Uplink

July/August 2001

TEST SCORES

Dispelling a myth

By Betsy Hammond
The Oregonian

Oregon education officials have emphatically touted our state's public schools as among the very best in the world, and parents in most Portland suburbs say they moved to their particular part of our metro area for the great schools.

But our newspaper was able to use computer-assisted reporting, supplemented with a lot of in-the-field reporting, to establish that Oregon's suburban schools are in fact overwhelmingly mediocre – and to name which schools have done the least to help students reach their academic potential. Our series – "Why Not the Best?" (*http://www.oregonlive.com/special/oregonian/education/*) – drew hundreds of calls and e-mails from outraged readers.

The Oregonian used SPSS to run regression analyses on student test scores, using family background statistics to "level the playing field" among schools and see how each local school performs, given the students who come through their doors.

Many newspapers issue "report cards" on schools in their communities. These typically pull together existing data, including test scores, and organize it in reader-friendly fashion. The problem is that, inevitably, schools with more educated and affluent parents get the top grades; schools with uneducated and poor parents the worst.

A few newspapers go beyond that by conducting regression analyses. Those efforts all are grounded in the same reality: A school's average test score tells you more about how wealthy and well-educated the parents are than how effective the school is, so it's essential to systematically discount the effects of family background to measure school performance.

The Oregonian added some new twists to its particular analysis, so that we can stake a real claim not to have let our suburban schools off the hook. We focused not on finding schools in impoverished neighborhoods that do well, a story that local educators were eager for us to tell. Instead, we took a hard look at schools in the most privileged neighborhoods and told readers about the ones that were coasting or failing. And we anchored our findings to national data, from the SAT and the National Assessment of Educational Progress, to provide a data-grounded reality check for parents who were hearing a relentless drumbeat of boosterism about Oregon schools being "world-class."

SPSS is easy to use, and we relied on NICAR's Advanced Bootcamp on Statistics to get the refresher we needed in how to use it right.

We gathered four key datasets, only one of which required a big battle. We ended up get-

Continued on page twenty

CRIMINAL TEACHERS

A new wrinkle

By Dianna Hunt
Fort Worth Star-Telegram

It started with a high-profile football coach and allegations of sexual assault.

The story made headlines for weeks in the Fort Worth area, as a local school district investigated the allegations against a popular high school coach.

The coach eventually resigned from the district after officials agreed to pay him $100,000 and drop the investigation. He was never charged with assaulting a female coach, but was charged a few months later with harassing another co-worker and her boyfriend.

It wasn't until *Star-Telegram* education reporter Matt Frazier called the state teacher licensing board, however, that the bigger story began to emerge.

The agency, the State Board for Educator Certification in Austin, Texas, acknowledged it was investigating the coach's teaching credentials. But beleaguered investigators also acknowledged that their workload was so high they had a backlog of cases that could take years to resolve.

Continued on page twenty-one

Inside Uplink

With students soon to return to the classroom, this issue of *Uplink* focuses on education. Among the stories discussed in this issue are: A *Fresno Bee* analysis that discovered underqualified teachers in low-income and rural schools (see page two), An *Akron Beacon Journal* project that analyzed the reasons behind failing schools (see page three), and a series by *The Columbus Dispatch* that uncovered "glaring inequities" among elementary schools (see page six).

Interesting datasets

NICAR outlines two datasets that can be useful in reporting on fires. The databases deal with forest fires and arsons.

SEE PAGE SIXTEEN

Zoot

Mark Schaver of *The Courier-Journal* (Louisville, Ky.) talks about the wonders of Zoot, software that can help reporters organize interviews, e-mails and other information.

SEE PAGE TWENTY-THREE

Uplink

July/August 2001

Volume 13, Number 4
A newsletter of the National Institute for Computer-Assisted Reporting

EDITOR
Brant Houston

MANAGING EDITOR
MaryJo Sylwester

ASSOCIATE EDITORS
Amy Sherrill
Mike Sherry

ART DIRECTOR
Yan Jin

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 bimonthly 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

TEACHERS

Law creates shortage

By Russell Clemings
Fresno Bee

With tax revenues rising and a state law requiring a fixed percentage to be spent on education, the California Legislature five years ago reduced class sizes in three primary grades from an average of more than 30 to a maximum of 20.

Almost overnight, California's elementary schools created more than 25,000 new classrooms and hired a like number of new teachers. The result was a statewide teacher shortage that persists to this day and which has burdened low-income and rural schools with large numbers of underqualified teachers.

In December, the *Fresno Bee* analyzed five years of teacher census data. We determined that, for the poorest one-quarter of California schools, the teacher shortage – as measured by the percentage of teachers working without a full credential, the basic teacher license – has grown worse each year since class-size reduction began.

The PAIF data is derived from an annual teacher census. Each record includes details about a single teacher, such as education level, credential status, experience and demographic background.

Throughout the state, schools that have trouble competing in a tight teacher market are relying on ever-larger numbers of teachers on temporary permits, internships and waivers – teachers who, in a normal market, would never be hired.

The data came from the California Department of Education's Professional Assignment Information Form and CalWORKS Children/Meal Programs databases.

The PAIF data is derived from an annual teacher census. Each record includes details about a single teacher, such as education level, credential status, experience and demographic background. Data on teacher assignments is also included. The CalWORKS/Meals database contains demographic information on each school's students, including the percentages on welfare or receiving free or reduced-price lunches.

Personal identifiers are either not collected or purged from the public data releases, so it's not possible, for example, to track an individual teacher's movements. But the data does provide rich detail on populations of teachers and students at the individual school and (for teachers) grade level.

Many states collect similar data, and the National Center for Education Statistics (*http://nces.ed.gov*) has nationwide datasets and surveys covering some of the same subject matter.

Much of the California data was available online at the department's Web site, *www.cde.ca.gov/demographics/files/paif.htm*, for the most recent PAIF data and *www.cde.ca.gov/demographics/files/afdc.htm* for the welfare/meals data. The California Department of Education provided the two earliest years of PAIF data at no charge on nine-track tape.

Collating and analyzing the data was not easy. Midway through the five-year period, the structure of the PAIF data changed from a flat file with room for multiple assignments for each teacher to a relational database with a flat file of teacher records linked to a table with one or more assignment records for each teacher.

Additional reference tables were needed to translate assignment codes into English terms such as "self-contained first-grade classroom." And in determining which assignment codes to use and how to define credentialed and uncredentialed teachers, studies from the state Class Size Reduction Consortium (*www.classize.org/*), the Public Policy Institute of California (*www.ppic.org/*), and the Center for the

Continued on page five

SCHOOLS

Teacher, student exodus

David Knox
Akron Beacon Journal

Computer-assisted reporting is no different than other journalism in an important way: Often the best stories start with a reporter's gut instinct.

When Reginald Fields took over the education beat at the *Akron Beacon Journal* in June 1998, it didn't take him long to learn that the city's public schools weren't equal in quality.

Among the district's eight clusters – each composed of a high school and its feeder middle and elementary schools – Buchtel High School was considered by parents, teachers and students far and away the worst.

The evidence

Evidence backing that poor reputation was easy to find. During the final grading period of the 1999-2000 school year, 36 percent of the students at Buchtel earned a D letter grade – or worse.

Comparisons to other schools were equally dismal. Scanning a sorted listing of Ohio's proficiency test scores found almost all the Buchtel cluster schools near the bottom.

But grades and test scores didn't explain why one group of schools performed so poorly.

Fields soon came to question the conventional answer – that Buchtel's students were too economically disadvantaged (read, too black) – to do better.

True, Buchtel schools were by far the most segregated in the city – 97 percent black compared to a district average of 47 percent. But the neighborhoods where they were located were much more racially diverse. Furthermore, visits to the Buchtel schools took Fields down streets lined with many tidy, attractive homes near well-kept parks – hardly a portrait of urban blight.

Fields also heard a steady drumbeat of complaints from parents and students about the low experience of the teachers at the Buchtel schools. He wondered whether the school district's liberal policy toward teacher transfers – by contract, teachers can ask to move after a single year of service – might contribute toward Buchtel's poor academic performance.

"Destined to Fail," a six-month project sparked by Fields' skepticism, provided a challenging opportunity to use multiple analytical tools and strategies to go behind the issue of poverty and search for other factors under the control of school administrators that can affect educational quality.

The Ohio Department of Education publishes reams of readily available, easily imported data – almost all useless for the Buchtel project.

As always, the first task was getting the data.

The Ohio Department of Education publishes reams of readily available, easily imported data – almost all useless for the Buchtel project because none of it was at the cluster level. With the exception of proficiency tests information, which included the number of students taking and passing each test in each school, all the other "building-level" data came in the form of percentages and averages. Those numbers couldn't be brought up to the cluster level because their denominators – the number of students and teachers in each school – varied widely.

What was needed was student- and teacher-level data that could be added together, and then calculated for each cluster.

That information was available from the Akron Public Schools – but mainly as paper records.

Calculating poverty

For the issue on student poverty, that wasn't a problem. All that was needed to calculate the percent of low-income students was the enrollment and number of students eligible for free or reduced lunch for each of Akron's 48 elementary and middle schools. (High schools were left out because of the

Continued on page four

Education tipsheets

The following education tipsheets are available from the IRE Resource Center, (573) 882-3364:

Tipsheet #957 outlines new angles for reporting on the same old education data. Also included are tips for putting together your own local education investigation.

The Herald-Leader (Lexington, Ky.) ran a four-part series that examined teacher quality by evaluating Kentucky's teaching programs and certification process. Tipsheet #892 contains the resources used in the series as well as a list of do's and don't's for tackling the education beat.

"Destined to Fail," a six-month project, by David Knox and Reginald Fields, can be found at *www.ohio.com/specials/buchtel_series*

The Akron Public Schools Web address is located at *www.akronschools.com*

Continued from page three:

Exodus

tendency of older teen-agers not to apply for free and reduced-cost lunches out of embarrassment. Two other schools were excluded because they are magnet schools that draw from all parts of the city.)

Measuring teacher inexperience took a little more work. The numbers of years of service for each of the 2,023 teachers in the schools were input into a Microsoft Excel spreadsheet and a percentage of those with less than six years calculated for each building and cluster.

The results showed that nearly half the 285 teachers in the Buchtel cluster's eight schools had less than six years experience. The cluster average was 9.7 years of experience, compared to the district average of about 14.4 years.

SPSS, a statistical analysis program, was used to gauge the possible impact of teacher inexperience on student performance.

A multiple regression analysis found that low teacher experience and poverty together could explain as much as 54 percent of the difference in the proficiency test results among Akron's elementary and middle schools. The probability of the relationship being a coincidence is less than 1 in 1,000.

A limitation of regression analysis is that, while it shows correlations, it doesn't prove cause-and-effect – a point stressed by several researchers at two local universities who were asked to review the data. Teacher inexperience may contribute to low test scores, but a lack of experienced teachers also could reflect an exodus of veterans who choose to transfer to better performing schools.

That scenario would be more likely if poverty is considered an invariable factor – beyond the influence of school policy. That's because a single-variable regression analysis showed poverty alone could account for 48.7 percent of the difference in the test scores, compared to 43.1 percent for teacher inexperience considered as the sole independent variable.

