GEOSPATIAL TOOL EVALUATING JOB LOCATION MISMATCH, BASED ON AVAILABLE WORKFORCE AND TRANSIT OPTIONS

Evaluating property location in a city using large-scale datasets.

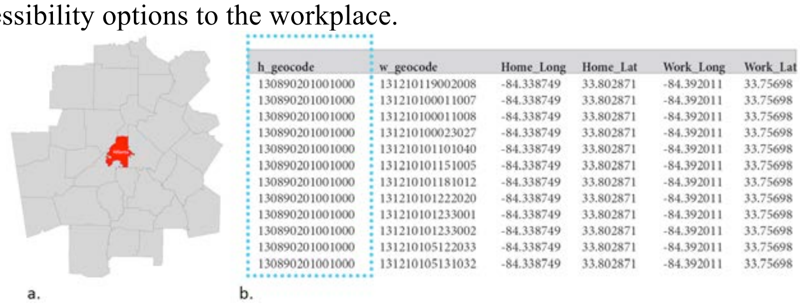
FLORINA DUTT AND SUBHAJIT DAS
Georgia Institute of Technology, Atlanta, USA
Email address: florina.design@gatech.edu, das@gatech.edu

Abstract. The paper addresses the issue of spatial mismatch of jobs and the accessibility to job locations based on different age, income and industry group. Taking Atlanta as a case study, we developed a geospatial analysis tool enabling developers, the city planning bureau and the residents to identify potential sites of redevelopment with better economic development opportunities. It also aids to find potential location to live with respect to user's choices for transit options, walkability, job location and proximity to chosen land use. We built our model on a block level in the city, imparting them a score, visualizing the data as a heat map. The metrics to compute the score included proximity to job, proximity to worker's residence, transit availability, walkability and number of landmark elements near the site. We worked with Longitudinal Employer-Household Dynamics (LEHD) Data along with residence area characteristics (RAC) and work place area characteristic (WAC) data sets, where the total number of data-points was over 3 million. It was challenging for us to optimize computation such that the prototype performs statistical analysis and updates visualization in real time. The research further is prototyped as a web application leveraging Leaflet's Open Street Maps API and D3 visualization plugin. The research showed that there is a high degree of spatial mismatch between home and job locations with very few jobs with driving distance within 5 -10 miles with limited transit options in Atlanta. Further, it showed that low-earning workers need to travel significantly larger distance for work compared to higher class.

1. Introduction

There has been rising anxiety in Atlanta related to spatial mismatch of available workforce and job accessibility. Atlanta region's growth is evidently unbalanced, creating a stark divide between the affluent North and disadvantaged South (Policy, 2000). The nature of the problem has either produced a ripple effect of economic growth in certain communities or completely confounded communities with unemployment and degradation. The various studies and empirical inquiry from various sources bolsters the fact that impoverished conditions of the neighborhoods in Atlanta are due to poor job accessibility geographically as well as due to job –education mismatch. Fewer jobs are available within walking distance of public transit stops (Ihlanfeldt, 1993).

The research aims to address the issue of spatial mismatch of jobs and study the accessibility to jobs locations based on different age groups, income group and industry category for any given city. The investigation has two parts. First part accounts for the socio-economic parameters and conditions of available workforce in a specific geographical area (blocks). The second part identifies the degree of current mismatch based on accessibility options to the workplace.



h_geocode	w_geocode	Home_Long	Home_Lat	Work_Long	Work_Lat
130890201001000	131210119002008	-84.338749	33.802871	-84.392011	33.75698
130890201001000	131210100011007	-84.338749	33.802871	-84.392011	33.75698
130890201001000	131210100011008	-84.338749	33.802871	-84.392011	33.75698
130890201001000	131210100023027	-84.338749	33.802871	-84.392011	33.75698
130890201001000	131210101101040	-84.338749	33.802871	-84.392011	33.75698
130890201001000	131210101151005	-84.338749	33.802871	-84.392011	33.75698
130890201001000	131210101181012	-84.338749	33.802871	-84.392011	33.75698
130890201001000	131210101222020	-84.338749	33.802871	-84.392011	33.75698
130890201001000	131210101233001	-84.338749	33.802871	-84.392011	33.75698
130890201001000	131210101233002	-84.338749	33.802871	-84.392011	33.75698
130890201001000	131210105122033	-84.338749	33.802871	-84.392011	33.75698
130890201001000	131210105131032	-84.338749	33.802871	-84.392011	33.75698

Figure 1 a. Red overlay represents the geographical extent of analysis, while the gray overlay represents the whole Atlanta Metropolitan Area. b. Latitude - Longitude table.

