Visualizing San Francisco 311 Data

Jason Hayes, Joseph Ledesma, and Garey Le Monte

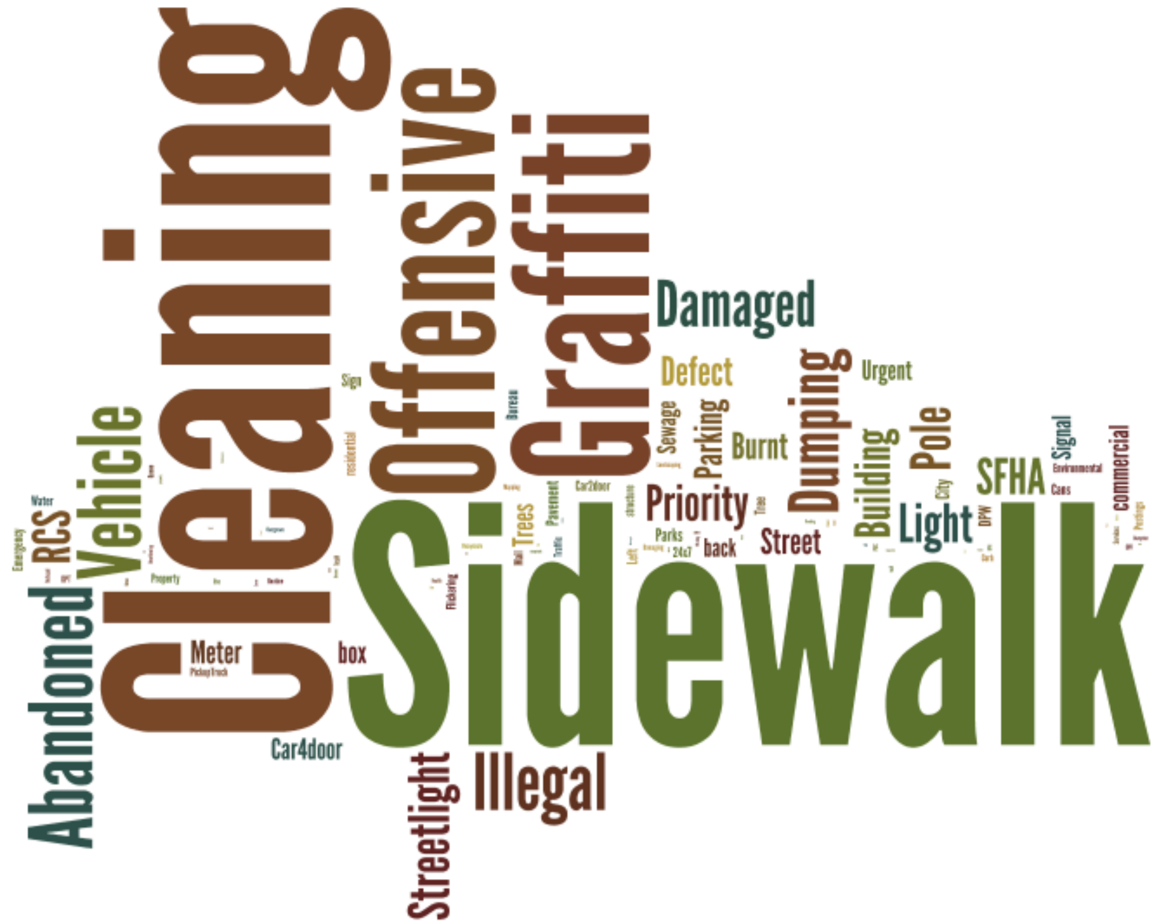

Fig. 1. Word cloud of Request Type of San Francisco 311 service requests.

Abstract— The goal of this paper is to show how visualizations can be used to see patterns in service requests to 311, a service available in some communities for interacting with local government. Utilizing open San Francisco Data, we used several information visualization tools to produce a word cloud, bubble chart, treemap, stacked bar graph, and line graphs.

