

Uplink

November 1998

A newsletter of the National Institute for Computer-Assisted Reporting

UNITED STATES OF AMERICA

Census preparation

By Paul Overberg
USA Today

Almost six months ago, I spoke on a panel at the IRE national conference about the wisdom of starting Census 2000 preparations immediately.

Remarkably, the shape of Census 2000 is even less clear now. That's why it's even more important to start planning now.

Much of the uncertainty stems from a battle that's erupted over the Census Bureau's plan to use sampling to help count the nation.

Republicans oppose it, and they control Congress. Democrats favor it, and they con-

trol the White House.

In Washington shorthand, sampling has been reduced to 'like polling.' It is, but what's planned would be much more sophisticated and complicated. There are actually two forms of sampling planned in the Census 2000 count, further complicating the task of explaining it to our editors and readers.

Sampling snafu

At this writing, two appellate courts have ruled that sampling can't be used in lieu of a count-to-the-last-person approach to produce the state totals that will reapportion the House of Representatives.

Neither court, however, has forbidden the use of sampling to produce totals for other purposes, such as distributing federal funds or redistricting state legislatures. The Supreme Court's final word is due by spring. Congress and the White House have agreed to make peace by mid-June or shut down a big chunk of the government.

The Census Bureau hasn't decided whether or how it would use sampling if it can't use it for everything. The bureau's overseers in the Commerce Department and the White House won't even discuss this prospect because they're still fighting to conduct just one census with built-in sampling.

Two head counts worse than one?

Just like in 1990, it's quite possible the bureau will produce two sets of numbers: one from the count, the other from the sampling-adjusted count. If this happens:

- President Clinton could order federal agencies to use adjusted numbers to dole out \$100 billion a year. He could also order that the adjusted numbers be used to gauge equal opportunity compliance by federal contrac-

Inside Uplink

Endless speculation has been waged on the computer crisis that *may* await as the clock ticks toward 2000, but a limited amount has considered the data treasure chest that *definitely* awaits — Census 2000. By January 2001, when the flood of data will rush forth, computer-assisted reporters need to be prepared.

Paul Overberg of *USA Today* discusses just that, including the politics surrounding sampling, changes in the data collection and release, and the logistics of getting your newsroom ready.

Steve Doig of Arizona State University focuses on the necessities of getting ready, from basic spreadsheet skills and data familiarity to advanced statistical analysis and programming practice.

Neil Reisner of *The Miami Herald* offers a sample of census stories that have been done and press releases that appear on the Census Bureau Web site.

Mike Diamond and John Froomjian of *The Press of Atlantic City* detail their uncovering of a fuel dealer overcharging state agencies.

PAGE 8

On the Internet

PAGE 15

First Ventures

CENSUS ACHIEVEMENT TEST 2000 tuneup

By Steve Doig
Arizona State University

The first release of data from the 2000 Decennial Census is more than two years away, but it's none too soon to start preparing. Here's a checklist from someone who wrestled a lot with the 1990 Census, covering what you need to know and what you should do to get ready.

The basics

Master your spreadsheet skills: The majority of census analysis can be done using a spreadsheet program such as Excel. Make sure you know how to import a text file, build a spreadsheet, write and copy formulas, sort and filter, export your results into a table that can be published, and other such tasks.

Master your math skills: The census is numbers, lots of 'em. Memorize my formula for percentage change: NEW divided by OLD, minus one. (You don't believe it's that easy? Try it.) Make sure

Continued on page five

Continued on page two

Uplink

November 1998
Volume 10, Number 10
A newsletter of the National
Institute for Computer-Assisted
Reporting

EDITOR

Brant Houston
brant@nicar.org

MANAGING EDITOR

Brent Johnson
bjohnson@nicar.org

SENIOR CONTRIBUTING EDITORS

Len Bruzzese
len@ire.org
Sarah Cohen
sarah@nicar.org
Richard Mullins
richard@nicar.org

COPY EDITOR

Jeanine M. Davis

ART DIRECTOR

Wendy Charron

STAFF

Britton Clapp
Dawn Fallik
Jason Grotto
Ben Lesser
Ted Peterson
Noemi Ramirez
Aaron Springer
Neil Warner
Alistair White

Uplink is published every month
by the National Institute for
Computer-Assisted Reporting,
138 Neff Hall Annex
Columbia, MO 65211.

(573) 882-0684. Subscriptions are
\$40 for IRE members, \$60 for
nonmembers.

Postmaster: Please send address
changes to NICAR.

Send email to jeanine@nicar.org

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.

NICAR is supported by grants
from The Freedom Forum and
other foundations intended to
help the institute deliver its
services nationwide to news
organizations and associations.

From page one:

Counting curricula

tors. (Congress, of course, could pass a law against this and try to override a veto.)

- Unless Congress forbids it, some states might use unadjusted numbers and others might use the adjusted numbers. Some might use both for different things. Unless a state forbids it, even adjoining cities or school districts could choose one set or the other as they redraw ward or district lines.

- Businesses will almost certainly use the adjusted numbers, which will be more accurate. The only thing that's certain is confusion and litigation.

Over or under

Much of the publicity over sampling focuses on how it corrects for the undercount, those missed by the mail survey and door-knockers — most frequently racial minorities, the young, the poor and illegal immigrants.

It's basic American fairness and law, at least in concept. But sampling also corrects for overcount, the less-publicized double-counting that happens, most often in affluent communities, when people get counted in two places. They may be snowbirds, college students, commuter spouses, newly divorced or migratory consultants.

So Somerset County, N.J., a fairly wealthy suburban area 30 miles west of Manhattan, would have seen its 1990 population adjusted *down* from 240,279 to 239,522. It's only three-tenths of one percent, but some towns would have seen their totals drop two percent. What politician would let that go unchallenged?

