2010

# Mapping the digital divide in neighborhoods: Wi-Fi access in Baton Rouge, Louisiana

Luke Driskell
*Louisiana State University and Agricultural and Mechanical College*

# MAPPING THE DIGITAL DIVIDE IN NEIGHBORHOODS: WI-FI ACCESS IN BATON ROUGE, LOUISIANA

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Science

in

The Department of Geography and Anthropology

by
Luke Driskell
B.S., Louisiana State University, 2008
May 2010

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

ABSTRACT

The communication made possible by the Internet has leveled the global playing field in some ways, but helped maintain traditional inequalities as well. The "digital divide" refers to disparities in telecommunication access and use from global to local scales. This study uses access point mapping to quantify local Internet access in Baton Rouge, Louisiana. A Wi-Fi access point (router) density was obtained and compared to various demographic and socioeconomic attributes in neighborhoods. Fieldwork confirmed the expectation that traditionally disadvantaged groups would have the lowest rates of Wi-Fi ownership, but median household income was unexpectedly less related than race, education, and single-mother households. Results from research following the access point mapping technique can help inform planners in implementing municipal Wi-Fi networks meant to redress the digital divide. It can also be used as a proxy measure for socioeconomic data that are not updated often or are expensive to collect.

# INTRODUCTION

Social scientists' discussions about the Internet tend to follow a broader movement toward critical theory. Outside of the social sciences, though, the Internet is talked about as an enabler; the technology to help level the playing field (Friedman 2005), and create opportunities where they otherwise never would have existed. It creates new relationships between individuals and businesses to the potential benefit of both. For example, workers have access to broader job markets, which at the same time gives businesses access to broader labor markets. Benefits like these are not limited to high-level or tech jobs; at the very least it makes it possible to browse online classifieds for regular jobs in other cities. Also, the Internet has lowered the barriers to starting a business or expanding into new markets, which creates competition, giving consumers more choices and lower prices. Critical research, however, usually examines the Internet as part of a critique of port-Fordism.

Manuel Castells is often cited in regard to what he called the *space of flows*; the places where data are moved and consumed to support ongoing social processes (Malecki 2002). If contemporary societies are a network of flows, then the space of flows is where symbols, technology, and capital are transmitted simultaneously (Hearn 2004). The places in the network society can be seen as being organized into a hierarchy – Castells' *spaces of place* – depending on their importance to the network. A place's cultural and physical characteristics affect its position in this hierarchy, and this position in turn affects those characteristics. Steven Graham, another author often cited by Internet geographers, talks about this complex interaction as an argument against technological determinism, insisting the Internet is not a unidirectional force, and humans and technology are engaged in a recursive relationship that can be likened to that of nature and society (Zook 2005).

The space of flows enables the constant exchange of information required by the post-Fordist informational mode of production (Warf 2001). Those who cannot afford to adapt to the information society will not be able to participate in it, leaving them at a disadvantage in terms of access to labor and political participation. The Internet is described as a tool of the professional class, which helps increase wealth for class members while creating barriers for outsiders (Warf 1997). Adoption of information technology by the professional class has brought new high-tech skill requirements upon some of the working class, widening the existing divide between those who have access to expensive products and education and those who do not.

The development of the Internet's infrastructure has largely mirrored previous development patterns in regard to where fiber-optic "back-bone" lines run and which cities act as major hubs. As Graham describes it, the telecommunication infrastructure is an archipelago of technology-rich islands. Technology like fiber optics allows these islands to be linked over long distances without including the places in between. This is the urban/rural component of the digital divide. And within the islands – the cities – there is another divide. Deregulation of the telecommunications industry led to infrastructure development in places with a lot business investment, and neglect of what Graham called "less powerful users and spaces" (Graham 2000, pp. 185). Awareness of this neglect grew after the National Telecommunication and Information Administration (1995) conducted a series of studies on the "have-nots" in rural and urban America. Debates over the possibility of government intervention helped politicize the issue, and studies on the digital divide became more common across a variety of disciplines. There is a strong spatial component in this kind of research, so the field of geography has the potential to make a significant contribution, particularly in empirical studies at the local level.

Tony Grubesic of Ohio State has been active in meeting this need. Grubesic and Murray (2002) did a study on DSL access in Franklin County using service hub coverage area to gauge what percent of the population lacks access to DSL. They found that 20% of residents are outside of any of the service hubs' 12,000 foot coverage radius; about a quarter of the white population and one-fifth of the black population would be outside, and 23% of the residents with college degrees would be without access. Median household income was actually *lower* in the coverage area. The study was meant to evaluate local-level access in terms of the *physical availability*, not socioeconomic constraints, so it lacks a detailed analysis of the relationship between household attributes and Internet adoption.

Paul Torrens (2008) of Arizona State mapped Wi-Fi networks in Salt Lake City, but his efforts were focused more on the nature of the infrastructure than specifically on the digital divide. He found the city to be blanketed in Wi-Fi and discussed the potential for a centralized networked being implemented across the existing infrastructure.

Other studies of Internet accessibility and adoption have mostly focused on international (Fife 2002) and interurban (Grubesic and O'Kelly 2002) scales. The exceptions lack a spatial component or do not thoroughly examine the attributes of users and nonusers (Horrigan 2008; Jones 2006). This study, however, offers a quantitative assessment at a local scale, focusing on the geographic pattern of adoption of Wi-Fi technology (as a proxy for Internet access) across neighborhoods and analyzing the association with various demographic and socioeconomic attributes.

The value of this research can be summarized as follows. (1) The access point mapping technique applied in this study is fairly inexpensive and minimally intrusive, and thus can be easily implemented in other regions. By doing so, one can assess the extent of the digital divide at a local scale measured in multiple dimensions including demographic and socioeconomic attributes across neighborhoods. (2) Results from research following the technique can help inform planners in implementing municipal Wi-Fi networks in order to achieve the highest overall connectivity of citizens. (3) Based on the statistical relationship between socioeconomic variables and Wi-Fi access density, the technique has the potential for wide usage as a proxy measure for socioeconomic attributes that are not frequently updated and are often expensive to collect.

## STUDY AREA AND SAMPLE SELECTION

The study area consists of 30 non-contiguous block groups in East Baton Rouge Parish (EBRP), Louisiana. A parish is a county unit in Louisiana. According to the 2000 census, EBRP has a total population of 430,770, with the capital city of Baton Rouge in the middle and a dozen satellite towns surrounding it along with unincorporated areas. The geographic unit for the study is census block group, also referred to here as a "neighborhood," which was the smallest unit with the available demographic and socioeconomic variables desired that could accommodate the access point mapping technique.

Most of the secondary data such as the census and corresponding GIS files were readily available from the Census Bureau. In preparing for fieldwork, the first task was to identify sample neighborhoods (block groups) that are representative of the 306 block groups in EBRP. EBRP is a racially and economically segregated area. Figure 1 shows that only 26 of 306 block groups have a black population roughly representative (±5%) of the parish average of 45%. Areas with a high black population percentage are concentrated in the north. A histogram (Figure 2) further demonstrates the division; most of the block groups have either a very high or a very low black population percentage and very few in the middle. Figure 3 shows the distribution of median household income. Medium- and high-income areas are concentrated in the southeast quadrant and scattered in distant suburban areas.

