NEIGHBOR VIS beta

Query Selection

Selected Data

sandybefore sandyduring

From 10/25/2012

To 11/03/2012

Keywords sandy

Category Input

Query Manager

Select_ID	Data_Name	Records	
query1	2 Data	....	🗑
data 1	sandybefore	75	🗑
data 2	sandyduring	224	
query2	2 Data	....	🗑
query3	2 Data	....	🗑

✓ sandyduring P H

Date:2012-10-30 00:49:00
Category:Weather and Environment
Text: Picture of #Sandy: Brooklyn Bridge Pa
much submerged we were here 6 hours ag

Data Manager ▾

Query Analysis

Select ID: data 2

Query Filter: 2012-10-25 2012-11-03 sandy

Data Filter: 2012/10/30 Brooklyn

Event Records

ID	Date	Category
19603	2012-10-30 04:58:00	Infrastructure/Util
19703	2012-10-30 05:23:00	Infrastructure/Util
17108	2012-10-30 01:11:00	Infrastructure/Util
15927	2012-10-30 00:11:00	Infrastructure/Util
16330	2012-10-30 00:33:00	Weather and Environment
21170	2012-10-30 14:54:00	Infrastructure/Util
18410	2012-10-30	Infrastructure/Util

> Time ✖2012/10/30

> Keywords ✖Brooklyn

> Category

NEIGHBOR VIS

A USEFUL TOOL TO STUDY GEOSPATIAL SEMANTIC EVENT DATA

Table of Contents

I. Introduction

1.1 Aims of NeighborVis

NeighborVis is a visual analytics system of **geospatial-semantic event data** in localized neighborhoods within a larger city, town, suburb or rural area. The geospatial-semantic event data (we name it **GSE data** in this document), consists of geospatial (longitude/latitude or locational address), time, and semantic dimensions, which forms a large part of the important information people deal with for business, personal and administrative use. Here are some GSE data examples:

- Social media data of geo-tagged twitter, facebook, Instagram posts;
- Real estate data with house price, buyer information, loan and other related documents;
- Police data of crime reports with crime locations, involved personnel, and case details;
- Public health data of epidemic disease cases with locations and descriptions;

NeighborVis is developed to allow users to study these GSE datasets in a community neighborhood. It is developed to answer the simple questions about a community, such as: what is happening at this location, is it good or bad, and can outcomes be improved? It may be a police department considering a blighted street, a planning department considering urban renewal and the "greening" of spaces, a community group wanting to add neighborhood gardens, or a hospital mining its child injury data. A large variety of users, such as real estate analyzers, public health and safety administrators, business planners, and social scientists, are enabled to conduct exploratory study through an integrated interface of maps, charts, semantic, and information visualization tools, while a set of fluent user interactions are incorporated around the way user think, feel and behave. The visual analysis is easily conducted over fine-grained geographical resolutions such as small neighborhoods around local restaurants, churches, or schools with the provisioning of temporal, semantic and geographical filters and constraints. NeighborVis can greatly help users from multiple sectors including business, finance, public services, social research, and local governments by leveraging the advanced computation and visualization capabilities to utilize their large, complex, and rapidly accumulating community data.

1.2 NeighborVis Features

NeighborVis is developed as a Web-based system so that users can use it through web access to remote server platforms. It can also be used solely as a localized version on desktops and laptops. As illustrated by Figure 1, it supports a full data processing pipeline with the functions of:

- *Data Loading and Processing: allow users to upload raw data including GSE data and geographical shape files. The loaded data NeighborVis supports users to upload multiple datasets.*

- *Database Storage: The loaded GSE datasets are transformed and stored into the NeighborVis database. The corresponding map with geocontext is also loaded. The database we currently used is designed based on MongoDB and it supports both local and server based operations.*

- *Data Management: allow users to interactively define active GSE data resources for visual analysis. Users can select more than one datasets at the same time which is important for comparable and integrative study of multiple data sources.*