The woe ahead

Even without disputes over sampling, Census 2000 would pose major new challenges to journalists. Two of the biggest:

- **Race and ethnicity.** The Census Bureau reported 1990 totals for four races (plus a tiny 'other'): white, black, American Indian/Eskimo/Aleut and Asian-Pacific Islander. This time, Asian/Pacific Islander will be split into Native Hawaiian/Pacific Islander and Asian.

But unlike previous censuses, people will be allowed to check more than one race — up to all five! So Tiger Woods, who calls himself 'cablinasian,' could check white, black, American Indian and Asian.

Theoretically, he'd be counted as a white-black-American Indian-Asian, not to be confused with a black-American Indian-Asian like his father or a white-Asian like his mother.

I'll save you from the math: There are 63 possible race categories.

Test surveys show that only a tiny percent will check off more than one race, but in many neighborhoods that will be higher. You could collapse race categories into their 1990 counterparts for comparisons. But you'll

Although some agencies and newspapers have begun reporting Hispanic totals as a separate race, they're not counted that way. Race and ethnicity are two separate census questions, so each racial total includes subtotals for Hispanics and for non-Hispanics. In all, there will be 126 possible categories for race and ethnicity!

have to deal with all the data if you want to report the news: the rich new details available on your area's diversity.

It's even more complicated than that. Although some agencies and newspapers have begun reporting Hispanic totals as a separate race, they're not counted that way. Race and ethnicity are two separate census questions, so each racial total includes subtotals for Hispanics and for non-Hispanics. In all, there will be 126 possible categories for race and ethnicity!

The Census Bureau hasn't decided how many it will track and publish. It's seeking guidance from the Office of Management and Budget, which made the new rule.

Continued on page three

From page two:

Competition upgrade

Imagine that the bureau settles on publishing 15 racial categories: five races, nine race combinations and a not-so-tiny 'other.' Times two ethnicities. Times the mandated two age groups – total and voting-age total. That's 60 columns of data. In 1990, the equivalent file had just 25 columns.

As if all that wasn't enough, the Census Bureau has also decided to add 'Latino,' the preferred term in the West, to the identifying terms on the Hispanic question. It has also moved the Hispanic question ahead of the race question on the census form. Tests show this is less confusing and makes people more likely to fill it out. The net effect of all these changes may be surprising and hard to analyze, especially in comparison to the 1990 Census across a decade of heavy Hispanic immigration.

• **Data release.** Who needs local newspapers or TV? The Census Bureau is building a powerful Web engine called American FactFinder to release Census 2000 data. If you want data, this probably will be the first place to grab it.

It will let anyone call up basic summaries or a profile of a town or neighborhood. It will let them do customized cross-tabulations, so that anyone might figure out how census tracts in their city stack up in the percentage of college-educated Hispanics with children younger than six. It would even let them map these statistics.

A working version I tested this summer was still confusing and buggy. But by next spring, the bureau plans to release its 1997 Test Census and 1997 Economic Census data this way.

By January 2001, when Census 2000 data starts streaming out, you might find that two-thirds of your customers own home PCs with Internet connections. They won't be very interested if you grab basic tabulations, do some percentages or sorts, and then publish stories. Even big tables and color maps may seem tame.

And if CNN, MSNBC or some other big news-driven Web site puts some brainpower to work, those customers could find a wealth of insight on their neighborhoods without ever turning to a local media outlet. You'll need better stuff to keep them satisfied.

Start planning now

First, you need a point person. Reporter or editor, it doesn't matter, but he or she must work well on teams and projects; like numbers; know your area well. (And plan to stick around through 2001!)

The point person needs to read as much background as possible. Then he or she should start planning, touching base with everyone in the building who can help or will be affected, especially folks in marketing, advertising and

By January 2001, when Census 2000 data starts streaming out, you might find that two-thirds of your customers own home PCs with Internet connections. They won't be very interested if you grab basic tabulations, do some percentages or sorts, and then publish stories...You'll need better stuff to keep them satisfied.

circulation. What do they want? What will they spend? What can you provide?

It helps to think about three plans: how to cover the run-up, including the controversy; how to get newsroom logistics ready; and finally, how to report results.

Keeping up with the run-up

This may be as easy as running wire briefs. But if your area includes a significant number of:

- Minority residents, you might want to do local stories as the sampling issue plays out.

- Hispanics, you might look at changes in the question wording and placement, which are designed to improve accuracy but may also boost the response of non-Hispanics.

Continued on page four

FOR EXAMPLES OF CENSUS DATA IN STORIES, SCAN THE RESOURCE CENTER'S STORY LIBRARY OF MORE THAN 12,000 INVESTIGATIONS AT WWW.IRE.ORG/RESOURCES/CENTER/SEARCH.HTML. THE CENTER CAN BE REACHED BY E-MAIL AT RESCNTR@IRE.ORG.

EXAMPLES EMPLOYING CENSUS DATA INCLUDE:

- **OSHKOSH NORTHWESTERN'S UNCOVERING OF ERRORS IN METROPOLITAN STATISTICS DATA IN 19 STATES AND THE RAMIFICATIONS FOR AGENICES AND BUSINESSES USING THEM. (#12773)**
- **THE WALL STREET JOURNAL'S ANALYSIS OF DECLINING AND SOON-TO-BE STATISTICALLY IRRELEVANT FARM OWNERSHIP BY BLACKS. (#14735)**
- **THE AUSTIN AMERICAN-STATESMAN'S COMBINATION OF CENSUS DATA WITH AUTO INSURANCE RECORDS SHOWING THE HIGHEST RATES IN PREDOMINANTLY MINORITY AND LOW-INCOME ZIP CODES. (#11065)**

THE WEB SITE FOR THE CAMPAIGN FINANCE INFORMATION CENTER (WWW.CAMPAIGNFINANCE.ORG) IS A VALUABLE RESOURCE FOR ANYONE INVESTIGATING CAMPAIGN FINANCES. IT INCLUDES:

- DOWNLOADABLE DATABASES OF CONTRIBUTION DATA AND LINKS TO SEARCH ENGINES MAINTAINED BY NON-PROFIT ORGANIZATIONS AND STATE BOARDS OF ELECTION
- STORIES FROM PRIOR ISSUES OF TRACKER
- STORIES EXPLORING CAMPAIGN FINANCE AT THE LOCAL, STATE AND FEDERAL LEVELS
- INFORMATION ON THE CFIC-L MAILING LIST
- TIPSHEETS ON COVERING CAMPAIGN FINANCE
- LINKS TO POTENTIAL SOURCES FOR CAMPAIGN FINANCE STORIES AND OTHER RELEVANT SITES

From page three:

Coverage outline

Budget bests

No excuses: We know when this thing is coming and how big it is. Take inventory. See what you can afford in 1999 and start building a list for newsroom budgeting for 2000. Some things to think about:

- Mapping software, plus local address and boundary files.
- A big, fast PC to handle big data files.
- A fast Internet connection to pull data from the Census Bureau's Web site.
- Fast graphics computers to handle big maps and graphics.
- Data from the 1990 Census or other sources.

Broader logistics

- Will we have a Web site by 2001? An Intranet? What could/should we do with customized local census information in each of those channels? What other local data could we integrate/link with census data in those channels?
- If you enhance local Census 2000 data, will you give it away or sell it?
- What access will your local government allow to its working census files, such as audits of Census Bureau address lists, small-area correlations of 1990 and 2000 and their boundaries, and so on?
- Should you seek a local media partner? Should you seek a statewide newspaper or media consortium?

Prize packages

Collect your paper's 1990 census coverage and any big demographic packages since then. Make copies and get staffers, especially editors, to brainstorm. Call in your paper's reader panel. Check with your Washington bureau for its plans. Start a subscription to American Demographics. Think.

It's a little hard to anticipate how much you'll be zoning or sending to your Web site in 2001-2002, but imagine the data will be coming out next year. What would you need to publish to be competitive?

How fast? How detailed? How sophisticated? Lay out an aggressive plan such as 30 Sunday packages over two years plus spot news as data appears. Work backward to see what resources you'd need. Do the same with

a modest plan, such as a dozen Sunday packages over two years plus spot news. Aim high -- it's easier to scale down.

A good way to cover Census 2000 trends is to jump the gun. Demographic trends don't change quickly, so what you see now won't change much by 2001. Cover it now

A good way to cover Census 2000 trends is to jump the gun. Demographic trends don't change quickly, so what you see now won't change much by 2001. Cover it now and you'll be much more sophisticated when you do so in 2001.

and you'll be much more sophisticated when you do so in 2001. There's plenty of current data for counties and metros from the Census Bureau's own county and city estimates; the INS annual immigrant data; the IRS county-to-county migration data; the Bureau of Economic Analysis' county personal income data; local school enrollment, bilingual education and school lunch data; local births and death data; and tax rolls.

You can also start reporting likely impacts of Census 2000 in your community:

- A political reporter can ask how the state legislature might reapportion your area, both for state and congressional districts. It's as much politics as demographics. What happened last time? Who controls the legislature? Who appoints any reapportionment commission? Who owes favors to whom?

- A city reporter can ask how ward boundaries might shift. Who has the power?

- An education reporter can ask what might happen to school district lines.

See Overberg's list of suggested Census Bureau links on page thirteen.

Paul Overberg can be reached by email at poverberg@usatoday.com

From page one:

Decennial duties

you know the difference between mean and median and how to calculate a weighted average. Soon you'll be ready for the Gini coefficient.

Understand census geography: ZIP codes are for delivering mail, not demographics. Learn that blocks tile into block groups, which tile into tracts, which tile into counties, then states and then the nation. Learn about Census Designated Places and MSAs and CMSAs and FIPS codes. Compare the 1990 census maps for your area with the ones for 2000. If tracts have been split or combined, you're going to have some work to do to make sure your numbers are comparable.

Understand census terminology: Well in advance, buy the technical documentation for the census files with which you will be working. Know the difference between a household and a family, between housing unit and group quarters, between industry, occupation and class of worker, between race and ethnicity. Know about summary levels and microdata and tenure and imputation.

Gather 1990 data: Change over time is a big part of most census stories. To measure that change you'll need to have 1990 data (and maybe earlier, if you're ambitious) on hand. If you don't have it already, get a collection of the 1990 Census CD-ROMs for your area. Not only will it give you something to practice building spreadsheets with, it'll be there when you're on deadline and can't get into that overloaded online census data server.

Learn a mapping program: The census is not only numbers, it's numbers that have locations. Geographical Information System software like ArcView or Atlas*GIS or MapInfo is essential for helping you see the stories in the data and for creating interesting graphics for your readers. A good GIS is worth the money and time you'll invest in learning. You'll soon be using it for all sorts of non-census stories, too.

Get local census geography map files: If you have or get a mapping program, you'll need the files that will draw those nice maps of your area. The 1990 census geography files are easy to acquire, either on the Internet or from your mapping software vendor. Start looking for the 2000 map files, which prob-

ably are being developed by your local planning department.

Get to know your local planning department: They'll be deeply involved in planning for the 2000 census. They'll have a library of local demographic studies, good data and map files, and a lot of expertise.

Get familiar with the Census Bureau: Scour its Web site. Get the list of contacts, not the PIO, but other various specialists. Subscribe to email press releases – and read them.

