Development of a community planning support system based on open data: Neighborhood quality of life and health in City of Atlanta

Ge Zhang, Wenwen Zhang, Subhrajit Guhathakurta and Nisha Botchwey

Abstract

Open data has come of age with many cities, states, and other jurisdictions joining the open data movement by offering all relevant information about their communities for free and easy access to the public. Yet, despite the growing volume of open data, their use has been limited in planning scholarship and practice. The bottleneck is often the format in which the data are available and the organization of such data, which may be difficult to incorporate in existing analytical tools. The overall goal of this research is to develop an open data-based community planning support system that can collect related open data, analyze the data for specific objectives, and visualize the results to improve usability. To accomplish this goal, this study undertakes three research tasks. First, it describes the current state of open data analysis efforts in the community planning field. Second, it examines the challenges analysts experience when using open data in planning analysis. Third, it develops a new tool for examining neighborhood quality of life and health for the City of Atlanta as a prototype, which addresses many of these open data challenges.

G. Zhang (Corresponding author) • W. Zhang • S. Guhathakurta • N. Botchwey
Center for Geographic Information Systems, Georgia Institute of Technology, Atlanta, GA, USA. Email: ge.zhang@coa.gatech.edu

W. Zhang. Email: wzhang300@gatech.edu

S. Guhathakurta. Email: subhro.guha@coa.gatech.edu

N. Botchwey. Email: nisha.botchwey@coa.gatech.edu

1. Introduction

Data collection has historically proven to be time consuming and expensive (deLeeuw and Collins, 1997; Krieger et al., 1997; Axinn et al., 2011; Bifulco et al., 2014). With the development of new technologies and advanced theories, several tools and algorithms have been developed together with new data sources for improving data collection techniques (Seo et al., 2008; Curtis et al., 2013). For example, some algorithms in geostatistics, such as bayesian maximum entropy, can produce maps of estimated future water usage based on historical data and projections of future population density (Lee et al., 2010). Similarly, with the help of satellite imagery and image interpretation technology, analysts can collect land cover data for several square kilometers in a few hours, rather than doing field investigation that could take several months (Karnieli et al., 2008; Vittek et al., 2014).

Open data is attracting increasing attention in urban planning as new innovative ways for using such data are being developed (Bonatsos et al., 2013; Balena et al., 2013). One good example is the water database provided by the United State Geological Survey (USGS). This database provides real-time and historic surface-water, groundwater and water quality data for all streams in the United States. It has been used to estimate nutrient and bacterial concentrations in water bodies and to help decision-makers efficiently manage watershed resources (Christensen et al., 2002). Another example is the census database provided by the U.S. Census Bureau. It provides population characteristics at different geographic levels, such as the census tract and the block group, and has been used in myriad applications, such as in urban and environmental modeling and in policy assessments (Wang, 2008; Mennis and Dayanim, 2013). However, in many cases, open data cannot be directly used because of the variation in accuracy, format and scale (Arentze et al., 2007; Diesner et al., 2012).

While the domains of marketing or ecology have existing frameworks based on open data formats, the field of urban planning does not currently have a protocol for utilizing open data, specifically, for the type of analysis that advances research (Reichman et al., 2011; Fleisher, 2008). Where such protocols or frameworks exist, they usually address questions about open data production rather than open data analysis. The users of such protocols are data providers, rather than data analysts. Additionally, how open data can be used efficiently has not been adequately explored in the urban planning literature. Therefore, in this study, we particularly focus on examining an open data-based community planning support tool to demonstrate how open data can be used in urban planning applications. We leverage prior work on the use of public participation geographic information systems (PPGIS) to develop a tool that people in the community can use to understand, modify and interact with spatial data about their community within a geographic information systems platform (GIS). Our goal is to offer a "proof-of-concept" for an understandable, easy-to-use, open data-based community planning support tool.

The rest of the paper is comprised of three sections. The next section discusses the current state of the use of open data in planning practice and scholarship. Section Three documents the challenges analysts experience when using open data in planning. In Section Four, a new open data based tool for examining neighborhood quality of life and health in the City of Atlanta is examined as a potential "proof of concept" for integrating open data in participatory planning. Finally, the article closes with a concluding section highlighting the main findings and indicating future directions for further research.

2. The current state of utilization of open data

There are numerous sources of open data across different geographic scales. Some sources are already widely known and intensively utilized by many planning practitioners, while some others are less accessible due to technical difficulties in both downloading and converting data given their unfamiliar structure and format. In this section, we review and summarize the existing sources for open data at different geographic scales from nation-wide datasets to state and city level data sources.

2.1 National Level Open Data

National level data provide information for most, or usually all, of the U.S. The resolution of the data can vary by data source. For instance, the U.S. Census Bureau prepare data for various topics for states, counties, census tracts, block groups and blocks. Because the datasets are available for all municipalities in the U.S., we consider it a national level data. A brief summary regarding the national level open data is tabulated in Table 1.

Among all data sources, the most frequently used data provider at the national level is the U.S. Census Bureau. Most planners utilize this national data source that contains demographic and socio-economic data aggregated to different geographic boundaries from national to blocks. Although the website of the U.S. Census Bureau provides a well-designed user interface for planners to download the targeted data, it remains quite labor intensive to download, maintain, and update the data for local planning purposes. Fortunately, the U.S. Census Bureau has already developed an application programming interface (API) for users to access the Census Bureau data automatically. However, not many planners are aware of this tool, and few planners are equipped with the programming skills necessary to take advantage of this application. The primary formats of this data source are .csv and .xls, rendering it quite user friendly, due to the availability of software such as Microsoft Excel, SPSS, and ArcGIS, that can manipulate and analyze the data.