- *Visual analysis: allow users to conduct interactive data visualization and analysis with a set of charts, filters, maps, and other visual functions through human-machine interactions.*
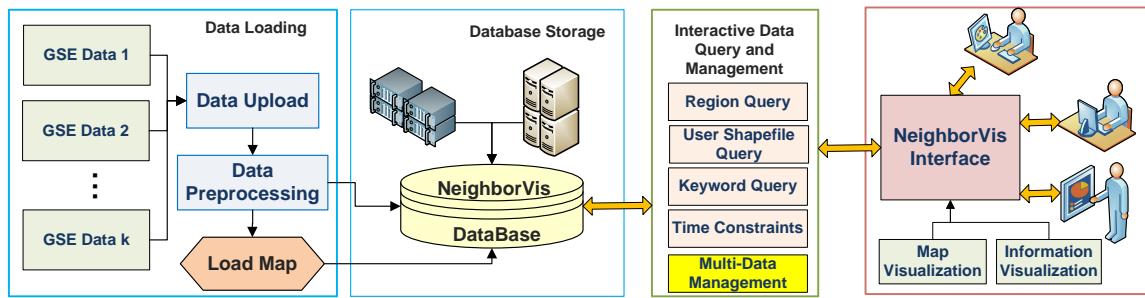
Figure 1: NeighborVis system.

II. GSE Datasets

GSE data consist of a set of geospatial semantic events. Each event at least contains three important attributes, location, time, and semantic description, while many of them have a categorical attribute (about classes, types, or kinds) as well. Then the data may have more fields related to specific fields but cannot be predefined in NeighborVis. In data loading function, we check and validate that each raw dataset has the following attributes:

(1) Event ID, a unique identification number for a GSE record.

(2) Date/Time, which indicates when this event happened. This attribute should have the format as YYYY-MM-DD HH:MM:SS.

(3) Geolocation of longitude and latitude, which indicates where an event happened.

(4) Text, which is the semantic description of this event.

(5) Category (optional), which is the category this event belongs to.

III. NeighborVis Overview

## 3.1 Overview of NeighborVis Data Operations
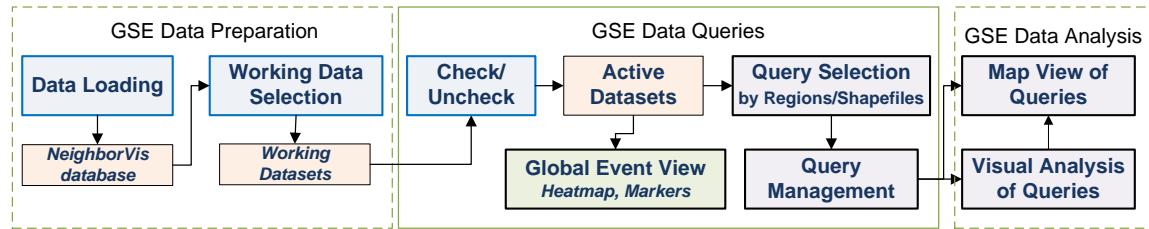
Figure 2: NeighborVis Data and Operation Flow.

Figure 2 shows the data and operation flow of NeighborVis system. The use of NeighborVis consists of three stages:

1. ***GSE Data Preparation***

   Raw data is first loaded into NeighborVis database. Then users can select a set of working datasets for their current task.

   Once the data preparation is done, the working datasets can be interactively checked to make as active datasets. Users can view events of active datasets on the map globally as heatmap or markers.

2. ***GSE Data Queries***

   A **query** is the fundamental operating units of NeighborVis analysis. One query refers to a set of events (may coming from multiple active datasets) inside a geospatial region and a given time period. Users can define a **query** by drawing regions on the map by mouse, or loading predefined regions from GIS shapefiles.

3. ***GSE Data Analysis***

   Once a group of queries are defined, users can manage them to visualize one or more queries on the map. Meanwhile, details of queries are shown to study their contents. A set of visual representations and tools are implemented to support interactive analysis of queries.