2. Geographical Context and Use case

The current exploration is restricted to the home census blocks within the city of Atlanta (6664 blocks under the study). The number of primary jobs in each census block is considered for each home block. Accessibility measures are obtained for home blocks within the city of Atlanta. The destination or work blocks spread across Atlanta MSA (Metropolitan Statistical Area) (See Fig 1a). Subsequently, we evaluated and ranked geographic locations (block groups) based on job suitability, accessibility, and consideration for future redevelopment. Taking the research further, we developed a prototype of a web application leveraging Leaflet's Open Street Maps API (JS, 2011) and D3 visualization plugin (Figure 2a and 2c). We foresee the use of the tool by residents or visitors to a city, trying to evaluate potential location for work, stay or other businesses. The web application has an input panel on the right, which takes user input i.e. the name of the city, user preference or weights on diverse subjective characteristics like, work location distance, transit time to work, driving distance, walking distance, nearby restaurants, hospitals, shopping malls, etc. Based on the user weights and on the computational model as described below the application computes a heat map at block level for the city visualizing it on the left panel. Each block gets a score, which decides the color overlay for it. This gives a quick overview of potential areas or location in the city, which meets the users need.

3. Related Works

Wang et al. researched to calculate job accessibilities by transportation modes using buffering and network analysis operations. One of the drawbacks of their model is that they did not consider disaggregating by job categories or by different social groups like race, income, etc. They also did not consider built environment qualities and land use types in their statistical model (Wang & Chen 2015). Work of Hadas et al. analyzed the performance of public transit networks with respect to coordination and their connectivity on a case study in Auckland, New Zealand. Attributes they considered were passenger transfers, ride, walk and wait times and type of transfers made. They leveraged Google transit data to develop a tool as a GIS package to evaluate pros and cons of defined zones of transit lines by comparing and analyzing transit network alternatives (Hadas & Ranjitkar 2012). Jianquan et al. showed a six step GIS-based methodological framework to measure urban job accessibility, validated via a case study on Amsterdam. Their work depicted a modified measurement to represent, measure and interpret job accessibility with respect to competition, distance decay, and job diversity. One of the limitations of their work is a disregard to segmented job data, travel modes like a car, public transport and cycling (Cheng & Bertolini, 2013). Kim et al. found that same demographic, socioeconomic and spatial conditions had varied effects on workers in complex ways (Kim *et al.*, 2012).

4. Data Sources and Structure

We used data from various sources. First, we used block data from Census and American community survey summary files. Then, we used Longitudinal Employer-Household Dynamics (LEHD) data along with residence area characteristic data (RAC) and workplace area characteristic (WAC) to find the origin-destination data for jobs, jobs in different age group, income category, and industry type(Census). LEHD is public use information which combines federal, state and U.S Census Bureau dada on Employers and Employees data. Primarily origin destination (OD) data has been used from the LEHD data set (http://lehd.ces.census.gov/). OD dataset constitutes jobs totals associated with both home census block and a work census block. Primary jobs data from LEHD has been used to assess distance traveled and travel time to the workplace. Noteworthy to mention here that primary jobs are the job that accounts for the most amount of income of an individual. To maintain consistency with the census block data from various sources, the year 2010's origin-destination data has been used. The distance traveled and travel time estimates are obtained using Bing's route API. This provided a real-time estimate of travel time and travel distance from each home and work geocode. We also tested Google Maps Distance Matrix API, but that only allowed queries of 2500 blocks per day. While Bing's API service allowed as much as 70000 queries a day, with no upper limit on total number of requests. Additionally, the quality of the data needed was same for both services.

5. Methodology

The following are the major steps of the workflow to develop the computational model delivering various heat maps pertaining to specific metrics as described further:

5.1. STEP 1: CLEANING DATA

LEHD data comes as a very large data set with over 3 million data entry for Georgia from which home blocks within Atlanta is selected and associated with their individual work blocks and some jobs in various category of age groups and income. To keep the data set manageable and reasonable for the purpose of analysis the geographical boundary for the workplaces are restricted to the Atlanta MSA. Processing the LEHD data is done in MS Access using inbuilt query design. MS Access proved to be adequately competent with a robust GUI to handle our big data challenges.