In other words, the conventional wisdom still works: poverty could be the dominant factor in explaining low test scores.

But a key finding of the Buchtel data analysis was that the school district's "open enrollment" policy, which encourages students to leave their neighborhood schools in search of a better education, might be altering the family-income demographics of the schools.

Such student flight might contribute to poor test scores if the students leaving a cluster are more likely to do better on the exams than the classmates they leave behind.

To investigate that possibility, another methodology – geocoding – was used to roughly gauge the performance potential of students leaving and staying in each school cluster by determining the household income of their neighborhoods.

The *Beacon Journal* obtained from the school district the addresses of more than 25,000 students – about 84 percent of all students enrolled in Akron's public schools. In accordance with Ohio law, the only addresses omitted were those of students whose parents signed forms requesting the information be withheld.

Microsoft Word's global search and replace functions were used to massage the text files into delimited columns and rows for import into a database program, Microsoft Access 97.

Initially, school officials said the data, which included which school each student attended and grade level, was available only on paper because the district couldn't afford the cost of programming. Officials relented after the paper pointed out that redirecting the ASCII text output of the report program to a floppy disk entailed no additional cost.

Microsoft Word's global search and replace functions were used to massage the text files into delimited columns and rows for import into a database program, Microsoft

Continued on page five

Continued from page four: **Exodus**

Access 97.

The addresses were then geocoded into census block groups that included 1999 estimates of median household income, as provided by Claritas, a national marketing research service used by the *Beacon Journal*'s advertising department.

The next step was to determine where the students would go to school in the absence of the open enrollment program by matching their addresses to the district's 317-page directory of school attendance zones.

The job was made manageable by sorting the database on street number and name – allowing a dozen or so records to be updated with their assigned schools in a single query.

The completed database had all the information needed to determine the impact of the district's open enrollment program: what school each student was assigned and the school actually attended.

The analysis found that 20 percent of the children in the Buchtel district were going to schools in other clusters, compared to a district average of 13 percent. At the high school level, one of every three Buchtel students had left the cluster.

Comparing the average neighborhood incomes of students staying and leaving provided evidence that the student flight might contribute to keeping test scores down. Students fleeing the Buchtel cluster were found to have neighborhood incomes more than $3,000 higher than those who remained. Only one other cluster – Central-Hower, the next poorest performing cluster – showed a net loss of income – about $1,800.

And what was the reaction of the school officials to the series? They didn't challenge the findings of the statistical analysis.

That changed soon after "Destined to Fail" was published, December 10-13.

Within a week, the school board president called for a panel of parents, school adminstrators and community leaders to address issues raised in the series.

Within a month, the school administration presented the board with over a dozen recommendations to better the Buchtel cluster, including several proposals for upgrading the quality of teaching in the schools.

David Knox can be reached by e-mail at Dknox@thebeaconjournal.com

Continued from page two: **Shortage**

Future of Teaching and Learning (*www.cftl.org/*) were invaluable.

It took a 290-line FoxPro script and a week of work to reduce the disparate databases, which totaled more than 400 megabytes, to a manageable flat file of 6,295 records, one for each California school that had at least one teacher in a self-contained elementary classroom.

Each record in that table contained five years of data, and for each year there were five data fields: (a) Percentage of students on welfare, (b) Percentage of students on free/reduced price lunches, (c) Total teachers, (d) Fully credentialed teachers, (e) Less-than-fully credentialed teachers, and (f) Teachers for whom no credentialing information was available.

We published tables with the local data, and our Web designers created a searchable interface for the entire state table at *www.fresnobee.com/man/projects/teachers.*

Statewide, a simple line graph told the story. For the wealthiest one-quarter of California schools, the proportion of uncredentialed teachers rose from zero in 1995-96 to 5 percent in 1999-2000. In the poorest schools, the proportion rose from 3 percent to 23 percent over the same five-year span.

Reaction

"No one intended for the poorest schools to have the least-qualified teachers, but that seems to have been the effect" of class-size reduction, one researcher said.

Class-size reductions have been an undeniable boon for most California schools. Test scores are rising, pupils are getting more individual attention, and parents are pleased. But for those schools that can't compete in the tight teacher market, the data shows a decidedly bleaker situation.

Russell Clemings can be reached by e-mail at Clemings@cris.com

The Fresno Bee story on the teacher shortage is available from the IRE Resource Center (Story # 17883). To order, call (573) 882-3364.

The National Center for Education Statistics (*http://nces.ed.gov*) has nationwide datasets and surveys covering some of the same subject matter.

SCHOOLS

Separate and equal?

By Doug Haddix and Bill Bush
The Columbus Dispatch

Like many government offices, the Columbus school district maintains its electronic records in a mainframe computer system that makes it nearly impossible for outsiders to analyze.

Information is kept on separate tables that can be linked only by a confidential field such as student identification number or teacher Social Security number.

The moment we saw the record layouts for the district's master student file and master employee file, we saw red flags. Those databases, we realized, would play a pivotal role in the newspaper's investigation into whether the return to racially segregated schools in Columbus had produced unequal opportunities for poor and minority students.

Glaring inequities

The four-day "Dividing Lines" series, published in June 2000, uncovered glaring inequities among the district's 88 elementary schools. Those findings came despite a school board policy approved four years earlier, promising that the end of cross-town busing would not produce separate and unequal schools.

Seven weeks after our request for the district's master files, we finally obtained the electronic data in a usable format – but only after finding a smoking gun in Ohio law.

At four separate meetings with the school district's data crunchers, we repeated our request that programmers randomly generate a unique identifier for each of the 65,000 students and 11,000 employees. We assured district officials that we had no interest in identifying particular students or obtaining Social Security numbers of any teachers, principals or administrators. We simply wanted to be able to link the district's 19 separate data tables.

Without a link, for instance, we had a table that included teacher salaries but not their names, which were kept in a separate file.

Six weeks after our initial data request, the district provided its master student and employee tables on a CD-ROM, complete with a nifty jewel case cover. The only problem: no unique identifiers. The tables were useless.

We scheduled a fifth meeting with top district officials to press our case. The day before the meeting, we discovered through a source in the Ohio Department of Education that all public school districts are required by law to generate a unique student identifier when they submit data to the state.

Armed with that ammunition, we met with the district's top brass. Immediately, they agreed to provide the data with a randomly generated identifier created just for *The Dispatch* – the missing link we needed. Without that discovery, we might have been forced into taking legal action. Ohio's public record law sets no firm deadline for completing requests; it merely requires public bodies to make records available in a "reasonable" period. It took the district another week, however, to turn over data that it had maintained all along.

Seven weeks after our request for the district's master files, we finally obtained the electronic data in a usable format - but only after finding a smoking gun in Ohio law.

Once the district's usable master student and employee tables arrived on a new CD-ROM, we did a bit of tedious work to move the 57 MB of data into a Microsoft Access database. The 19 tables had an unfamiliar ".JUN" extension, so we saved them as flat text files, then moved them into Access.

We used the advanced import function to save the import specifications. That saved a lot of time because we had requested master files for the last year of cross-town busing (1995-96) and the most current data for the return to neighborhood schools (1998-99). It took about 20 minutes to type in the im-

Continued on page seven

Resources

The *Columbus Dispatch's* series, "Dividing Lines," is available from the IRE Resource Center (Story # 17279). To order, call (573) 882-3364. The series can also be found at: *www.dispatch.com/news/special/schools/cpsmain.html*

For useful education statistics and background, try these sites:

Center on Education Policy: *www.ctredpol.org*

Institute for Urban and Minority Education: *http://iume.tc.columbia.edu*

U.S. Department of Education: *www.ed.gov*

Ohio Department of Education: *www.ode.state.oh.us*

Continued from page six: **Separate**

port specs for each table for 1995-96, so saving them slashed the import time for the 1998-99 tables. In addition, the saved specs will make it faster to import data for subsequent years, provided that the data fields don't change.

Because the Columbus district moves slowly on requests for information, we filed simultaneous data requests with the Ohio Department of Education, which maintains information on the more than 3,600 public schools across the state data.

The judge who originally ordered desegregation of Columbus schools in the 1970s termed the new zones "curious."

While not as extensive as the Columbus district's master files, the state data gave the project team helpful information on enrollment, per-pupil spending, proficiency test scores, median teacher salaries and other topics that we could query by individual schools and by district.

Unlike the Columbus district, the Education Department responded with amazing speed, sometimes e-mailing Microsoft Excel spreadsheet files to us within hours of a request. The school district maintained it didn't have the staffing to provide information that quickly.

To get the fullest picture possible of changes in Columbus elementary schools since the end of busing for integration, we obtained from the district and analyzed several other types of data:

• Student school assignment records.

We received one of those "you can't get there from here" responses from the district when we first asked for records showing Elementary Attendance Zones, which are geographic boundaries that determine each student's "home school."

First, the district provided a huge paper file that included an elementary-zone breakdown of the number of students living in each assignment area, their race, their gender and the school they attend. In Columbus, students can attend their home school or try to transfer to another school through a lottery.

With 65,000 students to track, the paper files were useless for analysis because of their size. In a bizarre twist, the school district official who created the paper report maintained that he did not keep it electronically; rather, he insisted, he printed the report and then deleted the electronic version from his PC. (Don't ask.)

After some wrangling, we were able to obtain the underlying records to recreate a version of this report ourselves using a Microsoft Access database. We obtained an electronic database in which each record represented a student; each record contained demographic information, as well as that student's "home school" and the school the student actually attended.

Because we had data for the last year of crosstown busing and for the most recent year of neighborhood schools, we were able to analyze changes in the assignment boundaries for elementary schools. Those attendance zones, it turned out, were being grouped together in many cases along racial lines.

The judge who originally ordered desegregation of Columbus schools in the 1970s termed the new zones "curious." One school assignment plan formed a sideways L, picking up predominantly white neighborhoods while avoiding minority neighborhoods. Another linked white neighborhoods that were separated by a railroad switchyard and a freeway, again avoiding adjacent black neighborhoods.

• Uniform school accounting system database.

This huge district database tracked dollars by various expenditure codes by building. That allowed us to analyze how much money was actually flowing to schools with different racial and economic compositions, and how those funds were being spent. Among other uses, we were able to gauge a school's ability to raise money through parent-teacher groups.

• Free and reduced-price lunch program participation by school.

For whatever reason, neither the state nor the district kept this information in computer form. The district's food services office maintained paper records, which we entered into a Microsoft Excel spreadsheet. We linked that table using the school building identification number to other

Continued on page nine

Another story on desegregation can be found at the IRE Resource Center:

"Deciding Desegregation," a 1999 series by the *Charlotte Observer*. (Story #15958)

To order this or other stories, call the IRE Resource, (573) 882-3364.

Software:
Microsoft Access, Excel, and mapping software.

Data:
Various school district databases, including student-assignment records and spending data.

Sources:
Columbus (OH) school district and the Ohio Department of Education.

