

## **Development of a flow-based planning support system based on open data for the City of Atlanta**

Ge Zhang<sup>1</sup>, Wenwen Zhang<sup>1</sup>, Subhrajit Guhathakurta<sup>1</sup> and Nisha Botchwey<sup>2</sup>

---

1. Center for Geographic Information Systems, Georgia Institute of  
Technology, Atlanta, GA, USA

2. School of City and Regional Planning, Georgia Institute of Technology,  
Atlanta, GA, USA

G. Zhang (Corresponding author)

Email: [ge.zhang@coa.gatech.edu](mailto:ge.zhang@coa.gatech.edu)

W. Zhang

Email: [wzhang300@gatech.edu](mailto:wzhang300@gatech.edu)

S. Guhathakurta

Email: [subhro.guha@coa.gatech.edu](mailto:subhro.guha@coa.gatech.edu)

N. Botchwey

Email: [nisha.botchwey@coa.gatech.edu](mailto:nisha.botchwey@coa.gatech.edu)

**Keywords:** Open data, Planning Support System, Flow-based, Web Application

## **Abstract**

Open data has come of age with many cities, states, and other jurisdictions joining the open data movement by offering relevant information about their communities for free and easy access to the public. 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 flow-based planning support system for examining neighborhood quality of life and health for the City of Atlanta as a prototype, which addresses many of these open data challenges.

## **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; French, Barchers, & Zhang, 2017). Open data is assumed to be non-privacy-restricted and non-confidential data, which can be used or distributed by public without any charge (Janssen et al., 2012). One good example is the water database provided by the United State Geological Survey (USGS). This database provides real-time and historical 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 offers characteristics regarding population and households at different geographic scales, such as the census tract and the block group, and has been used in myriad applications, such as in urban and environmental modeling and policy assessments (Wang, 2008; Mennis and Dayanim, 2013). However, the open data cannot be directly used without some preprocessing because of the variation in data 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, to date, 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. Planning support systems (PSS) have been widely used for data management, modeling and planning support in the urban planning (Geertman and Stillwell, 2003; Geertman et al., 2015). However, the data for PSS are usually hosted locally and highly customized; hence difficult to use for other purposes. 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 practices. 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 system platform (GIS). Additionally, the concept of flow-based programming has been introduced as a framework for the tool described in this paper. It is a programming paradigm, which we expect will make open data analysis more intuitive and user-friendly. The 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 and flow 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 directions that merit further efforts.

## **2. The current state of utilization of open data**

Several national and state level legislations in the U.S. have made non-confidential government data accessible for public use. The primary goal of these legislations is to engage citizens, businesses, and other organizations in public-oriented decision making processes via access to more transparent public datasets. The Open, Public, Electronic, and Necessary (OPEN) Government Data Act was enacted in 2016, which makes the open data policy signed by President Obama in 2013 into law (Schatz, 2016). To date, at least 36 states have passed various legislations requiring state websites to provide data across departments (National Conference of State Legislatures, 2016). Additionally, public agencies have also emphasized the need for large private firms, such as Uber and Google, to release some of their data for public use (Vaccaro, 2016). These mandates have led to abundant machine readable and easy to access open data at national, state, and even city levels.

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. The URL links for national and state level datasets can be found in Appendix A.

### 2.1 National Level Open Data

National level data provide information for all or most 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 and counties 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. 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, which can manipulate and analyze the data.

**Table 1.** National Level Open Data Source Summary

<b>Data Source</b>	<b>Data Content</b>	<b>Data Resolution</b>	<b>Obtaining Method</b>	<b>Data Format</b>
<b>Census Bureau</b>	Socio-economic Demographic	State	Manually	csv
		County	Download	xlsx
		Census Tract	API	shapefiles

		Block Group Block		
<b>Data Government</b>	Agriculture/ Business/ Climate/ Consumer/ Ecosystems/ Education/ Energy/ Finance/ Health/ Local Government/ Manufacturing/ Ocean/ Public Safety/ Science Research	State County Census Tract Zip code	Manually Download	csv xlsx pdf shapefiles
<b>Google Data</b>	Place/ Elevation/ Street Views/ GTFS transit	Most cities in U.S.	API	Json
<b>Zillow Data</b>	Housing	Household Units	API	xml
<b>Walk Score</b>	Walk Score/ Transit Score	Most cities in U.S.	API	xml
<b>Social Media</b>	Twitter/ Facebook/ Foursquare	Most cities in U.S.	API	xml
<b>Quandl</b>	Economic Earnings Housing	Most cities Company Aggregated	API	Json
<b>Wunderground</b>	Weather	City Zipcode	API	Json xml
<b>New York Times</b>	All types of News	All over U.S.	API	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 evaluate transit service quality.

Social media companies also provide a great deal of open source data. Posts on social media 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 random forest trees, 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, and impressions about the quality of particular places as expressed in social media. The data from social media accumulate and grow over time but may require some effort and programming 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 includes property sale records; house characteristics, such as year built, the number of rooms, lot size, and square footage, as well as some neighborhood-level information, including nearby school quality, crime, and risk of exposure to hazards. It is also possible to download historical 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 historical data about weather, which can be extremely useful when developing plans relevant to climate change. The New York Times offers archives of past news reports, making it easier to look into the important historical events in 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 assessing 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. Madaio et al. (2016) used Google Place data and machine learning tools to develop a model named Firebird to prioritize fire inspections in Atlanta.