Index Terms—311, service request, word cloud, bubble chart, treemap, San Francisco

1 INTRODUCTION

The issue presented is an attempt to grasp the scope of what types of service needs citizens contact their local government about, and the frequency and time frame of these events. As the reporting of incidents in a community are made more effortless with advances in technology, such as mobile apps, the data sets of these requests are ever increasing. Accessing the collective observations of community members has the potential to inform about the nature of the interactions between a local government and its citizens. This paper uses information visualization to identify patterns or trends.

1.1 Our Approach

The communications service known as 311 provides a central phone number as a point of contact for making non-emergency requests of local government services. The data related to these service requests have the potential to provide insight into the issues facing citizens and how the government responds to them.

Our approach is to visualize 311 service request data in a non-geographic manner that may lead to meaningful insights. We avoid geographic visualizations since we assume this is already the most common form of visualization and since the field of information visualization is concerned with visualizing abstract data [1]. Our challenge will be transforming the non-geographic data to intuitive and meaningful graphical representations. Therefore we used primarily two aspects of the data for service requests: time date data including opened and closed dates and category or type of service request information.

From a design perspective, the best information visualization is the one that can effectively convey the meaning to its users. From a user's point of view, it should be intuitive or easy enough to figure out the intended meaning. A main objective of any information visualization is for the user to gain insights, and in this paper we will search for hidden knowledge within the data.

1.2 The Data

The primary data used for this report was the public data on 311 service requests to the City and County of San Francisco provided by data.sfgov.org [2]. The City and County have made the data publicly available to allow individuals and organizations the ability to access and make use of the data. This data, Case Data from San Francisco 311, comprises the record headers of over half a million service requests since mid–2008 until early 2012 as of this writing, including service request opened and closed dates, category and type, and location information including street address, neighborhood, and geographic point. Since the records are only header data, no data on the person calling or other details of the request or its resolution are provided.

1.3 How Other Have Addressed the Same Issue

A visualization of complaints by time of day received in one week in New York City was published in 2010 [3]. The visualization, produced by Pitch Interactive, used a streamgraph, a type of stacked graph, or "complex layered graph" [4]. The visualization revealed, among other patterns: more noise complaints at night, taxi complaints in the middle of the night, and more parking complaints during lunch [5].

Our main data source, data.sfgov.org, runs its website using software provided by Socrata, Inc. [6]. The software allows the service request data to be browsed in tabular format and provides built-in visualization tools. However, most of the available visualizations are maps, calendars and conventional charts, and we found these were not as powerful as other tools.

2 Tools

Wordle: The word cloud was generated using Wordle (wordle.net). Wordle is a publicly accessible, web-based word cloud generator. A word cloud is a visualization of text data in which the words and phrases that occur at higher frequencies in the data set will receive higher significance in the

