

Home	Table of Contents	Titles & Subject Index	Authors Index
----------------------	-----------------------------------	--	-------------------------------

What is being done with open government data? An exploratory analysis of public uses of New York City open data

Karen Okamoto

Assistant Professor, John Jay College, 549 West 59th Street, New York, USA.
E-mail: kokamoto (at) jjay.cuny.edu

Received June 2, 2016; Accepted June 25, 2016

Abstract

In 2012, New York City Council passed legislation to make government data open and freely available to the public. By approving this legislation, City Council was attempting to make local government more transparent, accountable, and streamlined in its operations. It was also attempting to create economic opportunities and to encourage the public to identify ways in which to improve government and local communities. The purpose of this study is to explore public uses of New York City open data. Currently, more than 1300 datasets covering broad areas such as health, education, transportation, public safety, housing and business are available on the City's Open Data Portal. This study found a plethora of maps, visualizations, tools, apps and analyses made by the public using New York City open data. Indeed, open data is inspiring a productive range of creative reuses yet questions remain concerning how useable the data is for users without technical skills and resources.

Keywords

Government information; Open data; Citizen participation; New York City

Introduction

In 2012, New York City Council passed legislation to make government data open and freely available to the public ("Local Law 11," 2012). The law stipulates that City agencies are required

to publish the data they own and maintain on the City's Open Data Portal (Figure 1. nycopendata.socrata.com). By data, the law, in summary, is referring to the "final versions of statistical or factual information" in alphanumeric form "regularly created or maintained by or on behalf of and owned by an agency" ("Local Law 11," 2012). Among the over 1300 data sets currently available on the Portal, users can browse, for example, the following data sets: building permits, the leading causes of death in New York City, demographic statistics of school districts, and a list of legally operating businesses. The law states that it attempts to make New York City government more transparent, accountable, and more streamlined in its operations. It also permits "the public to assist in identifying efficient solutions for government, promote innovative strategies for social progress, and create economic opportunities." The purpose of this study is to explore precisely this motive to "promote innovative strategies for social progress" by gathering and analyzing reuses of New York City open data (hereafter referred to as NYC open data).



Figure 1. The New York City Open Data Portal (nycopendata.socrata.com)

Note: Data sets are available for browsing and downloading from the portal.

To date, the field of research on open government data has been wide-ranging and productive. Researchers have covered various challenges and issues surrounding open data such as ensuring citizen participation (Evans & Campos, 2013), maintaining privacy (Scassa, 2014), and overcoming barriers to releasing data (Conradie & Choenni, 2014). Others have developed frameworks and benchmarks (Veljkovic, Bogdanovic-Dinic, & Stoimenov, 2014) to evaluate and assess open data policies (Zuiderwijk & Janssen, 2014) and open data's efforts to support transparency and accountability (Lourenco, 2015). However studies specifically pertaining to the uses of NYC open data or open data from other cities has been limited (Kassen, 2013). Kassen (2013) provides one such local case study of Chicago's open data initiative by examining its

potential to empower local residents and by reviewing tools that use Chicago's open data. Similarly, Goldstein and Dyson's (2014) edited collection, *Beyond Transparency*, surveys the open data landscape from the perspective of practitioners. These practitioners share their stories working with open data for various government bodies and agencies. This study attempts to contribute to this growing and productive area of research by examining the uses of NYC open data.

Materials and Methods

This study searched for and analyzed web apps, mobile apps, and visualizations that use NYC open data. This study also considered websites and blogs that specialize in gathering and interpreting NYC open data. These apps, visualizations, websites and blogs were identified and culled from a number of sources. The first source was the official NYC Open Data Portal which allows users to search for and download open data. The Portal features notable apps, visualizations and interpretations of NYC open data. The Portal does not indicate the date these items were produced or posted on the Portal. Ten open data apps, visualization and websites from the Portal were analyzed and considered for this study. The second source for items for analysis was the NYC open data Tumblr blog (nycopendata.tumblr.com). The blog featured eleven posts about open data products and services that were considered for this study. A third source was Twitter (twitter.com/search?q=%23opendata+AND+nyc). Tweets were found using the search string “#OpenData and NYC”. Tweets announcing a new NYC open data visualization, app, blog or website were selected for analysis. Due to the volume of Tweets found containing the search string “OpenData and NYC”, Tweets from a six month period -- October 1, 2015 to March 31, 2016 -- were only considered for this study. A total of 37 items for analysis were culled from Twitter.

