.* To order databases, call 573-884-7711 or 573-884-7332.

To order story reprints contact the IRE Resource Center at 573-882-3364 or rescntr@ire.org.

NEW

Announcing the

Philip Meyer Journalism Award

Sponsored by

N · I · C · A · R

National Institute for Computer-Assisted Reporting

a joint program of IRE and the Missouri School of Journalism and

John S. and James L. Knight Foundation

the Knight Chair in Journalism at the Walter Cronkite School of Journalism and Mass Communication, Arizona State University

The contest recognizes the best journalism done using social science research methods.

Go to: **www.ire.org/meyeraward** for more contest information.

All entries are placed in IRE's Resource Center story library so IRE members can learn from others' triumphs and troubles.

POSTMARK DEADLINE: October 31, 2005

Deadly

continued from page 1

floorboards that had rusted through. Others had broken windows and lacked seatbelts.

Most of these vehicles belonged to third-party labor contractors who were hired by growers to put together work crews. The contractors were licensed and monitored by federal and state agencies, but neither seemed to be doing a good job in the area of transportation based on a cursory look at the vehicles.

Transportation is a central determinant in the lives of farmworkers because it is difficult for many to get driver's licenses because of their illegal status. Without driver's licenses or cars, the farmworkers in Immokalee are forced to live in substandard housing within blocks of La Mexicana so they can walk to the parking lot where the contractors pick them up for jobs. With this captive audience, the grocers and landlords in the area charge more for food and housing than in other surrounding communities.

Transportation, I would find during my research, also is the No. 1 cause of death for Florida's farmworkers.

Curious about how well the state regulated contractors, I asked the state Department of Business and Professional Regulation for its database about violations of state labor law made by Florida labor contractors since 1997. They didn't have one, even though Florida has one of the nation's largest farmworker populations. Those paper records were gathered in rubber bands in cardboard boxes at the state agency's headquarters in Tallahassee.

So I created a database.

I drove to Tallahassee, laptop computer in hand, with the expectation it would take me two days to enter the records into a Microsoft Excel spreadsheet where I could look for patterns, find the worst offenders and glean other information. By the second day, I wasn't even a quarter of the way through the four boxes of records, so I furiously photocopied the remaining records and took them back with me to Orlando.

After several weeks of data entry during spare moments, I finally had 878 entries, covering all types of violations, that spanned January 1997 to March 2004. The rows listed the name of the contractor, the violations, punishment, date, the farm where it occurred, county, the contractor's address and complaint number.

Once the spreadsheet was finished, I ran some basic calculations in Excel to find out what the leading violations were and how the violators had been punished.

While I was putting the database together, tragedy struck. A van jammed with 19 farmworkers rolled over on Interstate 95 near Florida's Atlantic coast, killing nine men. Almost everyone was ejected because nobody wore seatbelts, and bodies were scattered on the highway. One of the farmworkers killed was Jose Luis Garcia Pichardo, a Mexican farmworker whose widow I would later interview for my story.

About the same time, the Florida Legislature, at the urging of Gov. Jeb Bush, passed legislation increasing the fines for farm labor law violations and making it easier to report violations. The law was created in response to a handful of recent cases involving farmworkers made to work in slave-like conditions. The law, however, ignored transportation-related issues.

Once the spreadsheet was finished, I ran some basic calculations in Excel to find out what the leading violations were and how the violators had been punished. The database revealed that:

- More than one-third of the farm labor violations were related to transportation, such as using vehicles that had not been inspected by the state, lacking driver's insurance or employing an unlicensed driver.

- State regulators had not taken transportation violations as seriously as others, such as failing to pay delinquent taxes, something which often led to a contractor's license being revoked. Not one contractor had a license revoked for a driving-related violation, even though those violations could have deadly consequences. Repeat violators of farm labor transportation laws still got their contractor licenses renewed.

- Problems commonly found in the farm contractor vehicles were rusted-out holes in floorboards, cracked windows, doors that didn't open and seats that had been torn out. Many vehicles also lacked seatbelts, although Florida law exempts the mandatory use of seatbelts for buses and vehicles that weigh more than 5,000 pounds, which includes most farm labor vehicles.

To get a better sense of how widespread a problem the vehicle-related deaths were for Florida farmworkers, I checked injury records from the U.S. Bureau of Labor Statistics and the Florida Department of Business and Professional Regulation and found out that transportation incidents were the leading cause of work-related fatalities for Florida's farmworkers, causing as many as 83 deaths and more than 400 injuries since 1992.