Non-traditional licenses

By Alice Warchol and Philip Walzer
The Virginian-Pilot

In the scramble to hire teachers, school divisions in Virginia are now employing educators who do not have traditional teaching licenses.

Some of them have yet to take or pass the PRAXIS I – a national teaching exam. Others have not completed college-level coursework required by teacher-preparation programs. But they are allowed to teach for three years with varying types of alternative licenses.

While administrators defend their hiring practices, a slew of elected officials and teaching professionals say these temporarily licensed teachers should not be in the classroom.

While administrators defend their hiring practices, a slew of elected officials and teaching professionals say these temporarily licensed teachers should not be in the classroom.

Investigating dispute

That was one of the disputes we set out to answer in our story. We also wanted to know how many teachers being hired had some type of alternative license. Where did they work? Were they assigned randomly, as administrators claimed, or were they mostly working in poor schools or schools with low test scores?

To begin, we started with the Virginia Department of Education. We learned that there are two types of alternative licenses: provisional and conditional. We also found out that half of the new licenses issued last year were alternative ones. (We wanted to gauge if this was a growing trend, but the state just started tracking this information.)

Our next step was to ask each of the five local school systems that we cover how many provisional and conditional teachers they employ, and more importantly, where they worked. We used Virginia's Freedom of Information Act to request the numbers. We encountered little resistance. We were not charged.

Organizing data

To organize our data, we used Excel and created one worksheet for each city. Then, we downloaded from the Department of Education's Web site school enrollment numbers and free- and reduced-price lunch data – a national indicator of poverty. We were lucky because the state posts its enrollment and lunch program information in Excel.

For every school in each city, we had columns for enrollment, percent of children eligible for free- and reduced-price lunch and number of alternatively licensed teachers. Then we added another column for the test score information. That came from our colleague Matthew Bowers who had created an Excel spreadsheet when the state's test results were announced last year. Each school in each city was assigned a number between 1 and 4. The number 1 meant the school was fully accredited by the state; a 4 indicated the lowest rating: accredited with warning.

To determine whether one school had more alternatively licensed teachers than another, we inserted a column that gave us the rate of teachers per 1,000 students for each school. (We divided each school's number of alternatively-licensed teachers by its student enrollment and then multiplied that result by 1,000.) At this point, we sorted each worksheet by the rate per 1,000. That showed us the top 10 schools in each city with the largest number of alternative teachers taking into account their student enrollment.

Top 10 list

All of the hiring folks we interviewed said they placed alternatively licensed teachers wherever needed. True, we found poor and wealthier schools among our list of top 10. But when we looked at how well those schools did on the state's exams, our analysis showed that most of the schools did not meet Virginia's standards.

When we published our story, we arranged

Continued on page nine

The story "Many Teachers Carry 'Conditional' Licenses," is available from the IRE Resource Center (Story # 17798). To order, call (573) 882-3364.
It can also be found at *www.pilotonline.com/news/nw0426lic.html*

The database that contains information on how many alternatively licensed teachers each school has is posted for search at *http://home.hamptonroads.com/schoolguide/AccreditedTeachers/search.cfm*

For more information about how teacher licensing works in Virginia, check out the Web site *www.pilotonline.com/news/nw0426pro.html*

Continued from page eight: **Licenses**

to post part of our database online. Parents could log on to our Web site and search by city for their child's school. The online database shows enrollment, number of alternatively licensed teachers and the rate per school. We e-mailed the file to our online news coordinator, Kerry Sipe, a couple of days in advance to publication. He arranged to have it posted on the Web site.

Our advice to fellow reporters is to repeatedly proof your work. Have someone else look over your computations.

For the print version of our story, we ran a large graphic listing the same info for the top 10 schools.

Numbers give structure

The numbers gave us a structure for the story. But most of our reporting focused on why teachers weren't qualified for a full license and whether that was hurting kids. We found alternatively licensed teachers ready to defend themselves. We showed what it's like to be in their classroom. We also talked to education professionals who worry about the quality of education someone can bring to the classroom when they haven't student-taught or can't pass the PRAXIS I – a basic-skills test.

Our last step was to think like parents. Would a school tell them if their child's teacher had a temporary license? The conflicting yes and no answers we received from school divisions only made our story better.

Although the project was a lot of work and took several weeks to complete – we did it in between covering our beats – it was fairly simple as far as CAR endeavors can go. Our advice to fellow reporters is to repeatedly proof your work. Have someone else look over your computations. And when you're sending Freedom of Information Act requests to several school systems, it's imperative to ask for the same information.

Alice Warchol can be reached by e-mail at awarchol@pilotonline.com. Philip Walzer can be reached by e-mail at walzer@pilotonline.com

Continued from page seven: **Separate**

data tables so that we could explore the relationship between student poverty and items such as test scores, per-pupil spending and teacher experience.

In addition to basic spreadsheet and database analysis, the newspaper used GIS mapping software to look for geographic patterns involving the 88 elementary schools. Seeing the data on a map helped us spot several important changes, most notably the concentration of less-experienced teachers in the inner-city and near-downtown schools. Mapping also vividly showed how dramatically many of the schools changed in racial composition of the student body.

The reporting explored the key data findings:

• Schools with high concentrations of poor students generally have the least-experienced teachers.

• Elementary school assignment boundaries in some neighborhoods closely match the lines once deemed unconstitutional by a federal judge.

• Although schools with high numbers of poor students receive hundreds of thousands of federal dollars, overall spending by school bears little relation to the poor student population.

• Schools with high concentrations of poor students score far below other schools on the statewide fourth-grade proficiency test.

• Elementary schools with low numbers of poor students qualify for additional resources because their students fare better on the proficiency test.

The work on "Dividing Lines" continues to benefit *The Dispatch* because we all learned about data sources and techniques that we never knew existed.

Doug Haddix can be reached by e-mail at dhaddix@dispatch.com.
Bill Bush can be reached by e-mail at bbush@dispatch.com.

Other computer-assisted reporting stories on education can be ordered at the IRE Resource Center:

In a three-day series, *The Virginian-Pilot* examined special education. Among other findings, the paper found that many special education classrooms have a disproportionate number of black students. (Story #12162)

In late October and early November of 1999, The Pittsburgh *Post-Gazette* reported that educators who molest students escape recrimination because school districts allow them to quietly resign and go to another school district. (Story #16055)

Funding schools

By Holly Hacker
St. Louis Post-Dispatch

In 1993, the courts threw out Missouri's school funding system. The judge in the case declared that because of huge gaps in spending, public schools ranged from "golden" to "god-awful."

So the state created a new funding formula that was supposed to be more equitable. Still, over the years, many educators and lawmakers complained that disparities still existed.

The education reporters at the *St. Louis Post-Dispatch* decided to tackle the school funding issue last fall. We knew it was going to come up for debate in the next legislative session. Meanwhile, across the Mississippi River, Illinois lawmakers also grappled with the issue. So we wanted to do two separate analyses, one for each state.

Here was the perfect opportunity for me to test my skills as the education team's CAR specialist. I had used Microsoft Excel and Access extensively and recently learned SPSS. Still, I would suffer plenty of headaches and emit several primal screams before our three-part series ran, after about three months of work.

We wanted to know if school districts spent about the same amount on each child. Sure, one could just get official spending figures for each district and compare. But those numbers don't tell the whole story. Just as it costs more to live in Chicago than in Peoria, some districts have to pay more for teachers and resources based on their location. Plus, children with disabilities or other special needs cost more to educate. So, we wanted to compare districts by taking those differences into account.

I called various school finance experts for help and struck gold with the National Center for Education Statistics. A researcher in the NCES school finance division pointed me to a bunch of studies and sources. I learned about something called the geographic cost of education index, which lets you adjust for cost differences by school district. The index is based mostly upon salaries for teachers with comparable experience, plus other costs.

Through NCES I also found a way to adjust for differences in student costs. Generally, children with disabilities cost 2.3 times as much to educate, while poor children and children with limited English cost about 1.3 times as much. So if a district has 1,000 students and 200 of them are in special education, it's the equivalent of educating 1,260 "regular" students. That is, 800 + (200 x 2.3) = 1,260.

Getting the data

To do this project, I pulled data from several sources.

- I downloaded the cost-of- education index from the NCES Web site.
- The Illinois education department e-mailed me spreadsheets with finance information, student enrollment, and the number of special education, limited-English and poor students in each district. I defined "poor" students as those who get free or low-cost meals through the federal school lunch program.

Luckily, officials at the Missouri and Illinois education departments not only provided data willingly, but also helped me make sense of the numbers...

- Missouri's education department e-mailed me a monster (16.8 megabytes) text file with detailed financial information. For reasons discussed later, I needed Missouri school district spending broken down in ultra-specific categories.
- I pulled student information for Missouri districts from an Access database I had already created.

It seems that on every CAR project, I inevitably run into more problems than I ever expected. This one proved no different. For starters, we had to determine how to count students. Illinois gives districts money based on student attendance. Missouri uses "eligible pupils," which is like attendance, but with some differences. We picked student

Continued on page eleven

Resources:

The St. Louis Post-Dispatch story on school funding is available from the IRE Resource Center (Story #16960). To order call (573) 882-3364.

National Center for Education Statistics, *www.nces.ed.gov/edfin* – For more information on the NCES finance research, see "Inequalities in Public School District Revenues," July 1998, publication number NCES 98-210; and "Geographic Variations in Public Schools' Costs," March 1998, publication number NCES 9804.

You can also contact William J. Fowler, a program officer at the NCES education finance statistical center, at (202) 219-1921 or *William_Fowler@ed.gov*

Lawrence O. Picus, a professor in the School of Education at the University of Southern California and director of the Center for Research in Education Finance (CREF). (213) 740-2175 or *lpicus@usc.edu*

Continued from page ten: # Funding schools

enrollment as our measure, because even if kids don't come to school every day, the district still has to pay for their teachers, textbooks, utilities, etc. Plus, poor districts tend to have lower attendance, thereby inflating the amount spent per pupil.

We also had to define expenditures. We used operating expenditures, which includes costs to run the district, but not capital costs. To calculate Missouri district spending based on enrollment instead of "eligible pupils" (love that education jargon), we had to add some spending categories and subtract others. That's why we needed detailed spending files. I spent two weeks just importing and cleaning those files.

Luckily, officials at the Missouri and Illinois education departments not only provided data willingly, but also helped me make sense of the numbers and financial terms.

Number Crunching

I imported the various spreadsheets and tables into Microsoft Access, then joined them so I had a single table for each state. From there, I divided each district's operating expenditures by its cost-of-education index to get adjusted expenditures. Next, I adjusted student enrollment using the weights for special education, limited-English and poor students. Then I divided the adjusted expenditures by adjusted enrollment to get adjusted spending per pupil.

Actual expenditures = Adjusted expenditures
Cost-of-education index

Regular students + (2.3*special ed students) + (1.3*poor students) + (1.3*limited-English students) = Adjusted enrollment

Adjusted expenditures = Adjusted spending per pupil
Adjusted enrollment

So you might have two districts that spend $5,000 per student in actual dollars. But factor in location and student need, and one district could have more "buying power" than the other – say, $5,200 per student vs. $4,600 per student.