Several Tweets led to other websites that specialize in open data tools and analyses such as BetaNYC (projects.betanyc.us), I Quant NY (iquantny.tumblr.com) and The Open Bus (www.theopenbus.com). BetaNYC is a non-profit organization that helped pass New York City's open government legislation. It organizes civic-minded technologists and volunteers to create digital tools for community empowerment. BetaNYC maintains a list of civic technology projects being developed in NYC. For this study, open data projects updated within a six month period (October 1, 2015 to March 31, 2016) were analyzed. Eleven projects from this list were analyzed for this study. I Quant NY is a blog by Ben Wellington, a quantitative analyst, devoted to analyses and insights into specific NYC open data sets. Blog posts from I Quant NY published within a six month period (October 1, 2015 to March 31, 2016) were analyzed for this study. Six of these blog posts were analyzed for this study. Lastly, The Open Bus is a site specializing in the City's Metropolitan Transportation Authority (MTA) public transit bus data and data from the Citi Bike bicycle sharing program. These two data domains were included in this study's

analysis. The Open Bus includes a dashboard for individual bus routes and data on a route's overall schedule performance. The Citi Bike data focuses on the dependability of a bicycle docking station, that is the percentage of time the station is empty or full with bicycles. Technically, New York State is responsible for the operation of the MTA, but because the MTA ultimately concerns the daily lives of New Yorkers it will be considered City data for this study.

Content analysis was used to process and analyze the data gathered from the aforementioned sources. A total of 77 items were gathered from all sources and were coded according to the data type or subject of the data. A total of 31 categories were identified from the data. These categories are listed as follows in no particular ranking or order: Business; public transit; housing; taxis; trees; New York City facts; 311 data; landmarks; parking; health; crime; census data; homelessness; environment; public payphones; snow plowing; fire; legislation; schools; city services; buildings; demographics; baby names; land records; local government contacts; traffic; daycare; city salaries; information about politicians; Citi Bike bicycle sharing program; and property taxes. Each item for analysis was assigned one or more of these categories based on the type or subject of the data used. Also, as part of the analysis, the developers or creators of the items and the format of the items were considered for this study.

Results

Types of products and analyses

A diverse range of products, tools, and analyses were found among the 77 items considered for this study. The 31 subject categories identified from the 77 items, as mentioned above, demonstrate this diversity. Products or services include the Business Atlas (maps.nyc.gov/businessatlas), a website created by New York City government to provide vital business data to entrepreneurs and Citygram (www.citygram.nyc), a service developed by volunteers from the non-profit, civic technology organization Beta NYC, sends updates to users on selected city service activities and complaints within a specified neighborhood. An interactive visualization (www.cloudred.com/labprojects/nyctrees) of the variety and quantity of street trees in New York City illustrates the diversity of trees in the city. The NYC Cares mobile app (<http://apple.co/28Q7Bhp>) was created to help senior citizens find food pantries in the city. And lastly, Death Map (<http://deathmapnyc.com>) shows the location of homicides and traffic fatalities during a two-year period. These examples demonstrate the diversity of the products, tools, services, apps, analyses and visualizations created using NYC open data.

