https://doi.org/10.48009/2_iis_2007_484-488

VISUALIZING GENDER-BASED DIGITAL DIVIDE ISSUES IN THE UNITED STATES THROUGH DATA MINING AND GIS

Dr. Matthew A. North, Washington & Jefferson College, mnorth@washjeff.edu
Joan P. Downing, Washington & Jefferson College, downingjp@washjeff.edu

ABSTRACT

Since the invention of the digital computer, careers in computers have been dominated primarily by Caucasian men. This paper examines the current state of efforts to close the so-called Digital Divide, specifically addressing the gap drawn along gender lines in the United States. Using data from the 2000 U.S. Census and tools including the Jenks classification method and Geographic Information Systems (GIS) for visualization, we find that although women in the United States appear to be occupying an increasing share of Information Systems-related careers, substantial gaps still remain. Specifically, this study examines Computer Specialist jobs as a percentage of all jobs in each county in the U.S., and then examines the percentage of those jobs which are held by women. The results of this investigation show that both statistically and geo-spatially speaking, there are no correlations between the greatest concentration of computer jobs and the greatest concentration of women in those jobs. These findings form a foundation for conversation about the current state of the gender Digital Divide in the U.S., as well as a platform for a comparison study to be performed when the 2010 U.S. census data are released.

Keywords: Digital Divide, Women in Computing, IS Careers, Data Mining, GIS, U.S. Census

INTRODUCTION

The term *Digital Divide* refers to gaps which exist in a given society between those who have access to computerized technologies and those who do not [1]. These gaps are generally identified along divisions by gender, race or ethnicity, or socio-economic status. Historically, organizations including the Association for Computing Machinery (ACM) and the National Science Foundation (NSF), among many others, have actively involved themselves in programs to promote and support efforts to extend opportunities for Computer Specialist education and careers to women [2]. This paper represents an effort to identify and quantify progress resulting from concerted efforts by groups such as the ACM and NSF, and to prepare a benchmark for comparison to a future study of 2010 census data when such data become available in approximately 4-5 years.

According to the ACM, women represented approximately 28% of the computer-related workforce in the United States in 1990 [3, 4]. By the year 2000, involvement by women in computer-related education and careers had risen to approximately 40% [5]. In the year 2000, the U.S. Census Bureau added a new occupation description to the census form: Computer Specialist. The Census Bureau's documentation describes a Computer Specialist as one whose career deals primarily with computer information systems and technologies [6]. This addition to the census form allowed the collection of nationwide, comprehensive data on Computer Specialist careers for the first time.

Research Questions

This study is framed by three research questions:

1. Within the United States, where are the majority of Computer Specialist jobs found, and what is the gender distribution for those jobs?
2. In the year 2000, which areas of the country represented both a relatively high rate Computer Specialist jobs and a high rate of female occupancy in those jobs?
3. Is there a correlation between the number of computer jobs in a geographic region and the number of women in that region occupying Computer Specialist jobs?

Given the exploratory nature of the research conducted in this study, these questions are intentionally not stated as hypotheses. Rather, the preference here was to investigate the data without preconceived notions regarding what may be found, which is typical for an Exploratory Data Analysis approach [7].

Limitations

Because the U.S. Census form is a self-assessment, data drawn from the Census Bureau are reliant upon an accurate representation by each person who completes the form. The Census Bureau takes great care to validate and ensure reliability of the data it collects, and provides publicly available explanations of their data validation processes [6]. This study recognizes that some individuals, for example, an Information Systems professor, may classify their occupation as "Educator" as opposed to "Computer Specialist". While this may have some effect on the outcomes of this study, any effect is expected to be mitigated by the sheer size of the data set (129,719,719), which helps to ensure that although some variation may exist in the data, the results are still representative of reality.

RESEARCH METHODOLOGY

In order to address the three research questions posed above, data from the 2000 U.S. Census were compiled and analyzed quantitatively, through correlation and data mining, and spatially, through the use of a Geographic Information System (GIS).

Data Collection and Preparation

Data for this study were downloaded from the U.S. Census Bureau's web site in raw format [6] and then aggregated by county and variable. In all, 3,141 counties (including the District of Columbia) were represented in the data, and the final data set consisted of one observation for each county. Across each observation, four variables were then aggregated for each county: Total Number of Persons Employed (Tot_Emp), Total Number of Computer Specialists Employed (Tot_CS), Number of Male Computer Specialists Employed (M_CS), Number of Female Computer Specialists Employed (F_CS). Five random observations from the data set are provided in Table 1 as an example.