Now that we had adjusted spending for every district, I turned to SPSS for further analysis. I calculated the minimum, maximum and average spending per student for each state, in both actual and adjusted dollars, weighting for district size. One school finance expert suggested we also look at the coefficient of variation, which is the standard deviation divided by the average spending per pupil. This is the range that about two-thirds of students fall within.

In both Missouri and Illinois, the gap between the highest- and lowest- spending districts shrank using adjusted dollars, but not entirely. This table summarizes the results for Missouri:

Spending per pupil for Missouri school districts, 1998-99

	Actual dollars	Adjusted dollars
Minimum	$3,500	$3,250
Maximum	$10,500	$9,500
Average	$5,400	$4,400
Coeff. Of variation	20%	15%

In actual dollars, about two-thirds of Missouri students are in districts that spend within 20 percent of the average – that is, from $4,320 to $6,480 per child. In adjusted dollars, those students are within 15 percent of the average – from $3,740 to $5,060.

Answers

Ok, so we know the gaps still exist. Why? For answers, I ran some regressions in SPSS of school spending against district poverty rates, tax rates, property tax bases, and other variables.

We found that local tax rates explained about a third of the variation in Missouri district spending. The property tax base explained about a third of the variation in Illinois. In our story, we explained that Missouri rewards communities willing to tax themselves, while in Illinois, what a district spends depends more upon the property wealth of the community.

Continued on page thirteen

The Los Angeles Times also reported on state school funding in 2000. (Story # 17837)

On tipsheet #1103, Linda Shaw of the *Seattle Times* provides suggestions on how to obtain and use data to review the performance of a school system.

Tipsheets and stories can be ordered from the IRE Resource Center by calling (573) 882-3364.

Tipsheet #1287, by Hacker, outlines several steps to use when formulating and presenting an educational survey or "school report card" for your newspaper. Also, offers specific tips on adjusting free and reduced lunches as a means of measuring poverty in school.

SCHOOLS

Checking the grades

By Jeff Thomas
The Gazette

It took some computer power to pull this one off, but the crucial ingredient was a reporter refusing to let her sources control the story.

Colorado, like other states, has placed academic standards into statute.

And like many states, Colorado uses a test to determine if schools are meeting those standards.

Last year, lawmakers wanted to go a step further and translate the sometimes-murky test results into terms that millions of taxpayers statewide could rapidly understand. They agreed to issue each school an annual report card, complete with letter grades, A through F, based primarily on test scores.

Colorado drew up a formula that, for each school, blended the scores of the separate reading, writing, math and other tests taken by kids in the various grades. The result was a single number for the school that, in turn, would be translated into a letter grade.

Inside the statehouse committee rooms, debate revolved around the very concept of handing out As or Fs to schools. Outside the Capitol, the question was more visceral: What grade would my school get?

Inside the statehouse committee rooms, debate revolved around the very concept of handing out

The governor's office wasn't saying. It had drawn up the formula and had tried it out on some test results, but wasn't sharing the results. Here was a proposed law whose precise effects could be determined, yet those effects were closely guarded, even as the legislation moved toward a vote.

One of *The Gazette*'s Denver bureau reporters, Michele Ames (now with the *Denver Rocky Mountain News*), took matters into her own hands. She obtained a copy of the formula and reviewed it with the authors. Test scores were readily available from the state education department. She cracked open Excel, and began pouring test results into spreadsheets.

I'll spare you the specifics of the formula, but here were the basic steps:

1. Assemble test scores for each school in the state. Also for each school, assemble the number of pupils enrolled and the number taking each test. The raw data were available in flat files and were assembled in Excel.

2. The formula required adjusting each test score using the percentage of pupils taking the test. This also was done in Excel.

3. At this point, the formula wandered into statistics. Michele brought her disk to *The Gazette* newsroom in Colorado Springs, where we fed the numbers to SPSS, which is statistical software. The basic idea was to determine the z-score of each test score in each academic subject, in each grade level, at each school. A z-score tells you whether a test score falls within the average amount of variance around the mean test score, or if it falls somewhere above or below that average amount of variance.

4. Having produced a series of z-scores for each school, one for each of the tests, the formula blended the z-scores in such a way that produced a single numeric value for each school. These values themselves were compared to produce yet another set of z-scores, one for each school.

This was done in SPSS.

5. Under the proposed legislation, schools with the top 8 percent of the z-scores would receive A's. The bottom 2 percent would receive F's. B's, C's and D's would be distributed to the 90 percent of schools in between. We moved the final z-scores from SPSS back to Excel. We created a new column, into which we deposited a nested "IF" statement that read the z-score and issued the proper letter grade for each school.

Local schools

For our own story, we extracted the schools and grades in the school districts within our readership area, and published those in the paper. Meanwhile, we put the entire list of schools statewide on our Web site and asked the Associated Press

Continued on page thirteen

The Gazette's story on Colorado test scores can be found in the archive files at www.gazette.com

These stories on test scores are available from the IRE Resource Center:

"Too Good to be True," published in the *Times-Picayune* in 1997. (Story #14336)

The Times-Picayune found that, under intense pressure to improve test scores, principals and teachers in the New Orleans public school system used improper and sometimes illegal methods to improve their students' perfomance. The result: Overall scores rose dramatically, but so did the number of suspicious test-score swings.

In 1998, *the Detroit Free Press* published, "Testing MEAP Scores." (Story #14756)

In Michigan education, MEAP is the measure. But a *Free Press* analysis shows that a district vs. district comparison of MEAP test scores is inherently unfair.

Continued from page twelve:

Grades

bureau in Denver to send an alert to Colorado members, directing them to our Web site to find schools in their area. Newspapers across the state published portions of the list, along with localized stories of their own making. We also posted a nerd box that other papers were welcome to use.

A sidebar to Michele's story noted that 75 percent of the As would go to schools in just five affluent school districts. Most of the Fs, meanwhile, would go to urban schools in low-income parts of Denver.

Judging by the solid week of phone calls and e-mails that rained down upon Michele, and by the legislators who brought her story to the floor during debate, this could be called impact journalism. Computers made it possible, but a reporter's instincts made it happen.

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

Continued from page eleven:

Funding schools

We also found that in Missouri, poor children weren't necessarily trapped in the lowest-spending districts. But in Illinois, the lowest-spending districts had more poor students, even after adjusting for student need. Small districts in both states tended to spend more because transportation, administration and other costs eat more of the budget.

I also tried to see if differences in spending had any effect on test scores, but I didn't find a strong relationship.

I also tried to see if differences in spending had any effect on test scores, but I didn't find a strong relationship. Perhaps the relationship does exist, but one would need to break down spending into more specific categories – say, dollars spent on classroom instruction. Plus, because test scores can vary from school to school, next time I might try analyzing on the school level, not the district level.

Another issue we ran into was how to report our findings. Because I defined pupils differently, the actual spending did not match district figures. Would readers get the distinction? In the end, we decided to group districts as high, medium and low spenders based on actual spending, then on adjusted spending.

Advice

Here's some advice if you attempt a similar project:

- Follow what I call the Rule of 3.7 – Estimate the time you think the project will take, then multiply by 3.7. Or more.
- Be shameless in consulting with experts. Get their advice on how to design your analysis. What you're doing has been done before, so you can learn about pitfalls and caveats up front.
- Get those experts to review your findings.
- Work with graphics editors early and often. This is complicated stuff, and you should present your findings in a way that makes sense to readers but doesn't overwhelm them. CAR directors probably know this already, but as a reporter doing CAR, I made this crucial discovery too late.
- Once you've analyzed the numbers, get into the classroom. So one district spends $5,000 per child and another spends $10,000. How does that play out in the classroom? Two education reporters spent several days in some local schools getting vivid examples to illustrate the spending gaps.

Holly Hacker can be reached by e-mail at hhacker@post-dispatch.com

Computer-Assisted Reporting Boot Camps

Missouri School of Journalism
Columbia, Mo.

- July 15-20, 2001
- August 5-10, 2001
- Jan. 6-11, 2002

More information, including a registration form, is available at *www.ire.org/training/boot.html*, or call IRE at (573) 882-0684

"Session 2000: The best laws money can buy?" can be found at *www.people.vcu.edu/~jcsouth/on-the-ledge-2000/project* or ask for story #17382 available from IRE's Resource Center. (573) 882-3364.

The following stories also deal with campaign financing and gifts:

In 1997, the *Indianapolis Star* published a series documenting links between special interest money and the Indiana Legislature. The articles suggest that lawmakers could reduce their dependency on special interest contributions whithout harming their ability to reach voters. (Story #14392)

The Akron Beacon Journal was able to discover ways Ohio legislators have received hundreds of thousands of undisclosed campaign contributions through an almost untraceable pipeline of gifts and committees. (Story #10151)

Texas Lawyer reported that a large percentage of money raised by candidates was spent on personal gifts and services. (Story #8131)

STUDENTS

Legislator gifts

By Robb Crocker
Virginia Commonwealth University

Legislative reporting students at Virginia Commonwealth University took a close look this spring at the gifts state legislators receive. We found that Virginia's lawmakers got about $118,000 in gifts last year from special interests.

Add in the gifts to the governor, lieutenant governor and attorney general, and the haul totaled almost $185,000.

The 13 students in the class, myself included, culled the information from the conflict-of-interest forms state officials filed in January. We entered the information into a spreadsheet so we could analyze it and identify the biggest givers and recipients.

Online searchable database

We wrote more than a dozen stories and put them online – along with a searchable database detailing the gifts reported by each legislator. We highlighted the biggest gifts – including trips to hunt caribou in the Arctic Circle and quail in Georgia. And we raised ethical questions: Should lawmakers be accepting such freebies from businesses, interest groups and lobbyists seeking to influence the legislative process?

The project drew an enthusiastic response. AP picked it up, and we received several complimentary e-mails – especially from citizens thanking us for putting the records online.

Focus on General Assembly

We initially focused on the 140 members of the General Assembly. The project involved the following steps:

We obtained the forms, called Statements of Economic Interests, from the clerks of the House and Senate.

Each student was given the conflict-of-interest forms for 10 legislators. (Jeff South, who teaches the course, also took a batch of 10 forms.)

We entered the information for each legislator into Excel. For every gift, we recorded the description, the value, the giver's name and city and the recipient.

Using cut-and-paste, we combined the data from each student into one large spreadsheet.

We cleaned the data a bit – especially to standardize the names of givers. (The state's major utility company, for example, had been entered several ways: as Virginia Power, VA Power, Dominion Power, Dominion Resources and other variations.)

We analyzed the data, using Excel's PivotTable Report feature to subtotal it by givers and by recipients.

We highlighted the biggest gifts – including trips to hunt caribou in the Arctic Circle and quail in Georgia.

The class held a brainstorming session to generate story ideas. Some students decided to profile the most generous gift-givers – like Philip Morris, which gave legislators nearly $26,000 in trips, meals and other gifts. Other students chose to profile the top recipients. One student offered to write about quirky gifts, like WWF tickets; another, about legislators who refuse all gifts.