Given that only California had more transportation-related farmworker

deaths in the past dozen years, I called the California Highway Patrol's farm labor vehicle program and found that in 1999 California ended a seatbelt exemption for agricultural vehicles after 13 farmworkers were killed in an accident. The following year, there were no highway deaths resulting from farm labor vehicle accidents.

While interviewing lawyers who have represented farmworkers involved in vehicle accidents, I also discovered that many growers and contractors have farmworkers drive "goats," open-cab converted school buses used to haul citrus from groves. The vehicles are exempt from many requirements for road vehicles, such as the need for windshields, even though they can travel at the same speed as cars and trucks on the state's roads and highways.

State troopers and farmworker advocates have complained that the goats are too dangerous to be roadworthy, so I wrote a sidebar focusing on Juan Antonio Mondragon Lopez, a young Mexican farmworker who was killed in 2000 when he lost control of the goat he was driving.

When I interviewed the head of the Florida agency that regulates farm labor, he conceded that in years past the state had done a poor job of cracking down on contractors who regularly broke transportation laws. He blamed a lack of resources, inconsistent enforcement and the fact that oversight recently had been transferred from one state agency to another. He vowed the division was on a new path, and indeed, since I had begun researching the story, state regulators for the first time had taken away the licenses and issued stiff fines for three contractors who committed transportation violations.

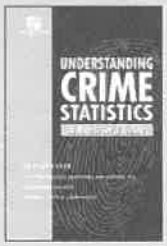
Contact Mike Schneider at mschneider@ap.org.

REQUIRED READING FOR YOUR NEWSROOM

HOME MORTGAGE LENDING:
How to Detect Disparities

Pulitzer Prize-winning journalist Jo Craven McGinty guides reporters through understanding and using Home Mortgage Disclosure Act data. Included are specific story ideas and lists of tipsheets and stories available through IRE.

UNDERSTANDING CRIME STATISTICS:
A Reporter's Guide



Covers using Uniform Crime Reports, National Crime Victimization Survey, National Incident-Based Reporting System, other major statistical sources, writing the crime statistics story and database analysis of crime statistics. Includes law enforcement contact information and stories and tipsheets available from IRE.

NUMBERS IN THE NEWSROOM:
Using Math and Statistics in News

Pulitzer Prize-winning reporter Sarah Cohen guides journalists through working with numbers, including fractions, rates, percents, per capitas, measuring change, making inflation adjustments, understanding averages, working with graphics, doing budget stories, questioning surveys and polls, and much more.

COVERING AVIATION SAFETY:
An Investigator's Guide

Learn to develop a crash plan for your newsroom, report from the scene of a crash, start an aviation beat, interpret aviation records, negotiate Web data and investigate planes and airlines on deadline. Includes related stories and tipsheets, as well as FAA contact information and useful Web sites.

COVERING POLLUTION:
An Investigative Reporter's Guide by IRE in cooperation with SEJ

This primer gives an overview of useful resources reporters can use for local investigations into environmental pollution. Filled with examples and references to stories, tipsheets and other resources available from IRE and SEJ.

MAPPING FOR STORIES:
A Reporter's Guide To Campaign Finance by Jennifer LaFleur and Andy Lehren

Learn step by step how to map data for daily news stories and larger projects. This practical introduction to mapping can be used alone or as a supplement to other books. It includes story examples and breaks down the elements needed to undertake analysis using mapping software. It is structured to help journalists complete better stories.

UNSTACKING THE DECK:
A Computer-Assisted Reporting Guide

Invaluable for pursuing stories about the impact of money on elections, political parties and candidates at the federal, state and local levels. Learn about loopholes in soft money restrictions, the use of nonprofits to funnel money to candidates, how to track where candidates spend the money they raise and how to obtain and use pertinent documents and electronic data

ORDER NOW!

IRE MEMBERS: $15 • NONMEMBERS: $25 • PLUS POSTAGE

BY MAIL
Send your check to IRE
138 Neff Annex
Missouri School of Journalism
Columbia, Mo. 65211

BY PHONE
Call
573-882-3364
with your VISA, MasterCard or American Express

BY WEB
Visit
www.ire.org/store
to order online or download an order form.

Borrowers

continued from page 1

- Of the 100 counties nationally with the highest rate of personal bankruptcy filings, 45 are in Georgia.