Table 1. National Level Open Data Source Summary

Data Source	Type of Data Provided	Data Resolution	Data Obtaining Methods	Source	Format
Census Bureau	Business Economy Education Employment Families & Living Arrangements Health Housing Income & Poverty International Trade Population Public Sector	State/ County/ Census Tract/ Block Group/ Block	Manually Download/Use API to query the data	http://www.census.gov	csv /xlsx
Data Government	Agriculture Business Climate Consumer Ecosystems Education Energy Finance Health Local Government Manufacturing Ocean Public Safety Sicence Research	State/County/Census Tract/Zip code	Manually Download	http://www.data.gov/	various
Google Data	Place Elevation Street Views GTFS transit	Most cities in U.S.	Use API to query the data	API License Key is required	Json
Zillow Data	Housing	Household Units	Use API to query the data	API License Key is required (Limited Query)	xml
Walk Score	Walk Score Transit Score	Most cities in U.S.	Use API to query the data	API License Key is required (Limited Query)	xml
Social Media	Twitter Facebook Foursquare	Most cities in U.S.	Use API to query the data & Data Mining	API License Key is required (Limited Query)	xml
Quandl	Economic Earnings Housing	Country Company Aggregated Level	Use API to query the data	API License Key is required (Unlimited Query)	Json
Wunderground	Weather	City/ Zipcode (also with historical data)	Use API to query the data	API License Key is required (Unlimited Query)	Json/ xml
New York Times	All types of News	All over U.S.	Use API to query the data	API License Key is required (Unlimited Query)	Json

In addition to the Census Bureau data, we found many other sources that provide datasets covering the majority of the nation. The largest data providers are Google and different social media platforms, including Twitter, Facebook and Foursquare. Google provides data by place, which includes hundreds of place categories such as elevation, street view, and a General Transit Feed Specification (GTFS). Among these resources, planning practitioners may find the place information resources to be the most useful. Place information resources offer location information about grocery stores, coffee shops, schools, banks, etc., which may be critical in analyzing community quality of life. For instance, it is quite easy to analyze how accessible fresh foods are in a certain community, using Google Place data. The Google GTFS data are also quite popular. The dataset is available in a standard format and provides machine readable transit data for most U.S. Cities as long as the local transit system has been uploaded to Google. The dataset includes information regarding transit lines, stations, stops and headways/ service frequency. Most recently Google worked together with some public transit agencies to develop a real-time extension for GTFS data, which provides real-time operation information to the general public. Planners may utilize these data to improve transit service quality evaluation.

Social media companies also provide a great deal of open source data. Posts on social media generally have timestamps and are tagged with geo locations, as long as the user does not turn off the geolocation function. Posts cover diversified topics about people's daily lives, their thoughts, and opinions. Data mining technologies, such as support vector machines, boosting and logistic regressions, can all be used to extract targeted information from tweets, check-ins and posts on various social media platforms. Planners may wish to examine attitudes toward plans, traffic conditions, the impressions regarding the communities, etc. The data from social media update continuously, is large in volume but may require some effort and computing knowledge to make use of the data.

Some other websites offer valuable data regarding real estate and the built environment. For instance, Zillow, a popular real estate listing website, offers updated property information for both sale and rental purposes. Zillow's downloadable data include property sale records; house characteristics, such as year built, number of rooms, lot size, and square footage, and some neighborhood level information, such as nearby school quality, crime and risk of exposure to hazards. It is also possible to download historic property data from Zillow, which makes longitudinal analysis easier. Walk Score is another source that provides facts about the built environment. This website estimates walk score based on accessibility to different types of nearby facilities, road networks and population density. In addition to the walk score, the website also offers transit scores based on GTFS data.

In addition to the above datasets, planners might use other national datasets such as Wunderground and the New York Times. Wunderground, for instance, has historic data about weather, which can be extremely useful when developing plans relevant to climate change. The New York Times offers historic news data, rendering it easier to look into the historic important events for certain communities.

Those data sources can provide information at a comparatively high spatial resolution. Many planning related studies have already attempted to use these datasets in innovative ways. Google Street View has been used successfully in a variety of studies. Street features, including traffic conditions, physical barriers, pedestrian safety, parking, active travel infrastructures, sidewalk amenities and presence of users. Rundle et al. (2011) audited the built environment using Google Street View. Odgers et al. (2012) captured neighborhood level characteristics that could influence the life of children using Google Street View. The results suggest that the measurements obtained from Google Street View are reliable and cost-effective. Kelly et al. (2013) hired graduate research assistants to manually extract a large amount of built environment information for both suburban and urban environments for public health analysis. The built environment data obtained from Google are highly acceptable according to statistical tests. In sum, the emerging studies suggest that Google Street View might serve as a reliable source for collecting a wide range of built environment data.

Social media data have also been widely used in many studies to describe and understand the social dynamics of the city. Cranshaw et al. (2012) combined data from Foursquare and Twitter to classify people's activity and behavior into different groups. The results were then compared with qualitative interviews and focus group results. The comparison between the quantitative and qualitative results revealed that social media could be a powerful tool to reflect subtle changes in neighborhoods. The analysis highlighted people's activity pattern changes as they responded to variations in policies, developments and resources. Frias-Martinez and Frias-Martinez (2014) proposed using geographically tagged tweets for urban land use detection. The classification results based on Twitter information were subsequently validated by official land use datasets.

Walk Score data and calculation methods have been validated by several empirical studies (Carr et al., 2010; Duncan et al., 2011; Hirsch et al., 2013). Some studies introduced Walk Score into the traditional hedonic model to argue that people are willing to pay for a walk friendly environment around their homes (Cortright, 2009; Pivo and Fisher, 2011). There is potential to conduct longitudinal real estate price change studies using data from both Walkscore and Zillow. However, so far little work has been done in this field. But some studies have used Zillow to obtain household characteristics (Balu and Hurin, 2013). Although there are currently heated debates over the quality of Zestimate, the property value estimates provided by Zillow, the general quality of household characteristic data that they provide are considered acceptable for analysis and can be used by planners (Wu et al., 2009).