One student decided to investigate the gifts received by Gov. Jim Gilmore, Lt. Gov. John Hager (who presides over the Senate) and Attorney General Mark Earley. As we did with the legislative gifts, the class combined forces to input the gifts records for Virginia's top state officials.

For our individual stories, we each did additional reporting, interviewing officials and tracking down more records. Then we started writing.

The stories were later published at *http://www.people.vcu.edu/~jcsouth/on-the-legel*. During the legislative session, we used the site to report on the daily happenings at the General Assembly.

Continued on page fifteen

Continued from page fourteen: **Gifts**

Learning spreadsheets

As a participant in the project, I learned quite a few skills that I plan to utilize in the future.

First, I learned how to compile data into spreadsheet – and why that's important: All told, we needed to tally up about 1,000 gifts, and you can't do that in your head. Excel helped us "interview" the data so we could discover things that weren't obvious, like the average haul in gifts for male legislators ($912) versus female legislators ($441).

In addition, I got my first exposure to working as part of a large group on a big project. That takes collaboration and teamwork. If any member of the group screwed up – made errors inputting data, for instance – it would undermine all our work. So reliability and attention to detail were critical.

As a participant in the project, I learned quite a few skills that I plan to utilize in the future.

I was honestly surprised at what our class found, and at the public's response. I think people are interested in gifts to politicians for several reasons:

• Many gifts have a cachet that campaign donations don't. A lobbyist may get more bang for the buck with a $100 steak dinner than with a $100, or even $1,000, check to a campaign committee. In Virginia, where $5,000 campaign contributions are pretty common, an interest group could spend less money but make a bigger impression by sending a key legislator on a vacation.

• Gifts often guarantee face time with lawmakers – giving lobbyists an opportunity to state their case for or against legislation. We found that legislators received $26,000 in meals and $24,000 in trips, with lobbyists usually serving as hosts.

• Constituents may view certain gifts more suspiciously than campaign contributions. People might sympathize with politicians' need for money to run for office. But they might think it's grubby for Virginia lawmakers to accept free Washington Redskins tickets, golf equipment and eyeglasses – things ordinary folks must buy.

• The rules governing gifts may be looser than the rules on campaign donations. In Virginia, lawmakers cannot accept campaign contributions during the legislative session – but they can accept gifts any time. (Virginia has no limits on the amount or size of gifts or campaign donations legislators can receive. Steve Calos, executive director of Common Cause of Virginia, describes the state as the "Cayman Islands" of campaign finance.)

• Campaign donations often are online (in Virginia, thanks to the Virginia Public Access Project *www.vpap.org)*. Not so for gifts; those records had been available only on paper in Richmond. Our project made the gifts data available to the public for the first time over the Internet.

Comparing regions

My story compared the regions of Virginia and the amount of money legislators received. To do this, I split the state into five regions: northern, eastern, central, southwest and western. Next, I determined which legislators go where, using Excel to tally my results.

I found that legislators from Eastern Virginia – including Norfolk and Virginia Beach – received the most overall gift money and had the highest average per legislator. They were followed by the legislators from central and Northern Virginia.

That was surprising because densely populated Northern Virginia has the most lawmakers – and a reputation for being highly courted by lobbyists.

From the results of my research, I also created a couple of bar graphs showing legislative gifts by region.

Experts' opinions

Next, I contacted Stephen Medvic, an assistant professor of political science at Old Dominion University in Norfolk, and Nelson Wikstrom, a political science professor at VCU. I told both experts about my findings and asked their opinions on why Eastern Virginia legislators received the most gifts.

Calos, the government watchdog, told me that political donors use gifts and campaign contributions to gain access to politicians. "Money does buy access," he said. "Access is the gateway to influence."

(Jeff South, an associate professor in the VCU School of Mass Communications, contributed to this article.)

Robb Crocker can be reached by e-mail at cobb9crock@msn.com

Other campaign finance stories include:

The Kansas City Star's 1996 investigation found Sen. Bob Dole's presidential campaign collected tens of thousands of dollars in illegal campaign contributions, most of which were engineered by a vice-chairman of his campaign. (Story #13901)

The Chicago Tribune published, "Ante Up: The Springfield Money Game," in 1997. Using a database detailing more than 100,000 campaign contributions and expenditures in 1995 and 1996, *The Tribune* painted a portrait of how Republican and Democratic leaders of the Illinois House and Senate used campaign cash to consolidate their power and manage the legislative process. (Story #14240)

Two useful datasets

By Raquel Dixon
IRE and NICAR

Each year, the U.S. Forest Service reports about 14,000 forest fires. Perhaps more startling, half-a-million arsons occur annually as well. These fires damage property, cost firefighters time and taxpayers money, and can sometimes kill.

The government is trying to understand how, why and when these fires occur. So they keep these incidents as records in databases. You can use the databases to find out if your city's firefighters are up to speed, if your state has to pay year after year to restore damage caused by forest fires or even who is to blame for the fire.

The National Interagency Fire Management Integrated Database (NIFMID) stores data on wildland fire occurrence and fire weather observations.

There are two databases that can help answer some of these questions and lead to stories that reveal the facts about your city and its fire protection.

Forest fires

First, the forest-fire data. It's important to understand that only fires that occur within national forests are included.

The National Interagency Fire Management Integrated Database (NIFMID) stores data on wildland fire occurrence and fire weather observations. Weather data is collected from federal, state, and local fire-management agencies from throughout the United States. Fire-occurrence data includes Forest Service fire reports from 1970 to the present.

With the weather data, observations are received hourly for approximately 800 stations. Another 800 or so stations observe weather once per day. The fire-occurrence data is updated periodically throughout the year. "Officially" the fire data is loaded "within 10-days of being declared out," but more often it is entered after the local fire season. So the agency tries to get all the data for the calendar year entered by the end of February, so it can run year-end reports.

All fires have an entry in the fire-occurrences table. Other tables offer more information about the events. However, not every fire in the occurrences table will be included in the report tables because filling out this information is optional. Most fires do have entries in the fire-resources-used, fire-events, and fire-causes tables. Other tables handle information generally only collected on large fires or for special purposes.

These tables join by the fire_id, which is unique to each fire. Some interesting information contained in fields is: land ownership, acres burned, agency responsible for reporting fire and working the fire, the cause of the fire, the intensity of the fire and the vegetation that it killed. All this information comes from fields that must be filled out.

Arson data

There is some incredibly detailed data found within three different tables, but there are some holes. Only 14,000 departments, in 42 states, out of 30,000 report. Also, the data is current only through 1998.

The data helps answer questions like: How many fires occurred on Fridays, Saturdays, etc.? How many fires occurred each hour of the day or month of the year? What was the average response time? How much did response time vary by station location? What was the average amount of time spent at the fire scene and how much did that time vary depending on the type of fire?

For more information, go to *www.usfa.fema.gov/pdf/fdah.pdf*

Obtaining the information may be costly. The Federal Fire Prevention and Control Act of 1974 (P.L. 93-498) authorizes the National Fire Data Center in the United States Fire Administration (USFA) to gather and analyze information on the magnitude of the nation's fire problem, as well as its detailed characteristics and trends. The system that gathers the data is called the National Fire Incident Re-

Continued on page seventeen

Arson information:

FOIA information is available at *www.fema.gov/library/foia01.htm*. Contact: Brad Pabody 301-447-1340 or by e-mail at *brad.pabody@fema.gov*

Forest-fire contacts:

The Department of Fire & Aviation Management within the USDA Forest Service maintains this data. Contact: Mike Barrowcliff within the National Interagency Fire Center by e-mail *mbarrowcliff@fs.fed.us*. The FOIA office can be reached at 202-205-1089.

A utility called KCFAST that retrieves NIFMID data in certain data formats allows access to updated data. The great benefit to this program is that you can extract information, whenever and however you want it. The costs depend on the amount of usage and data requested. To use KCFAST and access NIFMID you would need a NITC-logon ID. You can access the KC Fast User Guide at *www.fs.fed.us/fire/planning/nist*.

Continued from page sixteen: **Datasets**

porting System (NFIRS).

The NFIRS Master, Incident and Equipment/Casualty tables can be ordered from the National Technical Information Service on magnetic tape (9-track, EBCDIC character set) and CD-ROM.

File descriptions and prices are: Master File, which includes fire department identification, report of submitted incidents, incident, mobile property, hazardous materials incident, equipment, civilian casualty, and fire service casualty record types. Cost: $454. Incident File, which includes fire department identification, submitted incidents and basic incident records. Cost: $454. Equipment/Casualty File, which includes fire department identification, mobile property, equipment, civilian casualty, and fire service casualty records. Cost: $305.

CLASSROOM

Teaching CAR

By Jeff South
Virginia Commonwealth University

At a computer-assisted reporting conference several years ago, Stephen Miller, a technology expert at *The New York Times*, said he took an under-the-radar approach to CAR training in his newsroom.

Talking to reporters, he found out what tasks they were doing by hand – like calculating how a government budget had changed from one year to the next. Then Miller let slip that it would be easier and faster to do such computations with a spreadsheet, and would the reporter like to know how?

Miller called this approach "stealth CAR," and I have found it's a good model for classrooms as well as newsrooms.

At VCU

Virginia Commonwealth University doesn't offer stand-alone CAR courses. Instead, faculty members integrate CAR into existing newswriting and reporting courses. That approach has certain advantages: It signals to students that CAR should be part of everyday journalism – not an esoteric specialty.

Here are some tips on doing "stealth CAR" in J-schools:

• Begin with paper. Before crunching census data with Excel, for example, give students a sheet of paper with a few columns of population numbers, and have them compute the percentage change by hand or with a calculator. This will help students understand the data – and visualize what they want the computer to do.

• Keep it simple. Don't start with an outer join involving three tables and a nested query. Start with sorting data in a spreadsheet and doing simple calculations. It may be better to teach a kludge in a familiar program than to teach an elegant solution with more complicated software.

• Keep it real. Use real data – crime, traffic deaths and other stats – available online from government agencies. Taking advantage of student-power, obtain paper records (like health inspection reports for restaurants around campus), and have your class imput them into a computer.

• Relate CAR to other reporting skills. Describe CAR as another way of interviewing – but interviewing data instead of people. Your goal is to find bottom-line trends and then examples that illustrate the trends.

• Preview, present, review, practice. Tell students what you are going to teach them; tell them what they are learning as you teach them; then tell them what you just taught them. Have students practice a particular skill, like the AutoSum button and other bottom-line functions, on a variety of datasets.

• Emphasize people, not numbers. Reporters are in the business of telling stories, and data alone do not make a story. Discuss how to humanize the trends you discover with CAR. Then have students shut down their computers and go interview people.

Jeff South can be reached by e-mail at jcsouth@vcu.edu

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 associate editors Amy Sherrill and Mike Sherry. Amy may be reached at *amys@ire.org* and Mike can be reached at *mikes@nicar.org*. Both can be reached by phone at (573) 884-7711

The story, "Cycle of Influence" by Scott Finn is available from the IRE Resource Center (Story #17208). The story received an IRE Medal this year in the student category.