Popular data sets

Among the 77 items analyzed for this study, 4 of the 31 coding subject categories applied to 5 (6%) or more of the items analyzed. Table 1 lists these coding subject categories and the

percentage of the 77 items they represent. The 311 data category refers to data collected by New York City’s 311 service, a service which provides information and responds to non-emergency needs for residents, business owners and visitors (www1.nyc.gov/311). The crime data category was used for items that dealt with crime statistics and law enforcement. Items which used bus data such as bus stop locations and the punctuality of individual bus routes were denoted by the public transit category. The environment category was used to refer to items that used data covering energy use, greenhouse gases, waste, and air quality. Beyond these 4 coding categories, housing data such as rent regulated units and landlord violations constituted 5% of the 77 items for analysis. Also representing 5% of the items were city tree data, and city services data. Taxi data, New York City factoids, parking violations and traffic or vehicular data were represented 4% of the items analyzed. Business data, Citi Bike data and data about homelessness constituted 2% of the items analyzed. The remaining categories only represented 1% of the items analyzed: landmarks, health data, census data, public payphones, snow plowing, fire, legislation, schools, buildings, demographics, baby names, land records, local government data, daycare, city salaries, politicians and property taxes.

Table 1. The four coding categories that applied to five or more of the items analyzed for the study

Data category	Percentage
311 data	12%
Crime	12%
Public transit	10%
Environment	6%

Developers

The items analyzed for this study were developed by an assortment of individuals and groups. These developers are summarized in Table 2. The majority of items (61%) analyzed for this study were developed by individuals with technical backgrounds. These individuals include Chris Whong, Civic Technologist & Developer at NYC Department of City Planning, and Timothy Martin, a data scientist. BetaNYC is a non-profit civic tech organization, comprised of individuals with varying technical backgrounds. As such, projects developed by BetaNYC were included in the “individuals with technical and/or quantitative backgrounds” developer category. Government departments constituted 13% of the developers. Departments include City Planning, and Health and Mental Hygiene. Technology companies such as Enigma and CartoDB

constituted 9% of the developers, followed by non-profit organizations (8%) such as WNYC radio and Neighbors Allied for Good Growth. Some items did not list the name or names of a developer (6%), or did not provide information about the developer (3%).

Table 2. Summary of the types of people and groups who developed the 77 items analyzed for this study

Developer	Percentage
Individuals with technical and/or quantitative backgrounds	61%
Government departments	13%
Tech companies	9%
Non-profit civic organizations	8%
Unknown developers	6%
Individuals with unknown backgrounds	3%
Total	100%

Formats

The items analyzed for this study were presented in a range of formats and types. Most of the items were in the form of maps (48%). One example of these maps was developed by Timothy Martin (Figure 2). Martin displays crime data by community board districts across NYC. The second common format (19%) was web applications. Dangerous Roads (Figure 3) is one example of a web application or web app. It allows users to find the history of collisions for particular roads. Infographics were third in popularity for format types (9%). For example, HereHere is a cartoon that summarizes 311 complaint data by neighborhood (Figure 4). The remaining types of items analyzed for this study include: extractors (6%), which scrape open data from sources such as PDF files; mobile apps (5%); APIs (application programming interface) (2%); charts (2%); a web browser extension for displaying complaints filed against a rental property; an interactive visualization of trees in NYC; an extended analysis of City Council members' proposed salary increase; an email alerting service for 311 complaints; and one hardware item that detects apartment temperatures to report heating violations.