Social media data has also been widely used in many studies to describe and understand the social dynamics of a 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. Their 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). While there is potential to conduct longitudinal real estate price change studies using data from both Walkscore and Zillow, few studies have taken advantage of these datasets beyond simple derivation of 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, their use in addressing the real-world planning problems have been limited.

## *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 to comply with governmental transparency protocols. The government-related data usually includes information on funding distribution and financial reports. In this study, we did not consider states that only have limited data dealing with budgets and expenditures. A total of 20 states were identified to have more extensive open data portals, which are tabulated in Figure 1.

As may be expected, there is wide variation among states in the extent, resolution, and format of open data. Besides government and transparency related 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, as illustrated in Figure 1. 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. However, there is less data available for tracking urban utility services, business licenses and permits. The organization of data across states is not uniform. 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, as shown in the left column of Figure 1. For instance, States, such as California, Maryland, and Missouri, tend to have data available across various sectors. While, some other states only provide data covering less than five attributes. 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.

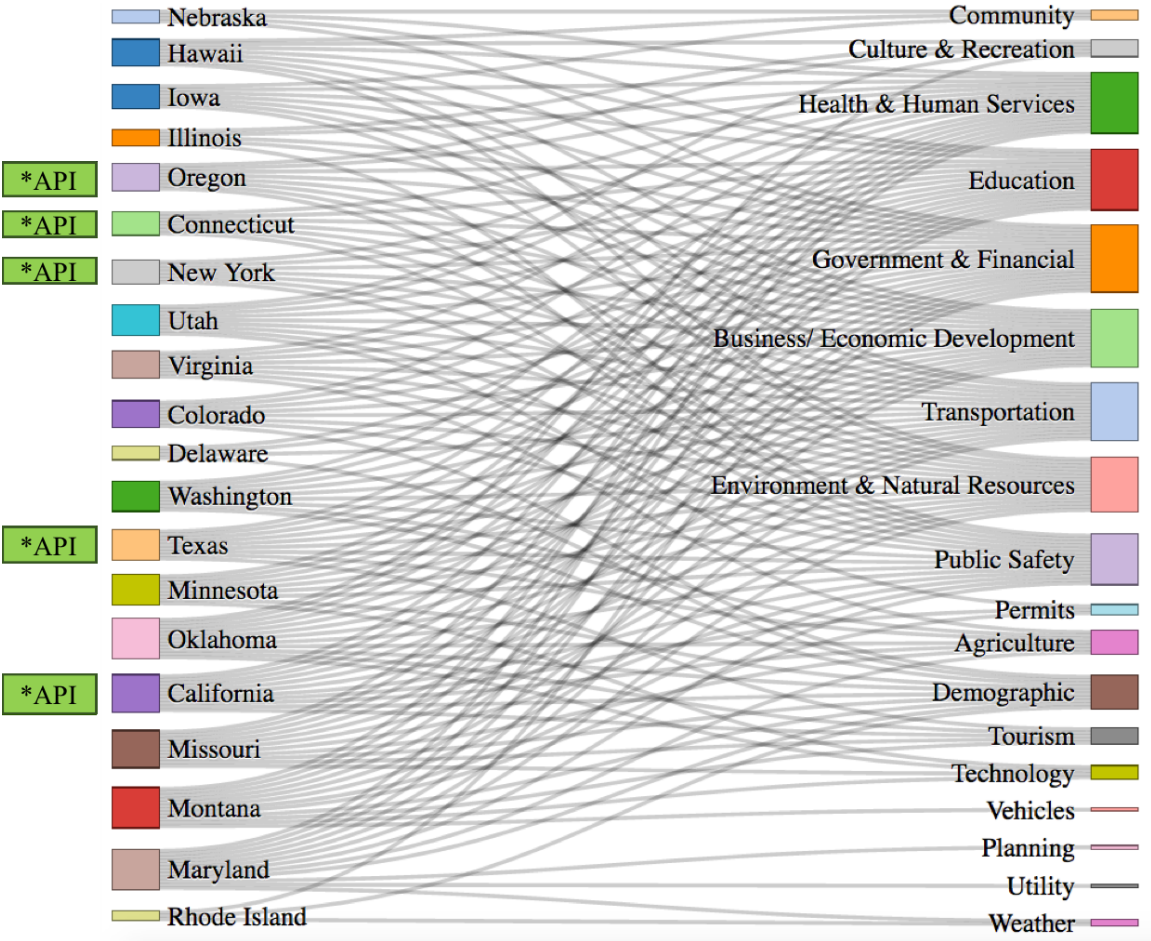


Figure 1. State Level Open Data Summary

Many state data portals provide 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 (<http://us-city.census.okfn.org/>), powered by the Open Knowledge Foundation, has a comprehensive list of open data available for various cities. The census calculates open data score for each city based on data quality, availability, and license conditions. The census evaluates data from 19 attributes, such as assets, budgets, crime, and zoning. Currently, this census provides information for up to 94 cities, with a total of 761 datasets. Approximately 25% of these datasets are open online. The census also provides a limited evaluation of the data quality available at the city level. For instance, some data on public safety are available online. 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 a map on the distribution of crime, while other places provide GIS shapefiles showing the type, location, and timestamp of crimes by type. 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, 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 (See Figure 2). However, it must be noted that this census is crowd-sourced, 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.

