The geospatial distribution of employment: a new visual asset

By combining geographic information with data from the Quarterly Census of Employment and Wages program, BLS provides analysts with a tool that will offer new insights into data that were previously unobserved

Sheryl Konigsberg

The advent of powerful computing capabilities and mapping software now allows more sophisticated analysis of new and existing problems through the visual display of information. The center point of these new features is the ability to provide pinpoint locations for geographic features; defined by precise latitude and longitude coordinates, called "geocodes." In any geocoding system involving businesses, the key is to have accurate physical location addresses.[1]

This article discusses the background of the Quarterly Census of Employment and Wages (QCEW) program, the definition of geocoding, and its current and potential uses. It provides examples of existing applications using labor market information and new ways of presenting these data. The article highlights an earlier pilot project that obtained and used geocodes from the Bureau of Labor Statistics business establishment list. Finally, the article profiles future uses, and explains how BLS plans to continue its efforts of geocoding business establishments from the QCEW.

The QCEW

The QCEW program is a by-product of the Unemployment Insurance (UI) system and is managed in a Federal/State cooperative environment. This program releases comprehensive tabulations of employment and wage information for workers covered by State UI laws and Federal workers covered by the Unemployment Compensation for Federal Employees (UCFE) program. BLS provides policies, standards, and funding, whereas States and the District of Columbia collect, edit, tabulate, and publish the data.

The QCEW program serves as a near business census and constitutes the only set of monthly employment and quarterly wage information. The QCEW program already provides economic data by the six-digit North American Industry Classification System (NAICS) at the national, State, Consolidated Metropolitan Statistical Area (CMSA), Metropolitan Statistical Area (MSA), and county levels in the Federal statistical system. This quarterly census is published within 6 months after each calendar quarter.

Every quarter, under the laws of each State, businesses are required to report the number of employees for all 3 months, total wages, taxable wages, UI taxes, and administrative data, such as physical location addresses. After these UI reports are collected and entered by the State UI department, they are passed to the State QCEW program for the reviewing, editing, and publishing stages. These data also are used for the QCEW business register.

In addition to the UI reports, BLS funds two other collections to support the needs of its users. The first is the Annual Refilling Survey (ARS) that, over a 3-year period, contacts all businesses to update or complete industry information (NAICS codes) and addresses. This is the primary method for updating

Sheryl Konigsberg is an economist in the Office of Employment and Unemployment Statistics, Bureau of Labor Statistics. E-mail: Konigsberg.Sheryl@bls.gov

physical location addresses within the QCEW business register. The second is the quarterly Multiple Worksite Report (MWR) that collects data for each individual establishment of a multi-unit business. The combination of information from these three sources makes up the resulting QCEW program. The program's comprehensiveness results in precise business and employment data with substantial industry and geographic detail.

Data from the QCEW serve as an important input to many BLS programs as well as other Federal and State programs. These data are used as a benchmark for the Current Employment Statistics and Occupational Employment Statistics. The QCEW also is used by the Bureau of Economic Analysis for gross domestic product (GDP) and personal income estimates.

Geocoding

Geocoding is the process of adding geographic information, such as latitude and longitude, to a file or database for use in a geographic information system (GIS). A GIS is a set of activities that involve the use of computer programs and staff to capture, store, update, manipulate, analyze, and display spatial information; often in the form of maps.

Geocoding uses either a point or polygon approach. In a point-based approach, business establishment information is linked to latitude and longitude coordinates. This information allows a user to plot locations on a map. In a polygon-based approach, business establishment information is linked to the center of a polygon that represents a reference layer such as census block group, census tract, ZIP Code or county. This information allows a user to identify and use all types of data that may be collected or available from other sources. The QCEW microdata file contains a rich set of geographic information, such as physical location address, city, State, ZIP Code, and county, that can be geocoded and applied to answer questions about the labor market.

There are two types of geocodes: address geocodes and ZIP Code centroid geocodes. The most precise is the address geocode. Address geocodes are derived using the physical location address. Addresses are geocoded by BLS using commercial software that accesses U.S. postal data files. This software estimates the location of each address record from an input file and standardizes the address. These standardized addresses are then matched against a Geographic Base File (GBF), which contains directories of street segment records. The second type, ZIP Code centroid geocodes, assigns the geographic center of each ZIP Code to an address. If the geocoding software is unable to match against an address, it will attempt to geocode to the ZIP Code centroid. These matching processes assign geographic codes to address records, establishing their spatial location.

Potential range of geospatial data

Geocoded data are used extensively in government, business, and research for a wide range of applications including environmental resource analysis, land-use planning, locational analysis, tax appraisal, utility and infrastructure planning, real estate analysis, marketing and demographic analysis, and habitat studies. At the most detailed levels, geocoded business addresses are valuable to transportation planning where approximate locations or higher level county aggregations are inadequate. For this purpose, the side of the street, the location along the block, and the exact corner of an intersection are critical to optimal planning of bus lines and other public transportation.