The fieldwork covered 30 block groups (Figure 4), considered a minimum number to ensure meaningful statistical analysis. These 30 block groups were chosen after carefully examining the geographic distributions of demographic and socioeconomic attributes as well as location. Specifically, census data on age, family structure, income, and race were first mapped at the block group level and divided into nine classes in ArcMap based on Jenks' natural breaks

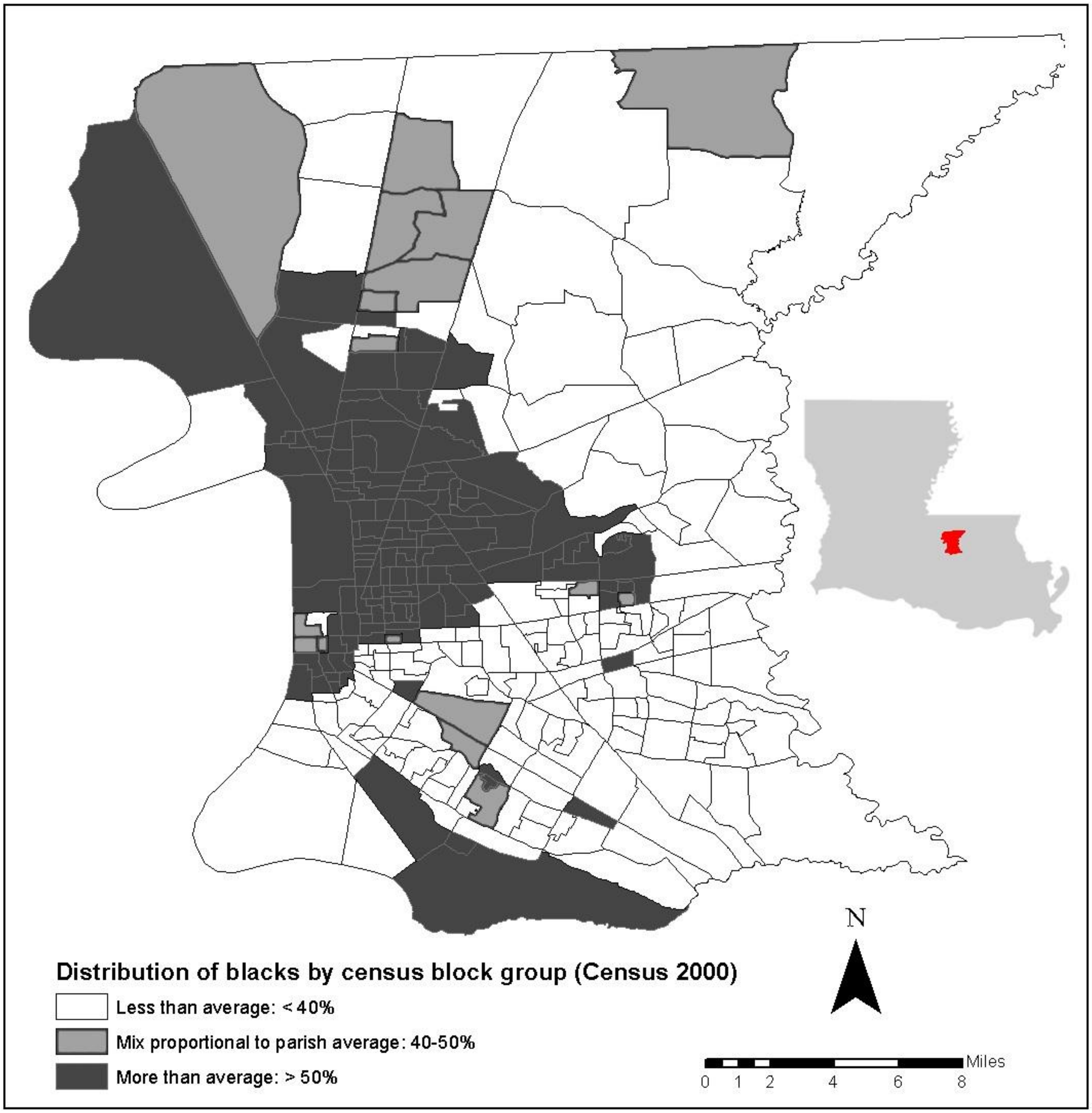


**Figure 1.** Black population in East Baton Rouge Parish, 2000

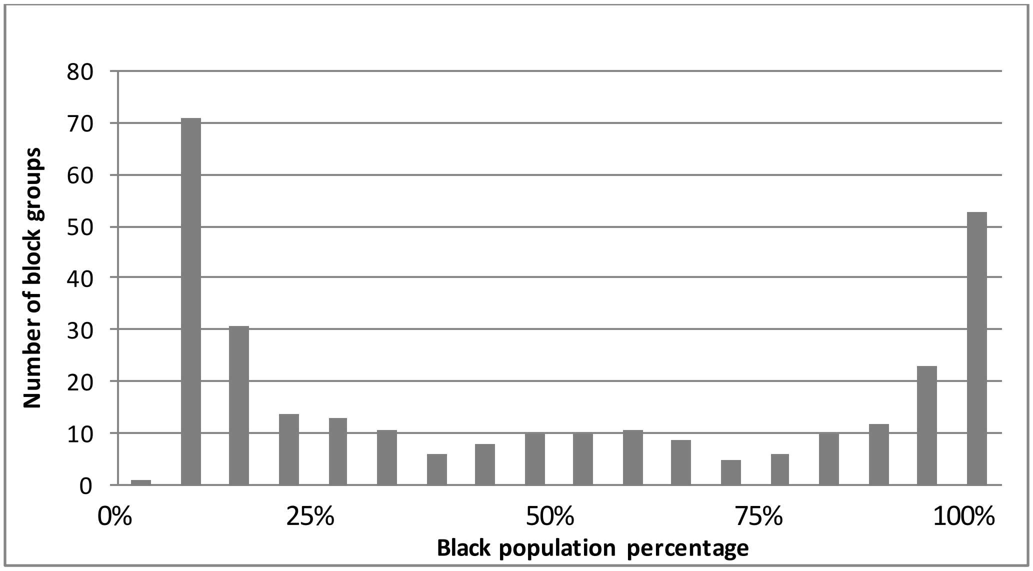


**Figure 2.** Histogram of black population in East Baton Rouge Parish, 2000

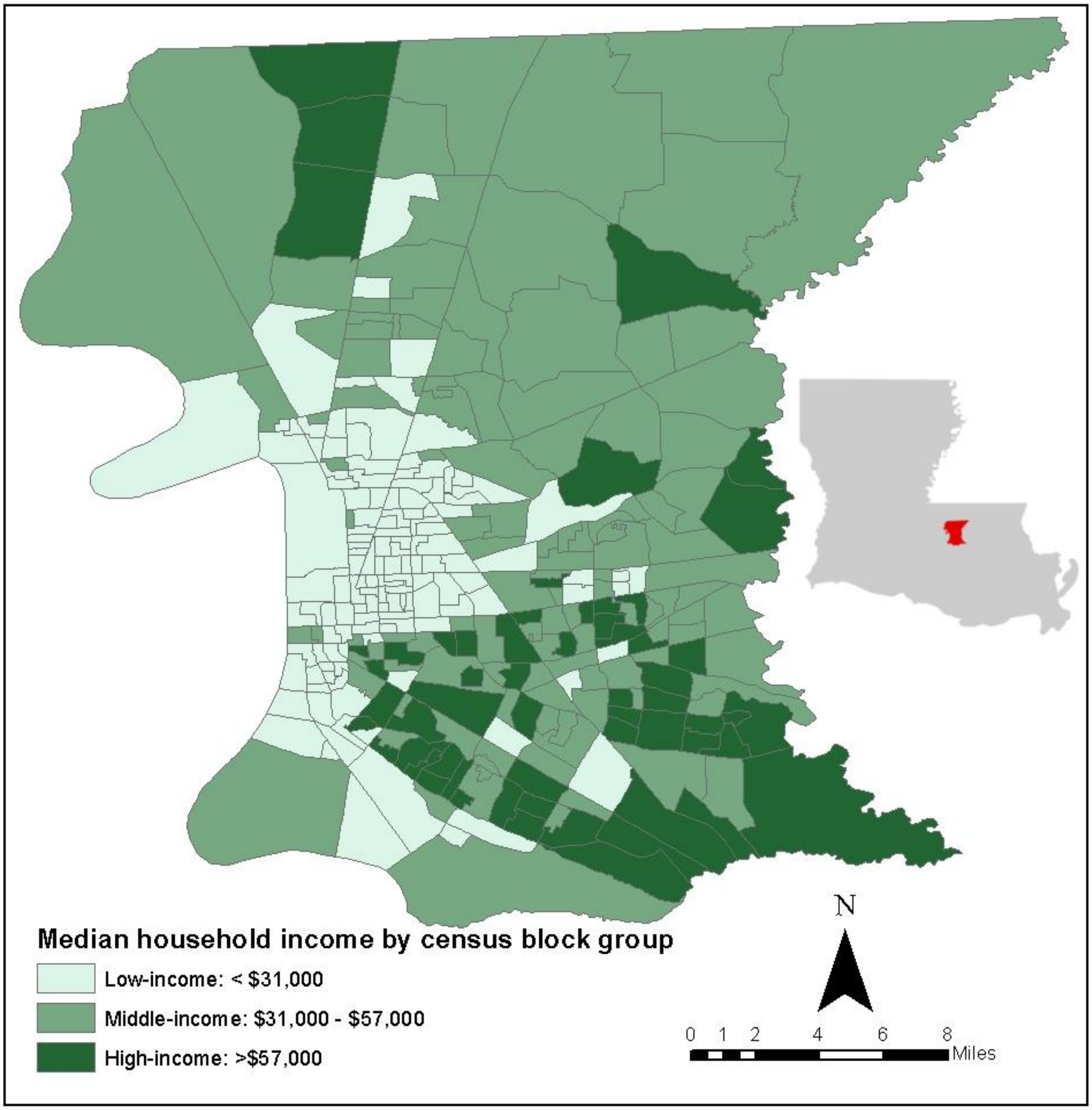


**Figure 3.** Income distribution in East Baton Rouge Parish, 2000

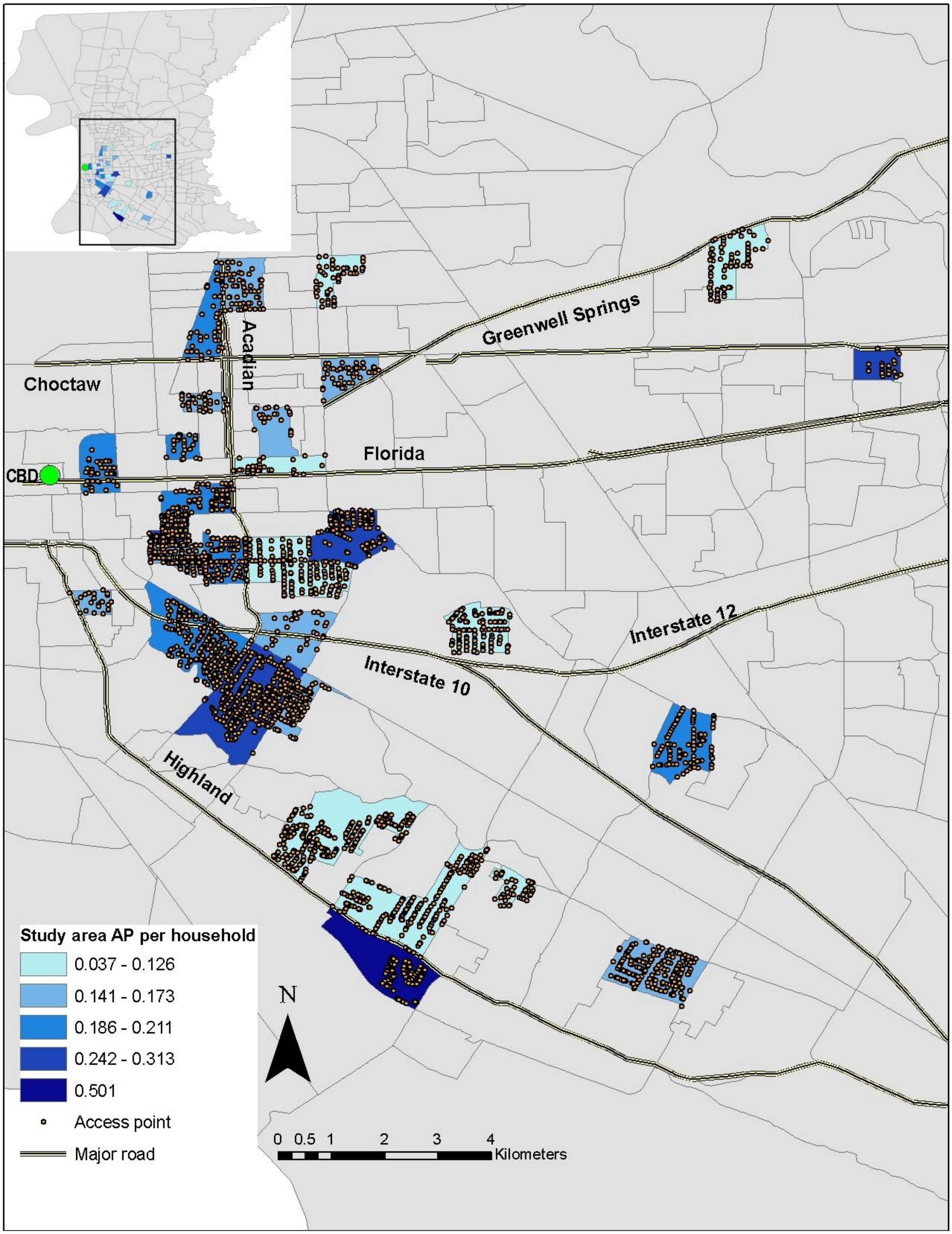


**Figure 4.** Wi-Fi access in 30 selected block groups

classification. From these nine classes, the second, fifth, and eighth classes were highlighted because they represented middle-lower, middle-middle, and middle-upper ranges of values for each variable. This was done to avoid selecting block groups at the extreme ends of the distributions. The selection was further narrowed by keeping block groups that would give a relatively even distribution in terms of distance to and orientation with the central business district. The CBD was defined as a point in downtown Baton Rouge, south of the Capitol building, where building heights are much greater than the rest of the city and land cover is almost 100% concrete. Distance and relative angle were calculated between this point and the centroids of census block groups. Also, Baton Rouge is home to two universities, so care was taken to ensure that selected block groups did not have a disproportionately high percentage of the population between the ages of 18 and 29. The final selections were then made by looking at aerial photos to verify