**Figure 2.** Top 10 Cities with the highest open data score (source: <http://us-city.census.okfn.org/>)



### 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 difficult for users to download entire datasets. Often, data providers do not allow the 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 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. Unlike the U.S. Census Bureau whose major purpose is to provide data to the general public, many national level data providers only provide an API for users to download the data. The use of an API for data downloads is more complicated than hitting a download button since a programming background is needed to write scripts for scraping data in large batches. This hurdle 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 an API, the download process can be automated, which make it easier 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 expertise in image processing and data mining machine learning models.

Third, most data downloaded with APIs are formatted into .json and .xml files. These data storage structures are not well-known to planners. The API accessible data are formatted using uniform code so that they can be transferred via the internet in a more efficient manner. Unlike the conventional Excel spreadsheet or ASCII data, these types of data tend to be highly machine readable but are not easily interpreted in the same way as raw text or numbers. Therefore, it is critical for planners to learn how to work with this form of data for their analysis and communication.

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.

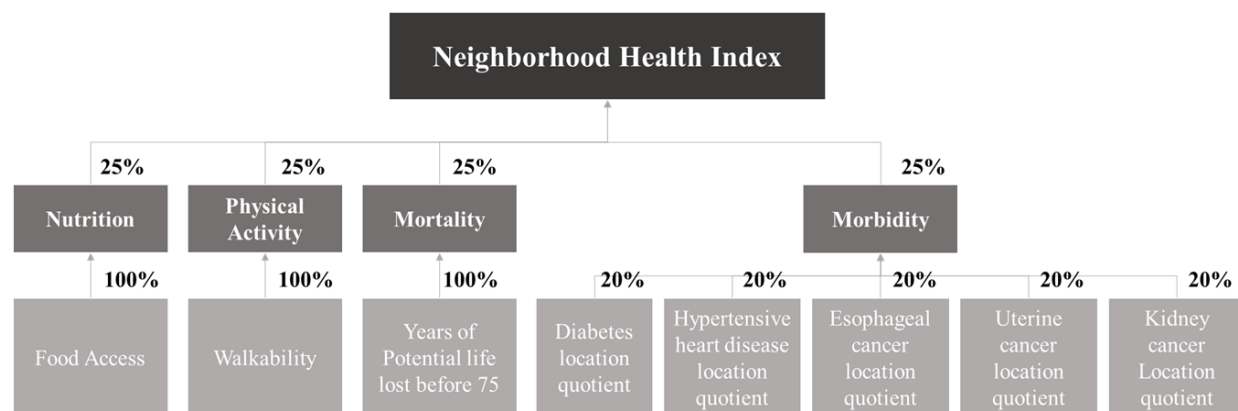
### 4. Finding solutions through open data-based tool development

We developed a web-based application named Flow-based Planning Support System (FPSS) to address some of the challenges in using open data and provide an intuitive tool for indicator creation and visualization. Currently, the pilot application can visualize neighborhood characteristics of the City of Atlanta at the neighborhood planning unit (NPU) level using two indices – the quality of life index and the quality of health index. The application demonstrates how the impact of planning decisions can be visualized with the help of a platform that offers several options for examining neighborhood characteristics that directly affect residents' health outcomes and quality of life.

The proposed FPSS system has two primary components: 1) a back-end, which helps to collect and refine data obtained from various open sources; and 2) a front-end platform, which includes a *virtual sandbox* for index calculation and a *visualizer* to visualize the index constructed or selected by users. In contrast to other planning support systems, the FPSS doesn't require users to input commonly available datasets. Instead, the system automatically collects, updates, manages, and cleans data from various open

databases in the backend. To be more specific, FPSS downloads data using either available APIs to existing data sources or from published URL links to the open data sources. The back-end processes then re-aggregate the open data to the desired geographic units selected by users. The reaggregated data/variables are then ready for use in the front-end *virtual sandbox*, thereby saving a considerable amount of time on data collection and cleaning processes. Given that, at the moment, many useful datasets are unavailable from open sources, the users have the option to upload their own datasets.