To conclude, many national level open data sources in addition to U.S. Census Bureau data have already been widely used in many academic studies. However, not many are utilized in the real-world planning problems.

2.2 State Level Open Data

Many states have already developed official open data websites for users to access state level open data. Almost all states provide open access to data regarding government transparency. The government-related data usually includes information on funding distribution and financial reports. In this study, we did not consider states that only have transparency data to be states with effective open data resources. Using this criterion, a total of 20 states were identified to have open data portals, which are tabulated in Table 2.

As may be expected, there is wide variation among states in the extent, resolution, and format of open data for each state. Besides government and transparency data, most of the 20 states with open data portals provide data for health and human services, education, transportation, business and economic development, environment and natural resources, and public safety. Meanwhile, some other types of data, such as demographic; agriculture; cultural, recreational and tourism; and technology also receive some attention at the state level, depending on the state. There is no uniform organization of data across states. Some states list items such as permits and weather as independent categories while other states tend to move this type of data under an umbrella category, requiring users to dig further into the websites. Additionally, some states provide a wider range of data, while other states are still at the initialization phase of website development. Yet, in all, there seems to be a trend for more transparent and data-driven government that is willing to open its data sources for public scrutiny.

Table 2. State Level Open Data Sources

State	Data Website	API Accessible	Government & Financial	Health & Human Services	Education	Transportation	Business/ Economic Development	Environment & Natural Resources	Public Safety	Demographic	Agriculture	Culture & Recreation	Tourism	Technology	Community	Permits	Weather	Utility
California	http://data.ca.gov/	9 API enabled	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes			Yes		
Colorado	http://opencolorado.org/		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes								
Connecticut	https://data.ct.gov/	API enabled	Yes	Yes	Yes	Yes	Yes	Yes	Yes									
Delaware	http://www.delaware.gov/topics/data.shtml		Yes		Yes	Yes				Yes								
Hawaii	https://data.hawaii.gov/		Yes	Yes	Yes	Yes	Yes	Yes				Yes			Yes			
Illinois	https://data.illinois.gov/		Yes	Yes			Yes		Yes			Yes						
Iowa	https://data.iowa.gov/		Yes	Yes	Yes	Yes	Yes	Yes							Yes			
Maryland	https://data.maryland.gov/		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes						Yes	Yes
Minnesota	http://www.state.mn.us/opendata/		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes				Yes				
Missouri	https://data.mo.gov/		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes	Yes				
Montana	http://data.mt.gov/		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes			Yes	Yes		Yes		
Nebraska	http://www.nebraska.gov/data.html		Yes	Yes	Yes										Yes			
New York	https://data.ny.gov/	28 API enabled	Yes	Yes	Yes	Yes	Yes	Yes	Yes									
Oklahoma	https://data.ok.gov/		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes				
Oregon	https://data.oregon.gov/	All API enabled	Yes	Yes	Yes	Yes	Yes	Yes	Yes			Yes						
Rhode Island	http://www.ri.gov/data/		Yes							Yes							Yes	
Texas	https://data.texas.gov/	1 API enabled	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes		Yes					
Utah	http://www.utah.gov/data/		Yes	Yes	Yes	Yes	Yes	Yes	Yes			Yes				Yes		
Virginia	https://data.virginia.gov/		Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes							
Washington	https://data.wa.gov/		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes							

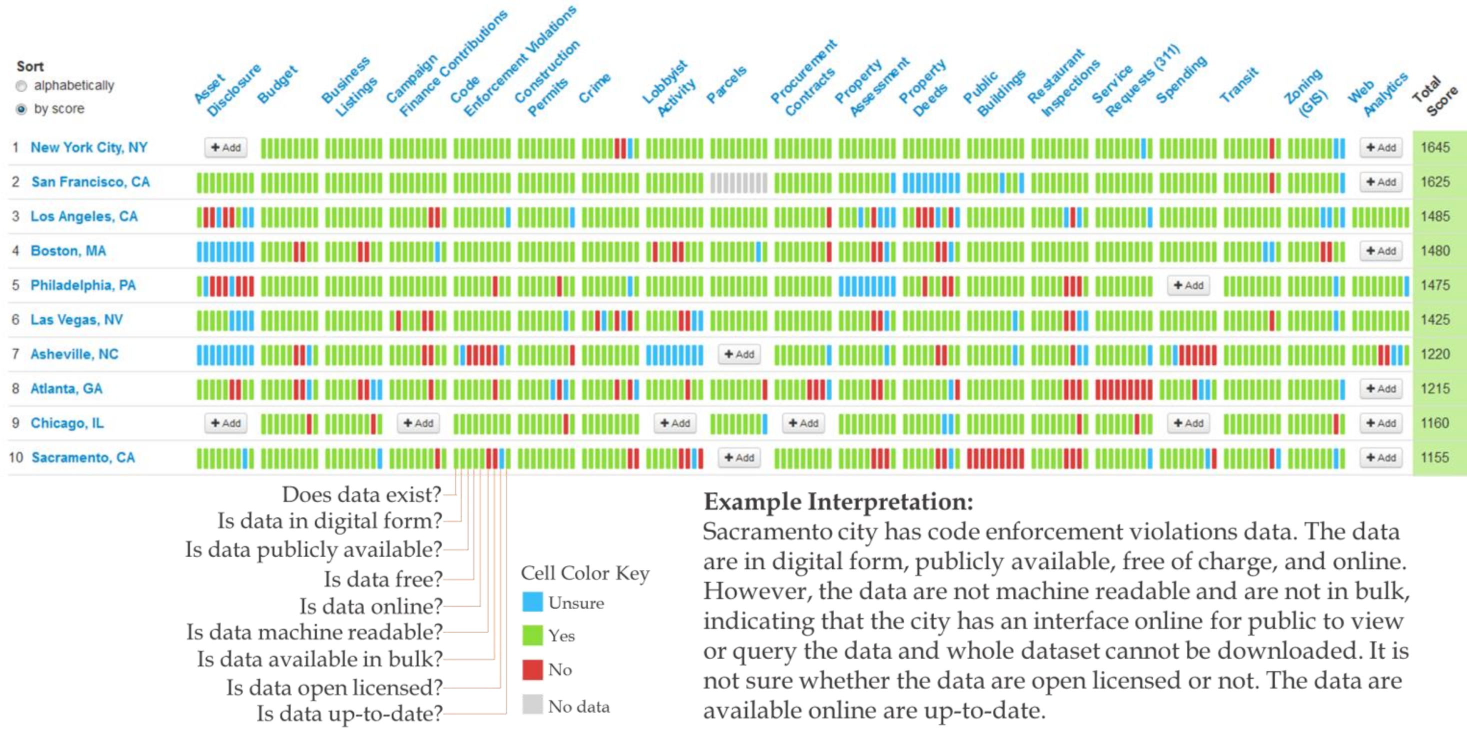
It is also impressive to notice that many state data portals have already provided APIs for developers to access data remotely. For instance, New York offers 28 APIs, and California provides nine APIs to help users navigate their datasets. These APIs makes it easier for users to track, maintain and update the data.