Get familiar with demographic research: Subscribe to American Demographics magazine. Check your local university to find out who specializes in population research. Discover nationally known researchers like William Frey of the Population Studies Center of the University of Michigan (<http://www.frey-demographer.org>)

Advance your coverage

The tips above will get you up to speed for doing most census stories. But if you want to be first with national census stories or beat your local competition, you'll also need to:

Learn to get data off tape: The Census Bureau has ambitious plans to release its data via the Internet, with a system called DADS (Data Access and Dissemination System). But don't count on DADS, or any other online delivery method, for the first release of new data. Nor should you count on getting the data on CD-ROM right away, either. As in 1990, the very first release almost certainly will be on 9-track tape or cartridges. If you aren't set up to work with tape, you won't be in the game. (Note: NICAR can process data off tapes and cartridges if your newsroom is not equipped to.)

License and learn the SAS programming language: Raw census data files are big and complex. The STF3 files for 1990, for instance, have more than 3,300 variables per record. But SAS is made for chewing up files like these, with all sorts of built-in tools that make it easy to import this kind of data and output it into something useful. Better yet, most serious census users in government, private industry and academia use SAS; chances are excellent you can find the necessary programs pre-written. Other data programs like SPSS and FoxPro can be made to

Continued on page thirteen

UNDER A GRANT FROM THE RUSSELL SAGE FOUNDATION, STEVE DOIG IS WRITING A CENSUS HOW-TO TOOLBOOK FOR JOURNALISTS.

SEE OVERBERG'S SUGGESTION OF CENSUS BUREAU LINKS ON PAGE 12 AND HIS EXAMPLES OF CENSUS DEFINITIONS TO KNOW ON PAGE 13.

Surfing data waves

By Jennifer Small
San Antonio Express-News

Why would a reporter want to access the Census on the Web? Easy. You can find data to add depth and context to your daily reporting or download large data sets for projects. You can keep yourself updated on developments at the Census Bureau and generate story ideas. It's a great resource, a site with statistical information ranging from population and housing to minority-owned businesses and international fertility rates.

Consisting of more than 40,000 pages, the Census Web site can be hard to navigate. Familiarize yourself with some of the tools mentioned below and you will be well on your way to utilizing its full potential.

Press releases, alerts and experts

Follow the News link off the main page to access tools the Bureau offers reporters. The Census Bureau posts all news releases in both chronological and subject order. News releases can also be sent directly to your e-mail. See www.census.gov/pubinfo/www/subscriptions.html. Each news release links to appropriate data sets, tables and reports. The press can also view embargoed news releases. Contact the Public Information Office.

If you need the names and phone numbers of Census Bureau employees who specialize in various areas of the Census, link to Ask the Experts at <http://www.census.gov/contacts/www/contacts.html>. Use State Data Centers to find more specifics for your state.

Understanding the data

Finding the information you need on deadline can be a challenge. If you are looking for demographic data, be aware of the two major summary data files: STF1 and STF3. The STF1 data are generated from responses to the more general Census short form. This form is sent to everyone and represents the 100 percent count of the population. The STF3 data are generated from the Census long form, which is sent to a sample of the population. The numbers provided are estimates, but the information tends to be more detailed.

The 1990 Census Lookup Tool <http://www.census.gov/cdrom/lookup> provides quick

access to every variable in the 1990 Census. From the Lookup, choose the appropriate data set — STF3A provides the greatest level of geographic detail.

Some of the most up-to-date yet overlooked data can be found in Census surveys. (See a complete listing at <http://www.census.gov/main/www/surveys.html>. Survey topics range from American Housing, Minority-Owned Business and Wildlife Recreation to Business Owners and Nationwide Truck Activity. The Current Population Survey (CPS) at <http://www.bls.census.gov/cps/cpsmain.htm> is one of the site's most useful surveys. However, because surveys occur more frequently, they tend to lack the geographic specificity of the census. Numbers are typically generated for the nation, states and, sometimes, metropolitan areas.

Quick hits

If you are on deadline, take advantage of the many lookup tools. In addition to the 1990 Census Lookup, tools are available for the Economic Census at <http://www.census.gov/epcd/econ/www/lookups.html>. Search County Business Patterns, the Census of Agriculture and U.S. Imports and Exports History.

As a member of the media, you have access to the great lookup tools in CenStats, located at <http://www.census.gov/apsd/www/censtats.html>. Call the PIO office for a user name and password. My personal favorite is the Census Tract Street Locator. Type in a street address and determine its census tract. Take that tract number back to the 1990 Census Lookup and get more information on a smaller corner of your city.

Use MapStats at <http://www.census.gov/datamap/www/index.html> to access county population data, county business pattern data, and Congressional district information.

The Census tries to provide quick access to the most popular Census variables (population, health insurance, marital status) at Hot Sites for the Media at http://www.census.gov/pubinfo/www/hot_frm.htm. Although you have to navigate through an unwieldy frames interface, this may be the fastest way to view comparative data broken down by race and ethnicity.

Continued on page twelve

SMALL ALSO SUGGESTS
SOME ADVANCED ACCESS
TOOLS (ALSO SEE PAGE 12):

MABLE ([HTTP://WWW.OSEDA.MISSOURI.EDU/PLUE/GEOCORR/](http://WWW.OSEDA.MISSOURI.EDU/PLUE/GEOCORR/)) IS A GEOGRAPHIC CORRESPONDENCE ENGINE. AT ITS MOST BASIC LEVEL, IT CAN BE USED TO BUILD A 'CORRELATION LIST' THAT SHOWS HOW ONE TYPE OF GEOGRAPHY RELATES TO ANOTHER. IF YOU NEED TO DETERMINE WHAT CENSUS TRACTS LIE WITHIN A PARTICULAR ZIP CODE, THIS IS THE TOOL TO USE. IF YOU ARE UNFAMILIAR WITH THE GEOGRAPHIC TERMINOLOGY OF THE CENSUS, CHECK OUT THE GEOGRAPHIC GLOSSARY **MAGGOT** ([HTTP://WWW.OSEDA.MISSOURI.EDU/PLUE/GECORR/HTMLS/MAGGOT.HTML](http://WWW.OSEDA.MISSOURI.EDU/PLUE/GECORR/HTMLS/MAGGOT.HTML)) USING **MABLE**.