Geocoding QCEW data allows labor market information to be presented in a new dimension. Demands for more local data give BLS an incentive to provide data for cities, towns, and even smaller areas. With the availability of geocoded data, BLS potentially can develop lower levels of aggregations, including cities, postal ZIP Codes, census tract, census block, and natural boundaries such as floodplains.

Data presentation

The conventional way of presenting economic data is two-dimensional, through tables and graphs. If tabular data are geocoded, they can be used to create a drawing illustrating the relationship among three data items. With the rise of Internet usage and improving technology, GIS has made it possible to plot economic data to create illustrations and publish in the form of maps. This can be done by using geographic information, computers and geographic software to read the information and create spatial data visually.

As an example, the QCEW program produces an annual bulletin with tabular data aggregated by State. Table 1, which shows establishment counts, employment, and wages by State, is a section from the 2002 QCEW publication. The data in table 1 are a standard way of presenting labor market information that has been in practice for many years. With this traditional way of displaying data, the lowest level of aggregation by boundaries is by county. This table can be challenging for an analyst to interpret

Table 1. **Establishments, employment, and wages in the private industry information sector, by State and 6-digit NAICS industry, 2002 annual averages**

Area	Average establishment	Annual average employment	Total annual wages (in thousands)	Annual wages per employee	Average weekly wage
Total United States	150,107	3,364,485	$188,758,526	$56,103	$1,079
Alabama	1,782	34,206	1,483,340	43,365	834
Alaska	365	7,076	317,971	44,937	864
Arizona	2,369	51,875	2,218,526	42,767	822
Arkansas	1,202	20,367	723,446	35,521	683
California	22,265	499,681	35,051,307	70,147	1,349
Colorado	3,877	93,397	5,900,532	63,177	1,215
Connecticut	1,871	41,145	2,310,682	56,159	1,080
Delaware	334	7,745	393,936	50,863	978
District of Columbia	1,129	25,448	1,934,773	76,029	1,462
Florida	8,751	177,973	8,212,392	46,144	887
Georgia	4,492	132,432	7,563,572	57,113	1,098
Hawaii	691	11,701	505,167	43,173	830
Idaho	713	9,162	305,019	33,292	640
Illinois	6,454	145,409	7,667,873	52,733	1,014
Indiana	2,178	42,528	1,657,356	38,971	749
Iowa	1,743	35,193	1,225,782	34,830	670
Kansas	1,485	50,745	2,377,331	46,849	901
Kentucky	1,767	31,745	1,120,354	35,292	679
Louisiana	1,659	29,018	1,098,531	37,857	728
Maine	736	11,546	429,314	37,183	715
Maryland	2,914	53,449	3,010,295	56,321	1,083
Massachusetts	4,521	99,989	6,645,535	66,463	1,278
Michigan	3,977	73,480	3,467,610	47,191	908
Minnesota	3,036	67,161	3,199,455	47,639	916
Mississippi	1,047	16,070	569,159	35,417	681
Missouri	3,174	70,899	3,177,280	44,814	862
Montana	830	7,780	255,185	32,800	631
Nebraska	1,013	24,690	1,053,470	42,668	821
Nevada	921	16,967	766,774	45,192	869
New Hampshire	917	12,821	701,327	54,701	1,052
New Jersey	4,058	112,163	7,602,398	67,780	1,303
New Mexico	971	16,864	537,844	31,893	613
New York	11,713	295,415	19,665,362	66,569	1,280
North Carolina	3,736	78,955	3,729,606	47,237	908
North Dakota	420	7,928	271,354	34,227	658
Ohio	4,202	101,279	4,650,075	45,914	883
Oklahoma	1,694	35,496	1,342,968	37,834	728
Oregon	2,244	36,211	1,704,070	47,059	905
Pennsylvania	5,980	128,315	6,311,853	49,190	946
Rhode Island	616	11,132	539,782	48,489	933
South Carolina	1,466	28,154	1,085,658	38,561	742
South Dakota	475	6,791	219,641	32,343	622
Tennessee	2,198	51,639	2,103,516	40,735	783
Texas	9,626	248,879	13,252,884	53,250	1,024
Utah	1,562	29,808	1,212,776	40,686	782
Vermont	(1)	(1)	(1)	(1)	(1)
Virginia	4,011	105,816	6,886,669	65,082	1,252
Washington	3,182	92,714	9,485,543	102,310	1,968
West Virginia	774	13,306	466,202	35,037	674
Wisconsin	2,096	51,123	1,990,237	38,930	749
Wyoming	(1)	(1)	(1)	(1)	(1)
Puerto Rico	508	21,273	684,425	32,173	619
Virgin Islands	45	935	32,420	34,674	667

[1] Data do not meet BLS or State agency disclosure standards.

what is being conveyed.