Scott Finn also reported on conflicts of interest among Missouri lawmakers in a Columbia Missourian story in May 2000. (Story #16567)

Money and politics

By Scott Finn
Charleston Gazette

I have a friend whose first newspaper job was covering West Virginia's Legislature for a conservative, family-owned newspaper. He sat down with the editor at the beginning of the session to talk about topics he'd like to cover, such as Medicaid funding and health care reform.

The editor soon cut him off, and said to go to the legislators with one question: "Why aren't you doing the will of the people?"

At first, I thought this was a ridiculous question, but the more I dug into the inner workings of our Legislature, the less paranoid this question seemed. Two examples:

• Even though a majority of West Virginians opposes any expansion of gambling, a bill legalizing coin-drop slot machines passed the Legislature.

• Even though a majority of state citizens supported a smokeless tobacco tax, it died in a Senate committee year after year.

Money influence

One obvious reason the Legislature didn't respond to what voters wanted was the influence of special interest money over the legislative process. But that's a vague answer. How can voters picture something like "influence" in their minds? I wanted to show how the money worked, and tell it to readers in a way that wouldn't make them sleepy.

I spent a lot of time stumbling around before I finally figured out how to tell the story of how money warps the legislative process. I completed the series "Cycle of Influence" for the *Charleston Gazette* as my master's degree project from the University of Missouri. I was new to reporting, and what I knew about the Legislature came from what I read in the newspaper and heard from friends.

I was fresh out of two Missouri classes – investigative reporting and a computer-assisted reporting "boot camp." I wanted to incorporate what I had learned into the series.

I read through the IRE archives and found a lot of stories about money's influence in state legislatures. The stories tended to fall into three categories: campaign contributions, lobbyist spending, and personal conflicts of interest.

In West Virginia, different groups already had computer databases that tracked all three of these types of money in our Legislature:

• Campaign Contributions: A non-profit group called the People's Election Reform Coalition (PERC) had campaign contributions to state legislators going back a decade. PERC includes the state chapter of Common Cause, the West Virginia Citizens' Action Group, and other reform-minded groups.

I had no trouble getting access to any data. [An ethics commission worker] said no one had ever asked for her databases before.

• Lobbyist spending: Two databases existed at the state Ethics Commission -- one of lobbyists and their registration, the other of lobbyist spending on meals, gifts, and campaign contributions.

• Conflicts of Interest: Fortunately, the 50 States Project of the Center for Public Integrity had just completed its nationwide searchable database of every personal disclosure form of every state legislator in America *(www.50statesonline.org)*.

The campaign contributions database was in Dbase and Microsoft Excel, and both translated easily into Microsoft Access, which I used to analyze the data. The lobbyist databases were in Paradox, but also were translated into Access.

I had no trouble getting access to any data. Lucy Suchy at the state Ethics Commission even gave me free disks. She said no one had ever asked for her databases before.

Continued on page nineteen

Continued from page eighteen: **Money**

Once I began to analyze these databases, I saw some disturbing trends. For example, gambling and tobacco industry donations were increasing – more than 200 percent for gambling and 800 percent for tobacco in just two years.

I struggled with how to tell these stories so readers would understand and not get bogged down in numbers. The best stories in the IRE archives tied one of these money issues into a "real" issue people cared about – like a bill about filling in wetlands, or locating a toxic dump near a low-income community.

Gambling and tobacco

I focused on two gambling and tobacco bills because of the disconnect between public opinion and what the Legislature actually did.

The Gazette's writing coach, Kate Long, helped me think of ways to show readers how things worked instead of telling them. I picked people on the losing side of the gambling and tobacco bills and reconstructed their experiences in fighting against the bills.

One anti-gambling activist, a minister, provided some great examples of how lobbyist money buys influence. A gambling lobbyist hangs his coat in a legislator's office and walks in without invitation, while the minister has to wait in the hall. Teams of lobbyists with cell phones tag- team different hearings and meetings with legislators, while the minister wears lace-up shoes instead of loafers so he literally can run from the House to Senate chambers during certain critical hearings.

The computer-assisted analysis gave flesh to the anecdotes. Pro-gambling lobbyists outspent anti-spending lobbyists by a ratio of 66-1. Legislators who voted to legalize slot machines in the state on average received more than four times as much in campaign contributions from the gambling industry as those who voted against the bill.

In a story about the failure of a smokeless tobacco tax, an anti-smoking activist told the story of her father's cancer death from chewing tobacco.

She then talked about her sense of betrayal as a legislator who knew both her and her father voted to keep the smokeless tobacco tax bottled up in the Senate Finance Committee.

The CAR analysis showed this committee had six of the top ten recipients of tobacco industry contributions in 1998. The committee chairman has a personal financial interest in several convenience stores that sell tobacco products.

Impact

Since the series ran, the Legislature passed a smokeless tobacco tax, after more than a decade of defeat. I would like to think the series helped put pressure on legislators to pass the bill.

The gambling industry suffered its first major defeat in a decade last fall when voters in Greenbrier County defeated a measure that would have allowed casino gambling at The Greenbrier, a four-star resort.

The results of "Cycle of Influence" and follow-up stories were part of the debate over gambling's increasing influence in the state.

I have continued to follow the gambling industry and its influence on politics in West Virginia. In October, I published a story showing that contrary to popular belief, the Republican governor who publicly opposed gambling quietly was raking in tens of thousands from the gambling industry in campaign contributions. He lost in a close election.

But gambling forces did help pass a bill to increase the minimum slot machine bet from $2 to $5 in this legislative session, showing their increasing clout among legislators, if not voters. Another bill passed that legalized, taxed and limited the number of video poker machines. During that debate on video poker, an anti-gambling legislator was almost censured by the House of Delegates for what he said during a debate: that there is a correlation between gambling contributions and how legislators voted.

As long as the Legislature ignores the will of the people, the *Gazette* will continue to show how money warps the political process.

Scott Finn can be reached by e-mail at sfinn@wvgazette.com

The Web site of IRE Compaign Finance Information Center (CFIC) at– *www.campaignfinance.org/*, provides information and stories about compaign finance databases.

To search for stories using compaign finance data, go to the Web page– *www.campaignfinance.org/stories.html*, and the stories can all be ordered from the IRE Resource Center.

The IRE Compaign Finance Information Center is building a list of sources specializing in campaign finance at– *www.campaignfinance.org/sources.html*

The tipsheets from reporters who have used campaign finance records or built contribution databases for their election coverage can be found at– *http://notes.ire.org/ireresources.nsf/webCFICtipsheets?OpenView*

Tipsheets that are helpful in reporting about education report cards can be found at the IRE Resource Center:

Tipsheet #890, by Jay Reeves of the *Associated Press*, provides basic ideas and suggestions for building solid education stories and projects. Reeves suggests obtaining state education laws, going beyond statistics and planning your project around the school calendar.

Continued from page one: **Myth**

ting all of this for free. Most of the datasets had fewer than 2,000 records, one for every school in the state. The datasets were:

• Two years of school-by-school test scores from our two states, Oregon and Washington. These databases, generally provided in an Excel or .dbf format, have become routine state education department output, and in the case of Washington, we we able to get a full multi-year dataset in a few hours. For our analysis, we looked at the percentage of students in each of several performance ranges set by the state – passing, excelling, nearly passing, and so on. But, lesson learned: Be sure and ask for the school mean scores too. The average scale score is a boring number your readers may not care about, but it's the best single mathematical summary of every student's achievement.

We gathered four key datasets, only one of which required a big battle.

• SAT scores for our state broken down by family income, parent-education levels and other demographic factors. This is readily available for the nation and every state at *www.collegeboard.org/sat/cbsenior/html/stat00c.html*. The College Board doesn't go out of its way to say which states get the best results with, say, the children of college-educated parents. The board provides the data in a .pdf format, not a spreadsheet. But you can download and compare the state data sets to find out. We were able to show that Oregon's seemingly sky-high results come from advantaged families, not spectacular schools.

• Scores on the National Assessment of Educational Progress, the only achievement test given in every state, again broken down by parent education levels, by rural-urban-suburban schools and by other demographic factors. The federal government is updating, and ostensibly improving, how it provides that data on line. So far, there is the same .pdf format challenge as with College Board data. You can get started at *nces.ed.gov/nationsreportcard/*

• Family background statistics for every school in our two states. In Washington, we could get only the most typical figure: the percentage of students qualifying for free-and reduced-price lunches. This is a standard educational statistic, widely used by schools and education agencies, and it's worth having. But it wasn't nearly as good as what we were able to get in Oregon (after a battle): Statistics about the education levels of parents in our middle and high schools. We also were able to get statistics about the mobility rate, the attendance rate, the percent exempted from the test and other demographics about each school.

A fight

We had to fight to get the parental education data, because Oregon Department of Education officials were worried about violating families' privacy... so much so that they had not ever reported the information back to the schools, even though the schools were required to help collect it. In the end, our open records arguments prevailed and we got the statistics for every school. We were able to tell schools more than they knew about their own demographic makeup and how it affected their performance. It's certainly worth inquiring whether your state collects any other demographic information about all students or, better, about specific test-takers (such as their status as English language learners, parent-education levels, race, free-lunch or Title 1 status, etc.). New federal rules requiring states to report their scores disaggregated by most of these factors should shake free a lot of new data like this in the next few years.

Using SAT and NAEP scores, we were easily able to show that the overall achievement results for Oregon and Washington schools are resoundingly ordinary. That is particularly true for our largest group of students: Middle-class and upper-middle-class kids from families with at least some college education. Because the state's average scores are generally viewed as good to very good when compared to other states, that alone was news. Oregon is overwhelmingly white, so we did not focus on racial breakdowns, but that would be an invaluable, and easy to conduct, addition in many media markets.

The heart of our analysis was the regression analysis: taking every Oregon and Washington school's achievement test average, weighing that against the school's family demographic profile and

Continued on page twenty-two

Continued from page one: **Wrinkle**

At that point, Frazier linked up with the *Star-Telegram*'s enterprise team to dig deeper into the story.

Before the year was out, the coach had moved on to work for another school district that was unaware of the allegations against him. And a six-month investigation by the *Star-Telegram* had found that convicted child molesters, rapists, drug users and thieves – in addition to the coach – had remained certified as Texas teachers, sometimes for years, while their credentials were under investigation.

State and school officials knew the teachers had problems – even convictions – but could not act quickly enough to keep them away from children.

The story was a new wrinkle on the standard teacher-as-felon story. In this case, state and school officials knew the teachers had problems – even convictions – but could not act quickly enough to keep them away from children.

Such information should be available routinely in other states, if you know where to look. In Texas, the information was stored at a relatively obscure government agency, and no one had ever asked for the data.

The records were fairly easy to get. State officials were eager to comply with the *Star-Telegram*'s request for information because they had been trying unsuccessfully to win additional funding from the Texas Legislature. They felt that a story detailing the problems might help their case.

An Excel spreadsheet was e-mailed by the agency to the newspaper, at no cost. It contained 9,130 records and 28 fields, or about 1.6 megabytes of information on all cases handled by the agency since it took over teacher licensing in 1995. The spreadsheet was moved into an Access table for analysis.