Census resources

The following excerpts are from handouts available in the IRE Resource Center.

Quick hits and the long term

By Jeff South

Virginia Commonwealth University

From quick-hit dailies to long-term projects, census data can be used in various ways:

- Every year, the bureau updates population estimates for states, counties and metropolitan areas. These updates are fodder for breaking news and trend pieces.

- The 1990 census and updated estimates are especially valuable when combined with other data, such as the number of crimes, gun permits, pawn shops, supermarkets, public swimming pools, billboards – you name it. Use census data to compute rates, like zoning violations per 1,000 people in different parts of your city. Then use the census to see if the rates are different between low-income and high-income areas or between predominantly minority and predominantly white areas.

- The bureau produces such reference materials as the Statistical Abstract of the United States, the County and City Data Book and USA Counties. These materials include data ranging from health care to the cost of living.

- The bureau continually cranks out studies and statistical briefs that spotlight trends.

- Beyond demographics, the Census Bureau tracks rental housing vacancies, government spending, business development and other trends.

Other demographic data

The Census Bureau is by no means the only source of demographic data on the Internet. The bureau's page has links to other sites, like the University of Missouri <http://www.oseda.missouri.edu>, census data at the Lawrence Berkeley National Laboratory http://www.lib.berkeley.edu/GSSI/uc_cdrom.html and the Consortium for International Earth Science Information Network <http://www.ciesin.org>. Also, check you state government's Web page, as well as sites operated by publications like American Demographics <http://www.demographics.com>, which often posts the data used in its articles.

Develop human sources, too – not just at the Census Bureau but at its regional office

nearest you, and at your state data center. The centers, usually affiliated with a state government agency or university, help distribute and analyze census information. You can find a list of the centers at <http://www.census.gov/fip/pub/sdc/www/>

If you need inspiration for census stories, read *Who We Are: A portrait of America based on the latest U.S. Census*. This 1993 book, by Sam Roberts of *The New York Times*, shows how to put a face on census data.

South's handout, #495 in the Resource Center, was provided when he was with the Austin American-Statesman.

Jeff South can be reached by email at jcsouth@vcu.edu

County population estimates

By D'Vera Cohn

The Washington Post

What makes county population estimates useful is not just the population totals, but underpinnings of those estimates. Included in the "components of change" are the bureau's year-by-year estimates of the number of people over 65, people under 65, births, deaths, immigrants and domestic migrants (which means people who moved to any county from any other county in the U.S.). This is good fodder for stories about the growth of the elderly and immigration populations, ups or downs in births (best to double-check those with actual birth figures, usually compiled by health departments), and the appeal of your region to people from other parts of the country.

Census has recently published 1996 estimates by county of the age, race and Hispanic origin population. Slap them into a spreadsheet and track change since 1990.

The numbers are not as accurate as the Census, because they are estimates from administrative records rather than results of a door-to-door survey. You should be cautious about using them for very small counties. But they sure beat figures that are nearly a decade old.

Another useful data set – and about to get somewhat more useful – is the International Revenue Service County Migration files. These records count the number of taxpayers and dependents who moved from one county to another in any given year and include their total and median income.

Continued on page thirteen

TO SEARCH ONLINE FOR HANDOUTS AVAILABLE FROM THE RESOURCE CENTER, DIRECT YOUR WEB BROWSER TO WWW.IRE.ORG/RESOURCES/CENTER/HANDSEARCH.HTML

CONFERENCE AUDIO

TAPES:

AUDIO TAPES FROM THE MARCH 1998 SESSIONS AT INDIANA CAR ARE NOW AVAILABLE FOR ORDERING FROM SOUND IMAGES, WHICH CAN BE REACHED AT (303) 649-1811. FOR MORE INFORMATION OR A DOWNLOADABLE ORDER FORM, POINT YOUR BROWSER TO

WWW.IRE.ORG/RESOURCES/INICAR/CONFERENCES/INDIANA/AUDIO.HTML

AUDIO TAPES FROM THE JUNE 1998 IRE NATIONAL CONFERENCE IN NEW ORLEANS ARE ALSO AVAILABLE. FOR MORE INFORMATION, POINT YOUR BROWSER TO WWW.IRE.ORG/RESOURCES/CONFERENCES/NEWORLEANS/AUDIO.HTML

Out-of-date data

By Paul D'Ambrosio
Asbury Park (NJ) Press

Note to the next Constitutional Convention: Please make the census an annual event. The Founding Fathers sure didn't have journalists in mind when they came up with the count-everyone-once-a-decade concept, confounding our obsessive need for total accuracy.

Try answering these questions on a municipal (or lower) level: How many immigrants live in town? How many toddlers are there (whose parents may not even have met before the last census)? What is the richest neighborhood in the county?

The 1990 census is almost a decade out of date, making such microdata as age woefully inadequate—if not downright paradoxical—to use: If there were 200 toddlers in town in 1989, does that mean there are none now, but there are lots of 10- to 15-year-olds? There are only two solutions. Ignore the 1990 census data by taking a sabbatical until 2000 or buy census updates from commercial vendors.

Educated guesses

Commercial updates have been the lifeblood of census information at the *Press* for the last five years. But keep in mind that the methodology for calculating estimates differs from one company to another. Most use algorithms based on death and birth lists, change-of-address data from the post office, and even mini-surveys.

In short, what you get with an estimate is a big ol' guess. Maybe a pretty good educated guess, but a guess nonetheless. Given the choice between using the 1990 census or an update, I would choose the update every time, with a note to readers that the information is based on estimates.

I use Claritas, which packages data mainly for businesses and marketers. Claritas comes with a mapping program and is hugely expensive (thousands of dollars for a stand-alone license). You don't have to spend that much money, though. Most newspapers have a marketing department that uses Claritas (or a competitor's product), which a newsroom should tap.