By contrast, table 2 is an example of tabular data that presents details on the number of establishments, average monthly employment, and total quarterly wage, by industry sector for the city of Cleveland, Ohio.

Table 2 demonstrates how geocoded data can be displayed at the subcounty level. Without the latitude and longitude information, these data could not have been aggregated at this level of fine detail. In addition, such data can be used in even richer applications, which this article illustrates later.

In September 2003, Hurricane Isabel, a category 5 hurricane in the Atlantic Ocean, made landfall on the east coast of North Carolina. Table 3 displays establishments within industries in the floodplain areas of Brunswick and New Hanover counties. North Carolina was able to show that approximately 11 percent of units or establishments and 10 percent of employment in Brunswick and New Hanover Counties are located in a floodplain. Some industries with a higher percentage of units and employment in affected areas of the hurricane might not be surprising. For example, some units might be in areas where boat rentals or other water recreational activities take place. These data help users determine the potential impact of this disaster.

These examples illustrate the traditional method of displaying data in a tabular format. A GIS-based presentation also provides a visual display that was previously unavailable. The following examples demonstrate the power of a GIS and how it conveys information visually.

The position of the business establishments (dots on the map) in map 1 conveys an immediate visual impression. (See page 57.) The dots on the map describe whether an establishment is within or outside a floodplain area. Most of the establishments lie outside of the floodplain. When this map is combined with the data from table 3, a user can see the distribution of the 11 percent of units located in the floodplain. The use of geocoding and mapping the QCEW data can help users understand the spatial distribution of employment, which can lead to better informed decisions about the local economy.

The hurricanes that hit central Florida in 2004—Charley, Frances, and Jeanne—are shown in map 2. (See page 58.) The State of Florida was able to track the path of each hurricane with a 20-mile radius to show the potential impact on employment within the affected areas. This map shows that all three hurricanes crossed through Polk County, Florida, where the density is 1 to 75 employers per square mile.

The impact of the October 2003 fires in San Diego, California, is shown in map 3. (See page 59.) The State of California was able to combine geographic information with QCEW data from the second quarter of 2003 to show establishments that were located within the fire areas and within a half mile of the fire areas. By looking at this map, one is able to see the areas where clusters of employment potentially were affected.

The State of Minnesota was able to display employment around major highways by using a thermal density map as shown in map 4. (See page 60.) With this type of map, States can show areas with a high concentration of employment without displaying confidential information.

It is apparent that maps show how "a picture paints a thousand words." Information displayed in a graphic format can allow a reader to process information more quickly, therefore, allowing for more timely conclusions about a particular set of information, such as employment density within a particular distance of a floodplain or fire area as shown in the previous maps.

Geocoding pilot project

In March 2003, the QCEW program completed a geocoding pilot project with the following 14 States and the District of Columbia: California, Connecticut, Florida, Hawaii, Maine, Maryland, Minnesota, Missouri, North Carolina, Ohio, Oregon, South Carolina, Texas, and West Virginia. These States published data based on the geocodes derived from the QCEW data. This study was used to help refine plans for implementing geocoding in all States.

The most important investment in the geocoding pilot project was the time State workers spent to improve the vast number of physical location addresses. Traditionally the States' primary resource for locating addresses was Internet sites such as company Web sites, online phonebook services, and online maps. They also used other sources such as telephone books and phone calls to employers to obtain addresses. These last two sources proved to be less reliable and more time consuming for most of the States in the pilot study.

Obtaining government physical location addresses was a major obstacle for all States that participated in the pilot project. Governments tend to provide county-wide reports and finding a geocodeable address can be difficult.

Lastly, nondisclosure is an issue. Many States were unsure if they could publish subcounty data and to what extent. Some questions that arose during the project were:

Table 2. City of Cleveland geocoded data on establishments, average employment and total wages paid by industrial sector, as covered under the Ohio and Federal unemployment compensation laws, first quarter 2002

Industrial sector	Number of establishments	Average monthly employment	Total wages (in thousands of dollars)
Total covered under Ohio unemployment compensation Law [1]	9,365	279,396	$2,958,645
Private sector	9,273	230,658	2,441,135
Agriculture, forestry, fishing and hunting	4	21	94
Mining	7	231	3,295
Utilities	14	1,042	15,761
Construction	511	6,198	73,192
Manufacturing	1,138	31,964	333,537
Wholesale trade	678	12,229	159,430
Retail trade	1,247	13,458	73,763
Transportation and warehousing	243	4,116	34,241
Information	163	7,151	99,972
Finance and insurance	500	23,046	420,040
Real estate and rental and leasing	311	2,696	20,157
Professional and technical services	1,162	21,367	312,719
Management of companies and enterprises	53	6,418	93,517
Administrative and waste services	446	15,624	99,983
Educational services	73	9,960	92,572
Health care and social assistance	768	46,598	445,038
Arts, entertainment, and recreation	108	5,150	56,995
Accommodation and food services	898	14,112	47,337
Other services, except public administration	949	9,276	59,492
State and local government	92	48,738	517,510
State government	21	3,940	46,608
Local government	71	44,798	470,902
Federal Government [2]	18	8,213	103,766