that the block groups were mostly (if not completely) residential, and that the neighborhoods were accessible by public roads and not gated. Since the intention was not to limit the study to clearly delineated subdivisions, some block groups were included in the study despite the occasional presence of businesses, schools, and churches. Even after taking steps to keep objectivity in the selection process, it was still fairly subjective because the driver must ultimately decide what is feasible to drive and what is not, both during preparation and in the field. For example, two particular block groups were rejected, despite meeting other criteria, after fieldwork revealed that they contained inaccessible, gated apartment communities. Admission of subjectivity in a study does not excuse bad data collection and analysis, but it helps make a more informed assessment of the results. With all that being said, the selection process did yield a fairly representative sample. Figure 5 shows the histograms of five selected variables, median household income, percent African American, distance to the CBD, percent with a college degree, and percent aged 18 to 29, from the entire parish versus the 30 sample block groups selected for fieldwork. To test their congruence statistically, the variables were first standardized to have a mean of zero and standard deviation of one. Observations for each variable were divided into bins and summed to make a histogram. The observation frequency of each bin was converted to a percent of total observations. This was done for the sample and population. The differences between percentages for the sample and population were averaged for each bin. For example, median household income for the sample was, on average, off by 2.4% across all bins. Percent African American averaged a difference of 1.3%, distance to CBD 7.1%, percent with a college degree 6.1%, and percent aged 18 to 29 1.4%.