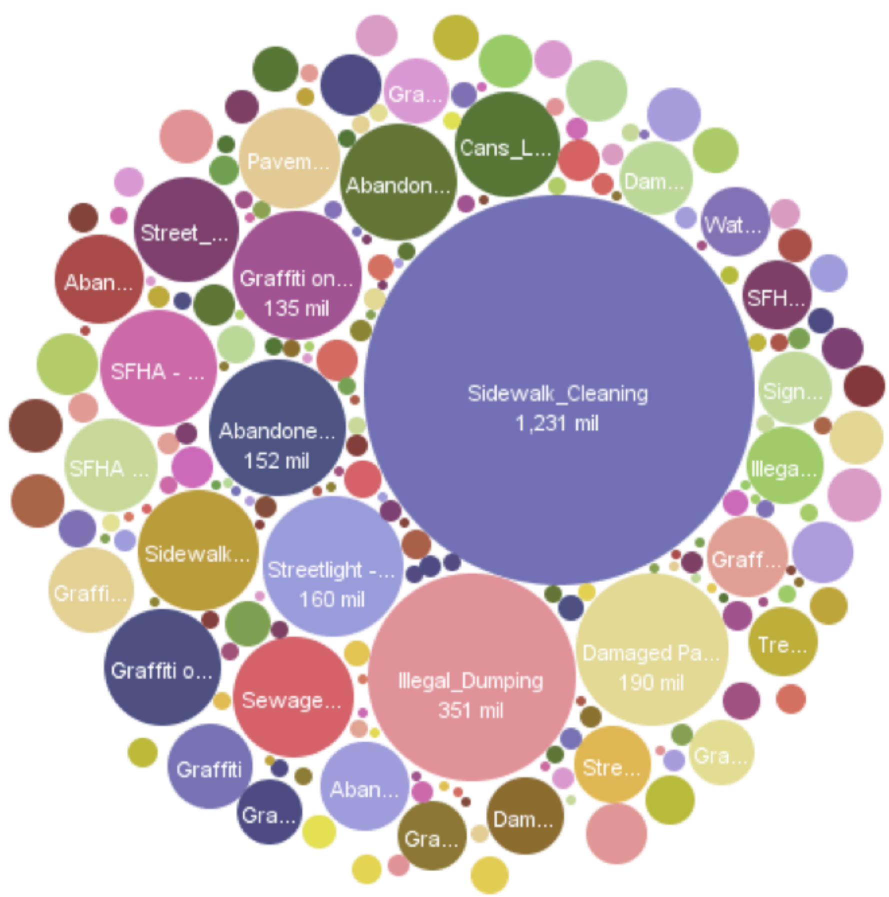

Fig. 2. Bubble chart of Request Type. The Many Eyes tool defaulted to using the Case ID field as a parameter for many of the visualizations. Here that resulted in the misleading number label below the Request Type label.

visualizations. Wordle allows users to enter text data and adjust the parameters of the resulting visualization.

Many Eyes: The tool Many Eyes was used to create the bubble chart and treemap for this project. The tools allows for user creation of visualizations. Users can either upload new data sets to Many Eyes or make use of existing data sets uploaded by other users. All data uploaded and all visualizations created are available for other users to access and make use of [7].

Tableau: Tableau was used to create graph visualizations of the data. Tableau is a local application that can analyze data and create a number of different visualizations. Tableau has a drag and drop interface that make creating and editing visualizations a straightforward task and allows the user to focus on the data and resulting information.

3 METHODS

Wordle: For the word cloud creation, only the text in the Request Details was used. Due to size restrictions of the application, a sample of this data from 1/1/2012 to 3/12/2012 was used. Upon entering the plain text into the text field, the word cloud generation was generated.

Wordle permits modifications to the word cloud based on a defined set of parameters. For this project, the font and layouts were changed.

Many Eyes: The bubble chart and treemap were generated in Many Eyes. For the bubble chart, the sampled data set was uploaded to the website. Request Type was used for the bubble field. The tool defaulted to also using the Case ID field as a parameter, resulting in additional label for the bubbles that could not be avoided. We represented each bubble as a different color to provide contrast.

For the treemap however, the data downloaded from San Francisco Data first had to be processed before uploading it to Many Eyes. We extracted the two fields of Category and Request Type and then performed an operation in SQL to count the duplicate rows of the same Category and Request Type combination, producing a rolled-up summary with the original two fields and the count. We then used this data set as the basis for the treemap. (The data set is available for inspection in Many Eyes and is linked to from the treemap, http://www-958.ibm.com/v/140716)

Tableau: Tableau was used to generate the bar graph and the line chart used in this project. The data set was imported into Tableau as a new workspace. The size limitation of Tableau meant that only the most recent 65,000 records were imported.

To create the bar chart, the category field was placed in the parameters for both the columns and rows, with the x axis representing the count of service requests in the categories. The status field was added to the color parameter to create the stacked bar graph.