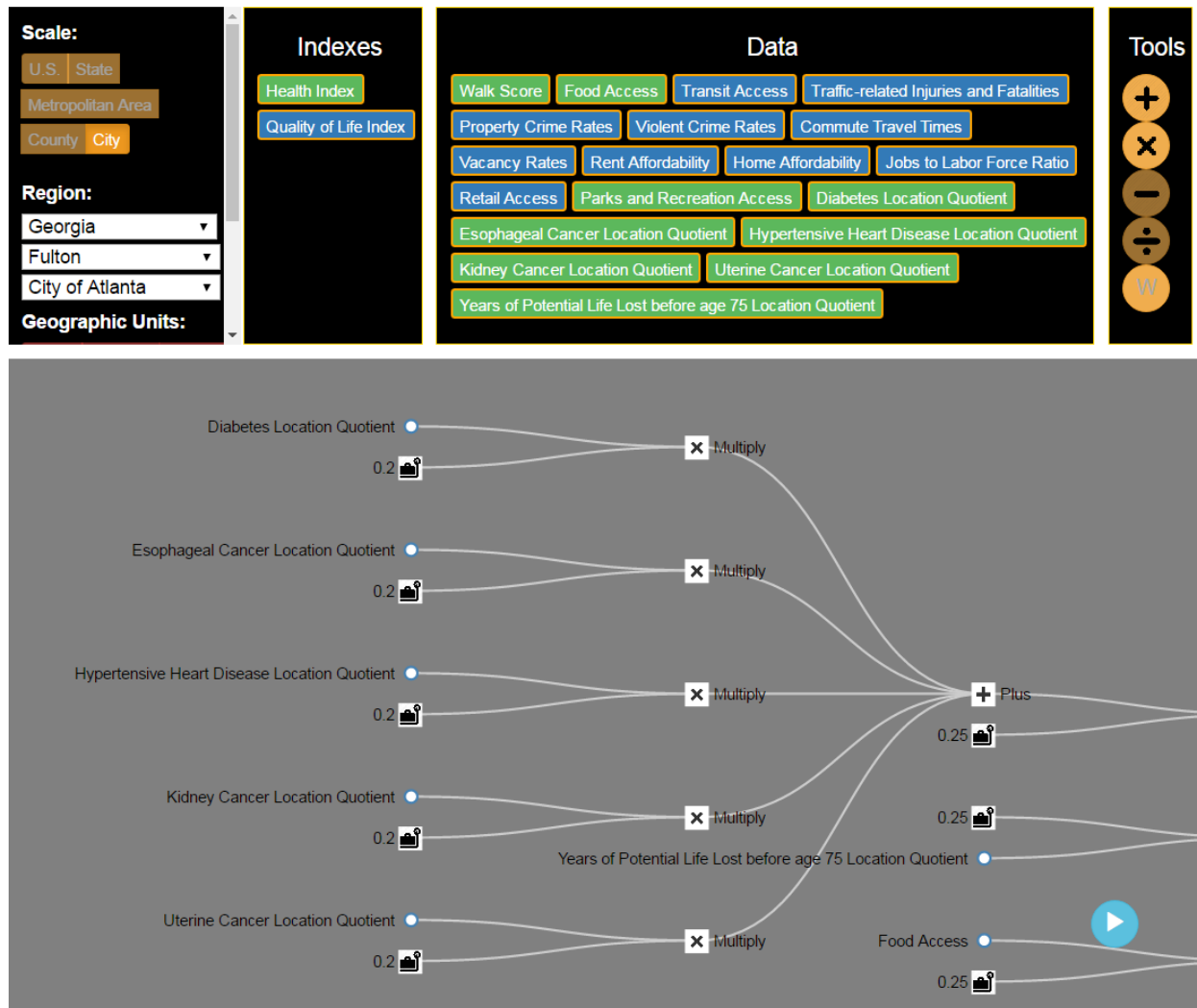
The FPSS *virtual sandbox* serves as a platform for users to define a study area, the geographical units, and select variables that are prepared and aggregated in the back-end. The selected data can be used to construct composite measurement indices for examining relationships and assessing effectiveness of potential planning decisions. The measurement indices are generally constructed from on a tree-based framework using different variables and aggregated indicators. For instance, the health index for the Atlanta's neighborhood planning unit level is calculated from a framework illustrated in Figure 3. The framework uses two levels of attributes and different weights to generate the aggregated health condition of the neighborhoods. The calculation process itself can be burdensome. When the lower level variables are updated, the calculation for higher levels in the branch of that tree needs to be computed again to update the final neighborhood index. Additionally, when the structure of the index is more complicated than the neighborhood health index example in Figure 1, the data collection effort will occupy significant amount of time and human resources, which often are luxuries in many public sector and planning offices.



**Figure 3.** Example Framework for Composited Neighborhood Health Index

To address the challenges mentioned above, the FPSS *virtual sandbox* is designed to help construct, manage and update the tree-based composite planning indices more efficiently and intuitively. The FPSS *virtual sandbox* adopts a flow and graphic based index construction platform, which visualizes the structure of the intended index in real time, as shown in Figure 4. The users can conveniently add/remove variables by dragging the variables and operators into/out of the sandbox. Before using the *virtual sandbox*, users need to first identify study area and geographic units, by selecting the corresponding states or cities from the drop down menu in the “Study Area” Box, shown on the top left corner of the interface. The geographic units available for selection includes state, county, zip code, neighborhood, neighborhood planning unit, census tract, block group, and block. After determining the study area, the corresponding available variables collected from open source data will be updated in the “Data” Box. The users can then start to construct indices by dragging available variables from “Data” Box and mathematical operators from “Tools” Box into the *virtual sandbox* area. All the variables and tools will be visualized as node icons in the *virtual sandbox*, with corresponding variable names and tool information. The current activated tools include addition, multiplication, subtraction, and assigning weights or constants. The variables and constant nodes can be joined together by an addition or multiplication tool node in the *virtual sandbox* by dragging the nodes close to other relevant ones. The successfully connected nodes will

be joined with edges. If two variables together with associated weights are joined using a multiply tool, they are multiplied together to form a new variable/index. Similar rules will also apply to the other tools that will be included over time. The users can update weights by clicking on the weight nodes and directly type in new values. Users can also hide or unhide substructure of the tree by clicking on the “parent” nodes of the branch. The structure of the index will remain unchanged, however, the users may find it easier to focus on unhidden branches of the variables.