[1]The first quarter 2002 covered employment and wage data for the city of Cleveland were developed as part of a special project conducted in cooperation with the U.S. Bureau of Labor Statistics. For this project, approximately 38,000 establishment records covering almost 764,000 employees in Cuyahoga County were processed for Geocoding using Geostan software by Sagent Technologies. A geocodable record contains a physical location address that can be assigned a longitude, latitude, and place code. In all, 87 percent of establishments, covering 97 percent of employment, were able to be geocoded at the subcounty level. The information presented in this table were those records identified as having the place code for the City of Cleveland (16,000) and are based upon employers' reports for first quarter 2002 received in the Bureau of Labor Market Information through January 1, 2003.

[2]Includes only Federal Government agencies.

NOTE: Summed totals and subtotals may not equal the sum of industrial divisions because of the exclusion of those industries with fewer than three employers or because of rounding. Includes the Private Sector and State Government entities, but excludes Federal Government agencies.

SOURCE: Ohio Department of Job and Family Services Office of Research, Assessment and Accountability Bureau of Labor Market Information Columbus 43266 03/28/03.

Does a point on a map disclose confidential data about a business establishment based on address, employment, or industry? Some States concluded that they could publish this type of information, whereas other States could not because nondisclosure laws vary from State to State.

Future uses

Since 2004, the Bureau's geocoding effort has provided insight into the techniques for improving the accuracy of QCEW physical location addresses. These techniques have involved extensive work, researching and updating the Bureau's existing business establishment list. With geocoded data, BLS is able to provide new economic information such as subcounty estimates, including city, census tract, or census block group for future research. There also is the potential to standardize addresses and reduce mailing costs for sample users.

Another use of geocoded QCEW data is to improve the Business Employment Dynamics (BED), a set of statistics generated from the QCEW. These quarterly data series consist of gross job gains and gross job losses statistics

Table 3. Geocoded industries and employment in North Carolina, Brunswick and New Hanover Counties floodplain, first quarter 2003

Sector	Units	Percent of units in floodplain	Percent of employment in floodplain
Total	8,478	11	10
Agriculture, forestry, fishing and hunting	36	11	2
Mining	([1])	([1])	([1])
Utilities	([1])	([1])	([1])
Construction	1,287	12	11
Manufacturing	288	5	12
Wholesale trade	420	8	7
Retail trade	1,313	9	6
Transportation and warehousing	217	11	4
Information	121	9	2
Finance and insurance	406	5	5
Real estate and rental and leasing	374	18	29
Professional, scientific, and technical services	841	11	12
Management of companies and enterprises	31	11	1
Administrative and support and waste management and remediation services	545	9	4
Educational services	124	10	17
Health care and social assistance	655	5	2
Arts, entertainment, and recreation	166	18	10
Accommodation and food services	698	20	18
Other services (except public administration)	692	8	7
Public administration	116	29	7
Unclassified	148	0	0

[1] This is a suppressed value, which is usually very small.

SOURCE: Labor Market Information Division, Employment Security Commission of North Carolina.

from 1992 forward. These data help to provide a picture of the dynamic state of the labor market. Most data in these series are linked across time, using a process that matches establishments by a unique number—the State Employment Security Agency identification numbers. Records that are not linked by this process are linked by various other means, one of which is a weighted match. The weighted match involves creating blocks such as name, address, and telephone number to match data in the current quarter with data in the previous quarter. With geocoded data, longitude and latitude information can be used in these blocks to create more accurate matches, thus allowing for better gross job gains and job losses data.