For inexpensive updates at the ZIP code level, there is CensusCD+Maps, which costs about \$300 and comes with its own mapping program (www.Geolytics.com).

In the city of Asbury Park, Claritas estimates the 1996 population at 15,469. The official county estimate is 16,804. The 1990 census says it was 16,799.

Claritas makes estimates down to the block group level, but it does not update every census item. Key items updated are population, race, residency, income and age. De-

The 1990 census is almost a decade out of date, making such microdata as age woefully inadequate – if not downright paradoxical – to use.

tailed business data, such as gross business income and types of businesses, are a welcomed inclusion to the dataset.

Series support

We used the business data in 1995 to show bank branching patterns and how banks shun minority and inner-city areas, even when the area proved to be economically vibrant. Without the business data, no story would have been possible. A series last year on mortgage fraud showed how the scam artists were almost exclusively preying on people in the poorest areas of town.

Recently, we ran a series on crime trends in each of our 80 towns and complimented it with population estimates. One police chief was ecstatic that our estimate showed his town had about 5,000 more people than town leaders thought. The chief wanted a copy of the data so he could bully the council into hiring more police officers.

I told him where we got our data but reiterated that it was just an estimate, depending on who was doing the math. I don't know what he did with the information, but none of us will know if the population estimate was on the mark or off by 5,000. We'll have to wait until 2000 to find out.

Paul D'Ambrosio can be reached at (732) 922-6000 x 4261 or by e-mail at pmd@app.com

UPCOMING TRAINING:

MAY 10-13, 1999 –

IRE AND NICAR AND THE UNIVERSITY OF NORTH CAROLINA SCHOOL OF JOURNALISM WILL HOST AN ADVANCED SEMINAR ON USING STATISTICS IN COMPUTER-ASSISTED REPORTING AT THE UNIVERSITY OF NORTH CAROLINA.

INSTRUCTORS WILL INCLUDE PHILIP MEYER, UNC PROFESSOR AND AUTHOR OF "THE NEW PRECISION JOURNALISM," AND SARAH COHEN, IRE AND NICAR TRAINING DIRECTOR.

AN OPTIONAL SESSION ON THE BASICS OF THE STATISTICAL SOFTWARE SPSS WILL TAKE PLACE SUNDAY, MAY 9 (MOTHER'S DAY).

MAY 16-21, 1999 – BASIC BOOT CAMP IN COLUMBIA, MO.

CALL (573) 882-2042 FOR MORE INFORMATION OR VISIT THE IRE WEB SITE AT WWW.IRE.ORG/RESOURCES/CONFERENCES/TRAINING/

Patterns to purchase

By Ronald Campbell

The Orange County Register

Last Aug. 27, when Wall Street began taking away what it had so kindly given millions of investors, several business reporters and editors at *The Register* sat in a room devising ways to cover the crash. My idea: Find out if Russia, the day's designated villain, really meant anything to the American economy.

Within a couple of hours, I had an answer. Russia was a minor trading partner, scarfing up less than one-half percent of all U.S. exports in 1997.

I found this information at the mother lode of business reporting sites on the Web, the Census Bureau. The census provides invaluable business information for a modest fee. Computer-literate business reporters owe it to themselves to buy "County Business Patterns" and "ZIP Code Business Patterns."

Land of the free

But first let's go window-shopping for freebies. Trolling this site for its variety and valuable data is well worth your time.

Our starting point for the Great Give-away Search is the current economic statistics page www.census.gov/econ/www/. The Census Economic Briefing Room contains a run-down of recent economic reports, together with links to press releases and charts. This is a logical starting point for a reporter commencing a thumb-sucker on the state of the economy.

The biggest limitation in the free data is that most of the data is national or, at best, statewide. For local data, you'll have to open your wallet.

SIC'em

The best single census resource for business writers is "County Business Patterns." The Census sells two years' worth of data, broken down by year and state, on CD for \$150. The data for 1996 should be available sometime in November or early December 1998. Census breaks down employment, payroll and number of business establishments by Standard Industrial Classification (SIC) code for each state and county. The CD includes a listing of SIC codes, though the format isn't (or wasn't as of 1995) identical to the format used in the main data file.

Each SIC code consists of four digits, arranged in a hierarchy. For example, codes beginning with a "2" or "3" signify manufacturing. A code beginning with "35" is for industrial machinery and equipment. The "357" code is for computers and office equipments. Finally, "3571" is for electronic computers.

The biggest limitation in the free data is that most of the data is national or, at best, statewide. For local data, you'll have to open your wallet.

(SIC codes are gradually being replaced by the North American Industry Classification System, a five-digit scheme for classifying businesses that will be used jointly by the United States, Canada and Mexico. You'll find details under the "NAICS" heading on the Census Bureau's current economic statistics Web page. But you probably already knew that.)

The great strength of SIC codes is that they're "nested." With a few mouse clicks in a database, you can bore down to a very specific industry or pull back for a broader look.

But take care: The Census is careful about the privacy of individual businesses. It suppresses data that might identify a particular business. For example, last February I used County Business Patterns to identify the nation's leading states and counties for high-tech employment. I was puzzled when my estimates of high-tech employment in North Carolina came in 16,000 below industry estimates. I found the missing employees when I pulled back to the three-digit SIC level.

Fortunately there is an easy way to detect this problem without cross-checking three- and four-digit codes. The "tflag" field discloses whether data is suppressed. All in all, it's a small price to pay for a mountain of data.