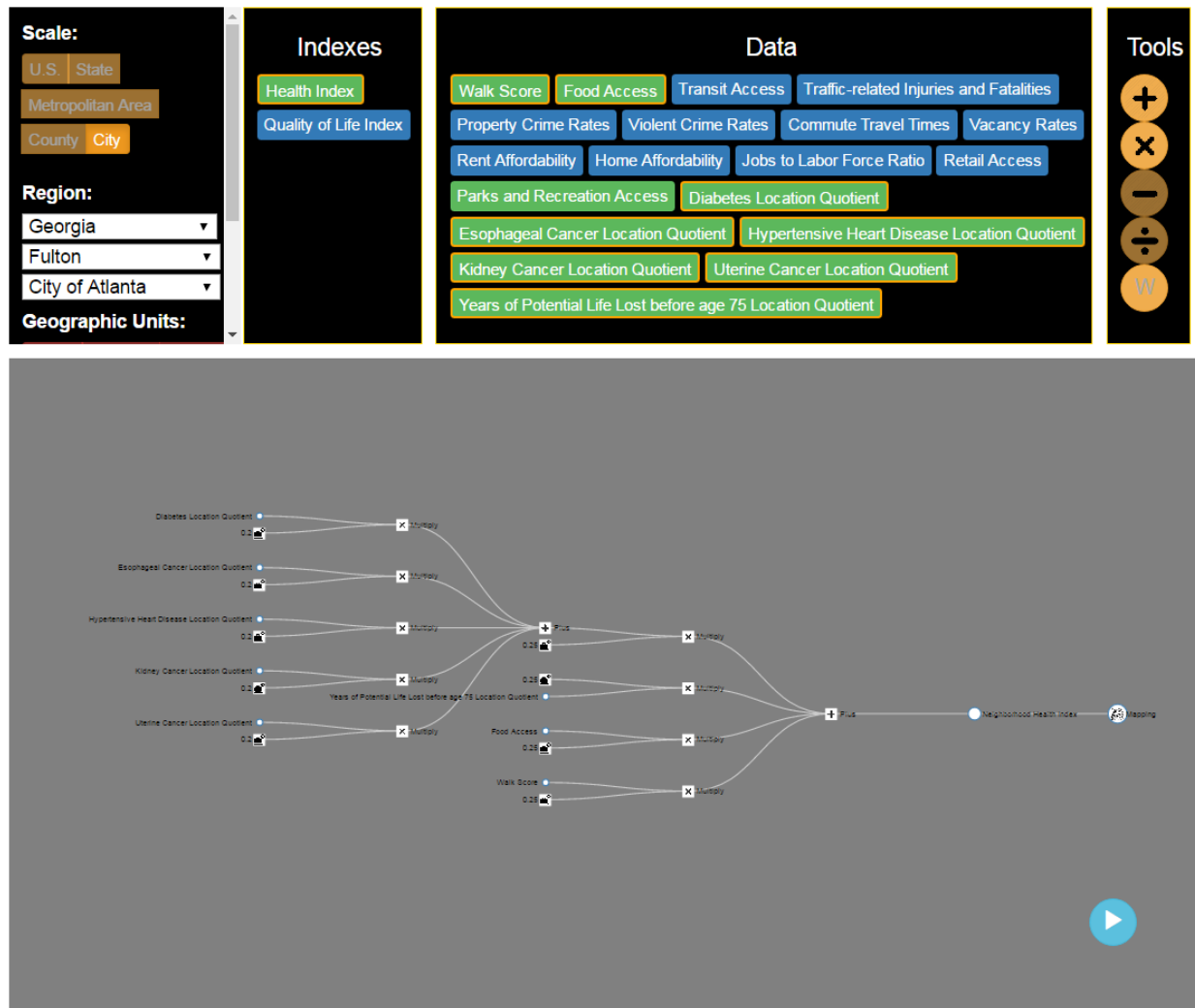


**Figure 4.** The interface of “virtual sandbox”

After constructing the composite index, the users can directly visualize the results by switching to the *visualizer* in the FPSS, which is activated by clicking on the blue circular button at the bottom right corner of the virtual sandbox. The major visualization component will be a thematic map at the selected geographic unit and scale. Based on the map, users can easily identify the spatial patterns of the constructed indices, as well as of other variables. Additionally, the *visualizer* also offers options for users to examine the descriptive statistics, scatter plots, and bar charts of the computed indices. The *visualizer* also provides the ability to compare among geographic units based on particular variables of interest or the final planning index. For instance, the *visualizer* can provide a bar chart to compare the health index for a selected neighborhood to that of the average for the city.

The front-end of the FPSS tool described above was developed using HTML and JavaScript. Several JavaScript libraries were used, such as jQuery and Google Map JavaScript API to support different functions. The back-end of the FPSS is designed to collect data from various open source data repositories, convert their format and geographic scale and perform calculations based on the structure designed by the users in virtual sandbox. The functionalities of the back-end were developed using Python 2.7 on Microsoft Azure.