The data contained information on both closed and open cases, including the date the case was opened, date closed, name of the teacher, the nature of the allegation and how it had been resolved.

First blush

A preliminary look at the data confirmed what the agency officials had said. Some cases had languished as long as 11 years before being resolved, and others had been pending for years.

The preliminary look also revealed a problem with the data. Although many of the fields were coded, no code-translation table had been included. In a series of phone calls to the agency, it became clear that there was no uniform coding system; the input clerks filled in the records as they saw fit, and the codes changed when the clerks changed. The agency was able to provide a general summary of what the various codes meant, but it was by no means comprehensive. I ultimately ended up standardizing the coding system myself.

Using Access, I then looked not just at how long the cases took to resolve, but what kinds of allegations were pending against teachers and who those teachers were.

The closed cases were separated from the open ones, and separate analyses were done on each set.

On the closed cases, the analyses included looks at how long the agency took to close cases on average, what percent of cases resulted in punitive action against the teacher, what percent of teachers were cleared of wrongdoing and what types of allegations were involved.

On the open cases, the analysis was aimed at finding how many teachers under investigation were convicted criminals and how many were still teaching. Because the SBEC data did not include birth dates or other identifiers except name, three other databases were needed to make a link. These were: a statewide roster of teachers working the previous year, which included a year of birth but not a full birth date; a statewide database of criminal records, which the newspaper already had obtained; and a smaller database of registered sex offenders, also already obtained by the *Star-Telegram*. The investigation was hampered by the lack of a current roster of school employees statewide.

Shoe-leather reporting

Once the data were analyzed, more traditional reporting methods were used for the most egregious criminals – sex offenders, child molesters, rapists and drug dealers.

Continued on page twenty-two

For a copy of Dianna Hunt's story, "Tardy Oversight," ask for story # 17451 at the IRE Resource Center

Another education story that can be found at IRE is the *New York Daily News*' 1999 investigation of how the New York School System failed to sufficiently educate elementary school children to their appropriate level. A computer-assisted reporting project revealed a correlation between low student test scores and high percentages of uncertified teachers in the paper's four-part series, "Everyone Let These Kids Down: The Definitive Story of How Five Children Made It Through Third Grade and Still Can't Read." (Story #16190)

The Oregonian series, "Why not the best?" is available online at *www.oregonlive.com/special/oregonian/education/*

Continued from page twenty-one: **Wrinkle**

Although the state files on pending cases were closed to us, we were able to obtain information on many open cases by turning to local school districts for their records and to a separate state agency that handled administrative hearings for those cases in which punitive action was recommended.

As is often the case in Texas, state lawmakers were not deeply moved by the agency's plight, or, apparently, its impact on children.

Armed with a list of registered sex offenders whose names matched those of teachers accused of misconduct, I called district attorneys across the state to confirm that their convicts were the same teachers now under investigation. Many prosecutors were shocked to learn the criminals they convicted were still licensed teachers.

Eventually, we were able to document that the backlog at the state agency had allowed teachers with serious criminal convictions to move to other districts or to move out of state and continue teaching. Several, we found, molested other children before the state could stop them.

Agency officials did not flinch at the findings. "Do bad apples slip through? Sure," said the agency's head of investigations. "We've got to be able to address cases on a more expeditious basis."

As is often the case in Texas, state lawmakers were not deeply moved by the agency's plight, or, apparently, its impact on children. "It's a problem all over," said a state senator who is now the acting lieutenant governor.. "We've got 30 or 40 agencies in Texas that oversee various professions, from architects to podiatrists, and virtually every one of them says they don't have enough staff to investigate or prosecute."

The stories ran over two days, with the main analysis running on Sunday, Jan. 14, 2001. On Monday, Jan. 15, the package featured a local assistant middle-school principal who preyed on female students and was ultimately convicted of indecency with a child.

The state agency, meanwhile, is back before the Texas Legislature seeking additional funding to pay for more investigators. It remains to be seen whether they will get it.

Dianna Hunt can be reached by e-mail at dlhunt@star-telegram.com

Continued from page twenty: **Myth**

looking at the residuals for every school.

Based on advice from education measurement experts, we used two years of test scores, not just one, and we used both reading and math achievement, not one or the other, to create the achievement index at each school. We also threw out any schools that didn't serve a cross-section of students from their attendance zone – magnet schools, alternative schools and the like.

We had to match records from different data sets (parent education, free lunch figures and test scores all were in different databases, for example) and from two different years. Most could be matched on a school ID number, with hand-matching required for a small subset of schools that opened, closed, changed names, changed grade configurations or were subject to data entry typos between the two years or data sets. We also had to customize our academic index and family demographics for schools that had unusual grade configurations.

The correlations we found between a school's family background profile and the school's average test scores ranged from high to very high. Our r-squared for Oregon's elementary schools was 0.59. That showed that about 59 percent of all the variation in schools' average test scores could be tied to the percent of students qualified for free lunch – the only family background information we could use for elementary schools, because elementary pupils are not asked to report their parents' education levels (and are not trustworthy reporters of that information when asked, according to a persuasive RAND analysis).

Betsy Hammond can be reached by e-mail at betsyhammond@news.oregonian.com

Information manager

By Mark Schaver
The Courier-Journal

A frequent question on journalism discussion groups is: What software will organize the flood of interviews, phone numbers, e-mail, documents, Web pages and other information I'm inundated with every day?

Some use the free-form database askSam. Others like Info Select. Still others just dump everything into Word files.

Tool of choice

For years, my tool of choice has been an obscure but powerful shareware program called Zoot.

Zoot's lone programmer, Tom Davis, whose nickname is "The Admiral," calls it an "information processor." Zoot makes it easy to vacuum up large quantities of information from a variety of sources, organize it automatically, and then find it fast days, weeks or years later.

Davis, who taught himself to write code, said he felt compelled to develop Zoot because the first computer program he used to gather information, Microsoft's cardfile, was so inept, and there was nothing else that served his needs.

"Information management is the only reason I use a computer," Davis said in an e-mail interview.

Zoot useful everywhere

Every day of my working life is spent in Zoot.

I use Zoot to keep my task list and my phone contacts. It's closely integrated with Microsoft Outlook, and information kept in one program can be synchronized with the other. You can also synchronize with Outlook's calendar, notes, journal and e-mail functions, although I personally don't use Zoot for those.

Zoot places a small Z – called the Zooter – on the Window's title bar. The Zooter gives you access to Zoot from any open application on your desktop. You can then rapidly grab text from other programs – either by selecting text or grabbing a whole page – and dump it in Zoot.

You can choose to either automatically add what you gather to a single folder within Zoot as you cut and paste while Web surfing, or you can open a window that allows you to assign your information to multiple folders, categorize it by subject or type notes. Folders can be further divided into separate "databases," which can then be grouped together as "projects."

You can also use Zoot in reverse, taking text already stored in Zoot, and inserting it into a document while using another program. This is great for adding boilerplate text to Freedom of Information Act requests and the like.

Zoot keeps track of every Internet site you visit. To return to a site takes one click on an Internet address in Zoot, which will open up your browser and return you to that site. Double click on an e-mail address, and it will open up your e-mail program.

I use Zoot to gather Web sites useful to reporters, quickly typing in descriptions with the Zooter. Later I import them into an Access database, which can be searched over our newsroom Intranet.

But Zoot's inner beauty is that you can customize it in limitless ways.

You can use what Zoot calls "rules" and "actions" to organize text once it's in Zoot.

Zoot makes it easy to vacuum up large quantities of information from a variety of sources, organize it automatically, and then find it fast days, weeks or years later.

If you are working on a long project, you can set up Zoot to automatically sort any notes, articles or other information you gather using what Zoot calls "smart folders." For example, you can create a folder for every major figure in an investigation, and any information you gathered would instantly be placed in the appropriate person's folder.

Tracking court cases

When I was a federal courts reporter, I used our court's Web version of PACER and Zoot to grab the court dockets for cases I was following. To check on whether there was any new action on a case, I would just click on the item in Zoot and it would return me to the case docket, where I could see if there was any new activity.

And I configured Zoot to automatically sort cases by prosecutor and judge, so I could see at a glance who was handling what.

Continued on page twenty-four

These tipsheets are available at the IRE Resource Center.

Tipsheet #421, "Personal Information Management," discussing how word processors, spreadsheets, database programs and personal information managers could be used to ease your load. The audio tape is available through Sound Images, Inc., (303)649-1811, for $10. Ask for tape #CAR95-18.

There are several tipsheets available about how to make the best of new technologies and publishing data on the Web.

Tipsheet #336 gives advice on how to computerize newsrooms reluctant about using computer-assisted reporting or unfamiliar with the technology. Included is information on computer systems and software at three different price levels and useful software programs.

Tipsheet #905, "Three Ways to Build Compelling Web Databases" explains that Web databases aren't CAR databases, and that you need different software to deliver them.

Tipsheet #903 provides advice for publishing data, tables, and queries on the Web.

Continued from page twenty-three:

Information

Each nugget of information in Zoot is called an item. Keyboard shortcuts make it easy to quickly read through them. You can also create links between items, so clicking on a button will take you to other, related items. It's also easy to create new items, combine them, or drag text from one item to create another. You can also configure Zoot to automatically change an item's color based on text contained within, such as turning all items with the word "urgent" in them red.

Searching tools

Zoot has multiple, fast search tools. A "quick query" box lets you type keywords and instantly see any items containing them. You can filter or search items by numerous criteria, such as the date it was created, the last date it was modified or its subject. A more advanced search will show you the first few lines of each match. There are also a number of buttons and shortcuts that make it easy to navigate through the program.

Zoot can be obtained at *http://www.zootsoftware.com* and costs $99.

That buys you the right to install it on both your home and work PCs, which is good because another Zoot feature is the ability to automatically synchronize information between your home and work PCs by e-mail or "sneaker net" – nerd slang for floppy disk.

Not always perfect

Zoot is not perfect. For the uninitiated, Zoot can be difficult to learn, and documentation for the latest version, 4.0, a radical redesign from previous versions, is incomplete at this writing, although Davis says he is working on it and will be finished soon.

Item sizes are limited, so large documents are split into multiple items that are given sequential numbers. You can print Zoot items to a printer or a file, or create a simple Web page out of them, but formatting options are few. Recently David added the option to print through Internet Explorer or Outlook.

Zoot also makes a poor substitute for a word processor because there is no text formatting. Zoot also doesn't handle graphic files, although you can configure it to use an external file viewer. You can also drag an Adobe Acrobat, PaperPort or similar file into Zoot, which will automatically create a link to them on your hard drive. You can then type notes and click on the file link, which will call up the associated program.

Zoot has been remarkably stable on my Windows 2000 and Windows 98 machines. But as you would expect from any program this complicated developed by one man, there are small bugs.

You can also configure Zoot to automatically change an item's color based on text contained within, such as turning all items with the word "urgent" in them red.

Nevertheless, Davis is exceptionally responsive and fixes them quickly. He is constantly tweaking the program, and often makes changes requested by users. He also maintains a Zooters mailing list to get feedback and post information about Zoot.