We obtained county-level bankruptcy data compiled by the Administrative Office of the U.S. Courts. The bankruptcy data is available through PACER, an online service that provides access to case and docket information from federal courts. (See (*http://pacer.uspci.uscourts.gov*). We then combined the bankruptcy figures with county-level population data from the Census Bureau and produced a personal bankruptcy rate for every U.S. county. We used Microsoft Access to perform the analysis.

Our database editor, David Milliron, had worked with this bankruptcy data previously and joined the team for this segment of the reporting. He produced a national county-by-county map, using MapInfo Professional, that ran on the first day of the series. We also produced a searchable online database that Web site visitors could use to compare their county rate to those in other counties.

We knew the state of Georgia had a relatively high bankruptcy rate, but by presenting the data at the county level we could see compelling information. Legislators who were in session at the state Capitol when the series was published passed around copies of the map.

We spent several days in bankruptcy court, interviewing people who had filed, along with bankruptcy attorneys, judges and trustees. We found that Georgia's lightning-fast foreclosure process had prompted many people to seek bankruptcy protection – they were simply trying to save their homes. We were able to establish that no state allows a faster foreclosure process than Georgia, where a house could be sold on the courthouse steps in as few as 37 days after a bank initiated the process.

- The number of foreclosures initiated against properties in metro Atlanta's 13 core counties in 2004 was more than double the number in 2000.

We went to a private contractor to acquire five years of data for every foreclosure initiated in metro Atlanta. The vendor, Equisystems LCC, provided the data to us in a Microsoft Excel file at no cost. In Georgia, foreclosures are nonjudicial, meaning the courts do not get involved. The only public record of a pending foreclosure in Georgia is a classified ad published in the county legal organ announcing the intent to foreclose. As a result, there was no database of foreclosures compiled by a government source. Equisystems compiles the data for clients, most of whom are real estate investors.

The data included the name, address, lender and other descriptive information for every property that had faced a foreclosure. We geocoded all the addresses in MapInfo to create a point map and used demographic data to calculate foreclosure rates and to determine the characteristics by Census block group. We used this data to look at the overall trends, for the metro area and individual counties, but we also mapped this data to look for spots with a high number of foreclosures. This helped us identify several streets in the metro area where multiple families had lost their homes.

- Georgia's "small loan companies" – old-time storefront lenders who offer loans to people with poor credit – are alive and well. Their interest rates are high and they increase the cost of the average loan by adding on with more than three questionable insurance policies.

Our look at Georgia's 1,000 small loan companies was the most laborious analysis of the series. The companies are required to file detailed annual reports to the state Insurance Commissioner. But the most recent reports were on paper. The commissioner's office eventually builds its own database, but it was a couple of years behind.

We hired a data entry contractor to key in the crucial data from the most recent reports.

And then we crunched the numbers for several years' worth of data, using Access. We did a variety of queries to determine the characteristics of the average loan – how much did the average customer borrow? How many credit insurance policies did they buy? What was the actual cost of borrowing? The findings were eye-opening: most of the 1 million loans taken out by poor Georgians every year came at an effective annual cost of more than 100 percent and the commissioner's office had permitted excessive rates for the credit insurance policies.

We estimated that in 2003 small loan customers had overpaid by $65 million for credit insurance. We arrived at that figure by calculating loss ratios on the insurance products. A loss ratio is a percentage, calculated by dividing the dollars spent on claims by the dollars collected in premiums.

The National Association of Insurance Commissioners, a professional association, has set the benchmark of a 60 percent loss ratio. That means rates are considered fair if a company pays out 60 percent of the premiums it collects in claims. In Georgia, two of the policies sold with small loans had loss ratios of 4 percent and 5 percent, indicating they were overpriced.

State insurance commissioners keep detailed credit insurance data. Summary data for states is available through the National Association of Insurance Commissioners at *www.naic.org*.

- Georgia was the first state to sanction car title lending – a loan that comes with an annual interest rate of 300 percent and results in many owners losing their cars.

Detailed research at the Legislature provided most of the news findings for this story, which featured a profile of Atlanta's Rod Aycox, one of the largest title pawn lenders in the nation. To understand how Aycox exported title pawn lending from Georgia to the rest of the states, we undertook a detailed analysis of his political giving. We turned to the Institute on Money in State Politics' (*www.followthemoney.org.*) online database of state-level campaign contributions. (This was helpful because Aycox has contributed money in many states and it otherwise would have been difficult to find all of his contributions)

We also looked up Federal Elections Commission information about contributions from Aycox.