To demonstrate the usability of the proposed platform, we used FPSS to generate quality of life and health condition indices for the neighborhoods in Atlanta. First, we designed the structure of neighborhood quality of life (NQOL) and neighborhoods health (NH) indices based on a review of the literature. The weights of variables that were included in the NQOL index were based on a citizen survey conducted by the City of Atlanta. The status of neighborhood health (NH) was based on measures of healthy food access, physical activity, mortality and morbidity in neighborhood populations, as shown in Figure 1. We programmed FPSS to collect data from several different open source databases, such as housing data from the U.S. Census database and public safety data from the Atlanta Police Department. We manually uploaded the datasets that cannot be collected automatically from public servers. Microsoft Azure was used as the back-end platform for data storage and cloud computing. To develop the two indices in the sandbox, we first selected “city” as the scale of the project, region as “Georgia -> Fulton County -> City of Atlanta”. The geographic units selected was the neighborhood planning unit. The available variables are then updated in the “Data Box”. We constructed NH and NQOL indices by dragging variables and calculation tools into the virtual sandbox and configuring the weights tool. The variable tags in the Data Box are colored green when they are used in the construction of the current open (unhidden) index. The specification of the constructed NH index is illustrated in Figure 5 (The NQOL index is hidden and the variables used in that index are colored blue in the “data” box).

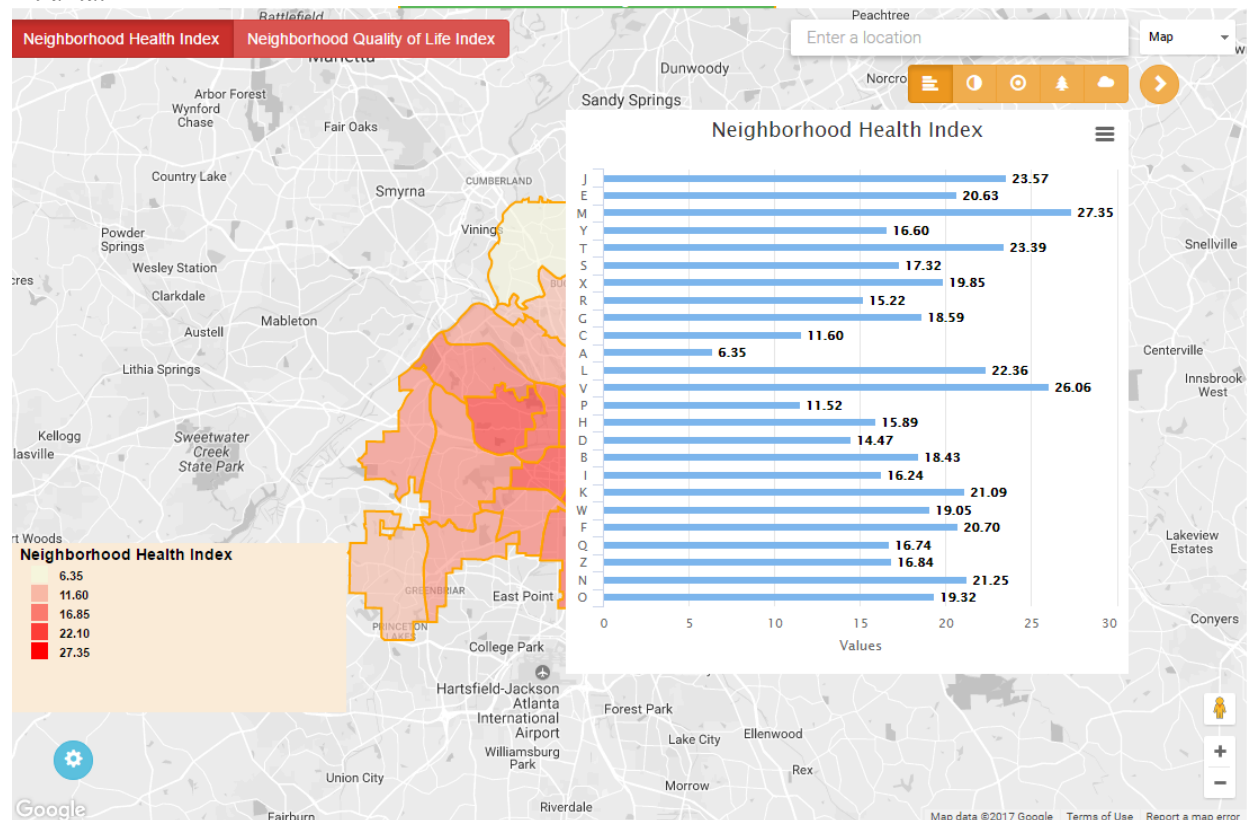


**Figure 5.** Construction of NH in FPSS *sandbox*

The constructed indices are visualized in the FPSS visualizer, as shown in Figure 6. The major mapping area displays a thematic neighborhood scale map for the neighborhood health index. The users can click on the index name tags on the top left of the map to switch between maps for different calculated indices. The thematic maps can help users identify particular clusters of interest in the study area. For instance, the map indicates that the health condition in urban center tends to be better than in the fringe neighborhoods. The visualizer also offers a search function for users to zoom into a specific address in the study area to further examine the values of the calculated indices at the place of interest. Moreover, the users can click on corresponding icons to activate bar charts (as shown in Figure 6), pie charts, scatter plots, tree charts, and spider charts to compare indices among different neighborhoods in



Atlanta.