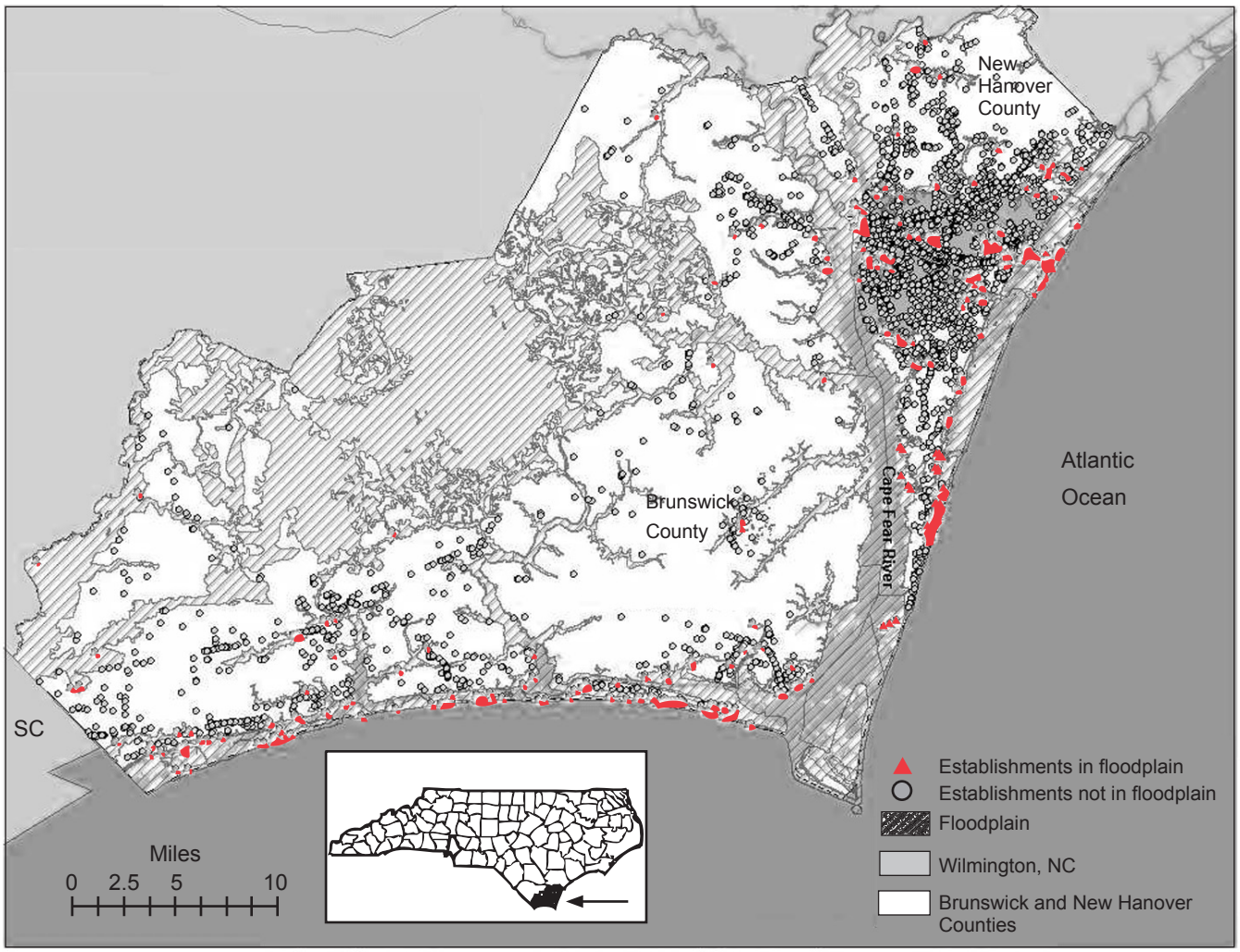
Not only does geocoded data improve the existing QCEW, but it enhances the uses of the data by improving the inputs to other programs within the Federal statistics system. Within BLS, these improvements benefit the Current Employment Statistics (CES), Occupational Employment Statistics (OES), and Local Area Unemployment Statistics, (LAUS) by creating more precise county and MSA data. The Bureau of Economic Analysis (BEA) and Bureau of Census (BOC) also benefit, as BLS maintains ongoing data-sharing agreements with these agencies that rely on the QCEW as primary inputs into key statistical products.

Lastly, GIS technology and spatial data play an important role in emergency response and preparedness. Large scale emergencies that have an impact on humans and land are unpredictable and hard to envision. Two types of hazards are natural disasters and human-induced disasters. Natural disasters include events such as hurricanes, earthquakes, volcanoes, landslides, wildfires, and floods. Human-induced disasters include events such as man-made fires, toxic spills, war, and bioterrorism.[2] A GIS saves a great deal of time in decisionmaking and in evaluating the impact of a disaster before and after it occurs.[3]

Getting the QCEW fully geocoded

The QCEW database contains approximately 8.8 million establishments with an employment level of approximately 135 million. By the third quarter of 2006, 83 percent of the QCEW records and 93 percent of the employment data had been geocoded. BLS considers this rate extremely good

Map 1. **Establishments in the Brunswick and New Hanover Counties, North Carolina floodplains, first quarter 2003**