- A profile of Life of the South, a credit insurance company, allowed us to show how business and politics mix in Georgia. The company has hosted dozens of elected officials who control its bottom line at its South Georgia hunting plantation – and it has grown and prospered as it turned almost everyone at the state Capitol into a personal friend.

To establish the business ties between this credit insurance company and state lawmakers, we obtained a database of all licensed insurance agents from the state Insurance Commissioner's office. The database included information about which companies each agent represents. We negotiated with the agency to provide the information, which contained thousands of records, at no cost. We were able to get it in Access format.

Using that data, and our examination of banking records and financial disclosure records, we were able to establish that two dozen members of the Georgia General Assembly had business ties to Life of the South. We also studied Life of the South's campaign contributions. While the story itself was a breezy tale of how a genteel Southern businessman prospered, the news of the story came largely through the analysis. And to bring the implications home, we featured a loan taken out by a rural Georgia woman who borrowed $450 to pay bills, but ended up also being charged $367 for Life of the South insurance premiums.

It took a massive group effort to crunch so many numbers, to find so many average Georgians for the stories, and to convince business executives to talk to us about themselves and their industries. But when it was all over, the reporting team felt great about a project where we found news through data and detailed records research – but conveyed the news to our readers through the rich personal stories of the people involved.

The "Borrower Beware" series is available online at *www.ajc.com/borrowerbeware.*

Contact Carrie Teegardin by e-mail at cteegardin@ajc.com.

Home mortgage resources

The Charlotte Observer, working with new Home Mortgage Disclosure Act data, reported in August that black borrowers from the largest U.S. lenders were four times more likely than whites to receive high-interest loans.

The *Observer* analyzed more than 2 million records of home purchase loans from 25 of the largest lenders in the United States. The *Observer* was able to do the story because federal banking regulators, starting in 2004, began requiring lenders to report whether a loan carries a high interest rate.

IRE and NICAR offer resources for journalists who are interested in using the HMDA data to detect racial disparities or report on housing trends.

A good start for beginning an examination of mortgages is the IRE Beat Book, "Home Mortgage Lending: How to Detect Disparities," by Jo Craven McGinty. The beat book explains how to access and analyze home mortgage data, what questions to ask, and what other resources are available. The Beat Book is available for purchase from the IRE Store at *www.ire.org/store* or the Resource Center at 573-882-3364.

The Home Mortgage Disclosure Act database, available from the IRE and NICAR Database Library, features more than 41 million records of loan applications for 2003, with archival data reaching back to 1992. The data includes demographic information about the loan applicants, as well as the reason for the loan and whether it was accepted or rejected. Journalists can purchase data for the country or a state by calling 573-884-7711. The Database Library hopes to make updated data available to journalists in October.

In addition to these two key resources, the IRE Resource Center offers stories and tipsheets:

"Burrowing into HMDA," by John Byczkowski of *The Cincinnati Enquirer* provides guidance for journalists looking to dig into the home mortgage data. This tipsheet offers cautions about interpreting the data. (Tipsheet No. 1774)

continued on page 22

Tech tip...

Making XML useful

By Jeff Porter, *IRE and NICAR*

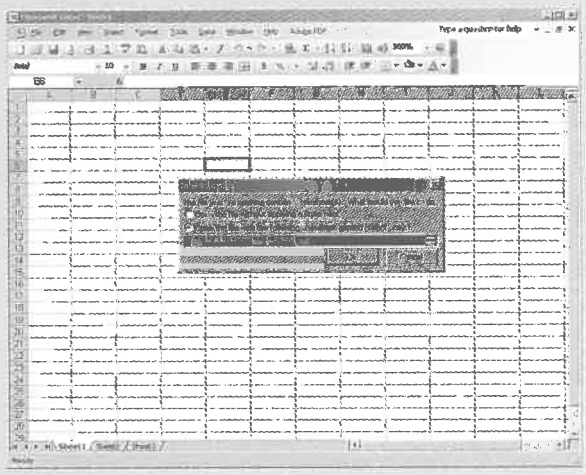
If you have not had to import XML data into a database manager, you're lucky so far – more and more government agencies use XML to store and transmit information. If the data is relatively small, that's good. But XML files can easily become bloated monsters, frustrating a user of off-the-shelf database software.

What is XML? Several definitions are floating around the Web.