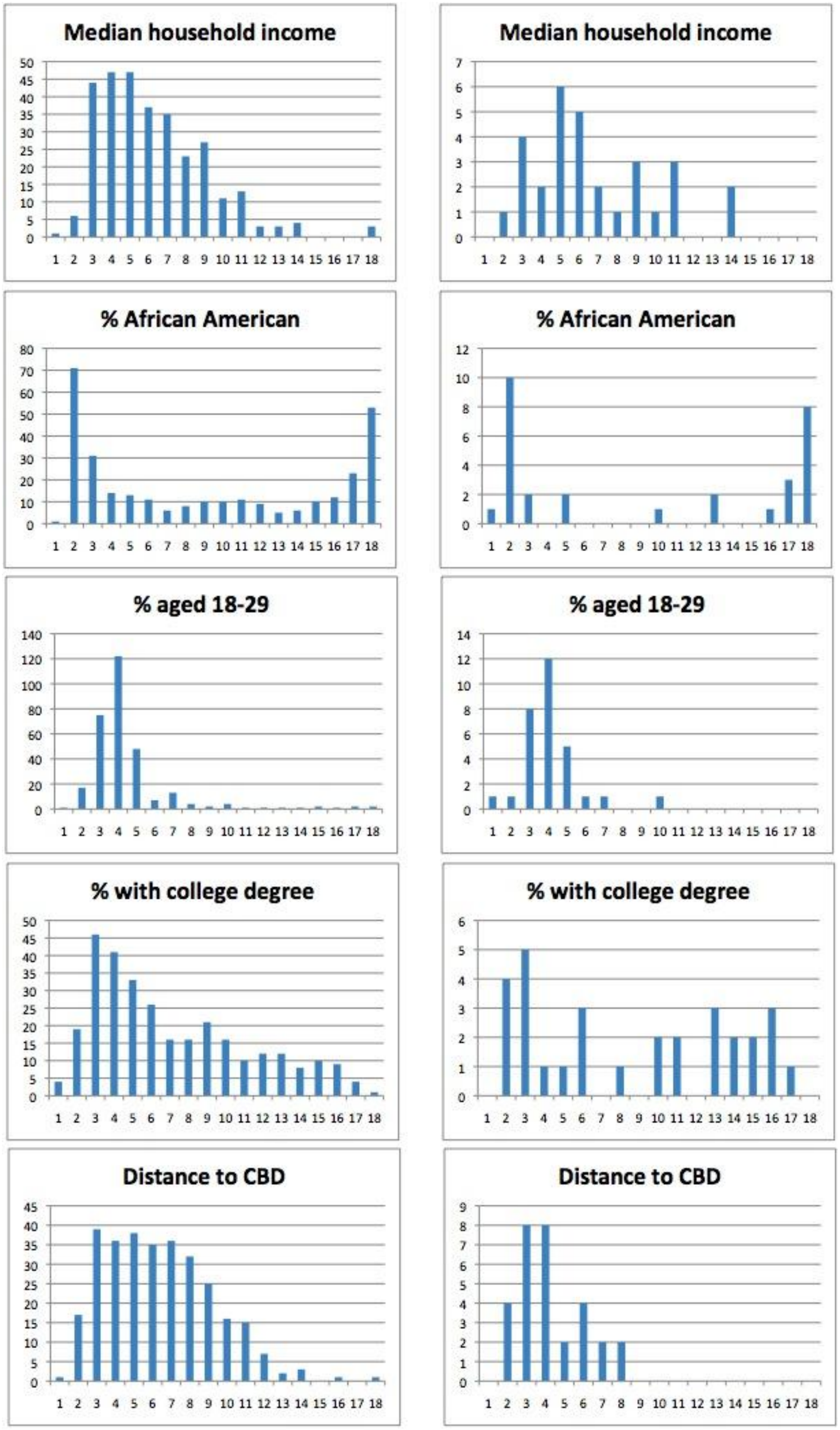


**Figure 5.** Histograms of variables for the entire parish (left) and sample block groups (right)

## DETECTING AND MAPPING WI-FI ACCESS

The presence of a Wi-Fi network in a neighborhood suggests that one of the residents probably has the following high-tech products: a Wi-Fi device (laptop, desktop, PDA, cell phone), a broadband Internet subscription (though it is possible to share a dial-up connection with a wireless router), and networking infrastructure (cable or DSL modem, Wi-Fi router, network cables). The barriers to acquiring these products must have been low enough for residents to obtain them through some means, and their lifestyles must have created sufficient demand to justify the resources (the very least of which could be time) necessary to use and maintain them. The Wi-Fi network density can be measured with a technique that originated with a hobby known as wardriving. Wardriving typically involves driving on the roads in a given area with a laptop and GPS, allowing special wardriving software to detect wireless networks and log corresponding GPS coordinates automatically as the vehicle moves along a route. There are numerous hobbyist websites explaining the process, and some have maps and databases made with point data uploaded by users. One website, Wigle.net (Wireless Geographic Logging Engine), has fairly extensive data for some cities, such as Chicago and Portland, but their data may be more of a reflection of the popularity of wardriving in certain areas than of Wi-Fi adoption. Nonetheless, for some studies it could serve as a useful proxy. There are also companies that collect and sell this kind of data for a number of cities. Researchers studying various aspects of urban Wi-Fi networks have also collected extensive data in some cities, like Torrens in Salt Lake City (2008) and Phillips *et al.* in Portland (2008). It may be possible to use the existing data from these sources to conduct similar studies in other cities without a need for wardriving, or as it is more formally known in the literature, *access point mapping*.

The access point mapping software used is "Kismet," an open-source network tool for Unix-based operating systems. It records coordinates of the vehicle every time a data packet is received from an access point (Wi-Fi router). While the vehicle is in range of an access point's signal, the software repeatedly logs the vehicle's coordinates (not the access point's), and it can record this information for many access points simultaneously. The result is several log files of different types containing coordinate data that follow the vehicle's path on the road system. One of the log files is a CSV containing one record for each detected access point and an averaged GPS coordinate. Another of the log files contains multiple GPS coordinates for each access point as the vehicle moves within the network's coverage area. If the software logs enough coordinates around an access point, GIS software could be used to find the weighted center of those coordinates as an estimation of the access point's actual location (rather than the vehicle's location on the road). This process, however, would be difficult to automate and impossible for access points without a sufficient number of logged coordinates. Furthermore, it is unnecessary here since the data is aggregated within block groups. Therefore, this study uses the CSV log file with a single, averaged coordinate for each detected access point.

After drawing out reasonably efficient routes, the roads in the study areas – about 270 km in combined distance – were driven. A laptop and USB GPS unit were used along with Kismet to automatically detect access points and record the coordinates of the vehicle. Kismet saved the point data to a CSV file that could be imported into ArcMap. The fieldwork was conducted in two rounds in November 2008 and February-March 2009.

As far as privacy is concerned, the otherwise unfortunate imprecision in access point mapping and the need to aggregate to census block groups makes it impossible to associate a detected access point with a specific home based on the point data. Another privacy issue that might come into question with this technique is whether personal data are captured. Since the software passively detects signals and does not actually connect to access points, data being transmitted across a network owner's private network is not accessible. The network name, however, could contain personal information if the owner has chosen to provide it. This could be something like "OurHomeNetwork" or could just be left blank or with the default value (e.g. "linksys"), but some users choose to enter family names, addresses, or other personal identifiers for this field. The access point broadcasts this information to any Wi-Fi enabled device, and the access point mapping software does not need to use any special operation to access the network name. For this study, names were not included in the reports.