3.2 Overview of NeighborVis Interface

Figure 3 shows the overview of NeighborVis interface. It includes a set of views to implement the functions of GSE data loading, query, and visual analysis:

A. **Data Loading**
B. **Datasets Management Panel**
C. **Query Selection Panel**
D. **Query Management Panel**
E. **Map view**

F. Query Analysis view
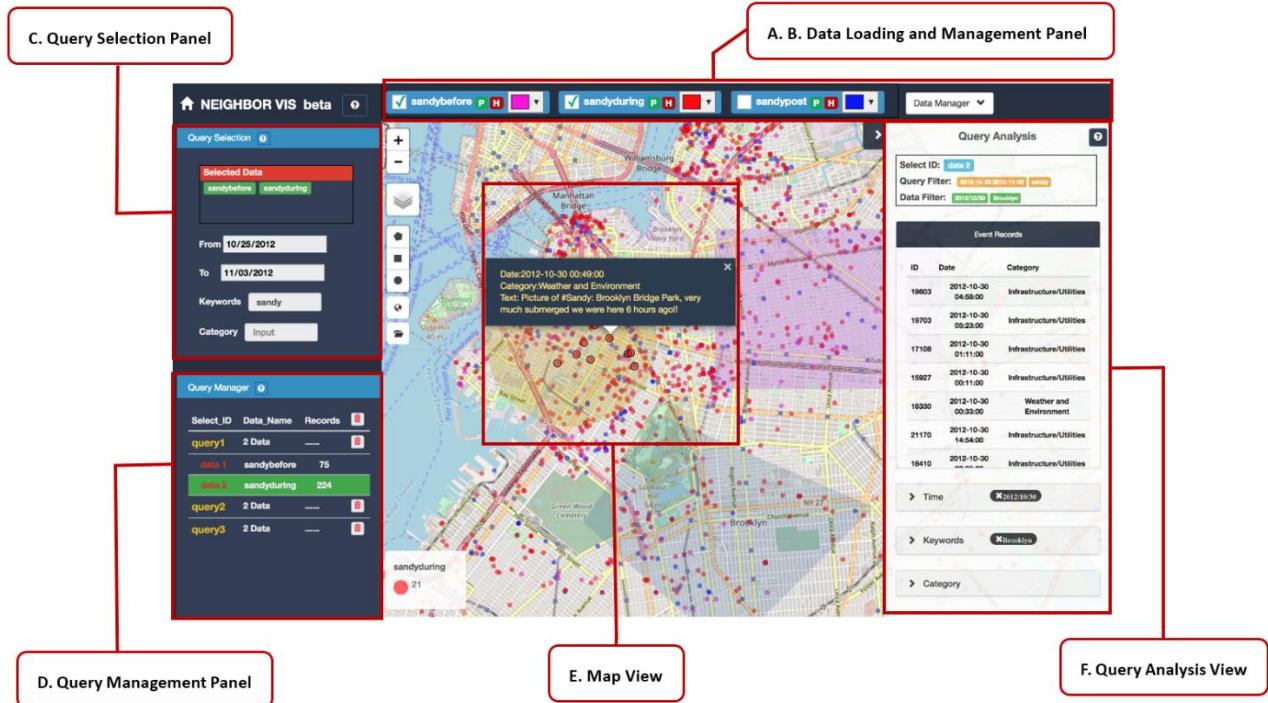
Figure 3: NeighborVis System Interface.

Next, we describe these functions with a full user guide of operations.

IV. NeighborVis Views and Functions

A. Data Loading

Users load raw data to NeighborVis database. NeighborVis supports users to upload multiple datasets in CSV format. A dataset is checked to make sure it includes the necessary data attributes (as explained in Chapter II). Users are required to specify the correct attributes in the CSV file. Then, the system automatically generates geoindex and stores the dataset in database. NeighborVis database is specifically designed and implemented over the MongoDB database.

Click Data Manager will start new data uploading process.

1. **Start data uploading**
   Click Data Manager, the popup window shows all the available datasets in NeighborVis database. Click Data Upload button to start uploading new data.

2. **Name your data**

   You are prompted to give a name of the new dataset. The name should be unique, which is different from existing datasets.