**Figure 6.** The Visualization of the Computed indices

In the Atlanta test case, the challenges for accessing, cleaning and using open data were resolved through various approaches and tools at the back-end and front-end of FPSS. First, FPSS used U.S. Census API and some Python libraries (i.e. json and urllib) to connect the back-end database on Microsoft Azure 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 data in real-time. Second, some python libraries for spatial analysis, such as shapefile and Fiona, were used in FPSS to transform data from census tract level to neighborhood planning unit level to resolve the geographic unit mismatching problem. The FPSS also used these libraries for conducting spatial analysis at the scale of neighborhood planning units. Additionally, Some Python libraries for data formatting, such as pyshape, json and csv, were used in FPSS to convert data with different formats, such as csv, json and shp. Finally, the use of sandbox and visualizer also demonstrates that FPSS front-end can help construct useful indicators, which can then be visualized, and updated in real-time.

## 5. Discussion

In this study, we demonstrated that the FPSS can be utilized to design, calculate, and visualize a set of composite planning indices at the neighborhood planning unit level for the City of Atlanta. This test case illustrates the conceptual and operational model of an open data-based planning support system that offers an user-friendly tool to design and visualize place based indicators.

The Atlanta test case demonstrates that the FPSS is effective in collecting data from various open sources, such as the Census Bureau and Atlanta police department, in automatically converting the data to desired formats, and in rescaling the geographic units to the preferred resolution. The FPSS *virtual sandbox* adopts a flow-based user interface, which is intuitive, flexible, and can facilitate designing, calculating, and updating the composite planning index. Finally, the FPSS *visualizer* can illustrate the calculated index using thematic maps, charts, and tables to help stakeholders from different domains, especially those without planning backgrounds, to better inform their decision-making process.

However, like most applications, the current version of FPSS has its limitations that merit further improvements. First, the existing open data cleaning module in the FPSS can only achieve conversion among file formats such as txt, csv, and shp files. Other data formats, such as json and xml files, cannot be easily incorporated into the platform. These conversion gadgets have already been developed and can be incorporated into the FPSS in the future. In addition, the aggregation function of the FPSS can be advanced by integrating more geographic unit boundary data into the platform. Currently, the FPSS only supports the scale transformation between census and planning neighborhood unit. Later version of FPSS will include a new module to easily convert data from one geographic scale to another, if users can provide both boundaries, in either shp or json format. Furthermore, historical data are still unavailable in FPSS, which means the system cannot support longitudinal analysis based on indices for different time periods. The current datasets, used by FPSS, are the latest version of the original database. We expect that FPSS will include data across multiple periods and years in the future.

The fourth limitation of the FPSS system is that users, to date, cannot import local data into the platform, rendering it less flexible. Most planning decision-making processes need to account for data from various stakeholders in diverse sectors such as healthcare, education, private sector businesses, and municipal governments. Therefore, it is necessary for the platform to accept user configured data, which can then be incorporated into composite indices. Such functionality is currently under construction.

A fifth limitation is that open data from different sources are updated at different intervals. When such data are comingled, there could be a mismatch in the temporal aspects of the indicators that are created from the data. 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, reducing the temporal lags among different datasets. However, in this application, some of the open data have long update intervals while some data are static.

The sixth limitation is that there has been no user testing of this platform to check for usability and potential bugs. Usability testing is important because, as previously reported, PSS implementation has faced challenges related to insufficient diffusion in planning practice and a lack of acceptance among targeted users (Vonk and Geertman, 2008). These factors may also affect the user acceptability and functionality of the developed FPSS platform. By adopting the flow-based user interface, we expect to reduce some of these implementation concerns about PSS discussed in the prior studies. However, it remains critical to collect user feedback to validate and improve the system.

A final limitation of the FPSS is that the current computational tools in the system cannot support more complicated spatial analysis. For example, spatial queries like selecting data samples within a certain study area is unavailable in the latest version of the system. More advanced spatial query and joining modules will be included soon to make the system more powerful for practitioners.

## **6. Conclusion**

This paper presents a novel open data-based community planning support system that provides a platform for computing, mapping, and visualizing spatial data. To develop this application, we analyzed the current state of open data efforts in community planning related domains; examined potential challenges associated with effectively using open data in the planning process; developed several tools based on Python at the back-end of the application to address some of these challenges; and designed an interactive graphical user interface based on HTML and Javascript at the front-end to help in constructing, updating, and visualizing composite planning indices to better inform decision-making processes.

The flow-based application was applied to assess the quality of life and health in the City of Atlanta at the neighborhood planning unit scale. The two composite variables, quality of life and health, were created by aggregating information from 19 variables in nine categories. The users can explore the spatial variation of each variable in different ways, including a map with geographic locations, a chart with trend estimation and a table with real numbers. The Atlanta test case demonstrates that this application can help various users, including planners, researchers, government officials, and the general public, better understand community trends and be more critically engaged in the community planning process. Despite its promise, there is much scope for improvement, such as including a larger variety of supported data formats, incorporating longitudinal spatial data, improving data resolution, enhancing interface design based on user feedback and developing more comprehensive spatial query and joining tools.