- Extensible Markup Language is a simple, very flexible text format ... Originally designed to meet the challenges of large-scale electronic publishing, XML is also playing an increasingly important role in the exchange of a wide variety of data on the Web and elsewhere. (*www.advancinginsights.com/mybiz/*)

- A widely used system for defining data formats. XML provides a very rich system to define complex documents and data structures such as invoices, molecular data, news feeds, glossaries, inventory descriptions, real estate properties, etc. As long as a programmer has the XML definition for a collection of data (often called a "schema") then they can create a program to reliably process any data formatted according to those rules. (*www.unitedyellowpages.com/internet/terminology.html*)

- ...Capable of describing many different kinds of data. Its primary purpose is to facilitate the sharing of structured text and information across the Internet. Languages based on XML (for example, RDF, RSS, MathML, XSIL and SVG) are themselves described in a formal way, allowing programs to modify and validate documents in these languages ... (*http://en.wikipedia.org/wiki/XML*)



If those definitions leave your head spinning (as they did mine), we will focus on the practicality of dissecting XML files – making them useful for data analysis in your software of choice. An XML data file does not have the usual feel of a flat file, such as ASCII and the like. There are no headers, no fixed fields, no columns. Instead, each element is defined line by line, as in this example from the Microsoft support Web site (*http://support.microsoft.com*).

```
<ROOT>
  <Customers>
    <CustomerId>1111</CustomerId>
    <CompanyName>Sean Chai</CompanyName>
    <City>NY</City>
  </Customers>
  <Customers>
    <CustomerId>1112</CustomerId>
    <CompanyName>Tom Johnston</CompanyName>
    <City>LA</City>
  </Customers>
  <Customers>
    <CustomerId>1113</CustomerId>
    <CompanyName>Institute of Art</CompanyName>
  </Customers>
</ROOT>
```

A bit of translation is in order here.

The tag <Customers> (and its counterpart, </Customers>) identifies each record that pertains to a given customer. It's repeated for each new record. You can trace the same syntax for CustomerId, CompanyName and City fields. The ROOT tags surround it all. The design can be simple, but in some ways, it's reminiscent of hierarchical data – bringing back the term "variable length" – in that each field does not necessarily have to show up in each record.

Microsoft Excel or Access can import and use many of the XML online files. For practice, download U.S. House roll call voting data that you can get from *http://clerk.house.gov/evs/2005/index.asp*. For this example, we chose to save this page: *http://clerk.house.gov/evs/2005/roll464.xml*.

Excel 2003's File > Open ... menu item has an option to choose XML as file type. After you choose this, it will say "The file you are opening contains stylesheet(s). What would

you like to do:" Your options are to use an existing stylesheet, ending with the suffix "xsl" – in essence, a technical record layout – or opening the file without applying a stylesheet – in other words, relying on the general structure of tags in the file.

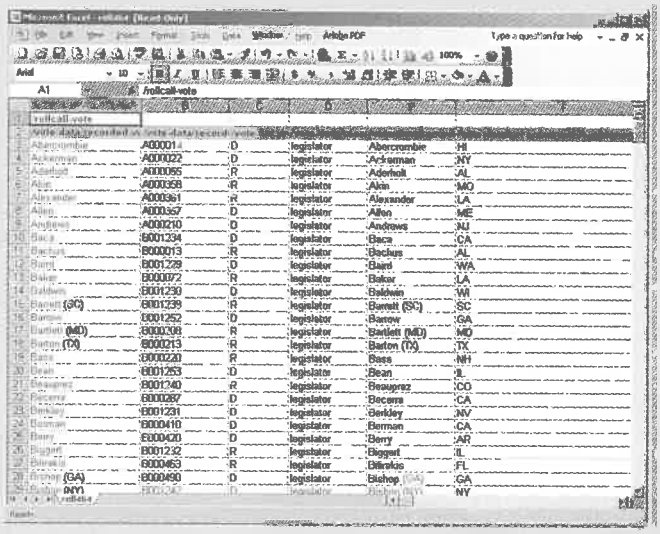
Where does the stylesheet come from? In this instance, it's on the House Web site. Online, the XML file looks nice and pretty. But viewing it with a text editor or using your browser's View Source function, you can easily find a reference for the stylesheet:

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE rollcall-vote PUBLIC "-//US Congress//DTDs/vote v1.0 20031119 //EN" "http://clerk.house.gov/evs/vote.dtd">
<?xml-stylesheet type="text/xsl" href="http://clerk.house.gov/evs/vote.xsl"?>
```

If you go to http://clerk.house.gov/evs/vote.xsl, you can view the entire document – a mixture of record layout, code sheet, and tips about the internal data the House uses to create the XML pages.

That brings us back to the question about whether to use the stylesheet or the file's tags. Which is better? It varies from case to case, so it is a good idea to try both during your first attempt at importing the data.

In this particular instance, using the stylesheet generates an Excel file that looks suspiciously looks like the Web page. Go with the other option, pick as an XML list or a read-only workbook and you get what I believe is a better-looking spreadsheet. You actually gain a bit more information, too: party affiliation. Using the stylesheet, parties are identified only by font: Roman type is Republican, italic, Democrat. Independents are underlined. Without stylesheet, the parties are identified with a column containing D for Democrat, R for Republican and I for Independent. And as an added bonus, you get the unique House ID number for each legislator – helpful in joining tables later.



The loser in this test was Access 2003. You can import the XML using the Import Wizard. Just open or create a database, go to File > Get External Data > Import ... and choose XML as file type.

But in this case, the results are disappointing. Like many dataset examples, the XML is structured into several separate "tables," so our results include five tables. The most vital one lists names of legislators and individual votes. Party affiliation and the House IDs are missing.

Other applications are available for working with XML:

- XMLSpy from Altova. This program was suggested by several people in and out of journalism. You can download a 30-day evaluation copy from the company Web site at *www.altova.com*.

- Navicat from PremiumSoft. Navicat is a popular graphical interface for MySQL relational database manager that provides a high level of control during importing. A 30-day evaluation copy is available at *www.navicat.com*.

- SAS. This powerhouse stats and data management program (*www.sas.com*) includes the SAS XML Mapper tool that, in a point-and-click interface, lets the user create a text file of instructions for field names, format and location. You can also edit the file for more control.

- Perl. Users of this programming language can download modules to translate XML files.

XML was never designed (in my humble opinion) as a format to store mammoth amounts of data, but increasingly it is being used for large and complex sets of information. Chances are that journalists doing computer-assisted reporting will probably run into a database in this format. Whether it's a huge file, or a small one, there are plenty of tools for CAR practitioners to try.

Contact Jeff Porter by e-mail at jeff@ire.org.

Looking

continued from page 4

which helped make a nice graphic.

Some lessons learned on my first CAR story:

1. Every reporter could easily do a story about the condition of local dams.
2. Look online for databases kept locally.
3. Ask local officials for databases they maintain for internal use.

Continuing the theme of updating old stories, I also decided to look online to identify local officials who fail to pay property taxes on time. Twenty years ago, I did that by sorting through paper lists at the county assessor's office and hoped I could recognize names of prominent people. Occasionally, I got lucky.

I found this time that Salt Lake County lists all tax delinquents online. Again, I copied and pasted information from the online database into an Excel spreadsheet, and then imported it to Microsoft Access database manager.

Next, I made another table of information about all known elected officials living in Salt Lake County by searching government Web sites. I included, when available, their addresses, birth dates, spouses and businesses. I then joined the databases by name to find possible matches. After making phone calls to weed out bad matches, I found three officials, including U.S. Sen. Bob Bennett, R-Utah, who were delinquents.

But I found a much better story once I started playing with the data. I developed lists of businesses and individuals that were the biggest delinquents, often owing hundreds of thousands of dollars. I mostly used Excel, creating pivot tables for amounts owed, and sorted to see the highest amounts.