2015 NYC Major Felony Offenses by Community Board

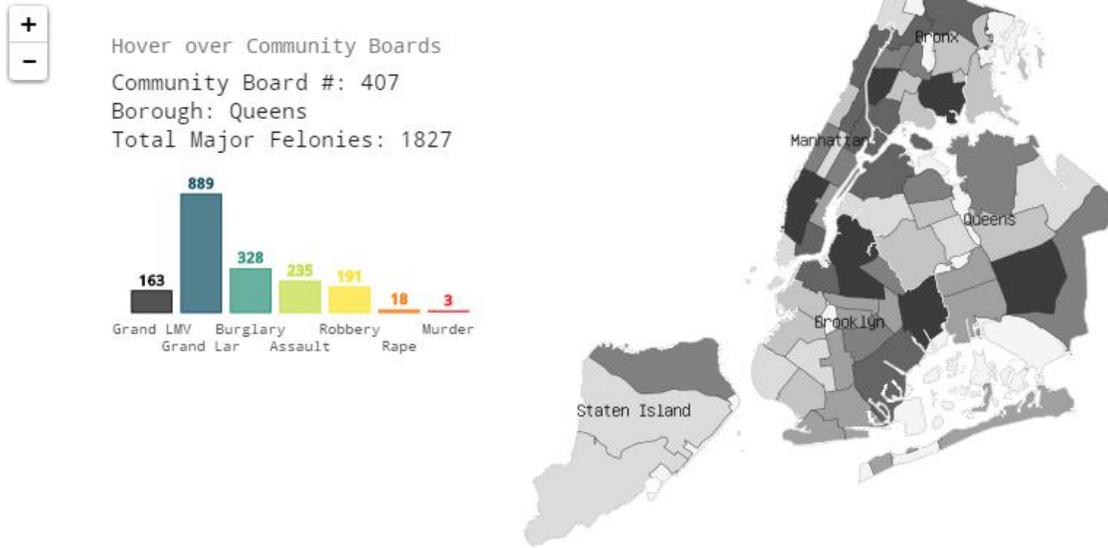


Figure 2. Crime data mapped: Major felony offenses (2015) by community board districts (timothymartin76.github.io/CMBD_Felony)

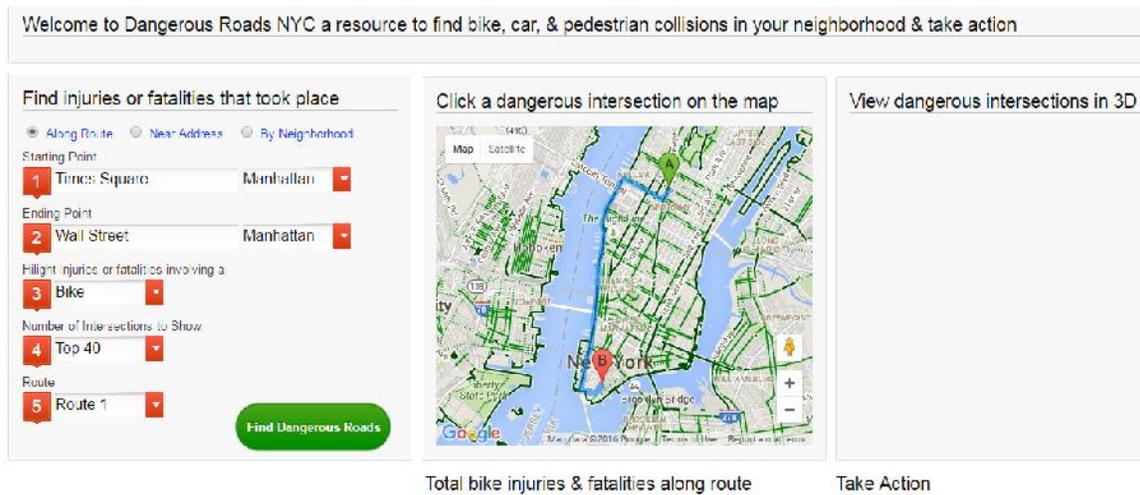


Figure 3. Example of a web app: Dangerous Roads NYC displays roads with collisions (dangerousroadsnyc.com)



Figure 4. Example of an infographic: HereHere summarizes 311 complaints by neighborhood in cartoon form (herehere.co/nabes)

Discussion

The datasets used to create the items analyzed for this study may inform City administrators about the types of datasets that the public is potentially interested in. Among the 77 items analyzed for this study, the following types of datasets were used by a number of the items analyzed: 311 data, crime, public transit, environmental quality, housing, trees, and city services. By monitoring or surveying existing open data usage, governments can gauge which datasets are of interest to users and gain a sense of the types of datasets that need to be complete, archived, updated regularly, standardized, and made easily reusable. Certainly, not all datasets can be released as government resources are limited and privacy issues are a concern. Dawes and Helbig (2010) point out that governments should analyze stakeholder needs to determine which datasets to prioritize for release. Acknowledging the importance of public input, the NYC Open Data Portal includes a dataset suggestion section (nycopendata.socrata.com/nominate) where users can suggest a dataset and approve a suggestion. Direct public input is undoubtedly important, but also analyzing which datasets are being transformed into tools and products for public consumption, as seen in this study, can help governments prioritize which datasets to release.