Cult-like following

Davis doesn't advertise Zoot, but he has developed a cult-like following. One of his biggest boosters has been the writer James Fallows, who has touted Zoot in *The Atlantic Monthly* and the *Industry Standard*.

Fallows was one of many participants in a forum on CompuServe for the now-defunct program Agenda, another information manager whose loyalists turned to Zoot after it was abandoned by Lotus in the early 1990s. On that forum Davis picked up his nickname as The Admiral, but he can't explain why, except that it has something to do with his favorite book, A Confederacy of Dunces.

Zoot is named after the 1976 Frank Zappa album, Zoot Allures, the meaning of which remains equally obscure.

Some fans say Zappa's album paid homage to the Zoot suit. Others say it is derivative of the French expression, Zut Alors, which translates as a mild exclamation like "Dammit!" – which just happens to be the typical expletive of reporters overwhelmed with information on deadline.

Mark Schaver can be reached by e-mail at mschaver@courier-journal.com

Internet health data

By Devon Hammes
IRE and NICAR

The Internet is host to many different options when looking for information quickly, and one of those options is health data.

The Health Care Financing Administration (HCFA) has several public-use files that can be helpful for looking into Medicare or useful supplements to other health care articles.

There is a catalogue of all public data Medicare provides available for downloading at *www.hcfa.gov*. The catalogue includes information on the different types of data available, pricing and how the data is available.

Several of the datasets would be a great tool for the general public and reporters alike to compare service, statistics and overall operations of Medicare. Of course, you can't write a story on data alone, but some Web sites can be a great starting point and can be used as tipsheets to prepare for interviews.

Some Web sites can be a great starting point and can be used as tipsheets to prepare for interviews.

Some of the HCFA data includes information about providers, cost limits and payment rates. The data can be accessed at *www.hcfa.gov/stats/pufiles.htm*. Some of the provider data includes the End Stage Renal Disease (ESRD) Renal Provider File, which contains information about Medicare-approved providers who supply kidney dialysis and/or kidney transplant services. The data includes the location of the providers and the range of renal services available at those providers.

HCFA also offers the (ESRD) Renal Facility Survey. This data is collected annually by HCFA from all facilities certified to provide Medicare-covered renal dialysis and transplantation. The survey includes the entire U.S. and lists the facilities by zip code. Each record contains information about the facility and the number of patients that have been treated, the number of dialysis treatments provided and the number of kidney transplants performed. The data includes services to both Medicare and non-Medicare patients.

The Medicare home page at *www.medicare.gov* provides a lookup function that will compare nursing homes. It is possible to search for a specific Medicare- and Medicaid-certified nursing home in the U.S. You can compare nursing homes in an area or find information on a specific one.

The database contains information like the percent of residents with behavioral symptoms and other information about the specific nursing home. There is also an option of comparing it to the average of residents with behavioral symptoms in the U.S. and in the state where the nursing home is located.

This look-up page also lets you view recent inspection results, with a column that measures the level of harm the deficiency could have caused. This table breaks it down into categories such as: mistreatment deficiencies, quality care deficiencies and resident assessment deficiencies.

The future

HCFA has implemented new regulations for all home healthcare agencies. The agencies will have to obtain information from all adult, non-maternity patients and submit the data to HCFA. Home-health agencies use a standard core assessment dataset, the Outcome and Assessment Information Set (OASIS). OASIS contains reports on all home-health agency patients, and HCFA uses it to monitor home healthcare quality and to determine payments for individual patients.

HCFA is still debating whether to release the data to the public. It would offer valuable information on how home-health agencies measure up. This data would be a great tool to compare service and statistics. The agencies, however, are also required to ensure the privacy of many pieces of information kept within this data, so many of the fields will be redacted to ensure the anonymity of the patients.

Devon Hammes can be reached by e-mail at devon@ire.org

The tipsheets for reporting on health can be searched at the Web page of IRE's -- *http://www.ire.org/resourcecenter/initial-search-tipsheets.html*, and can be ordered at the IRE Resource Center:

Tipsheet #1313 lists a number of Web sites that have useful health care data.

Tipsheet #1237 contains a list of Web sites to help reporters get information about the backgrounds of doctors who practice in the "vanity and anti-aging medicine" field.

Tipsheet #1236 contains a list of information helpful for investigating the health care system, such as court data.

Tipsheet #1135 is a listing of contacts and Web sites useful for stories on children's health. Sources are from both government and private institutions.

Tipsheet #1314 explains how to obtain data about nursing homes as well as how to interpret the information.

Tipsheet #1219 explains how to cross databases to uncover truths in the health care industry.

CENSUS

Dealing with race

By Erin McCormick
San Francisco Chronicle

Editor's note: This is a tipsheet from the 2001 IRE National Conference.

There is no way to measure exactly how much racial groups changed between 1990 and 2000 because of new categories in the census. The fact that respondents can choose more than one race creates a whole new playing field.

The best hope is to identify clear trends, then convey them in a way that doesn't mislead readers into thinking that it's an exact science.

Here are a few simple techniques for looking at change:

• Compare the percentage of "minorities."

Look at how many people list themselves as something other than non-Hispanic white to get a sense of whether the community has gotten more or less diverse.

Let readers compare the apples and oranges for themselves.

Look at how many people list themselves as something other than non-Hispanic white to get a sense of whether the community has gotten more or less diverse.

Put multiracial respondents in their own category. Then, run pie charts with the new 2000 categories and the old 1990 categories side by side. Readers will see that things have changed, but they will also get a sense of the general trends.

• When studying a single race, look at the range of potential change.

Say you want to look for changes in the American Indian population alone. Look at both the number of people who described themselves as being American Indian "alone" and the larger group of those who describe themselves as being partly or wholly of American Indian heritage. Compare that to the 1990 population. If both comparisons show an increase or a decrease, you have a clear trend. If not, it's impossible to conclude what's really happening.

• Look at Hispanics as if they were a race.

Build 1990 and 2000 pie charts showing Hispanics displayed as a race, alongside non-Hispanics in each race (non-Hispanic whites, non-Hispanic blacks, non-Hispanic Asians, etc.). This works well in areas like the West where most Hispanics are white. If there are many Hispanic blacks, it will result in an undercount of blacks.

Pitfalls to avoid:

Sometimes it will look like the population of a race has decreased when you compare the number of people who describe themselves as being of that race alone to 1990 figures. But, the population will appear to have increased when you compare those who chose that race alone or in combination with another race. This is a statistical tie. Because of the sometimes huge margin of error brought in by the new mixed-race categories, it's impossible to tell what's really going on.

Census tract boundary changes

When comparing 1990 census tracts to 2000, make sure the Census Bureau hasn't quietly rejiggered the tract boundaries. Usually, when the Census Bureau splits or combines tracts, it changes the tract number, so it's easy to tell there has been a change. But, we found many cases where the bureau shaved off a tiny sliver of a tract with no announcement. It often seemed to be a sliver that had a lot of the population in it, making it easy to draw some pretty wild conclusions about unexplained population declines.

Erin McCormick can be reached by e-mail at emccormick@sfchronicle.com

Several university research sites have already done the calculations that allow you to compare segregation in metropolitan areas, including:
University of Albany *mumford1.dyndns.org/cen2000/* or University of California Los Angeles *www.sppsr.ucla.edu/lewis/*

You can learn how to gauge diversity and segregation by building your own indexes at: *cronkite.pp.asu.edu/census/*

Tipsheet #1364 from the IRE Resource Center has more information on the 2000 census racial breakdowns – "Using the New Racial Categories in the 2000 Census" by Sharon M. Lee of Portland State University.

To register, please complete this form.
Visit our Web site at **www.ire.org or call 573-882-2042** for the latest details.
Please write carefully! This information will be used to make your nametag.

REGISTRATION FORM

Name: _____

Employer/Affiliation/School: _____

Address: _____

_____

City, State: _____

Zip Code: _____–_____ E-mail: _____

Phone: _____ Fax: _____

To register, mail this form and a check to IRE, 138 Neff Annex, Missouri School of Journalism, Columbia, MO, 65211. To register by credit card, you must have a Visa or MasterCard. We cannot accept American Express. You may fax your credit card registration to (573) 882-5431 or register online at www.ire.org./training/philly.

Cancellations need to be sent via e-mail to jgreen@ire.org. There is a $50 processing fee for all cancellations until October 10, 2001. Refunds will not be given for cancellations after October 10th.

Card Number: _____ Expiration Date: _____

Card Holder Name: _____

Card Holder Signature: _____

To attend this conference, you must be a current IRE member through 11/1/01. Memberships are non-refundable.

____$150 I'm an IRE professional member and would like to attend the conference October 11-14.

____$100 I'm an IRE student member and would like to attend the conference October 11-14.

____$200 I would like to attend the conference October 11-14 and need to join or renew my U.S. membership.

____$205 I would like to attend the conference October 11-14 and need to join or renew my international membership.

____$125 I would like to attend the conference October 11-14 and need to join or renew my student membership.

____$25 Late fee for registrations postmarked or faxed after September 17th.

_____Total

Bits, Bytes and Barks

NICAR Data Updates

The summer boating season has started and that means it's time for accidents. You can get a better picture of accidents, injuries and the boats in your area by analyzing the U.S. Coast Guard's boating accident and boat registration databases. The NICAR Database Library recently updated both datasets.

The boat registration data is current as of December 2000. The boat accidents data runs from 1969 through 1999, the most current information available from the Coast Guard.

The accidents database includes all recreational boat accidents involving: death, injury requiring more than first aid, vessel damage greater than $500, or the disappearance of a person from the vessel under circumstances that indicate death or injury.

The boat registration database has owner information and other details about currently registered commercial vessels and large recreational vessels (26 net tons or greater). A second table in the database includes all vessels, expired registrations and smaller recreational vessels.

More information about both datasets, including sample tables and record layouts, is available at *www.ire.org/datalibrary/databases*

Prices for the databases are:

Boat accidents. Entire U.S., all years:

- Circulation below 50,000 or market 50-200 (university freelance): $25
- Circulation 50,000 to 100,000 or market 25-50: $50
- Circulation above 100,000 or top 25 market: $75

Boat registration. Entire U.S.:

- Circulation below 50,000 or market 50-200: $40
- Circulation 50,000 to 100,000 or market 25-50: $80
- Circulation above 100,000 or top 25 market: $105

To order, call (573) 884-7332. Or you can download an order form from *www.ire.org/datalibrary/orderform/orderform.pdf*

Math for journalists

The fourth volume in the IRE Beat Book Series is titled "Numbers in the Newsroom: Using Math and Statistics in News." The guide, written by *Washington Post* database editor and former IRE training director Sarah Cohen, focuses on putting numbers into perspective for stories. It's also a great deadline guide.

Other books in the series: "Understanding Crime Statistics," "Covering Aviation Safety," and "Home Mortgage Lending: Detecting Disparities."

Additional information and ordering details can be found at: *www.ire.org/store/books/math.html*

Tipsheets from IRE National Conference in Chicago

Tipsheets from panels at the June IRE conference are available at *www.ire.org/resourcecenter/initial-search-tipsheets.html*

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