Ron Campbell can be reached by e-mail at ron_campbell@link.freedom.com

**CAMPBELL WRITES:
CENSUS DEMOGRAPHERS
CONSTANTLY ARE
STUDYING THE ECONOMY
AND RELATED TOPICS. THE
CENSUS THEN PUBLISHES
THEIR RESEARCH ON THE
WEB. SOME OF THEIR
WORK IS OF WIDE
INTEREST. FOR EXAMPLE, I
FOUND PAPERS ON THE
DECLINE AND REBIRTH OF
THE RUST BELT AND
STUDIES OF THE COST OF
ENVIRONMENTAL
REGULATION. I ALSO
DOWNLOADED A 43-PAGE
PAPER SUGGESTING THAT
THE EFFECT OF
ENVIRONMENTAL
REGULATION ON
PRODUCTION COSTS HAS
BEEN OVERSTATED.**

**A FURTHER SUGGESTION
FROM SMALL:**

**FERRET AND DES ARE
DATA EXTRACTION TOOLS
THAT YOU CAN USE TO
PULL DATA FROM THE
1990 CENSUS DATA
(PUMS), THE CPS AND
THE SURVEY OF INCOME
AND PROGRAM
PARTICIPATION (SIPP).
YOU CAN DO UNIQUE
CROSS-TABS WITH THESE
TOOLS, BUT NOVICE USERS
OF CENSUS INFORMATION
SHOULD BEWARE.**

From page eight:

40,000-pages strong

Use the U.S. Gazetteer at <http://www.census.gov/cgi-bin/gazetteer> to search for a place, county or ZIP code. Each search retrieves a record with population information, geographic coordinates, and a map.

Finally, don't neglect the international data available. The International Data Base at <http://www.census.gov/ipc/www/idbnew.html> contains demographic and socioeconomic data for all countries.

Jennifer Small can be reached by e-mail at jsmall@express-news.net

Linking to the Census Bureau Web site

By Paul Overberg
USA Today

Start with the Census Bureau's own information on Census 2000, <http://www.census.gov/dmd/www/2khome.htm>. Then read these pages:

- the goal and strategy at <http://www.census.gov/dmd/www/plan2.html>
- the detailed operational plan at <http://www.census.gov/dmd/www/newplan3.pdf>
- explanations of sampling for non-statisticians at <http://www.census.gov/dmd/www/insight.html> and <http://www.census.gov/dmd/www/amscien.pdf>
- news releases at <http://www.census.gov/Press-Release/www/2000.html>

Counting correction: A foundation-funded group lobbying for sampling to correct the undercount is located at <http://www.census2000.org/>. It offers e-mail news alerts and tables showing the effect of the 1990 undercount on various states, cities and groups. To see how the Census Bureau would have adjusted the 1990 totals with sampling for your county, town, neighborhood or even election district, see <http://tier2.census.gov/pl94171/pl94index.htm>.

Race and ethnicity: To understand the big changes in how race and ethnicity will be classified and counted, read the decision from the White House Office of Management and Budget at <http://www.whitehouse.gov/WH/EOP/OMB/html/fedreg/Ombdir15.html>.

Local context: One big part of getting ready is setting up your local context: 1990 Census data. What does it say, and is it ready

to be computer-matched with the new data? You can pull 1990 data from the Census Bureau's 1990 Lookup at <http://venus.census.gov/cdrom/lookup>. Or buy a CD at <http://www.census.gov/mp/www/rom/msrom.html#Decennial>. Also check GeoLytics' Census CD+Maps, http://www.censuscd.com/cdmaps/censuscd_maps.htm. To do publishable mapping, many newsrooms use ArcView, the PC version of powerful mapping software by ESRI, <http://www.esri.com/software/arcview/index.html>. The same company has taken over another popular mapping software, Atlas GIS, <http://www.esri.com/software/atlas/index.html>. Others like MapInfo Desktop, <http://www.mapinfo.com/software/sandd/desktop.html>. You should also check with your marketing department, which may have the powerful proprietary demographic data and mapping software sold by Claritas Inc. at <http://spider.claritas.com/desktop.htm>

Earlier data: You might want 1980 and 1970 data, too. There's not much online, but you can get some 1980 data at the county level at <http://www.lib.virginia.edu/socsci/ccdb/>. Check with your state's data center, <http://www.census.gov/sdc/www/>. Also, the Census Bureau will dump a 1980 Census tape to a custom CD for no extra charge. See <http://www.census.gov/mp/www/tape/oneoff.html#CD> and the list of tapes at <http://www.census.gov/mp/www/tape/mstape.html#Decennial80>.

Geography lesson: To cover the census you need to understand its geographical language, which includes half a dozen separate hierarchies. For a glossary on the Census Bureau's geographical terms, see <http://www.census.gov/ftp/pub/geol/www/GARM/Ch2GARM.pdf> and <http://www.oseda.missouri.edu/pluel/geocorr/html/maggot.html>.

Trend watchers: For demographic features and trends, check American Demographics magazine at your library or subscribe by visiting <http://demographics.com/publications/ad/index.htm>. Also watch American City Business Journals' Demographics Journal at <http://www.amcity.com/journals/demographics/>

Paul Overberg can be reached by email at poverberg@usatoday.com

From page nine:

Lingo look-up

The IRS files include at least 80 percent of taxpayers but do not include the poor and others who do not file taxes. They also leave out people who got divorced or otherwise changed tax-filing status. The money figure currently does not include biggies like capital gains/losses, dividends and IRA distributions, so it's not as reliable as it could be.

But it can be really useful to tell you where people are coming from who move into your area, where people go who leave, and how much money is accompanying them.

Information about this file is available at the IRS Web site (www.irs.gov) or by calling the Statistics of Income office, 202-874-0410.

Cohn's handout, #796 in the Resource Center, was provided at the 1998 IRE and NICAR National Computer-Assisted Reporting Conference.