Table 1. Example of data set

County	Tot_Emp	Tot_CS	M_CS	F_CS
Lancaster, SC	28,110	200	128	72
Tuolumne, CA	20,419	230	151	79
Logan, ND	964	0	0	0
Kiowa, KS	1,581	4	2	2
Daviess, IN	13,305	49	25	24

. . . 3,141 total observations in the data set

Using these variables, we then calculated Computer Specialist jobs as a percentage of all jobs for each county, and Computer Specialist jobs held by women as a percentage of all Computer Specialist positions.

Statistical and Data Mining Analysis

With the data prepared and organized by county, statistical and algorithmic formulae could be applied to the data in order to address portions of each of the research questions posed in this study. In order to equalize across the wide population variations from one county to the next (e.g. the range of persons employed in each county ranged from 149 to 3,953,415), only percentages were compared to one another. A standard Pearson correlation was performed between the percent of jobs classified as Computer Specialist, and the percentage of Computer Specialist jobs held by women. A positive and statistically significant coefficient would indicate that as the concentration of Computer Specialist jobs in a county increases, the number of women in those positions would also increase. Although such an outcome would not necessarily explain why women were enjoying a higher degree of success in computer careers, it would indicate whether or not an increase in opportunities in general also represented an increase in opportunities for women. In addition, an association rule induction was performed using the standard Apriori algorithm, in order to identify a possible association between the existence of Computer Specialist jobs and the occupation of those jobs by women.

In addition to correlation and Apriori analysis, the Jenks Natural Breaks method was used to classify the data into five categories along two variables: percentage of computer specialist jobs held by women, and percent of all jobs that are computer specialist jobs. The Jenks method [7] reduces the squared deviations of the means of each category generated from the data, with an objective of achieving a maximum goodness of variance fit (GFV). Jenks' formula begins by specifying a random grouping of the numeric data, in this study, the job percentage variables mentioned above. The mean of each category is calculated, along with the sum of squared deviations from each category's mean (SDCM). Observations in the data set are then shuffled from one category to another in order to reduce the SDCM, thereby increasing the GVF. The categorizatoin process continues until the GVF ceases to grow. The results of the application of Jenks' method to the census data in this study are visualized via GIS in Figures 2 and 3 below.

RESULTS

There are three research questions posed in the introduction to this paper. The first two questions, which relate to geographic locations of jobs, are best addressed spatially and are thus illustrated in the maps (Figures 2 & 3) included below. The second question is evaluated using a statistical test for correlation and the Apriori data mining algorithm.

In simply describing the data set, a weighted average calculation was used to determine the nationwide percentage of employed men and women who described their occupation as “Computer Specialist” on the 2000 census form. This calculation yielded 41.4% women and 58.6% men, which supports the claim that women had increased their roles in Computer Specialist careers to approximately 40% by the year 2000, as cited earlier in this paper [5, 8]. This statistic alone could be pointed to as evidence of success of some of the aforementioned gender-based opportunity programs supported by the ACM or NSF. This paper does not seek to detract from these claims; in fact we recognize this finding as a positive step toward women bridging the gender-based Digital Divide in the United States. Beyond this finding however, this paper does seek to further investigate additional dimensions of the data and results yielded through statistical, data mining and GIS analysis.