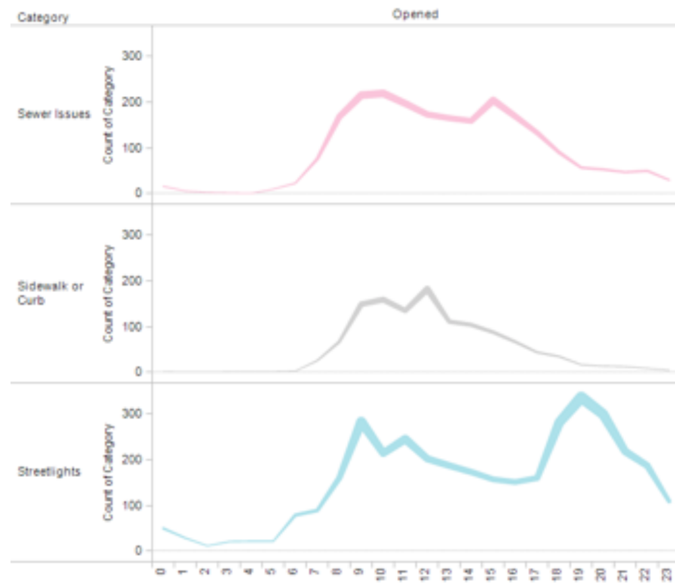

Fig. 3. When service requests in specific categories were opened by hour, for categories Sewer Issues, Sidewalk or Curb and Streetlights. Based on 65,000 service requests.

Within the same workbook, the line graphs were created by using the Hour part of the Opened field for the column and the Category and count of Category as nested rows. Count of Category was also used to determine the size (thickness) of the lines.

4 RESULTS

Based on the 311 data set previously introduced, five information visualizations were created. The word cloud visualization, Figure 1, presented at the beginning of this article, is an amalgamation of the most common words found in our sample data set.

Next we created a bubble chart in Many Eyes, Figure 2. This visualization proved to be unsatisfactory due to the tool insisting on using the Case ID field as a label for the bubbles.

We then created a set of line graphs, Figure 3, that visualize the daily flow of requests within a selection of three categories: sewer issues, sidewalk or curb, and streetlights.

The fourth visualization is a treemap hierarchy, Figure 4, organized by 311 service request case category and request type. It displays shades of the color orange to show variations in frequencies among the request types. The interactive version allows for zooming in and out, text searching, and adjusting the upper and lower number range for the coloring.

Finally, the fifth visualization, Figure 5, is a stacked bar chart that depicts the total requests for service by category. Each category bar is segmented to show the status, open or closed, of requests. This reveals the ratio of open to closed requests for each category.