Indeed, the NYC Open Data Portal provides its own site analytics, displaying the number of datasets available and other figures (nycopendata.socrata.com/analytics). Users can select the date range for the site analytics they want to be displayed. The top five datasets according to the

analytics page for the time period of October 1, 2015 to March 31, 2016 -- the period for which items were culled from Twitter, the I Quant NY blog and the BetaNYC project list -- are summarized in Table 3. The Portal does not explain how the top datasets are measured, that is, whether it is measured by the number of views or the number of times the data was exported by a unique user. Interestingly, the top search terms are different from the top datasets. Table 4 lists the top five search terms for the period in question. The first and third search terms from this list are the same except for differences in capitalization. "Street tree census" was the top search term for the time period covered by this study. Comparing the results of this study with the top datasets and search terms listed on the NYC Open Data Portal, the more popular datasets among the items analyzed for this study -- that is, 311 data and crime data -- are nowhere to be found in the top datasets and search terms. Further research could look into the difference between datasets that are ultimately developed into public-facing tools and datasets that do not.

Table 3. Top datasets listed on the NYC Open Data Portal's analytics page (nycopendata.socrata.com/analytics) for the period October 1, 2015 to March 31, 2016

Dataset	Count
Department of Buildings (DOB) Job Application Filings	750,495
For Hire Vehicles (FHV) - Active Drivers	101,528
For Hire Vehicles (FHV) - Active and Inactive Vehicles	55,265
Medallion Drivers - Active	27,884
DOB Permit Issuance	25,630

Table 4. Top search terms listed on the NYC Open Data Portal's analytics page (nycopendata.socrata.com/analytics) for the period October 1, 2015 to March 31, 2016

Search terms	Count
Street Tree Census	738,691
Flood	57,127
Street tree census	52,430
Lifelong Learning	47,211
Jobs and Economic Mobility	9,329

In addition to gauging which datasets are of interest to the public, the question of who can actually use, manipulate, and interpret open government data becomes an issue. The results for this study found that 61% of the items analyzed were developed by individuals with technical and/or quantitative backgrounds. Ranking a distant second were items developed by government entities (13%). These figures might suggest that those with technical backgrounds have the skills and resources to use, manipulate, interpret and create tools using open government data. Several researchers have noted that not everyone can use open government data (Calzada Prado & Marzal, 2013; Conradie & Choenni, 2014; Evans, 2013; Gurstein, 2011; Janssen, Charalabidis & Zuiderwijk, 2012; Larrick, 2015). Technical and quantitative skills and resources are required to process the sheer volume and complexity of the data. Gurstein (2011) aptly warns that open government data may create a new digital divide where data can be used to “further empower and enrich the already empowered and the well provided for.” Indeed, as Gurstein further points out, there appears “to be some confusion between movements to enhance citizen ‘access’ to data and the related issues concerning enhancing citizen ‘use’ of this data.” Citizen outreach, data literacy training, and tools for easy data manipulation might address these concerns. To address this inequality, the Mayor of New York City introduced the “Open Data For All” initiative in 2015 to provide tools for people without programming experience to find and use the data they need (“New York City launches,” 2015). As well, civic technology organizations, such as BetaNYC, are organizing volunteers to share their skills to engage communities and governments in the name of public service. Further studies could examine effective ways to broaden open data usage.

Lastly, the various formats found among the 77 items analyzed for this study illustrate the creative possibilities that open government data can inspire. Again, a large percentage of the items analyzed was maps (48%). This could be due to the nature of city data, which can be more visually compelling, accessible and understandable when mapped. It could also be an indicator of the various mapping tools available such as OpenStreetMap (www.openstreetmap.org) and CartoDB (cartodb.com). The 37 maps that were analyzed for this study used a combination of mapping tools: 16 used CartoDB and 11 used OpenStreetMap, sometimes in combination with each other. As more of these tools to process and display open government data become available and accessible beyond tech communities there may be an increasing number of visualizations, tools, services and products for a broader audience.