5.2. STEP 2: LINKING DATA

The block shapefiles and basic demographic information as obtained from the census is used to extract the latitude-longitude information for each of the census blocks included in the study area (Figure 1b). An optimized script written in Python is used to extract each block group's latitude-longitude information from the large data set as aforementioned. Each home geocode is associated with multiple work geocode forming a one to many relationships in the origin-destination data set. To match the respective home-geocode and work-geocode MS Access query design is used. Each row represented the work census blocks where people go to work from a particular home block. This is represented as a matrix in the Python script where each row represented a home-geocode and every column entry of the row represented work-geocodes.

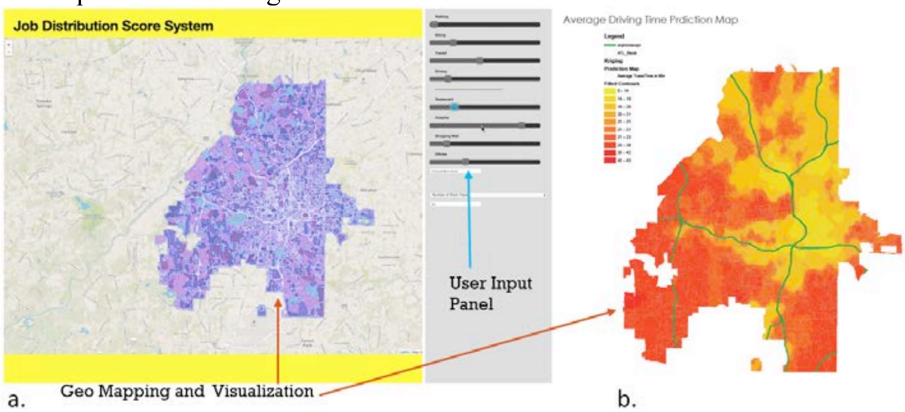


Figure 2. a. Prototype web application showing user input panel and geo mapping panel with visualization of heat maps. b. Visualization: Average driving time prediction map

5.3. STEP 3: WEB DATA MINING

Next, we make periodic queries to Microsoft Bing's Route API service, with source and destination longitude and latitude value as obtained from the matrix so formed. This service was free allowing us to make up to 70000 queries per day. The data returned from Bing's API included Driving distance and time, Walking distance and time, Transit distance and time, from home geocode location to any work geocode location. Likewise, a distance time matrix is computed for each home geocodes (for driving, transit and walking) as shown in Figure 3. We also made queries to Google Places API, for all census blocks, which essentially returned the number of hospitals, restaurants, shopping malls and hotels within a given threshold distance or walking time. Data so obtained was further used to build heat maps for the user, quantifying the suitability of a block for potential stay of the user.

5.4. STEP 4: VISUAL ANALYTICS

Subsequently, the next step was to aggregate the data retrieved and provide useful statistics for each block. The main data components obtained were total vehicle miles, total walking miles; total distance traveled by transit from each census blocks to their job location; total travel time to reach each location from census location. Average miles traveled per job was found by normalizing the number of jobs for the census block. Input threshold distance and threshold times are used to identify how many jobs are above or below certain distance threshold or travel time threshold.

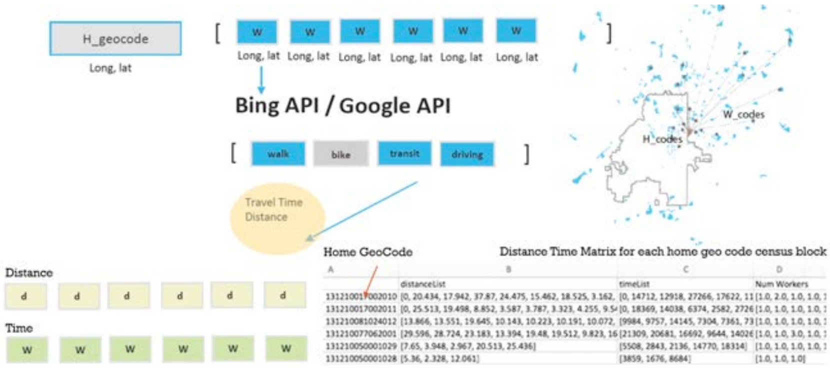


Figure 3. Distance, time, weight (number of jobs for each home –work geocode pair) Matrix

6. Analysis and Results

A wide range of parameters and relationships are studied primarily focusing on two things from the investigation. First, if there is any discernable trend or pattern in the job and home location relationship based on three different travel modes driving, transit and walking. Second, if the mode of travel, let's say driving is the control variable, then is there any difference observed in the travel distance, or travel time among different socio-economic variables i.e. age, income, job sector or industry.