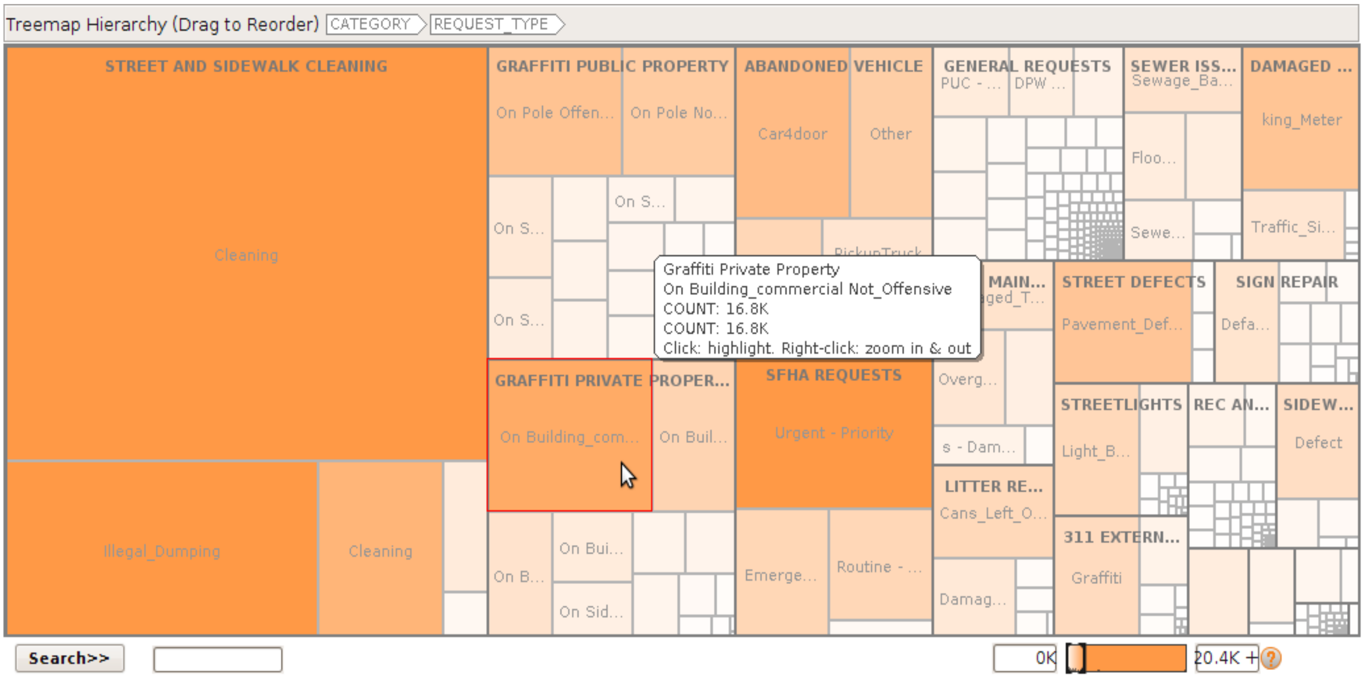

Fig. 4. Treemap using hierarchy of Category and Request Type. Screenshot of interactive applet using a pop-up to show the details of a selected rectangle for non-offensive graffiti on private, commercial buildings.

5 INTERPRETATION

The word cloud visualization for the 311 data set provides an interesting mass of words that when read, conjure an image of things that either occur in or are related to a street. Words like *cleaning, sidewalk, offensive*, and *graffiti* are the most pronounced. Other words catch your attention, such as *abandoned, vehicle, illegal, dumping* and *damage*. These words convey to the viewer things that may be a nuisance to residents, or things that are illegal, or those that need repair.

A theoretical issue concerning information visualization is differentiating between the art of inspiring users for new connections and the scientific side of information visualization, which requires the information in visualizations to be portrayed accurately and faithfully. Connections between the scientific and artistic components of visualizations are described as functional and aesthetic [8]. Functional information visualization is meant to convey a message to the user. Our word cloud is an aesthetic information visualization, which presents a subjective impression of the 311 data set by eliciting a visceral or emotive response from the viewer, which cannot be quantitatively measured.

The data set we used for this project also contained a Case ID field for record identification. Some of the less flexible tools used this field as a default parameter for the visualizations. In the case of the bubble chart created in Many Eyes, this caused the field to be displayed as a label in the bubbles. The result is an irrelevant number which is distracting to the visualization. Given the large number of records in the data set, representing individual records is not what we are after. Through the use of visualization we intend to group records to reveal patterns that may not be evident by simply looking at individual records.

The treemap displays a hierarchical organization of all the 311 data categories, in easily defined rectangles, showing labels for some of the most prevalent service request types. It is evident that the category of *Street and Sidewalk Cleaning* by far has the most service requests, and at a glance one can see for example that among graffiti incidents, most graffiti that is on public property is on poles while most graffiti on private property is on buildings. Next, one can see how category *General Requests* is a sort of miscellaneous category with many different request types since there is a fine grid of small to minuscule rectangles, too small to have their labels appear (unless one zooms in using the interactive version).

An interesting discovery that came from creating the visualizations related to the rates of open and closed requests per category: *Street and Sidewalk Cleaning* requests are by far the largest category, and they are also handled, or closed, at a high rate. While the total number of Housing Authority requests are consistent with a number of other categories, a very low percentage of them are closed. This could speak to the city's priorities in terms of areas to address. It also could provide insight to the resources appropriated to issues relative to the scale of the problem in that area.