To correct for variations in GPS accuracy, the points were automatically snapped to the nearest road using ArcMap. Another problem that needed correction was the detection of networks from neighboring block groups. One solution was to drive some roads just outside of a study area, likely causing Kismet to log the neighbor's points on those exterior roads instead of inside the study area. Some points were removed after being mapped if there was an obvious cluster along the border. Additionally, points were deleted based on non-spatial irregularities. For example, in a low-income block group where there were few networks, there was a string of access points on a road that ran along the back of a shopping center. The access points all had matching names and their manufacturer-assigned unique identifiers (BSSID) were the same up until the last few digits, suggesting the routers were part of a unified, commercial system in the shopping center and their signals were bleeding over into the residential area.

The final set of point data consisted of 4,111 access points across 30 block groups, which were shown in Figure 4. Access point density (Wi-Fi router ownership rate) was calculated by dividing the number of detected access points in a block group by the number of households. A per-household measurement was preferred over per-capita because multiple users in one household can share a network. In other words, five people sharing a connection do not necessarily have less accessibility than two people sharing a connection.

The access point density is used as a proxy for Internet access since survey data was not available for the percent of Internet subscribers in Baton Rouge also owning a Wi-Fi router. The next section will examine the statistical relationship between socioeconomic variables and Wi-Fi access density.

## ASSESSING THE DIGITAL DIVIDE

The issue of the digital divide may be applied to population groups defined by age, racial and ethnic identity, income, and other socioeconomic attributes. There is a rich set of variables available from the decennial census. First, 19 variables were selected from the 2000 census, most of which are commonly used in the literature to identify possible "disadvantaged population groups" (Wang 2008). The initial assessment was to simply obtain the correlation coefficient between each of these variables and Wi-Fi access point density as shown in Table 1. Bear in mind that all neighborhood attributes were from the 2000 census and the Wi-Fi access data were collected in 2008-09. The significant time gap may affect the statistical fitting power of the analysis. Population characteristics may have changed as people moved into and out of neighborhoods in the pursuit of jobs, comfort, or lower land prices. Areas where residential mobility was affected by gentrification, for example, may translate into a dataset with observations having an artificially high number of access points (from the new, wealthy residents) for a given income level (median household income from the 2000 census). This would decrease the two variables' measured correlation.

From Table 1, a low access point density was most closely associated with a high black population percentage (with a correlation coefficient of -0.926). The white population percentage was not included in the analysis because percents of white and black have a correlation close to (-1.0). Other minority groups (e.g., Hispanic and Asian) were not considered in the study due to their low percentages (< 5%) in the study area. As expected, percentage of the population with a college degree is highly correlated with access point density (with a correlation coefficient of 0.904). Percent of single parent households was significantly correlated (coefficient -0.853), but more so for single female households (-0.843) than single male households (-0.702). The

**Table 1.** Correlation coefficients between an extended set of variables and access point density

| | Correlation coefficient | *t* value (* significant at .001) |
|---|---|---|
| black pop. % | -0.926 | -42.53* |
| college degree % | 0.904 | 36.79* |
| single parent % | -0.853 | -28.44* |
| single female parent % | -0.843 | -27.24* |
| median household income | 0.787 | 22.20* |
| average family size | -0.776 | -21.39* |
| single male parent % | -0.702 | -17.14* |
| age 5 to 17 % | -0.647 | -14.74* |
| households married no child | 0.631 | 14.14* |
| home ownership % | 0.577 | 12.26* |
| age under 5 % | -0.572 | -12.11* |
| households married w/ child | 0.524 | 10.69* |
| average household size | -0.508 | -10.25* |
| household renter % | -0.497 | -9.95* |
| median age | 0.459 | 8.98* |
| age 30 to 64 % | 0.373 | 6.99* |
| age over 64 % | 0.295 | 5.38* |
| age 18 to 29 % | 0.049 | 0.85 |
| single occupant household % | 0.008 | 0.13 |

correlation coefficient for median household income was high (0.787), but lower than expected since income is the ultimate inhibitor for in-home technology adoption. Large family size was negatively correlated with access point density (-0.776). Age variables had relatively low correlation coefficients. Areas with many primary and secondary school-aged children had the highest at -0.572, and the 18 to 29 group had the lowest at 0.049 (not significant). And finally, a variable not taken from the census data, distance to the CBD, had no statistically significant correlation (not presented in Table 1).

The above bivariate correlation analysis is useful, but it only reveals some preliminary assessments of the relationship between Wi-Fi access point density and neighborhood attributes.

Since several variables contained duplicated information, the number of variables was reduced to nine (Table 2), but as shown in Table 3, some of the remaining demographic and socioeconomic variables were highly correlated. A multivariate regression using these variables directly would be misleading due to multicollinearity (Hamilton 1992). A large number of independent variables is not desirable with such a small number of observations, and the additional inputs will artificially inflate the $R^2$ value, making it a less meaningful measure of the model's fitting power. Principal components factor analysis was used to consolidate these variables into a small number of independent factors and uncover latent variables for easy interpretation (Wang 2006). According to the eigenvalues from the principal components analysis, two factors accounted for almost 80% of the variance. In other words, the two factors preserved nearly four-fifths of the information contained in the original nine variables. Therefore, two factors were used in the subsequent factor analysis.

Table 4 shows the factor loadings of each variable on the two factors. The variables in Table 4 are reordered so that the variable with the highest loading is placed first and so on. Factor 1 is labeled "socioeconomic attributes." It captures variables related to socioeconomic status such as percentage of female-headed households, larger families, higher black population percentage, and lower educational attainment. A higher factor 1 value indicates lower socioeconomic status. Factor 2 is labeled "household attributes" because age, children, occupancy, and home ownership are loaded heavily on it. A higher factor 2 value represents stronger presence of young nuclear families (highly negatively correlated with single occupant households) and higher home ownership. Note that the loadings of variables "median age" and "median household income" are split between the two factors. The wording used in each factor's label is largely irrelevant and is meant to loosely describe the factor with a meaningful heading, rather than simply calling them "factor 1" and "factor 2."

**Table 2.** Basic statistics for nine variables

| | Mean | Std. dev | Min | Max |
|---|---|---|---|---|
| A. median household income (dollars) | $38,804 | $21,400 | $3,594 | $129,133 |
| B. college degree (%) | 27.44 | 20.50 | 0 | 83.46 |
| C. black population (%) | 45.05 | 37.66 | 0.15 | 99.62 |
| D. average family size | 3.14 | 0.31 | 2.23 | 3.97 |
| E. median age | 33.82 | 6.62 | 18.1 | 51.5 |
| F. single occupant households (%) | 26.55 | 12.14 | 4.01 | 76.12 |
| G. married households w/ child (%) | 19.65 | 10.02 | 0 | 54.42 |
| H. single female parent household (%) | 11.03 | 8.66 | 0 | 50.15 |
| I. home ownership (%) | 59.19 | 25.55 | 0.29 | 97.82 |

**Table 3.** Correlation coefficients for nine variables (See Table 2 for full variable names)

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| B | 0.581 | | | | | | | |
| C | -0.695 | -0.652 | | | | | | |
| D | -0.234 | -0.637 | 0.655 | | | | | |
| E | 0.593 | 0.263 | -0.446 | -0.327 | | | | |
| F | -0.411 | 0.203 | 0.012 | -0.576 | -0.041 | | | |
| G | 0.729 | 0.162 | -0.458 | 0.201 | 0.261 | -0.736 | | |
| H | -0.566 | -0.625 | 0.748 | 0.637 | -0.595 | -0.180 | -0.306 | |
| I | 0.705 | 0.125 | -0.403 | 0.104 | 0.614 | -0.661 | 0.745 | -0.399 |