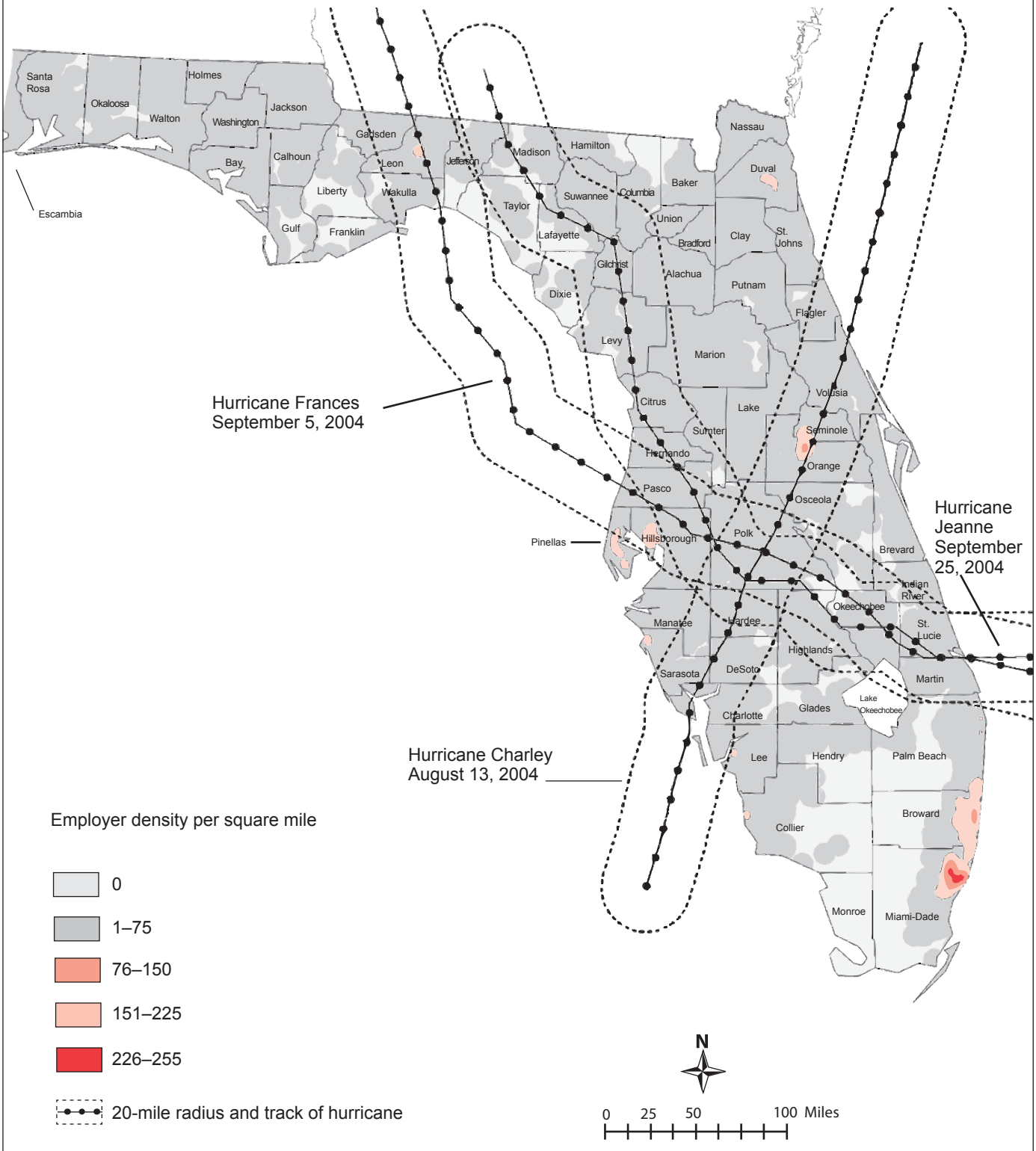


SOURCE: Labor Market Information Division, North Carolina Employment Security Commission.

and sufficient to proceed with developing a range of products such as maps and subcounty research data. The remainder of the units is mostly new small firms or Federal, State and local government units that do not provide QCEW data by worksite. A small number of large units also do not provide QCEW data by worksite. BLS continues to work with these firms to obtain accurate data by county and industry to allow for geocoding these areas.