We also found that we could not use the entire set of categories as the input to the line graphs. The initial version included all 21 categories, however, that hampered the ability to view the desired level of detail, rendering it unintelligible. It was decided to utilize selected categories so as to increase the likelihood of finding something interesting amongst the visualizations [9].

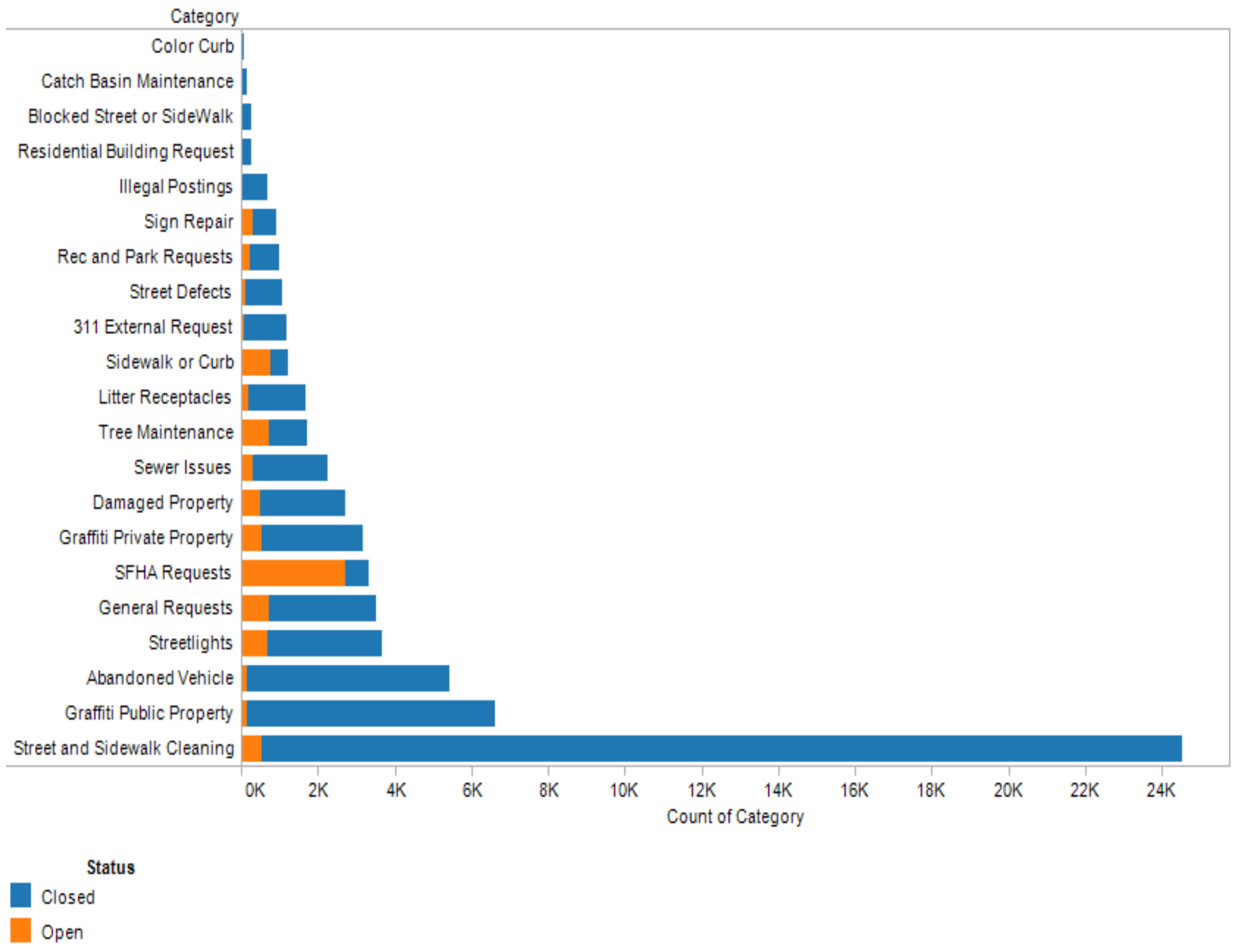

Fig. 5. Stacked bar graph showing count of Category for each Category. Color represents the Status of the requests.