6.1. DIFFERENT TRAVEL MODE

6.1.1. Driving

The result of the analysis (Figure 4a) indicates that the average driving distances are above 16 miles in most blocks in south Atlanta. Additionally, they show high clustering (marked red), positive spatial autocorrelation through hotspot analysis compared to the blocks in the center city. The results of the analysis for people with driving distance above 25 miles (threshold distance as a user input) also corroborates the finding that blocks further south has more number of people with driving distance more than 25 miles, and they show positive spatial autocorrelation. Also, the analysis (Figure 4b.) indicates that the average driving time is above 25 minutes in most blocks in the south. However, the number of census blocks above 25 minutes driving distance are concentrated at the southwest corner and at the northwest corner of the city.

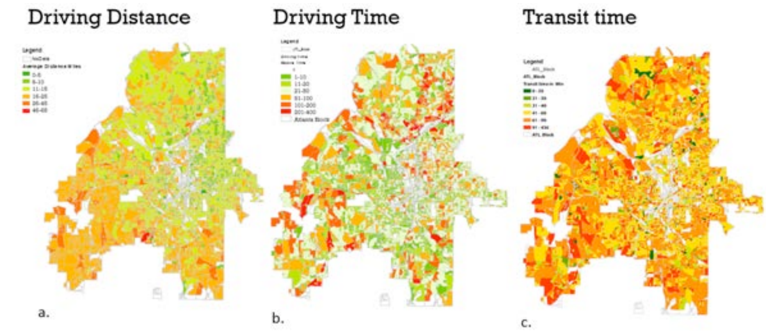


Figure 4. a. Average driving distance from each census blocks b. Average driving time from each census blocks. c. Average commute time for each block using transit.

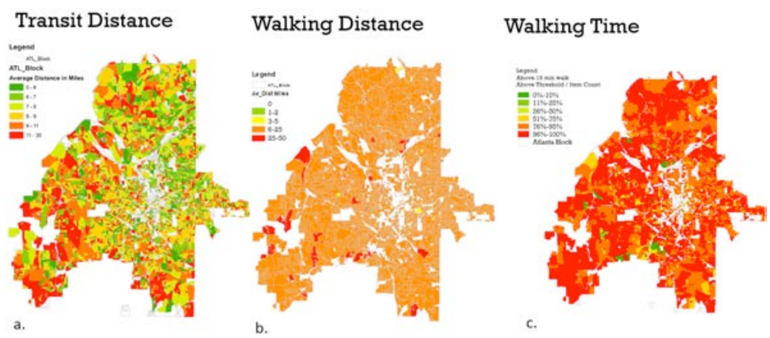


Figure 5. a. Average distance of commute for each block using transit. b. Average walking distance of each block. c. Percentage of people above 15 minutes walking time

6.1.2. Transit

For transit from home to workplace, the results showed the average transit time for the most block in Atlanta is above 40 minutes and those marked in red are above 90 minutes' transit time indicating no transit options to the workplace for the most number of people in that block. Figure 4c and 5a shows a clear trend that most blocks in the southwest and further south have higher transit distance (30 miles).

6.1.3. Walking

The walking distance and time study indicate the blocks that have an average feasible walking distance to jobs are very few blocks where only 0-10 % of the people are above 15 minutes walking time from their workplace. Most block shows almost 90 -100% of the people are located above 15 minutes of walking distance from the job locations. The average walking distance of all blocks is above 6 miles. Only very few blocks have an average walking distance of 2 miles to their workplace. (Figure 5b and 5c)

6.2. DIFFERENT SOCIO-ECONOMIC PARAMETERS

6.2.1. Age

The average driving distance traveled by different age groups are studied creating weight matrix with jobs in different age group categories. The average driving distance for age group below 29years is significantly higher than those in the age group 30 -54 years. This indicates that younger age group people travel significantly more compared to those in higher age group. Greater spatial mismatch is observed for age group 29years and below. (Figure 6a and 6b)