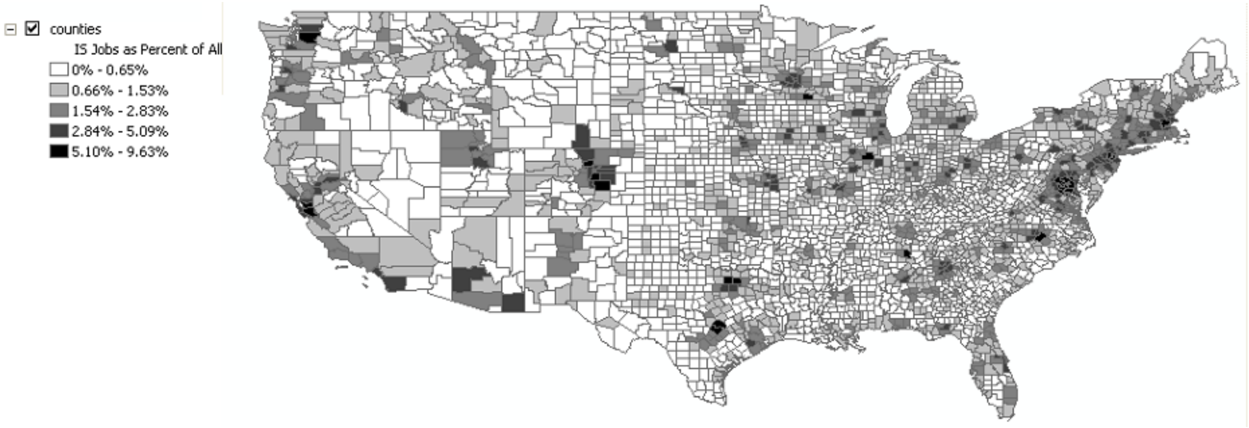

Figure 2. Computer Specialist Jobs as a Percent of All Jobs by County

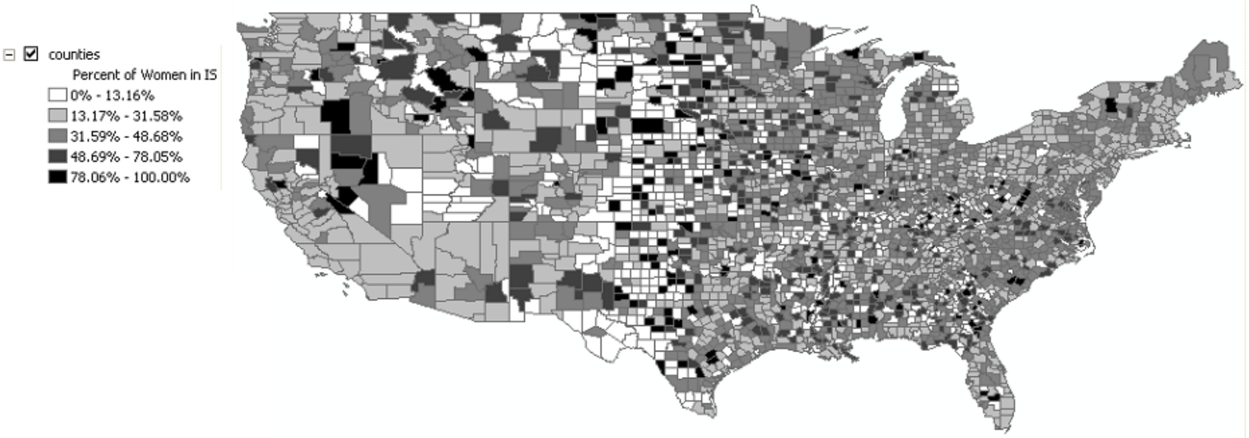

Figure 3. Percent of Computer Specialist Jobs Held by Women by County

Figures 2 and 3 are presented immediately adjacent to one another in order to facilitate comparison. Figure 2 uses graduated colors to represent the five categories of Computer Specialist jobs as a percentage of all jobs in each county, as identified through the use of the Jenks classification method. We see that the counties surrounding Silicon Valley, California; Seattle, Washington; Denver, Colorado; and the eastern metropolitan areas of Boston, New York, Philadelphia and Washington D.C. have the highest rates of Computer Specialist jobs in the country, as depicted by their dark or completely black shades. Given the historical connections of these regions to computer technology innovation, this finding is not surprising. What is a bit surprising however, is a comparison with Figure 3, which depicts the Jenks classifications of percent of Computer Specialists who are women, using the same graduated color scheme. It is interesting to note in this comparison, that the geographic regions where women hold larger portions of the Computer Specialists jobs do not necessarily correspond to the geographic regions that possess the largest proportions of Computer Specialist jobs as a percent of all jobs in each county.