2.3 City Level Open Data

Many cities have already taken the first step to make datasets open and available online, regardless of whether they have an open data policy. The U.S. Open Data Census, powered by the Open Knowledge Foundation, has a comprehensive list of open data available for various cities. Currently, this census provides information for up to 94 cities, with a total of 761 datasets. Approximately 25% of these datasets are actually open online. The census also provides a limited evaluation of the data quality available at the city level. For instance, some data are open online for public safety. However, the quality of data that are uploaded may turn out to be dramatically different from one city to the next. For example, some places publish image data regarding the distribution of crime, while other places provide GIS shapefiles showing the type, location and timestamp of crimes. The machine-readable data, like GIS files, provide more information than the static image file, and therefore, more useful for data analysis. In addition to machine readability, the U.S. Open Data Census also assesses data quality based on data cost (some places charge for parcel data and zoning data) and whether the data is up to date.

Based on the open data census, the top ten cities with the most open data with comparatively high data quality are shown in Table 3. New York City and San Francisco are top tier cities for open data, as their scores are much higher than the rest of the top 10 cities. However, it must be noted that this census is cloud based, indicating that the results depend heavily on people who provide the information about data resources to the census. If the information provider is not familiar with certain datasets provided by a city, then the results of the census may not be completely up to date. However, the census offers a platform to encourage the open data movement at the city level.

Table 3. Top 10 Cities with the highest open data score



3. The challenges in open data-based tool development

Although open data across geographic scales have already been widely used in several academic fields, some barriers still exist that limit planning practitioners from utilizing open data. Open data accessibility is a big concern, particularly for some national level datasets. High resolution national data are extremely large, rendering it impossible for users to download entire datasets. Often, data providers do not allow manual download. For instance, the U.S. Census Bureau does not offer a link to let users download all the data they have. For this type of large dataset, users generally need to provide more specific information regarding the specific scope of data they are looking for. The U.S. Census Bureau offers a user friendly interface for planners to create tables by topic and geographic location. However, most data providers are not like the U.S. Census Bureau whose major purpose is to provide data to the general public. Many national level data providers only provide API for users to download data, which is more complicated than hitting a download button on a website. A programming background is needed to write scripts to scrape data in large batches from websites using API. This limits the use of open data that are only accessible via APIs. However, it is important to note that once a script is developed to download data using API, the download process can be automated, which will make it easier in the future to update and maintain the data. Therefore, it will be of great value for planners if an application can be developed for accessing open data using APIs.

The second challenge in using open data is that most planners are not equipped to process some open data sources and formats. For instance, it is quite hard to process Google Street View and social media data without data mining knowledge and expertise in image processing machine learning models.

Third, some formats of data, such as .json and.xml files, are not well-known to planners. Data downloaded with APIs are generally formatted as .json and .xml files. These data formats are designed to simplify data transfer on the internet using uniform code across languages, as data are often text based data. Unlike the conventional Excel spreadsheet or csv data, these types of data tend to be highly machine readable but are not easily read and comprehended as raw numbers by humans. Therefore, it is critical to help planners transform this type of data to more conventionally structured datasets.

Finally, there are always some geographic unit mismatch problems. For instance, some data are available at the census tract level while others are available at the zip code level. However planners are usually interested in community and neighborhood level statistics. Reconciling data among two different geographic units requires simplifying assumptions about how the attributes of the data are distributed within each geographic entity (such as homogeneous density within the spatial unit). Additionally, some data are non-spatial and cannot be geo-located for neighborhood-based analysis.

4. Neighborhood quality of life and health in the City of Atlanta

Neighborhood characteristics directly affect one's health outcomes and quality of life and are critical considerations in determining neighborhood planning and health promotion strategies. A web-based application named Neighborhood quality of life and health (NQOLH) was developed to demonstrate the status of quality of life and health in Atlanta at the neighborhood planning unit level. The application was composed of a front-end user interface and a back-end server. The front-end was used to visualize the results and provide interactive functions for users. The program running in the front-end was written in HTML and JavaScript. Several JavaScript libraries were used, such as jQuery and Google Map JavaScript API. The back-end was used to collect the data from a public server, convert the format and scale of the data and customize an index that represents particular aggregate metrics based on the data. The program running in the back-end was written in Python. Some specific Python libraries as ArcPy were used to process the data.

This application was developed using a four-step process that included data collection, database establishment, index customization and interactive user-interface design. Some of the data collected in Step One came from open source databases, such as housing data from the U.S. Census database and public safety data from the Atlanta Police Department (Table 1). Some challenges were identified in using open data. First, cleaning and paring down the open data were necessary and time intensive. Second, the census data were collected based on smaller geographic units that required aggregation based on particular assumptions. Third, some open data were saved in a specific format, which could not be directly used for spatial analysis and visualization. Finally, a limited set of open data provided customized indexes and vector features, which incorporated the raw information we needed but required additional steps to extract that relevant information.