**Table 4.** Factor loadings for nine variables

| | Socioeconomic attributes (factor 1) | Household attributes (factor 2) |
|---|---|---|
| single female parent household % | **0.8769** | -0.1792 |
| average family size | **0.8603** | 0.3715 |
| black population % | **0.8532** | -0.2579 |
| college degree % | **-0.8205** | 0.0103 |
| median age | **-0.5696** | 0.3949 |
| married households w/ child % | -0.1528 | **0.9049** |
| home ownership % | -0.2320 | **0.8886** |
| single occupant households % | -0.3442 | **-0.8781** |
| median household income $ | -0.6005 | **0.7104** |

The socioeconomic attributes factor scores were highly correlated with access point density ($R^2$ = .81), as shown in Figure 6. The scatter plot also shows points in two clusters, high scores with low density and low scores with high density. This partially reflects the residential segregation pattern discussed earlier as most neighborhoods are highly concentrated with either white or black populations and very few in between. The household attributes factor had a relatively lower correlation with access point density ($R^2$ = .13, marginally significant at $p$ = .05). Its scatter plot shows two clusters of points running roughly parallel, with the bottom cluster being shifted to higher access point density (Figure 7). As with the scatter plot of the first factor, the clusters are a result of the racial distribution. Consider one of the components of the household attribute factor, household ownership. Its scatter plot with access point density also has two clusters that roughly fit two parallel lines. When the data are plotted for only the twelve block groups with the highest black population percentages (lowest white percentages), the points follow a single, linear pattern. The poor fit of the linear regression model to the household attribute factor may therefore be misleading since it is trying to use a single model to describe what are basically two sets of data for the two major race groups.

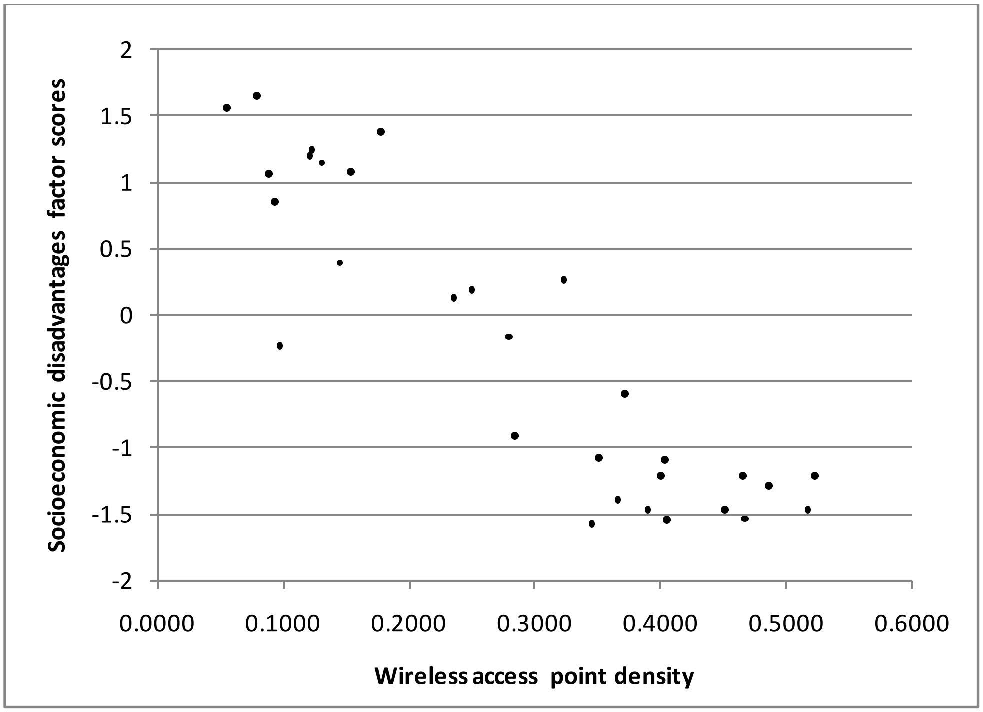


**Figure 6.** Scatter plot of Wi-Fi access point density vs. "socioeconomic attributes factor"

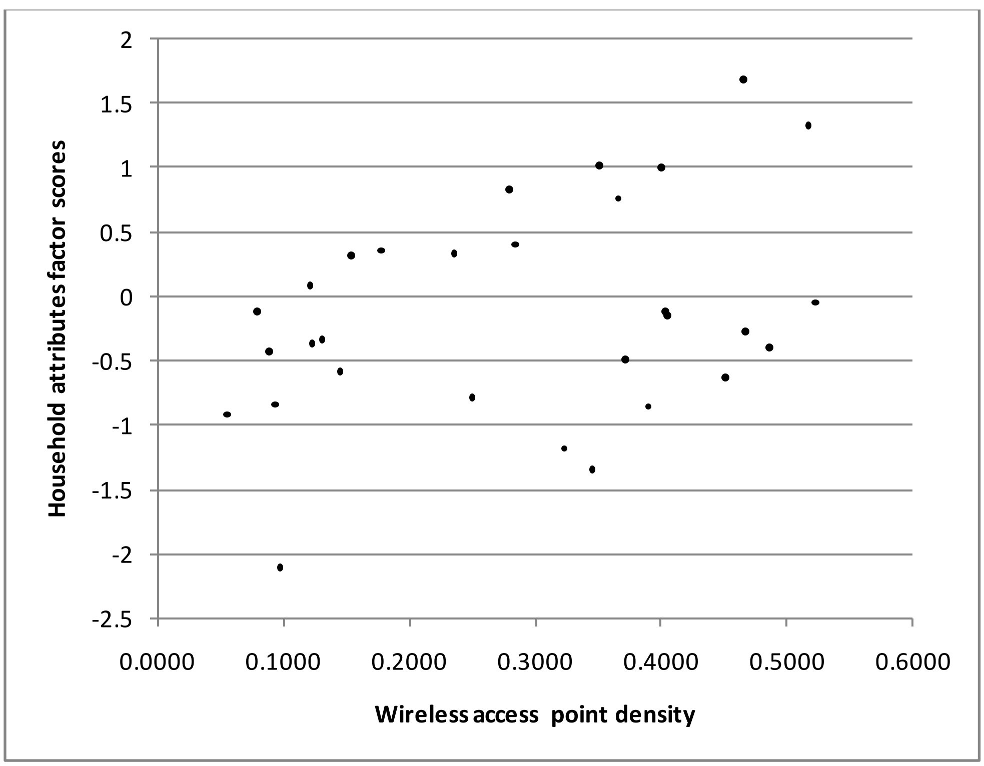


**Figure 7.** Scatter plot of Wi-Fi access point density vs. "household attributes factor"

A multiple linear regression model and a spatial lag model are used to assess the combined effects of both factors on access point density. Ordinary least square (OLS) regression assumes that all observations are spatially independent, and the spatial lag model accounts for spatial autocorrelation of the dependent variable (Wang 2006). The OLS regression including both factors yields the following model:

$$Y = 0.2624 - 0.1136X_1 + 0.0326X_2 \quad (1)$$

where $Y$ is the access point density, $X_1$ is the socioeconomic attributes factor, and $X_2$ is the household attributes factor for each block group. Note that the spatial lag term is not statistically significant in the spatial lag model. That is to say, the access point density is spatially independent across the samples in the study area. One may expect areas of low or high density values to be clustered together. However, most census block groups in the sample are not contiguous with one another and are spatially independent. This analysis used rook contiguity based on shared borders. Using a distance-based contiguity could have affected the spatial lag coefficient. The two models yield very similar results with an almost identical $R^2$ value of 0.844. Table 5 presents the results of both models.