This lack of correspondence is supported by the statistical correlation and the Apriori rule induction conducted in response to the third research question posed in this study. A Pearson correlation was run comparing the percentage of jobs that are Computer Specialist jobs to the percentage of women in the Computer Specialist jobs. The resulting coefficient of *r*=.0061 indicates that there is no statistical reason to believe that simply because the concentration of Computer Specialist jobs in a given county rises or falls, the occupation of those jobs by women will also rise or fall. Similarly, even with very low threshold values (Minimum Support=1%, Minimum Confidence=10%), the data mining algorithm failed to find a single association between the two variables.

It is important to recognize that Figure 3 may be a bit misleading. Figure 3 does not account for the broad variation in county populations, and may give the false impression that women are enjoying greater access to Computer Specialist careers in some areas than in others. This is illustrated by a close examination of Pendleton County, West Virginia, as one example. In Pendleton County, only 27 individuals indicated their occupation as Computer Specialist on their 2000 census form. All of them were women. As a result, Pendleton County is classified with the darkest possible shade in Figure 3, even though relatively few Computer Specialist jobs—for men *or* women—exist in this county. The state of Nevada in the western U.S. is also illustrative of this situation—the entire state has light counties in Figure 2, but many dark counties in Figure 3.

In order to address and equalize the variation in county populations, a spatial query was created and executed against the map depicted in Figure 3. The criteria for this query were set by the indicators of good Computer Specialist job occupancy by women identified through this study. The first indicator is that on average, in the year 2000, 41.4% of Computer Specialist jobs were occupied by women. Therefore in the spatial query, only counties where at least 41.4% of their Computer Specialist positions were held by women were considered to be regions where women have made headway in bridging the Digital Divide. In addition to that criterion, one other was also set. Counties with extremely high percentages of women in Computer Specialist jobs, but with very few total Computer Specialist jobs overall, can skew the results to appear as if good opportunities for women exist in a geographic region, when in fact they do not. By calculating quartiles by number of Computer Specialist jobs in each county, we find that the top 25% of counties in the United States in the year 2000 had at least 275 Computer Specialist jobs. Therefore, the second criterion set forth in the spatial query required that any county selected as a location demonstrating good Computer Specialist job opportunities for women must have had at least 275 persons employed as Computer Specialists in the year 2000. Figure 4 depicts the results of this query.

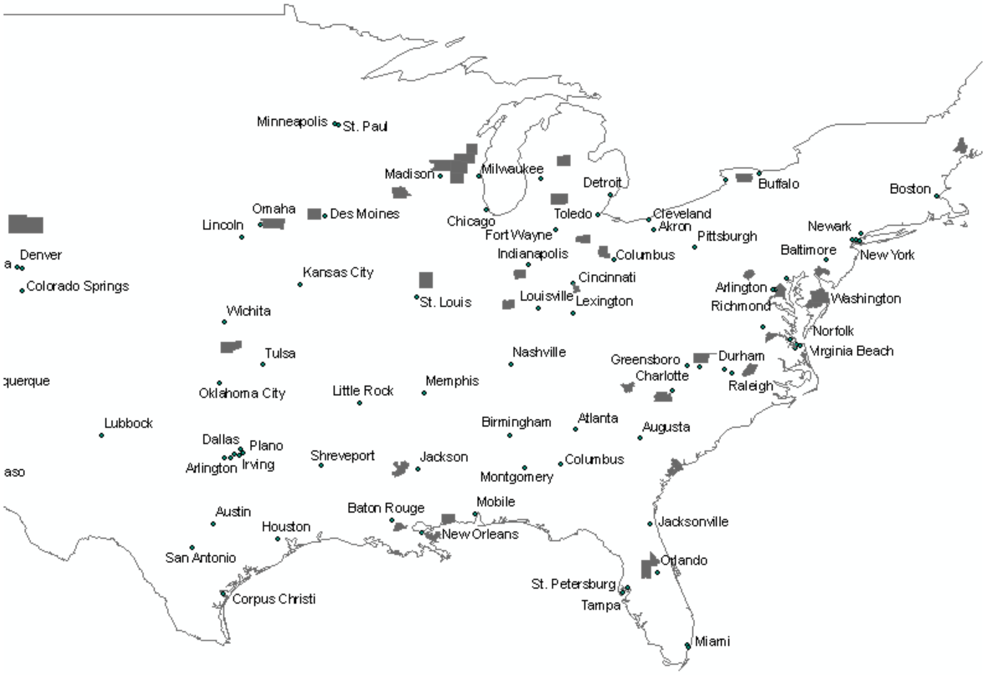

Figure 4. Counties with at least 275 Persons Employed as Computer Specialists in the year 2000; 41.4% or more of which were women.