In this application, the challenges were resolved through various approaches and tools. US Census API and some Python libraries (i.e. json and urllib) were used to connect the database to the public US Census server. This API offered an efficient way to download the data based on the census tract ID and to update the local database as the data source is updated. ArcPy was used to transform data from census tract level to neighborhood planning unit level. We also used ArcPy for conducting spatial analysis based on the neighborhood planning unit. Some Python libraries, such as pyshape, json and csv, were used to convert data from txt to csv and csv to shp. Python script was used to customize the new index developed for the project. Several common related indices were introduced to describe each aspect of neighborhood quality of life (NQOL) (Table 4). To create the aggregated NQOL index, the indicators were weighted with results from a citizen survey conducted by the City of Atlanta for their 2011 Comprehensive Development Plan update (Table 5). The status of neighborhood health (NH) was based on measures of healthy food access, physical activity, mortality and morbidity in neighborhood populations (Table 6). The NH measures were aggregated into indicators based on the County Health Rankings methodology (Figure 1).

Table 4. Indicators for neighborhood quality of life measurement

Indicator	Measure	Data Source	Calculation
Public Safety	Violent crime rates	City of Atlanta, Atlanta Police Department	The number of violent crimes occurring in an NPU, standardized by its residential population
	Property crime rates	City of Atlanta, Atlanta Police Department	The number of property crimes occurring in an NPU, standardized by its residential population
	Traffic-related injuries and fatalities	Georgia Department of Transportation	The number of traffic accidents resulting in physical harm occurring in an NPU, standardized by its residential population
Economy	Jobs to labor force ratio	U.S. Census Bureau	The ratio of the number of jobs to the number of workers in an NPU
Transportation	Commute travel times	U.S. Census Bureau	The mean average of travel minutes for a one-way commute

Amenities	Transit access	Metropolitan Atlanta Rapid Transit Authority transit locations shapefile	Percentage of residents living within 0.25 mile radial distance of a bus stop or 0.5 mile radial distance of a transit stop
	Parks and recreation access	Atlanta parks shapefile	Percentage of residents living within 0.25 mile radial distance of a park or recreational facility
	Retail access	Reference USA business database	Percentage of residents living within 0.25 mile radial distance of retail and restaurant clusters
Housing	Home affordability	U.S. Census Bureau	The ratio of the citywide median household income to median home values in the NPU
	Rent affordability	U.S. Census Bureau	The ratio of the citywide median household income to mean rental rates in the NPU
	Vacancy rates	U.S. Census Bureau	The percentage of vacant housing in the NPU

Table 5. Survey Scores and Indicator Weights

Attribute	Residents' Scores	NQOL Indicator	Weight
Police Service	4.3	Public Safety	25%
Fire Service	4.2		
Economic Development	4.2	Economy	22%
Transportation	4.1	Transportation	20%
Land Use	4.0	Amenities	18%
Parks, Greenspace & Recreation	4.0		
Housing	3.7	Housing	15%

Table 6. Indicators for neighborhood quality of health measurement

Indicator	Measure	Data Source	Calculation
Nutrition	Food Access	U.S. Census Bureau; Reference USA Database	The percentage of no vehicle households living beyond 0.9 mile radial distance of a supermarket or superstore

Physical Activity	Walkability	Walk Score	The mean average of Walk Scores of each NPU' s constituent neighborhoods
Mortality	Years of potential life lost before 75 rate location quotient	Georgia Department of Public Health	(YPLL75 rate in NPU*median age in Fulton County)/(median age of NPU* YPLL75 rate in Fulton County)
	Diabetes location quotient	Georgia Department of Public Health	(DIabetes rate in NPU*median age in Fulton County)/(median age of NPU* DIabetes rate in Fulton County)
Morbidity	Hypertensive heart disease location quotient	Georgia Department of Public Health	(Hypertensive heart disease rate in NPU*median age in Fulton County)/(median age of NPU* Hypertensive heart disease rate in Fulton County)
	Esophageal cancer location quotient	Georgia Department of Public Health	(Esophageal cancer rate in NPU*median age in Fulton County)/(median age of NPU* Esophageal cancer rate in Fulton County)
	Uterine cancer location quotient	Georgia Department of Public Health	(Uterine cancer rate in NPU*median age in Fulton County)/(median age of NPU* Uterine cancer rate in Fulton County)
	Kidney cancer location quotient	Georgia Department of Public Health	(Kidney cancer rate in NPU*median age in Fulton County)/(median age of NPU* Kidney cancer rate in Fulton County)