**Table 5.** Regression models for explaining access point density

| | OLS regression | Spatial lag model |
|---|---|---|
| Intercept coefficient | 0.2624 (22.4)*** | 0.2631 (17.43)*** |
| Socioeconomic attributes factor coefficient | -0.1136 (-11.1)*** | -0.1140 (-10.22)*** |
| Household attributes factor coefficient | 0.0326 (2.33)* | 0.0323 (2.35)* |
| Spatial lag coefficient | - | -0.0046 (-0.07) |
| $R^2$ (squared correlation coefficient) | 0.844 | 0.844 |
| Note: t values for the OLS model and z values for the spatial lag model are in parenthesis; *** significant at 0.001, * significant at 0.05. | | |

Aside from the $R^2$, mean absolute percent error (MAPE) is another measure of a model's goodness of fit. It takes the difference between predicted and observed values as a percent of the observed value and averages these differences for all observations. A lower percentage means there was less error in the prediction. The two-factor OLS regression model's MAPE for the 30 observations was 20.03%. If regression is run with the nine variables used to compute the two factors, the MAPE is 21.86% and the $R^2$ is .907. This means that even though using nine variables results in a higher $R^2$, it does not help make more accurate predictions, as reflected by the higher MAPE. To further test the accuracy of the two-factor OLS regression model's predictions, the observations were split into training and test sets. Five random observations were taken out of the set of 30, and the remaining 25 were used to build new two-factor and nine-factor OLS regression models. The models were then used to make predictions for the five observations that had been removed. The two-factor model had a MAPE of 8.62%, while the nine-factor model had a much higher error of 40.7% for the five-observation test set. The large difference between the two could be attributable to the specific observations used for the training and test sets if the error values are showing sensitivity to some outlier present in the subset of the five test records. For this study, only one split was tested, but any future studies that focus on building accurate prediction models could split the data several times, or test every possible combination of split/train sets. Finally, with the two-factor approach to OLS regression appearing to be more accurate, it was used to make predictions for all census block groups in the parish. Figure 8 shows East Baton Rouge Parish block groups with predicted access points per household based on the regression model defined in equation (1).

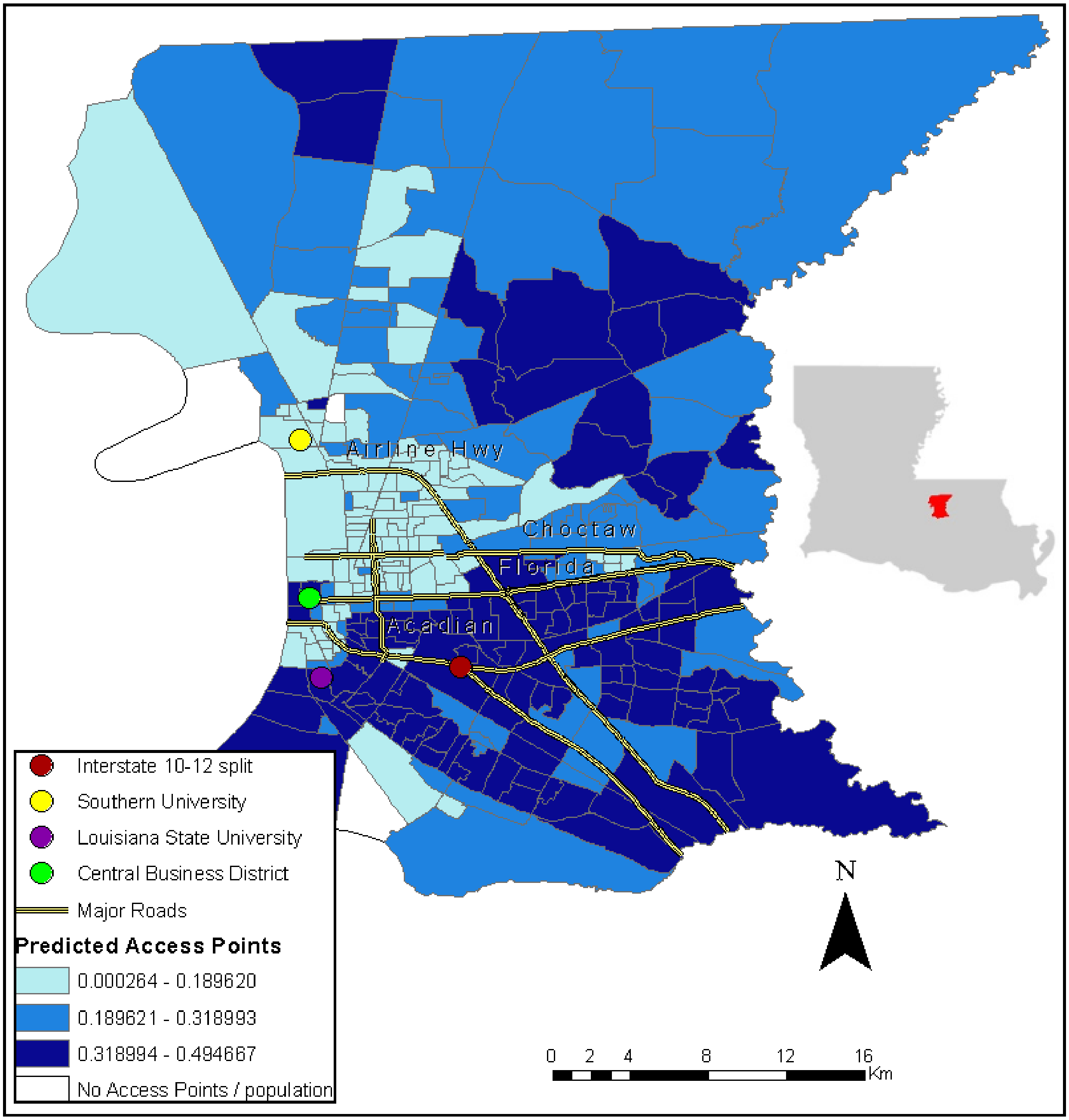


**Figure 8.** Predicted access points in census block groups in East Baton Rouge Parish

## DISCUSSION

Broadly, the results indicate that low-income households headed by a black, single, female parent with multiple children and no college degree will probably have the lowest rates of Wi-Fi ownership. While the general results may not be surprising, the varying correlations of each of these attributes with Wi-Fi ownership are more intriguing. Besides cases of donations, loan programs, or other exceptions, lack of money is the ultimate inhibitor for purchasing a Wi-Fi router and related technology products, yet income was less significant than most of the other socioeconomic and demographic variables. Instead, race, an attribute with no inherent inhibitors to technology adoption, is more closely associated with Wi-Fi ownership. To get in-home access to the Internet, a potential subscriber will of course need to have enough income to buy the necessary technology products, but also a lifestyle that creates sufficient demand for the products to justify purchasing them. The study area with the lowest access point density had a median household income of $16,935. An income this low may not make purchasing technology feasible, but it is still – numerically – enough money to buy the basic technology needed for in-home Internet access. So actually having the money is less of a barrier to entry into the digital world than is having justification for spending the money. The barrier is lower for those with high income and higher for those with low income, and a household that can easily afford technology with on-hand cash will not have to give themselves much justification for purchasing it. On the other hand, a household who has to save to safely afford technology, or must finance the purchase, will need much more justification for spending a significant portion of income. But even if a household has the funds to make the purchase, it still may not be worth it if the technology is not demanded by everyday life. This justification comes from their culture. This means that two households with the same income may have different levels of high-technology adoption because their respective cultures create varying demand. This is why it is important to consider variables other than income when studying Internet adoption. The high correlation between race and Wi-Fi ownership seems to indicate that intrasocietal differences between white and black populations lead to disparate demands for Wi-Fi and associated technologies.