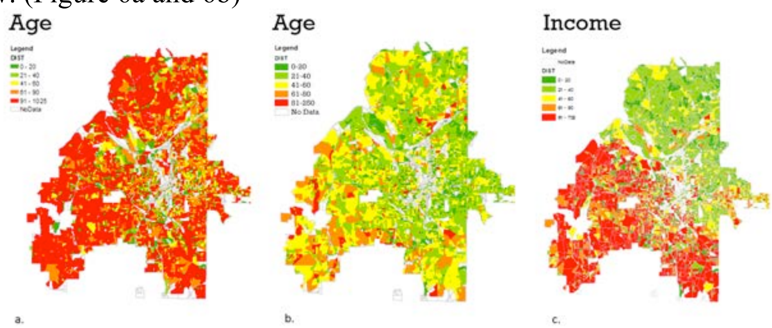


Figure 6. a. Average driving distance for workers 29 years and below. b. Average driving distance for workers age 30 -54 years. c. Average driving distance for workers with earning above \$3333 /month.

6.2.2. Income

The average vehicle distance miles for low-income jobs (\$1250 –or less/ month) is much higher compared to jobs in higher income bracket (Above \$3333/month). Which substantiates that low-income jobs are far away from the city center, and disadvantaged groups need to travel significantly more to their jobs. The northeast side of Atlanta reflects least spatial mismatch of home and job location for high-income workers. (Figure 6c)

6.2.3. Industry Type

The third factor studied is if there is any variation between workers in various industries. The average distance maps show that significantly high average driving distance for workers in Good producing sector, Trade, Transportation, and Utility sector compared to other Services industry sector. The goods-producing industry or trade transportation related business like warehouses etc. substantiates the fact that these industries are located away from city centers increasing vehicle miles, indicating worker's in this industry needs to significantly more to reach their workplace. Workers in services industry are located close to worker's home locations and accounts less travel distance.

7. Conclusion

The use of block level census geography and combining the dynamic data set from Microsoft Bing's API and Google Places API helped to create the possibility of a very high-resolution analysis of socio-economic factors. The results obtained clearly shows that there is a high degree of spatial mismatch between home and job locations. There are very few jobs with driving distance within 5 -10 miles. The transit options from most block groups are inadequate thus driving long distances above 16 – 20miles is the most viable option for travel. There are very few jobs within comfortable walking distance (less than 15 mins.) for low-income workers. Further research is under works to implement the same model in a different city with totally different demographics and socio-economic conditions. Using proximity analysis from real time / dynamic data sources using large scale geocoding and computing can help analyze block and parcel level data using a similar methodology. It can also inform various policy and design decisions, which might be overlooked by over-reliance on Census as the only source of data analysis and decision making.

References

CENSUS, U. (n.d.). *lehd.ces.census.gov.* Retrieved from http://lehd.ces.census.gov/data/lodes/LODES7/LODESTechDoc7.2.pdf

CHENG, J., & BERTOLINI, L. (2013, JUNE). Measuring urban job accessibility with distance decay, competition and diversity. *Journal of Transport Geography, 30*, 100-109.

HADAS, Y., & RANJITKAR, P. (2012, May). Modeling public-transit connectivity with spatial quality-of-transfer measurements. *Journal of Transport Geography, 22*, 137-147.

HTTP://LEHD.CES.CENSUS.GOV/. (N.D.). Lehd Data Source Weblink.

IHLANFELDT, K. (1993). The Spatial Mismatch Between Jobs and Residential Locations within Urban Areas. *Cityscape. 1*, pp. 219-244. Washington, D.C.: US Department of Housing and Urban Development.

JS, L. (2011). http://leafletjs.com/.

KIM, C., Sang, S., Chun, Y., & Lee, W. (2012, March). Exploring urban commuting imbalance by jobs and gender. *Applied Geography, 32*(2), 532-545.

POLICY, C. O. (2000). *Moving Beyond Sprawl: The Challenge for Metropolitan Atlanta.* Atlanta: Brookings.

WANG, C.-H., & CHEN, N. (2015, May). A GIS-based spatial statistical approach to modeling job accessibility by transportation mode: case study of Columbus, Ohio. *Journal of Transport Geography, 45*, 1-11.