#### Appendix A

Table 2. National Level Data Sources URLs

<b>Data Sources</b>	<b>URL</b>
Census Bureau	<a href="https://www.census.gov/">https://www.census.gov/</a>
Data Government	<a href="https://www.data.gov/">https://www.data.gov/</a>
Google Data	<a href="https://developers.google.com/apis-explorer/#p/">https://developers.google.com/apis-explorer/#p/</a>
Zillow Data	<a href="http://www.zillow.com/howto/api/APIOverview.htm">http://www.zillow.com/howto/api/APIOverview.htm</a>
Walk Score	<a href="https://www.walkscore.com/professional/api.php">https://www.walkscore.com/professional/api.php</a>
Social Media	<a href="https://dev.twitter.com/docs;">https://dev.twitter.com/docs;</a> <a href="https://www.yelp.com/developers/documentation/v2/overview">https://www.yelp.com/developers/documentation/v2/overview</a>
Quandl	<a href="https://www.quandl.com/">https://www.quandl.com/</a>
Wunderground	<a href="https://www.wunderground.com/weather/api/">https://www.wunderground.com/weather/api/</a>
New York Times	<a href="https://developer.nytimes.com/">https://developer.nytimes.com/</a>

Table 3. State Level Data Sources URLs

<b>Data Sources</b>	<b>URL</b>
California	<a href="https://data.ca.gov/">https://data.ca.gov/</a>
Colorado	<a href="https://data.colorado.gov/">https://data.colorado.gov/</a>
Connecticut	<a href="https://data.ct.gov/">https://data.ct.gov/</a>
Delaware	<a href="https://data.delaware.gov/">https://data.delaware.gov/</a>
Hawaii	<a href="https://data.hawaii.gov/">https://data.hawaii.gov/</a>
Illinois	<a href="https://data.illinois.gov/">https://data.illinois.gov/</a>
Iowa	<a href="https://data.iowa.gov/">https://data.iowa.gov/</a>
Maryland	<a href="https://data.maryland.gov/">https://data.maryland.gov/</a>
Minnesota	<a href="http://www.state.mn.us/opendata/">http://www.state.mn.us/opendata/</a>
Missouri	<a href="https://data.mo.gov/">https://data.mo.gov/</a>

Montana	<a href="https://data.mt.gov/">https://data.mt.gov/</a>
Nebraska	<a href="http://www.nebraska.gov/government/open-data/">http://www.nebraska.gov/government/open-data/</a>
New York	<a href="https://data.ny.gov/">https://data.ny.gov/</a>
Oklahoma	<a href="https://data.ok.gov/">https://data.ok.gov/</a>
Oregon	<a href="https://data.oregon.gov/">https://data.oregon.gov/</a>
Rhode Island	<a href="http://www.ri.gov/data/">http://www.ri.gov/data/</a>
Texas	<a href="https://data.texas.gov/">https://data.texas.gov/</a>
Utah	<a href="https://opendata.utah.gov/">https://opendata.utah.gov/</a>
Virginia	<a href="https://data.virginia.gov/">https://data.virginia.gov/</a>
Washington	<a href="https://data.wa.gov/">https://data.wa.gov/</a>

---

## 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.

French, S. P., Barchers, C., & Zhang, W. (2017). How Should Urban Planners Be Trained to Handle Big Data? In *Seeing Cities Through Big Data* (pp. 209)

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.

Geertman, S., Ferreira, J., Goodspeed, R., & Stillwell, J. C. H. (2015). *Planning support systems and smart cities*. Springer.

Geertman, S., & Stillwell, J. (2003). Planning support systems: an introduction. In *Planning support systems in practice* (pp. 3–22). Springer.

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.

Madaio, M., Chen, S.-T., Haimson, O. L., Zhang, W., Cheng, X., Hinds-Aldrich, M., Chau., D. H., Dilkina, B. (2016). Firebird: Predicting Fire Risk and Prioritizing Fire Inspections in Atlanta. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. 13-17-August-2016 (2016)

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.

National Conference of State Legislatures. (2016). Statewide Transparency and Spending Websites and Legislation. Retrieved January 24, 2017, from <http://www.ncsl.org/research/telecommunications-and-information-technology/statewide-transparency-spending-websites-and-legis.aspx>

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.

Schatz, B. (2016, December 12). S.2852 - 114th Congress (2015-2016): OPEN Government Data Act [legislation]. Retrieved January 24, 2017, from <https://www.congress.gov/bill/114th-congress/senate-bill/2852>.

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.

Vaccaro, A. (2016). Boston wants better data from Uber, and is taking a roundabout route to try and get it. Retrieved January 24, 2017, from <https://www.boston.com/news/business/2016/06/28/uber-data-boston-wants>.

Vittekk, 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.

Vonk, G., & Geertman, S. (2008). Improving the adoption and use of planning support systems in practice. *Applied Spatial Analysis and Policy*, 1(3), 153-173.

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.