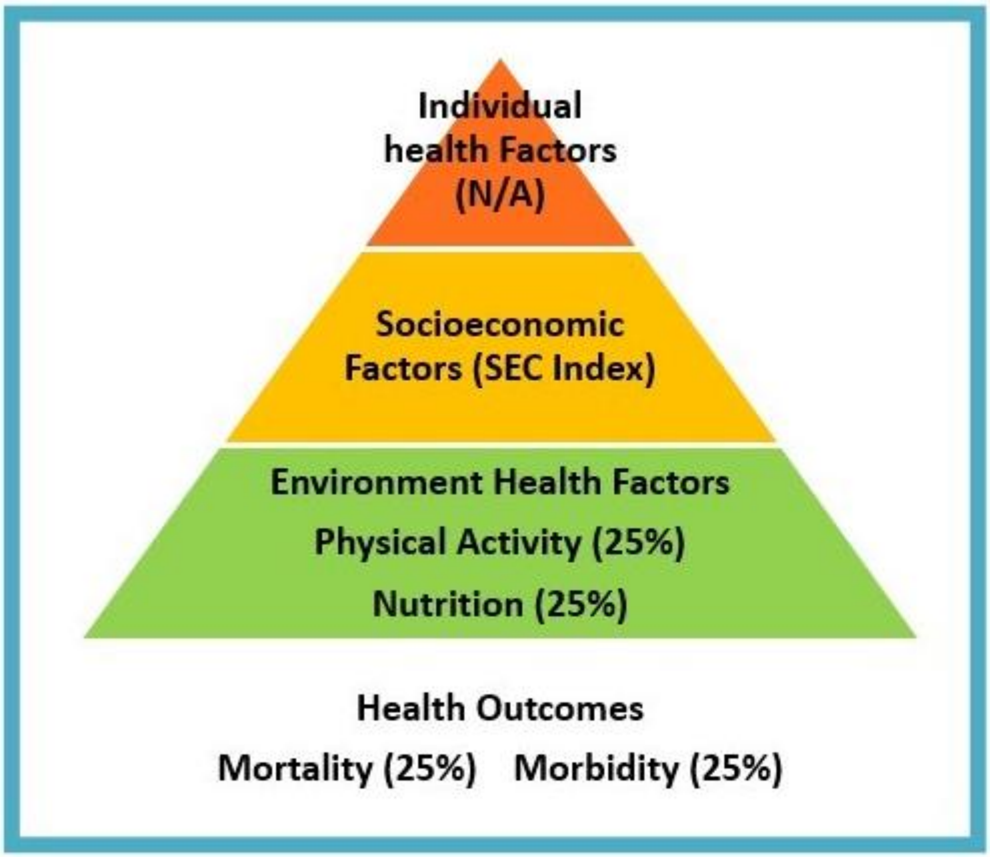


Figure 1. Neighborhood Health Weighting Scheme

Cloud technology was used to establish the database and store the data used by the application. In this application, we used Google Drive as our cloud database. We carefully considered how to design the interactive functions to improve the user experience. The website's functionality provided various methods to query the data from the database (Figure 1). These methods include selecting the variable, socioeconomic status and neighborhood name from drop down boxes, selecting the specific neighborhood planning unit from Atlanta's neighborhood planning units map, selecting the records from the table, selecting the dot from the scatter plot chart and entering an address or neighborhood's name. Several methods were designed to visualize the data to make users easily understand the relationships among different attributes of the data (Figure 2). In this tool, tables were used to show detailed information about all measurements for all neighborhood planning units. Maps were used to show the locations of the neighborhood planning units, the differences among them in each measure and the geographic pattern of each measure in Atlanta. The charts were used to show the statistical pattern and trend of each measure in Atlanta. Additionally, data export function was provided to share the data generated by this tool and provide a connection with other potential tools. The data can be exported in html and csv formats. The data comparison function between two specific neighborhood planning units or more was also developed as an additional user tool.

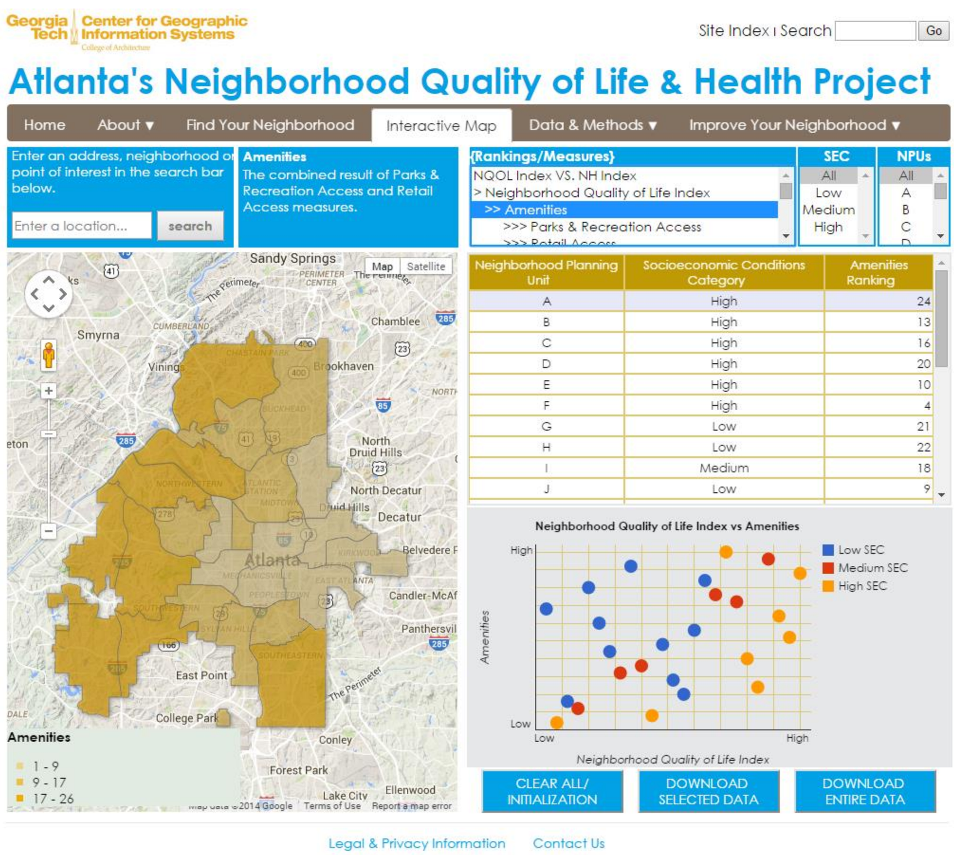


Figure 2. The interface of interactive map

5. Discussion

The application for assessing the quality of life and health in the City of Atlanta offers a test case for the use of open data in planning. It highlights the issues that developers of open data-based platforms might face and shows how such issues can be addressed with the help of existing tools and procedures.