3. **Select your data file in CSV format. The file will be uploaded.**

Data Uploading

⬇

Drag and Drop File Here

Or Select File Here

Uploading currently support: .csv and .geojson files

Data Uploading

Reading file successful

Now we have some required attributes we need you to specify

File name: **NYCsandy_post.csv**

File size: **93.47 KB**

Last Modified: **Tue Jul 10 2018 23:06:20 GMT-0400 (EDT)**

Please click next to continue

Next

---
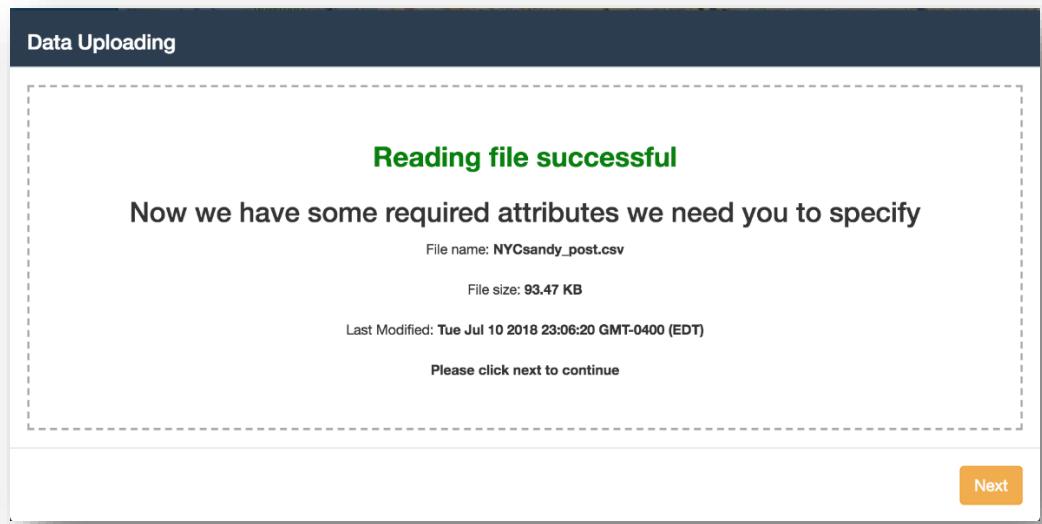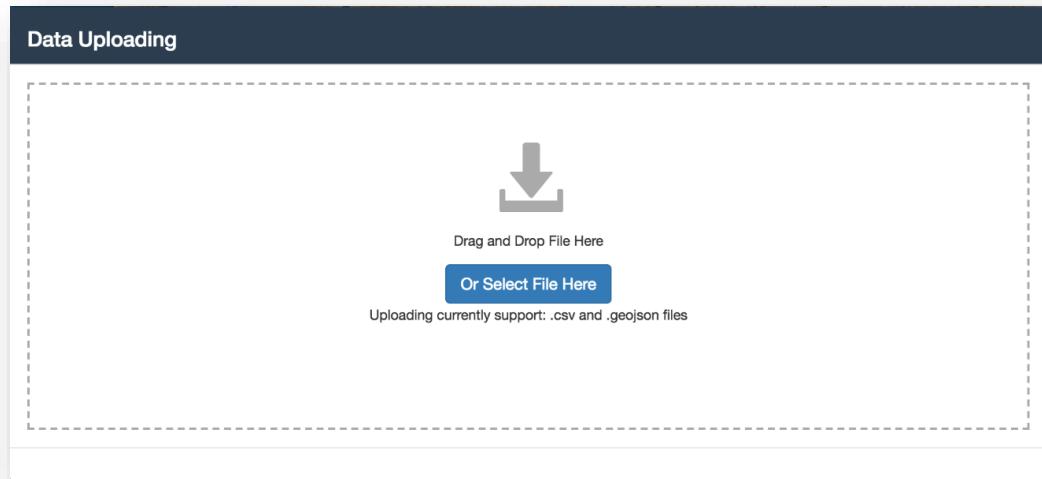