Figure 4 intentionally depicts only the Midwest and Eastern portions of the United States. This is because the query did not find any counties west of Laramie County, Wyoming, which satisfied the requirements as outlined above, therefore regions which did not return any data have been excluded. In interpreting Figure 4 it is important to remember that the physical size of the county does not indicate greater numbers of Computer Specialist jobs. In this figure, all counties should be interpreted, regardless of physical size, as areas where significant numbers of Computer Specialist jobs exist and where a meaningful portion of those jobs are held by women. In comparison to the metropolitan areas mentioned in the discussion of Figure 2, it is interesting to note that in Figure 4, some areas where women appear to be enjoying a closing of the Digital Divide include Indianapolis, Indiana; Madison/Milwaukee, Wisconsin; Columbus, Ohio; Charlotte/Greensboro and Raleigh/Durham, North Carolina; and New Orleans, Louisiana. In the case of New Orleans, these data were collected prior to Hurricane Katrina, and thus it will be interesting to compare the 2010 U.S. Census data to determine how that disaster might have impacted Computer Specialist careers for both men and women in that particular region.

CONCLUSIONS

It is encouraging to find that headway is being made in computer-related career opportunities for women in the United States. It will be interesting to compare the results of this study to the next census data in order to see if a trend toward greater balance will develop.

Clearly, this study raises more questions than it answers, and in this regard, it should be considered both a work-in-progress and springboard for additional research. For example, one might investigate the apparent relationship between the Indianapolis metropolitan area and the relatively high incidence of women in computing professions in this area. Is Indianapolis poised to become a new national—or even international—technology center? What forces are driving the closing of Digital Divide gap in the region? Are ACM or NSF programs in the area succeeding and if so, why? Going further, much additional research of this type could be conducted along racial or economic boundaries, or for societies outside the United States. It should be noted that it is considered outside the scope of this study to pass judgment as to the *quality* of the Computer Specialist jobs examined in terms of salary equity or access to manager-level positions, although this consideration would also be relevant for further research.

A study into issues of Digital Divide is a study of sociological phenomena. As we begin to understand where events are occurring, and what connections are made to other events, we can begin to understand, react to, and perhaps influence those formative forces for society's good [9]. In this study we have identified areas where women appear to have made headway in closing the gender gap in the Digital Divide in the United States. We look forward to additional research into the causes of this progress, which may result in new programs or revised efforts in other geographic regions in order to realize more broad-based success in developing opportunities and career paths for women in computing.

REFERENCES

1. Olsen, F. (2000). Institute for Women and Technology Works to Bridge the Other Digital Divide. *Chronicle of Higher Education, 46*(31), A47.
2. Cooper, J. & Weaver, K. D. (2003). *Gender and Computers: Understanding the Digital Divide.* Lawrence Erlbaum Associates: Mahwah, NJ.
3. Crombie, G., Abarbanel, T. & Anderson, C. (2001). Getting Girls into Tech Classes. *Education Digest, 66*(5), 42-48.
4. Gurer, D. & Camp, T. (2002). *Annual Report of the ACM Committee on Women in Computing.* ACM-W Annual Report: NewYork.
5. Blum, L. (2001). *Women In Computer Science: The Carnegie Mellon Experience.* Women@SCS 2.2: Pittsburgh.
6. Census 2000 Summary File 3 (SF 3) - Sample Data. (2001). U.S. Census Bureau: Washington D.C. [Electronic Version]. Retrieved on 24 January, 2007 from: http://factfinder.census.gov/servlet/DatasetMainPageServlet?_program=DEC&_submenuId=&_lang=en&_ts=
7. Jenks, G. F. & Caspall, F. C. (1971). Error on Choroplethic Maps: Definition, Measurement, Reduction. *Annals of the Association of American Geographers, 61*(2), 217-244.
8. Oder, N. (2001). Is Digital Divide Narrowing? *Library Journal, 126*(5), 13-16.
9. Epodoi, Rita. M. (2004). Bridging the Gender Gap. *UN Chronicle, 40*(4), 36-42.