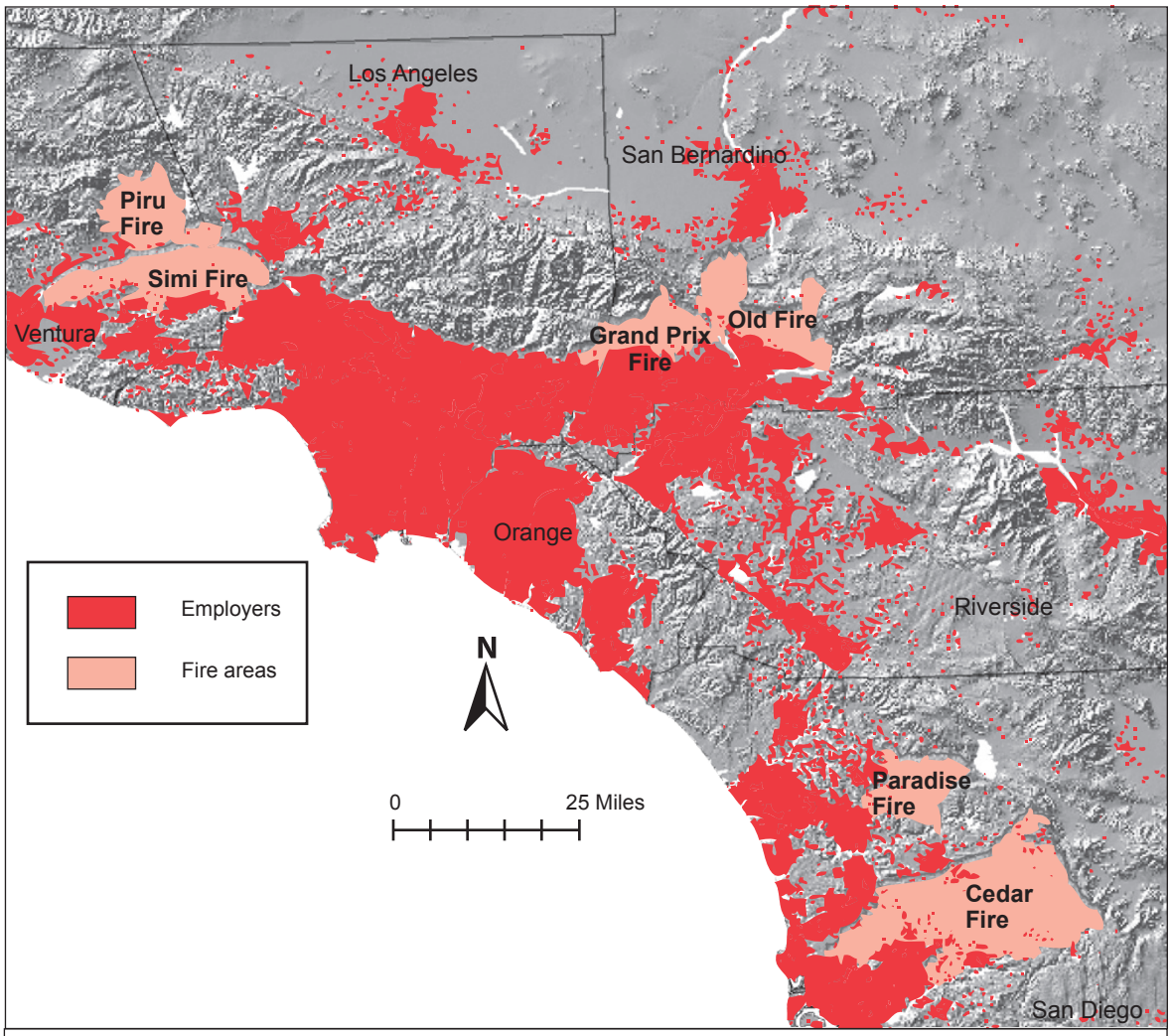
GEOGRAPHIC DATA ARE AN ASSET in data analysis, especially with the QCEW. Since BLS has implemented this new feature, the original tabular data can be combined or used to create an in-depth way of viewing data. Using a geographic information system such as geocoding and mapping software, many datasets can be combined into one picture, thus saving time in reviewing data results and providing new insights that previously were unobserved. This article has provided just a few examples of how data users can benefit from the use of QCEW geocoded data. □

Map 2. Employer density and hurricane tracks in Central Florida, 2004



SOURCE: Agency for Workforce Innovation, Florida Labor Market Statistics; Quarterly Census of Employment and Wages; and Florida Department of Environmental Protection.

Map 3. **Employment within fire affected areas, Southern California, 2003**



Employers within fire areas:

Fire area	*Employers*	*Employment*	*Wages*
Total	1,274	24,775	$224,465,788
Cedar and Paradise	790	13,831	120,724,115
Grand Prix and Old Fire	300	4,648	41,796,288
Piru and Simi	184	6,297	61,945,386