4. **Specify required data attributes (column names) in the uploaded CSV file, including ID, longitude, latitude, time/data, text description, and categories(classes).**
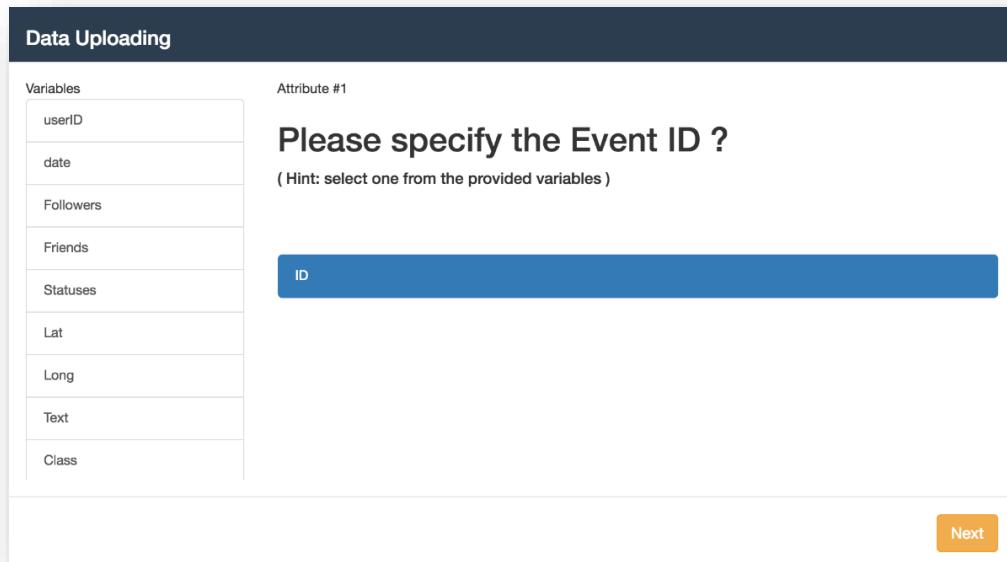
5. **Convert the CSV data into NeighborVis GSE tables. The system automatically completes geoindexing**
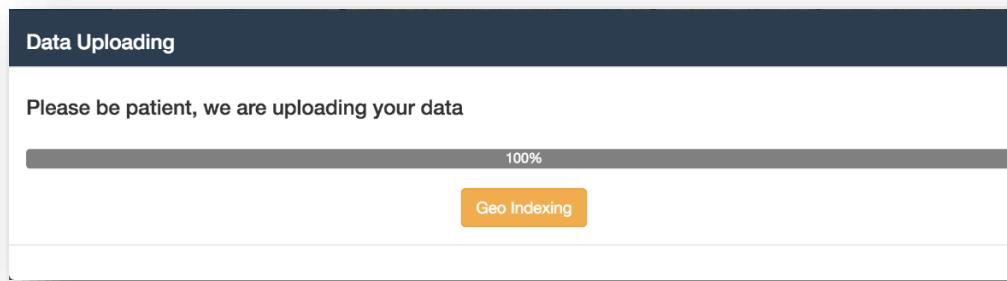
6. **Finish data uploading process**
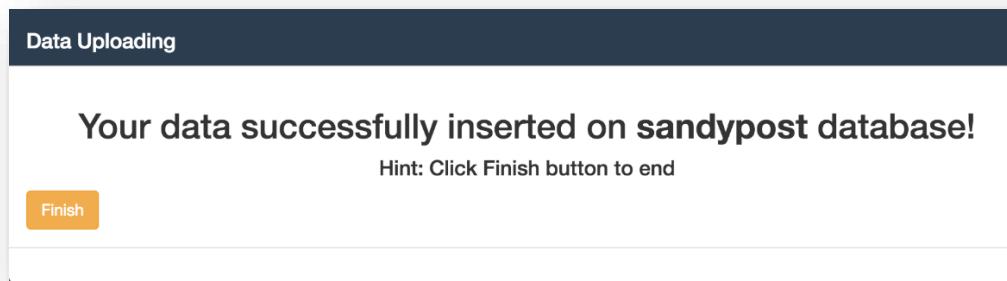
7. **After uploading, the newly added dataset automatically appear on the Datasets Management Panel.**