D'vera Cohn can be reached by e-mail at cohnd@washpost.com

Glossary glimpse

By Paul Overberg

USA Today

If you don't know "Censpeak," you'll drown trying to cover Census 2000. The more you learn, the more the players will understand what you want and the more willing they'll be to help. For a 35-page glossary on the Census Bureau's geographical terms, see <http://www.census.gov/ftp/pub/geol/www/GARM/Ch2GARM.pdf>.

- CCC: Complete Count Committee, local advisory groups assembled to get the word out and cue the Census Bureau to local circumstances. *Good local sources.*

- DADS: Data Access and Dissemination

System. A Web-based point-and-click tool to let people get most of the 2000 Census data they need. It's slated to make its debut early next year with results of the 1997 Economic Census.

- ICM: Integrated Coverage Measurement. A national survey of 750,000 households that will be done just after and matched to the census, duplicating its core in miniature. It will be used to double-check the accuracy of the data and—Congress willing—to adjust totals.

- NRFU: Non-response follow-up. If you don't mail back your census form, a live human comes to interview you. Or, eventually, your neighbors, if you can't be found. Very expensive, especially if Congress decides that "sampling" shouldn't be used as planned for the most elusive 10% households.

- One-number census: The Census Bureau's Grail. Congress willing, it plans to blend "enumerating" and sampling" to produce one number for a place or characteristic. Recent censuses have produced an "enumerated" count and "sampling" estimates of the undercount, creating two numbers and pressure to adjust totals.

- Title 13: Of the U.S. Code requires individuals' data be kept confidential. It's the Census Bureau religion. Journalists hear about it when they're told they can't tag along with census takers, get the address lists it shares with cities or do anything that might appear to breach confidentiality.

Overberg's handout is #809 in the Resource Center.

Paul Overberg can be reached by e-mail at poverberg@usatoday.com

OVERBERG SUGGESTS MORE "CENSPEAK" TO MASTER:

- **FAMILY: TWO OR MORE RELATED PEOPLE LIVING IN THE SAME HOUSING UNIT. ALL FAMILIES ARE HOUSEHOLDS, BUT NOT VICE VERSA.**

- **HOUSEHOLD: THE PERSON OR PEOPLE LIVING IN A HOUSING UNIT.**

CONTRAST WITH FAMILY.

- **STF: SUMMARY TAPE FILE, THE MAIN PACKAGES OF DATA THAT COME OUT OF THE CENSUS. STF-1 AND ITS COUSINS COME FROM THE "SHORT FORM," STF-3 AND ITS COUSINS, FROM THE "LONG FORM."**

- **SHORT FORM: THE QUESTIONNAIRE THAT EVERY HOUSEHOLD GETS. SEVEN QUESTIONS, FEWEST SINCE 1820.**

- **PRESS RUN: 118 MILLION. TOTALS WILL BE PUBLISHED ALL THE WAY DOWN TO "BLOCKS."**

- **LONG FORM: THE QUESTIONNAIRE THAT ONE HOUSEHOLD IN SIX GETS. IT'S 52 QUESTIONS. SOURCE OF "SAMPLE DATA," FROM DETAILED QUESTIONS ON HOUSING, INCOME, OCCUPATION, EDUCATION.**

From page five:

Faster turnaround

work with census files—but not as easily.

Write and test programs in advance using practice data: If you're ambitious, you can program well in advance of the actual data release. The Census Bureau did a practice census earlier this year. This data will be in the same format as the actual data, so you can test your programs.

If you've planned ahead well, sometime in the spring of 2001 you'll be able to pull a tape

out of the overnight delivery box and have a story, complete with maps and tables, in the paper or on the air the next day.

For the 1980 Census, it was months after the data came out before the first serious stories started appearing. For 1990, it was only a few weeks. This time around, I bet it'll be hours.

Steve Doig can be reached by email at steve.doig@asu.edu

Bits, Bytes and Barks

Broadcast award winners on video

The video compilation of the 1997 IRE Award broadcast winners is available for sale. The 2 1/2-hour VHS tape costs \$10. Call the Resource Center at (573) 882-3364.

New! Editor Boot Camp

These new computer-assisted reporting boot camps for editors are intensive three-day workshops tailored to the needs of newsroom managers. This includes top editors or news directors, managing editors, AMEs, assignment editors and other editors directing reporters. These boot camps will teach editors the things they need to know to make CAR successful in their newsrooms. They'll experience just enough hands-on work to understand what their reporters are tackling – and what more is possible. The first of these boot camps is scheduled for Feb. 5-7, 1999. Registration information will be available online soon at www.ire.org

Campaign Finance Information Center

Downloadable campaign finance data from 15 states is free from the CFIC at www.campaignfinance.org. We also have links to 16 online search engines hosted by other non-profits and state boards of election. We will soon release an online search engine where you can track a contributors money trail across states. Contact the CFIC coordinators, Jack Dolan (jack@nicar.org) or Cindy Eberting (cindy@nicar.org) for more information. They can both be reached at (573) 884-1802.

Conferences abound

Information on upcoming IRE and NICAR national and regional conferences is now available on the IRE Web site. Visit the IRE Web site at www.ire.org/resources/conferences/

The National Computer-Assisted Reporting Conference, presented by IRE and NICAR along with *The Boston Globe*, will be held March 11-14, 1999 in Boston. You may register online or download an early registration form at www.ire.org/resources/conferences/boston

The schedule for the conference is being finalized. We hope to offer new classes in SAS basics and data management and PERL scripting. We plan to reinstate basic Fox training using Visual Fox and continue popular basic classes, a special session for broadcasters, and Intranet training.

Contact Sarah Cohen, IRE and NICAR training director, with suggestions for panel topics and hands-on classes, plus ideas for who might speak or teach. Cohen can be reached by e-mail at sarah@nicar.org

The Investigative Reporters and Editors National Conference, co-sponsored by *The Kansas City Star* and KCTV, will be held June 3-6, 1999 in Kansas City. You may also register online or download an early registration form at www.ire.org/resources/conferences/kansascity

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

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