The platform for access to open data described here provides a user friendly platform to categorize information so that users can understand data without being overwhelmed by it. Most open data in planning are multi-layered. Applications such as the Neighborhood quality of life and health project provide simplified analysis of the complex relationships found in much community planning data. The demonstration and comparison tools used in this application, such as tables, charts and maps, were designed to meet the needs of a variety of user types. Quality data evaluation is a key component of community planning data analysis, as it is used to shape reliability and rationale for planning decisions. The initial evaluation of the open data should be done before using that kind of data. It can be made based on some experience about the data if the spatial or temporal pattern of the data can be demonstrated. Additionally, let public understand the planning results and participate in the planning process are also important parts of planning process. Different people has different kinds of background knowledge and also has different preferences for information understanding.

However, like most applications, the Neighborhood quality of life and health project has its limitations. For one, the format and scale conversion used for this project was developed specifically for this project and cannot currently be easily replicated across other projects. The open data prepared by different institutions can appear in a variety of formats, such as shp, json, csv and txt formats. To effectively analyze data or create composite results from these data, format conversion is necessary. The application should be designed to easily convert between different formats. Additionally, open data are often recorded at some geographic unit or scale, such as block, block group, census tract or site. If the scale of the required data does not match to the scale of the open data, the open data cannot be used. Therefore, the function of the conversion between different scales should be developed to aggregate the information from smaller scale to larger scale and distribute the information from larger scale to smaller scale. In the Neighborhood quality of life and health project, the functions for scale transformation and format conversion were designed, but the scale transformation is currently limited to transforming from tract scale to neighborhood scale and the format conversion is limited to converting from txt to csv and csv to shp. How to transform data from larger scale to smaller scale and easily convert data into other formats are still unsolved in this application.

As a second limitation, this project did not provide a platform for users to upload personal information that could serve as supplementary data to what is currently available. Community planners bring together a variety of stakeholders in fields such as healthcare, education, private sector business and municipal government. However, a full range of data for all of these fields is always unavailable at some scale as planners deal with high costs of maintaining a public database and the legal issues of protecting certain private data. . Therefore, the system should provide a platform to let people share their own data easily. It would establish another “source” for open data. However, in this application, the function for uploading personal data was not designed which may limit the usage of the application in some areas where there is limited official open data.

A third limitation of this tool is that it is not linked to other frequently used tools that can be other “sources” for open data. Some tools have successfully been proven to effectively simulate some aspect of human activity in the community; these tools can provide invaluable information for community planners as they make decisions about their neighborhoods and municipalities. Some of the input data of these tools can be found in open data. Therefore, the system should provide ports for well-known or frequently-used tools to make the planning process more efficient. In this application, some simple algorithms for producing new index were used at the back-end, but these algorithms are based on the authors’ experience. Process based or agent based simulation with higher accuracy was not used in this project.

As a fourth limitation, this tool did not display open data in the most accessible format. Community planning tools utilize open data that comes in a variety of structures, styles, dimensions and scales. Arranging these data as user-accessible is a very big challenge. In the traditional design process for data queries, the most common ways for users to reach data is through a dropdown menu, Input box or button. However, in this application those methods would be time intensive and confusing to users who are not familiar with the dataset. Hierarchy is an effective way to simplify complex data queries. The data in this system is organized as a list rather than in a hierarchy. However, the original dataset should be aggregated into one hierarchy system with different levels. The system should provide a multi-level way to query the data based on the process from top-level to low-level. Additionally, graphics are more intuitive and easily understandable than words. Considering that open data for community planning is usually geographically related, selecting data from maps with different geographic units would be an easier way to access the target data. Icons implying specific meanings should be widely used in the system rather than just words. In this application, the users can query the data by clicking on the map, but there are limited icons available in the program.

A fifth limitation of this project is the inflexibility of the current indexes on the site. A good index can be used to monitor changes of social service in a community and provide direct evidence for decision making on a community plan. Many indexes using open data have been created in the community planning field, showing they can successfully and effectively present the characteristics of a planning service. But they are still limited, and many do not fit with a specific research topic or planning objective. Therefore, the system should provide the function of letting users customize variables by themselves. It would extend the domain of the system and make it more flexible for different projects or objectives. However, in this application, the indexes were customized only at the back-end which may reduce the usage of the application because some users may want to use other variables based on the current open data sources.

As a sixth limitation, this application does not provide space for evaluation or feedback from users. Performance evaluation is an important part of system development. The comments from user can provide valuable suggestions to improve a system and recognize the failures the developer did not notice. Therefore, the system should provide connection between users and developers. Comment and editor function should be developed. However, this application lacks these functions.

A seventh limitation is about open data updates for this project. Open data is time sensitive. With the development of technology, open data in real-time is becoming a reality, and more and more new data will be produced as dynamic open data. Open data without real-time updates may have low accuracy. Thus, the system should automatically update data sets in the existing database. However, in this application, some of the open data have long update intervals while some data are static.

6. Conclusion

This paper presents a new open data based community planning support system. To develop this application, the current state of open data analysis efforts in the community planning field was described; the challenges analysts are experiencing associated with effectively using open data in the planning process were examined; several tools based on Python were developed at the back-end of the application to solve these challenges; and an interactive graphic user interface was designed at the front-end of the application to help understanding and decision-making based on html and Javascript.

The new application was used to assess the quality of life and health of the City of Atlanta at the neighborhood planning unit scale. The application output includes two composite variables, quality of life and health, created by aggregating information from 19 variables in nine categories. The users can explore the spatial variation of each variable in different ways, including on a map with geographic locations, in a chart with trend estimation and in a table with real numbers. This application can help multiple kind of users, including research scientists, government managers, students and the general public better understand the relationship between quality of life and health in communities where they live, work or play.

References

Arentze, T., H. Timmermans and F. Hofman (2007). "Creating synthetic household populations - Problems and approach." Transportation Research Record (2014): 85-91.

Axinn, W. G., C. F. Link and R. M. Groves (2011). "Responsive Survey Design, Demographic Data Collection, and Models of Demographic Behavior." Demography 48(3): 1127-1149.