6 Discussion

The overall goal was to identity and analyze the scope of service requests for which San Franciscans contact their local government and the timing of these requests. This goal was accomplished by identifying patterns and trends with the aid of information visualizations.

With that said, one significant problem was the overall size of the data set. The roughly 540,000 (and counting) records in the data set proved to be too much to handle for the visualization tools we used. This necessitated sampling or summarizing the data to a reasonable size to create the visualization. Each of the team members did this following their own methods, resulting in slightly different samples used for the various visualizations. Standardizing this step prior to creating the visualizations would be an improvement to the process.

To assist in the gaining of insight from all of our visualizations (Figures 1-5) depicting either service request categories or types with rates and time of day, Shneiderman's taxonomy [10] provided a basic framework for the steps needed to properly interact with the graphically presented information visualizations. Following Shneiderman's methodology, the entire data set was reviewed to gain some initial insight as to the entire collection. After the overview, items of interest were zoomed in on, which included service categories, request type, status (opened/closed), and time of day. Items of little to no interest were filtered out to allow for a quick focus on relevant interests by eliminating unwanted items. As exemplified by the treemap visualization, a details-on-demand function was employed to allow the selection of an item or group of items to retrieve details when needed. One component of the methodology not used was the keeping of a history of our actions.

6.1 Future Work

While visualization of service request data can reveal patterns and changes, it does not reveal much about how these service requests are handled or about the operations of local government. If data were available on how a service request was processed, visualizations could reveal bottlenecks in the processes. However, if this data continues to only be available internally, applications of information visualization could be developed for use within government.

7 Conclusion

In this paper, we attempted to grasp the scope of service needs that citizens contact their local government about using information visualizations. We looked at category types, time frames, and closure status. These visualizations contributed to the identification of patterns through new and interesting ways to display 311 data, which provided insights.

REFERENCES

[1] C. Chen. *Information visualization*. Wiley Interdisciplinary Review: Computational Statistics, vol. 2, no 4, pp. 387-403, July/August 2010.

[2] San Francisco Data. Case Data from San Francisco 311. Mar. 13, 2012. Retrieved from http://data.sfgov.org/Service-Requests-311-/Case-Data-from-San-Francisco-311/vw6y-z8j6

[3] S. Johnson. What a Hundred Million Calls to 311 Reveal About New York. *Wired*. Nov. 2010. Retrieved from http://www.wired.com/magazine/2010/11/ff_311_new_york/all/1

[4] L. Byron & M. Wattenberg. Stacked Graphs – Geometry & Aesthetics. Feb. 2008. Retrieved from http://www.leebyron.com/else/streamgraph/

[5] R. Hand. What a Week of Calls to 311 Reveals About New York. *Vizworld*. Nov. 5, 2010. Retrieved from http://www.vizworld.com/2010/11/week-calls-311-reveals-york/

[6] Socrata Open Data API 1.0 Beta, used in data.sfgov.org. Retrieved from http://data.sfgov.org/api/docs/

[7] F. Viégas, M. Wattenberg, F. Van Ham., J. Kriss, M. McKeon. Many Eyes: A Site for Visualization at Internet Scale. 2007. Retrieved from http://www.research.ibm.com/visual/papers/viegasinfovis07.pdf

[8] A.R. Gaviria. When is Information Visualization Art? Determining the Critical Criteria. *Leonardo*, 41:479–482, 2008.

[9] M. Davis. That's Interesting! Towards a Phenomenology of Sociology and a Sociology of Phenomenology. *Phil. Soc. Sci* 1:309-344, 1971.

[10] B. Shneiderman. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualization. In *Proceedings of IEEE Workshop on Visual Language*, Boulder, CO, pp. 336–343, 1996.