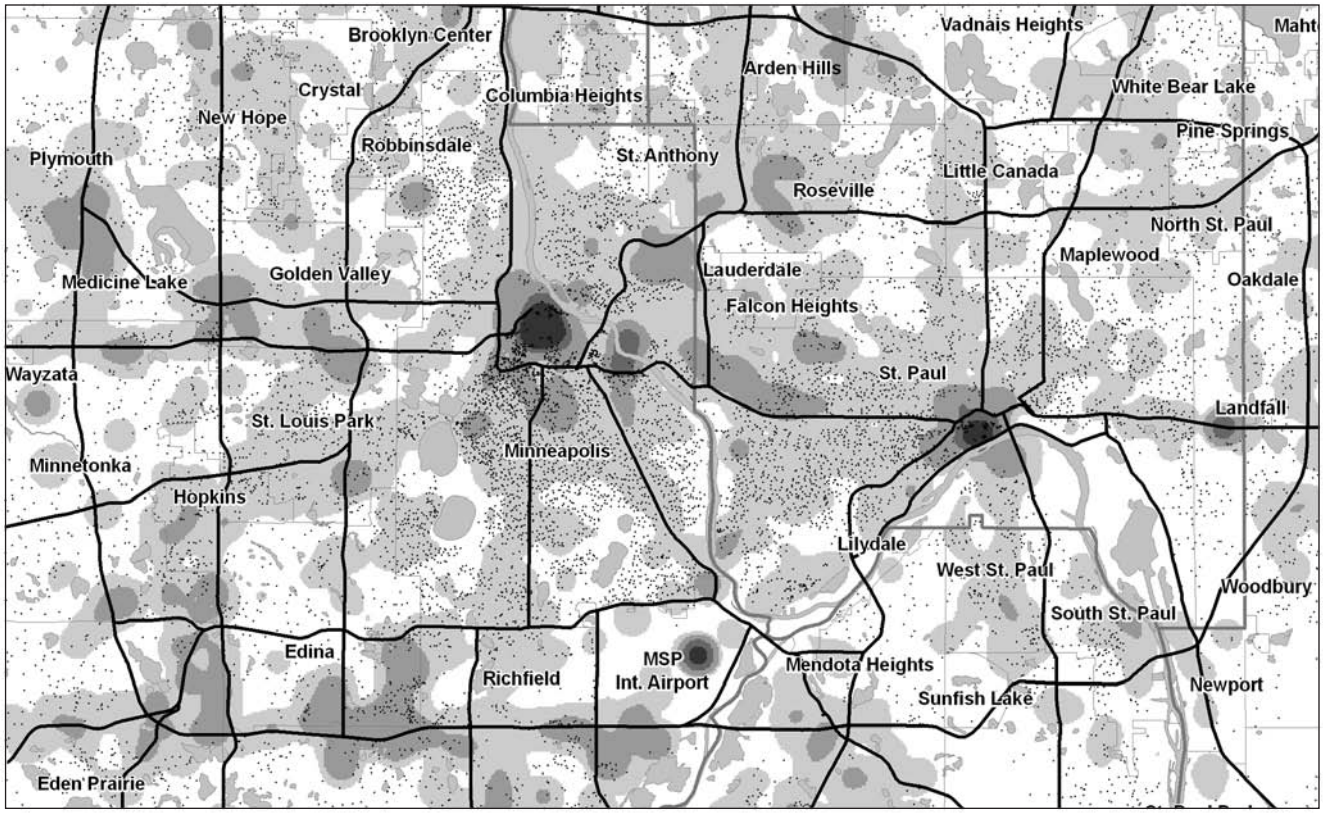
Employers within fire areas plus those within 1/2 mile of fire perimeter:

Fire area	*Employers*	*Employment*	*Wages*
Total	5,654	91,991	$867,769,086
Cedar and Paradise	2,948	52,447	514,078,146
Grand Prix and Old Fire	976	10,358	80,496,437
Piru and Simi	1,730	29,186	273,194,503

NOTE: Employment is from June 2003. Wages are the total paid for the second quarter of 2003.

SOURCE: California Quarterly Census of Employment and Wages, second quarter 2003; fire perimeters from Geospatial Multi-Agency Coordination Group (GeoMAC), Nov. 3, 2003; cartography from Current Economic Statistics Group, Labor Market Division, California Employment Development Department, November 2003, on the Internet at **www.calmis.ca.gov**.

Map 4. **Employment and housing density, Minnesota, Twin Cities Metropolitan Area, 2000**



SOURCE: Metropolitan Council, based on Quarterly Census of Employment and Wages (QCEW) data from the Minnesota Department of Employment and Economic Development.

Notes

[1] Richard Clayton, "Geocoding the Business Register at the Bureau of Labor Statistics," Paper presented at 15th International Rountable on Business Survey Frames, Washington, DC, Oct. 22–25, 2001.

[2] "Challenges for GIS in Emergency Preparedness and Response," An ESRI White Paper, *Environmental System Research Institute*, 2000, on the Internet at **www.esri.com/library/whitepapers/pdfs/challenges.pdf**.

[3] "GIS Aids Emergency Response," *ArcUser*, July–September 2001, on the Internet at **www.esri.com/news/arcuser/0701/umbrella15.html**.