I realized that even healthy businesses, especially real estate developers, were often choosing not to pay property taxes because penalties and interest are so low in Utah. Several delinquents confirmed that on the record, including the manager of two hotels owned by Bennett that had racked up nearly $400,000 in back taxes.

I also wrote a story about the condition of bridges in Utah, another story that I had done some 20 years ago with paper records.

The Federal Highway Administration provides bridge summary data by state online in Excel format at *www.fhwa.dot.gov/bridge/britab.htm*. After downloading and sorting the data, I asked the Utah Department of Transportation for its more detailed database of the 506 "deficient" bridges in the state. The data included a deficiency rating from 0 to 100 for each bridge, which allowed quick identification of the worst bridges by Excel sorting.

The big lesson is, don't be a dinosaur. Dare to do some CAR work. Start with the basics and improve your skills. Maybe update some of your old stories. If an ancient like me can do it, anyone can – just ask my children who will verify that quickly and firmly.

Contact Lee Davidson by e-mail at lee@desnews.com.

CALL FOR ENTRIES

2005 IRE AWARDS

The annual contest of Investigative Reporters and Editors, Inc.

The contest recognizes the best investigative reporting in print, broadcast and online media, and helps identify techniques and resources used by entrants.

Postmark DEADLINE January 9, 2006

CATEGORIES
Newspaper
Television
Online
Other Media:
Magazine/Speciality Publication
Book
Radio
Special Categories:
Renner (crime reporting)
FOI
Student

For entry forms and detailed information, visit our Web site at www.ire.org/contest

Uneven

continued from page 5

to create the larger polygons. Bailey named these larger polygons with the street name and two cross streets. So the block on Cleveland, intersecting Elmwood and Delaware, was called ClevelandElmwoodDelaware.

Bailey exported the resulting table and then imported it in Access. He ran some basic queries in Access to determine the address range of each block.

Then Bailey queried the database for the 35 or so streets in the heavily ticketed Upper West Side area. This resulted in about 60,000 tickets. Using the address ranges from the block table, Bailey assigned a specific block to each ticket.

A query grouping by the block was used to create a table with the results, and then was imported back into ArcView for the final block map.

We verified addresses of some corner properties in question by walking along blocks in the neighborhood and recording house addresses in each block. The violation locations we could not assign to a block, and thus excluded from this map, represented less than 1 percent of the tickets.

We shaded the map by number of tickets issued in each block. But, of course, not all blocks are created equal. A typical block on Lafayette Avenue, for instance, covers the same distance as three on Ashland Avenue. Our first thought was to group blocks for the streets with smaller blocks, such as those on Ashland. We quickly abandoned this notion because we felt that it would provide an inaccurate depiction of the data by glossing over the variations.

Digging deeper

Most of our maps were not published, but were pivotal in our analysis.

We created maps for each type of parking violation. This showed us the obvious, that meter overtime violations were mostly downtown, but it also alerted us to ask why there were so many more alternate-side-of-street parking violations in a North Park neighborhood and so many more no-standing violations in a South Buffalo neighborhood.

We also mapped violations for certain times of the day. This allowed us to see where the parking officers started their days, and which residents had less time to get away with an alternate-side parking violation than others. We mapped morning violations in 20-minute increments.

The map revealed Ashland Avenue to be the block with the most violations right after 9 a.m., so we knew to take a closer look at that street's data.

Our findings did not surprise those who start their day at the coffee shop in the city's Upper West Side. While downtown and the relatively affluent Upper West Side were prime targets for the city's parking officers, they spent a lot less time in Buffalo's crime-ridden East Side streets. The all-white ticketing force is afraid to go there, we were told.

The data prompted us to ask about deployment. With eight parking officers working the day shift, one covers South Buffalo and most of the East Side. Another works North Buffalo. The rest work downtown or the Upper West Side.

So in Buffalo, your fortune with parking tickets hinges on where you park, we reported in a special report called "Ticketing the city."

Contact Patrick Lakamp by e-mail at plakamp@buffnews.com.

CAR Workshops