Bifulco, G. N., F. Galante, L. Pariota, M. R. Spena and P. Del Gais (2014). "Data Collection for Traffic and Drivers' Behaviour Studies: a large-scale survey." Transportation: Can We Do More with Less Resources? - 16th Meeting of the Euro Working Group on Transportation - Porto 2013 111: 721-730.

Bonatsos, A., L. Middleton, P. Melas and Z. Sabeur (2013). "Crime Open Data Aggregation and Management for the Design of Safer Spaces in Urban Environments." Environmental Software Systems: Fostering Information Sharing 413: 311-320.

Blau, D., & Haurin, D. (2013). The Impact of Housing on the Wellbeing of Children and Youths (No. eres2013_73). European Real Estate Society (ERES).

Balena, P., A. Bonifazi and G. Mangialardi (2013). "Smart Communities Meet Urban Management: Harnessing the Potential of Open Data and Public/Private Partnerships through Innovative E-Governance Applications." Computational Science and Its Applications - Iccsa 2013, Pt Iv 7974: 528-540.

Carr, L. J., Dunsiger, S. I., & Marcus, B. H. (2010). Validation of Walk Score for estimating access to walkable amenities. British Journal of Sports Medicine, bjsports69609.

Christensen, V. G., P. P. Rasmussen and A. C. Ziegler (2002). "Real-time water quality monitoring and regression analysis to estimate nutrient and bacteria concentrations in Kansas streams." Water Science and Technology 45(9): 205-211.

Cortright, J. (2009). Walking the walk: How walkability raises home values in US cities.

Cranshaw, J., Schwartz, R., Hong, J. I., & Sadeh, N. M. (2012, June). The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. In ICWSM.

Curtis, A., J. K. Blackburn, J. M. Widmer and J. G. Morris (2013). "A ubiquitous method for street scale spatial data collection and analysis in challenging urban environments: mapping health risks using spatial video in Haiti." International Journal of Health Geographics 12.

DeLeeuw, E. and M. Collins (1997). "Data collection methods and survey quality: An overview." Survey Measurement and Process Quality: 199-220.

Duncan, D. T., Aldstadt, J., Whalen, J., Melly, S. J., & Gortmaker, S. L. (2011). Validation of Walk Score® for estimating neighborhood walkability: an analysis of four US metropolitan areas. International journal of environmental research and public health, 8(11), 4160-4179.

Diesner, J., K. M. Carley and L. Tambayong (2012). "Extracting socio-cultural networks of the Sudan from open-source, large-scale text data." Computational and Mathematical Organization Theory 18(3): 328-339.

Fleisher, C. S. (2008). "Using open source data in developing competitive and marketing intelligence." European Journal of Marketing 42(7-8): 852-866.

Frias-Martinez, V., & Frias-Martinez, E. (2014). Spectral clustering for sensing urban land use using Twitter activity. Engineering Applications of Artificial Intelligence, 35, 237-245.

Hirsch, J. A., Moore, K. A., Evenson, K. R., Rodriguez, D. A., & Roux, A. V. D. (2013). Walk Score® and Transit Score® and walking in the multi-ethnic study of atherosclerosis. American journal of preventive medicine, 45(2), 158-166.

Karnieli, A., U. Gilad, M. Ponzet, T. Svoray, R. Mirzadinov and O. Fedorina (2008). "Assessing land-cover change and degradation in the Central Asian deserts using satellite image processing and geostatistical methods." Journal of Arid Environments 72(11): 2093-2105.

Kelly, C. M., Wilson, J. S., Baker, E. A., Miller, D. K., & Schootman, M. (2013). Using Google Street View to audit the built environment: inter-rater reliability results. Annals of Behavioral Medicine, 45(1), 108-112.

Krieger, N., J. T. Chen and G. Ebel (1997). "Can we monitor socioeconomic inequalities in health? A survey of US health departments' data collection and reporting practices." Public Health Reports 112(6): 481-491.

Lee, S. J., E. A. Wentz and P. Gober (2010). "Space-time forecasting using soft geostatistics: a case study in forecasting municipal water demand for Phoenix, Arizona." Stochastic Environmental Research and Risk Assessment 24(2): 283-295.

Mennis, J. and S. L. Dayanim (2013). "Neighborhood collective efficacy and dimensions of diversity: a multilevel analysis." Environment and Planning A 45(9): 2176-2193.

Odgers, C. L., Caspi, A., Bates, C. J., Sampson, R. J., & Moffitt, T. E. (2012). Systematic social observation of children’s neighborhoods using Google Street View: a reliable and cost-effective method. Journal of Child Psychology and Psychiatry, 53(10), 1009-1017.

Pivo, G., & Fisher, J. D. (2011). The walkability premium in commercial real estate investments. Real Estate Economics, 39(2), 185-219.

Reichman, O. J., M. B. Jones and M. P. Schildhauer (2011). "Challenges and Opportunities of Open Data in Ecology." Science 331(6018): 703-705.

Rundle, A. G., Bader, M. D., Richards, C. A., Neckerman, K. M., & Teitler, J. O. (2011). Using Google Street View to audit neighborhood environments.American journal of preventive medicine, 40(1), 94-100.

Seo, G. D., A. Yazici, U. Ozguner and J. H. Cho (2008). "An approach for data collection and Traffic Signal Control in the futuristic city." 10th International Conference on Advanced Communication Technology, Vols I-Iii: 667-672.

Vittek, M., A. Brink, F. Donnay, D. Simonetti and B. Desclee (2014). "Land Cover Change Monitoring Using Landsat MSS/TM Satellite Image Data over West Africa between 1975 and 1990." Remote Sensing 6(1): 658-676.

Wang, Q. (2008). "Race/ethnicity, gender and job earnings across metropolitan areas in the United States: A multilevel analysis." Urban Studies 45(4): 825-843.