Conclusion

Open government data is inspiring a wide range of creative reuses, visualizations, maps, apps, services, tools and analyses. This exploratory study examined a range of these items as featured on the New York City Open Data Portal and its Tumblr blog. Items were also culled from Twitter posts with the hashtag #opendata and the keyword “NYC”, the BetaNYC project list, the

I Quant NY blog, and the Open Bus website covering the time period from October 1, 2015 to March 31, 2016. A total of 77 items using NYC open data were found, analyzed and coded according to the data subject areas covered such as crime data and 311 complaint and service data. These results illustrate the creative products that open data can inspire. Further research could examine the impact of these products, for example, by examining the number of users. This exploratory study provides a snapshot of the creative and varied reuses of open data and seeks to contribute to the growing literature on the uses of open data from various cities.

References

- Calzada Prado, J., & Marzal, M.A. (2013), Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri: International Journal of Libraries & Information Services*, 63(2), 123-134.
- Conradie, P., & Choenni, S. (2014). On the barriers for local government releasing open data. *Government Information Quarterly*, 31(Sup 1), S10-S17.
- Dawes, S.S., & Helbig, N. (2010). Information strategies for open government: Challenges and prospects for deriving public value from government transparency. In M.A. Wimmer, J.L. Chappelet, M. Janssen, & H.J. Scholl (Eds.), *Electronic Government* (pp. 50-60). Berlin: Springer-Verlag Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-14799-9_5
- Evans, A.M., & Campos, A. (2013). Open government initiatives: Challenges of citizen participation. *Journal of Policy Analysis and Management*, 32(1), 172–185.
- Ganapati, S., & Reddick, C.G. (2012). Open e-government in U.S. state governments: Survey evidence from Chief Information Officers. *Government Information Quarterly*, 29(1), 115-122.
- Goldstein, B., & Dyson, L. (Eds.). (2014). *Beyond transparency: Open data and the future of civic innovation*. Retrieved from <http://beyondtransparency.org>
- Gurstein, M. (2011, February). Open data: Empowering the empowered or effective data use for everyone? *First Monday*, 16(2). Retrieved from <http://firstmonday.org/ojs/index.php/fm/article/view/3316/2764>
- Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, 29(4), 258–268.
- Kassen, M. (2013). A promising phenomenon of open data: A case study of the Chicago open data project. *Government Information Quarterly*, 30(4), 508-513.
- Larrick, S. (2015). Why should cities have an open data policy? Retrieved from <http://sunlightfoundation.com/blog/2015/10/01/why-should-cities-have-an-open-data-policy/>
- Local law 11 of 2012. (2012). Retrieved from <http://www1.nyc.gov/site/doitt/initiatives/open-data-law.page>

- Lourenco, R.P. (2015). An analysis of open government portals: A perspective of transparency for accountability. *Government Information Quarterly*, 32(3), 323-332.
- New York City launches "Open Data for All" initiative. (2015, July 16). Retrieved from <http://statescoop.com/new-york-city-launches-open-data-initiative/>
- Scassa, T. (2014). Privacy and open government. *Future Internet*, 6(2), 397-413.
- Veljkovic, N., Bogdanovic-Dinic, S., & Stoimenov, L. (2014). Benchmarking open government: An open data perspective. *Government Information Quarterly*, 31(2), 278-290.
- Zuiderwijk, A., & Janssen, M. (2014). Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly*, 31(1), 17-29.

Bibliographic information of this paper for citing:

Okamoto, Karen (2016). "What is being done with open government data? An exploratory analysis of public uses of New York City open data." *Webology*, 13(1), Article 142. Available at: <http://www.webology.org/2016/v13n1/a142.pdf>

Copyright © 2016, Karen Okamoto.