These unique seminars, taught by IRE and NICAR's experts, train journalists to acquire electronic information, use spreadsheets and databases to analyze the information and to translate that information into high-impact stories. In addition, IRE and NICAR provide follow-up help when participants return to their news organizations.

2006 Boot Camps:
Columbia, Mo.
- Jan. 8-13, 2006
- March 26-31, 2006
- May 21-26, 2006
- Aug. 6-11, 2006

Mapping Data for News Stories:
Columbia, Mo.
Learn how to uncover news stories by mapping data with geographic information system software.
- Jan. 6-8, 2006

Advanced CAR Statistics
Tempe, Ariz.
Aimed at strengthening the skills of reporters who want to move beyond basic CAR and use statistical analysis in their work.
- Feb. 17-19, 2006

www.ire.org/training/bootcamps.php

What previous Boot Campers have said about the experience:

"They were able to talk in a real way about issues that reporters face when practicing computer-assisted reporting skills ..."
– Analisa Nazareno, *San Antonio Express-News*

"I realized I hadn't been fully tapping my beats, and you don't have to be a computer nerd to do computer-assisted reporting."
– Teresa Taylor Williams, *Muskegon Chronicle*

Fellowships
A limited number of minority and small-news organization fellowships are available for IRE and NICAR workshops. Visit *www.ire.org/training/fellowships.html* for more information and an application.

Homes

continued from page 7

in the past 40 years.

By increasing the transparency of the fire-zone shapefiles, we were able to see clearly the building permit locations, spread across the land like so many freckles. Adding fire-protection district boundaries to the map demonstrated that significant construction had occurred in places not only at risk of burning, but also outside fire protection districts, and rapid assistance from fire crews.

I knew many of the fire-prone areas were in forested mountain foothills. Some quick Web browsing led to digital images that I could open in my ArcView mapping project to accurately demonstrate topography. These images brought some on-the-ground reality to our analysis.

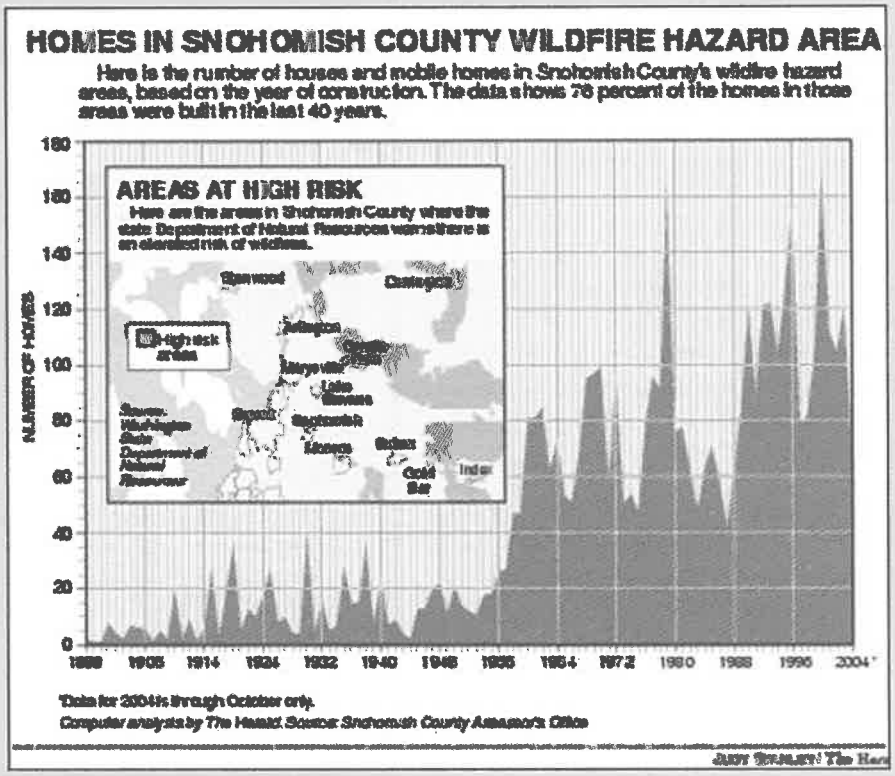
I then turned to the 30 years of fire data at my fingertips. It showed that people had sparked most of the fires, primarily by burning trash and other debris around their homes. I used a land survey shapefile of townships, with its big, uniform-sized blocks, to section off the county. I calculated the number of fires in each section and prepared a graduated color map that showed how many fires had occurred in each part of the community over the years. The map showed the state had been accurate in its concerns about the fire-prone areas.

Hefley and I reached out to the experts, sharing Adobe PDF copies of our data maps. We found fire officials and county planners not only aware of the problem, but in some cases, frustrated with their inability to spark attention. There is no law in Washington against building a home in a place that is prone to fire. One former fire marshal told us she feared that Washington – damp, green and growing in population – is just an ill-fated spark away from the devastation and loss of life that has hit other places in the west, especially California.

Her fears seem founded. One of the people we talked with is a state expert on wildfire. The land shows scars from devastating fires that have swept through the forests in these parts every 150 to 300 years, he said. People now call those areas home. We are in the midst of one of the worst dry spells in more than a century. Do the math. You may find a similar story developing in your neck of the woods.

Contact Scott North by e-mail at snorth@heraldnet.com.

Would you be willing to share a mapping example with fellow journalists? Send an electronic copy of the map along with details to David Herzog at dherzog@nicar.org



Mortgage

continued from page 17

In "Seeing Red: How to Expose Redlining Practices," Penny Loeb of *U.S. News & World Report* provides useful background information for journalists working with the HMDA data. (No. 410)

The (Fremont, Calif.) *Argus*, analyzed home mortgage data and found stark disparities in rejection rates between minority and nonminority applicants. They found that Hispanics and blacks were twice as likely to be turned down in the various counties in California. The article also concentrated on possible ways of solving the problem. (Story No. 21008)

The (Nashville) *Tennessean* focused on how race played into home mortgage lending in the Nashville area, showing that black home owners were denied for mortgage loans "up to three times more often than whites even when incomes were similar." (No. 21136)

Call 573-882-3364 or e-mail rescntr@ire.org to order copies of these tipsheets or stories.