Aside from the quantitative analysis of Wi-Fi adoption in Baton Rouge, this study offers an example of how to apply access point mapping as a tool of geographic inquiry. The methods in this study allow researchers to gather data on technology adoption in specific, delineated geographic areas for which other data already exist, meaning the collected data can be easily compared to the results of other studies. The scale on which the data can be collected is only dependent on how thorough the researcher wishes to be. Walking through apartment complexes, shopping centers, and yards could produce very accurate point data down to the census block level, while driving is more feasible for the block group level. Looking at differences in technology adoption in small areas makes the concept of a digital divide more tangible, and demonstrates how there is actually a physical separation of people with and without access to certain technology. It applies Castells and Graham's high-level theories derived from large-scale data to the local. One particularly useful application of access point mapping is in planning municipal Wi-Fi networks. When a city government decides to subsidize low-cost or free public Wi-Fi, the goal is to increase overall connectivity of citizens and spur economic growth (Bar 2005). To meet the first part of that goal, the wireless access points should be selectively placed in areas with low rates of Wi-Fi coverage, rather than in an indiscriminate grid centered on the CBD, which may satisfy the second part. Municipal Wi-Fi ventures in some U.S. cities, including Baton Rouge, have failed to attract users (Gautreau 2007). Understanding the existing Wi-Fi infrastructure would help inform planners before implementing municipal Wi-Fi (Torrens

2008). If a city does decide to build a Wi-Fi network, it is in the interest of the beneficiaries and tax payers that it be implemented efficiently in areas that will attract enough users to justify its existence.

Another interesting aspect of access point mapping is its possible use as a proxy measure of other variables. Just as vehicle and telephone ownership have served as socioeconomic variables, information and communication technology ownership could be an easily obtainable supplement to incomplete datasets. The data are passively observed and thus not dependent on the availability, willingness, and honesty of people. The cost of acquisition is low, consuming more time than anything else. And the actual data collection – the wardriving – involves little more than starting a program on a laptop and driving a vehicle, so there are few restrictions on who can do the work.

Similarly, the regression models used to assess correlation between access point density and other variables can be used to predict access point density outside of the study areas. Because Baton Rouge's black and white populations are largely separated, it might be more appropriate to use two prediction models for each population, or even have a third for mixed areas. Having two or three models would make this study more applicable to other cities, perhaps those looking to implement municipal Wi-Fi but first need to make inferences about the city's current access. Regression methods beyond OLS could also help make more accurate predictions. Because access point density is merely a proxy for Internet access, obtaining precise values is not all that important. Independent variables could be discretized and used with Bayesian classification to predict access point density in simple terms of "high," "medium," and "low" categories, for example.

## CONCLUSION

This paper mostly references theoretical perspectives put forth by geographers, which are based on literature from other disciplines, but a more comprehensive, multidisciplinary approach could bring together theories from anthropology, sociology, political science, and economics to expand our understanding of how technology is adopted by individuals and institutions. Geography is especially useful, though, because the virtual landscapes of high technology do not escape spatial variability. As the Internet spreads through countries, cities, neighborhoods, and households, it seems to be following a path that aligns with existing socioeconomic inequalities. The extent and variability of this diffusion can be quantified and mapped if the presence of a Wi-Fi network is accepted as a proxy for high-technology adoption. Regression results from the small dataset showed some correlation between Wi-Fi presence and income, but other factors like race and family type were more closely correlated. Certain cultures create a demand for technology that outweighs simple affordability when users are considering adopting technology. Identifying the characteristics of these cultures will help determine why people end up on different sides of the digital divide. However, quantitative research needs to be supplemented with interviews with people in the study areas. Stories of why people have or have not adopted technology will help build a more complete description of users and non-users, which will help researchers make informed conclusions on the relationships between people and information technology.

## REFERENCES


Bar, François and Namkee Park. 2005. Municipal Wi-Fi networks: the goals, practices, and policy implications of the U.S. Case. Paper presented at the First Transatlantic Telecom Forum, November 22, Montpellier, France.

Byers, Simon, and Dave Kormann. 2003. 802.11b access point mapping. *Communications of the ACM.* Vol. 46, No. 5: 41-46.

Fife, Elizabeth and Francis Pereira. 2002. Socio-Economic and Cultural Factors Affecting Adoption of Broadband Access: A Cross-Country Analysis. Paper presented at the 41st European Telecommunications Congress, September 11-14, Genoa, Italy.

Friedman, Thomas L. 2005. The World Is Flat: A Brief History of the Twenty-first Century (New York: Farrar, Straus, and Giroux).

Gautreau, Chris. 2007. JoVoGo to shut downtown Wi-Fi. *The Advocate*. Baton Rouge, Louisiana, June 30, sec. D.

Graham, Steven. 2000. Constructing Premium Network Spaces: Reflections on Infrastructure Networks and Contemporary Urban Development. *International Journal of Urban and Regional Research*. Vol. 24, No. 1: 183-200.

Grubesic, Tony and Alan Murray. 2002. Constructing the divide: Spatial disparities in broadband access. *Papers in Regional Science*. Vol. 81, No. 2: 197-221.

Grubesic, Tony and Morton E. O'Kelly. 2002. Using points of presence to measure accessibility to the commercial Internet. *The Professional Geographer*. Vol. 54, No. 2: 259-278.

Hamilton, Lawrence C. 1992. *Regression with Graphics* (Belmont, CA: Duxbury).

Hearn, Jeff. 2004. *Information Society and the Workplace: Spaces, Boundaries and Agency*. London: Routledge.

Horrigan, John B. 2008. Home broadband adoption 2008. *Pew Internet and American Life Project*. Report prepared by Pew Internet.

Jones, Kipp and Ling Liu. 2006. What, where, Wi: An analysis of millions of Wi-Fi access points. Atlanta: Division of Computer Science and Systems, Georgia Institute of Technology.

Malecki, Edward. J. 2002. The economic geography of the internet's infrastructure. *Economic Geography*. Vol. 78, pp. 399-424.

National Telecommunications and Information Administration (NTIA). 1995. *Falling Through the Net: A Survey of the 'Have Nots' in Rural and Urban America*. http://www.ntia.doc.gov/ntiahome/fallingthru.html (last accessed May 21, 2009).

Phillips, Caleb *et al*. 2008. Robust coverage and performance testing for large-area wireless networks. Paper presented at the third International Conference on Access Networks, October 15-17, in Las Vegas, Nevada, USA.

Torrens, Paul M. 2008. Wi-Fi Geographies. *Annals of the Association of American Geographers*. Vol. 98, pp. 59-84.

Wang, Fahui. 2006. *Quantitative Methods and Applications in GIS*. Boca Raton, FL: Taylor & Francis.

Wang, Fahui *et al*. 2008. Late-Stage Breast Cancer Diagnosis and Health Care Access in Illinois. *The Professional Geographer*. Vol. 60, No. 1:54-69.

Warf, Barney. 2001. Segueways into cyberspace: multiple geographies of the digital divide. *Environment and Planning B: Planning and Design*. Vol. 28, pp. 3-19.

Warf, Barney and John Grimes. 1997. Counterhegemonic Discourses and the Internet. *The Geographical Review*. Vol. 87, No. 2: 259-274.

Zook, Matthew. 2005. Geographies of the Internet. *Annual Review of Information Science and Technology*. Vol. 40, pp. 53-78.

# APPENDIX

Source code to calculate factor scores in SAS.

```
/*By Fahui Wang on 2-4-05                                */

/* read the attribute data */
proc import datafile="C:\projects\wifi2\wifipca.csv"
     out=wf1 dbms=dlm replace;
     delimiter=', ';
     getnames=yes;
proc means;

/* Run factor analysis */
proc factor out=fscore(replace=yes)
  nfact=3 rotate=varimax; /* 3 factors used */
  var x1-x9;
/*export factor score data */
proc export data=fscore dbms=csv
     outfile="C:\projects\wifi2\factscore.csv";
```

# VITA

Luke Driskell studies information technology and geography at Louisiana State University. He completed the undergraduate program in information systems and decision sciences in 2008, and is expected to graduate with a master's in geography in 2010. His research interests include Internet accessibility and Wi-Fi mapping.