B. Datasets Management Panel

1. Show and hide datasets from the panel

Datasets management panel shows a set of data objects, which are NeighborVis datasets users are currently working on. By clicking Data Manager button, in the popup window of Figure 4, users can select which datasets (already uploaded to NeighborVis) are shown in this panel, which are **working datasets**. Other datasets are hidden from this panel, if it is not being used in current analysis tasks.
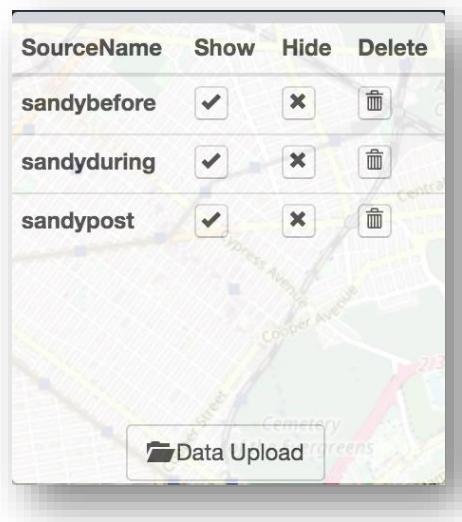
Figure 4: Data manager interface

2. Delete datasets

Users can click delete icon, which **fully delete** the corresponding dataset from NeighborVis database and system. You will need to upload a dataset again once fully deleted.
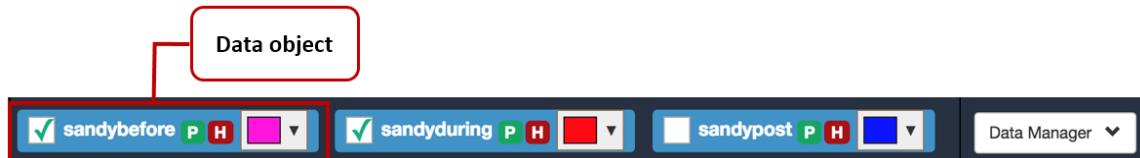
Figure 5: Datasets Management Panel.

3. Check/uncheck active datasets

On the panel shown in Figure 5, user can check or uncheck a working dataset to make it an **active dataset**. The query selection (see next section) is effective on all active datasets.

4. Show events of active datasets

Users can operate on the panel to globally show all events on the map view of specified datasets.

P Click to show/hide all events from the corresponding dataset as markers on the map

Click to select color of the markers shown on the map

H Click to show/hide all events from the corresponding dataset as heatmap on the map

C. Query Selection Panel

Figure 6 shows the Query Selection Panel interface. It allows users to set up multiple conditions of query generation.

First, the selected data list shows the names of active datasets, which user checked on the Datasets Management Panel.

Then, users set up filters of
(1) Time period
(2) Keywords in text description field: The keywords can be combined by Boolean operators: AND OR, match such as "sandy" and "floor", "sandy" or "floor".
    a. Fuzzy matching is also supported such as "sand" -> "sandy"; "floo" ->"floor".
(3) Category name: The input must be one of the names in the origin datasets.
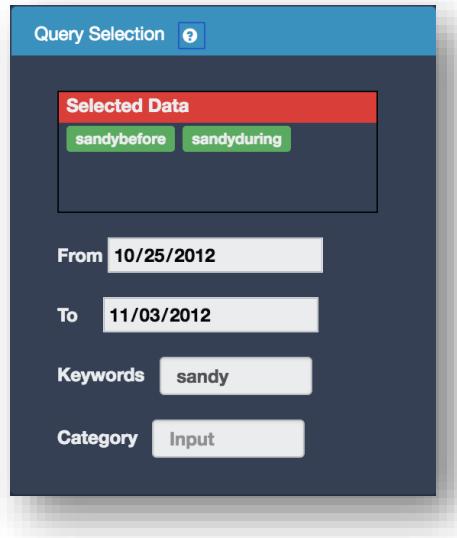
Figure 6: Query Selection Panel

After these query filters are set, users can define a region of interest by drawing on the map. Click the icons on the map will allow users to select regions (See IV.E for details), the system will then generate a query. The query will include GSE data based on the region and the filtering conditions specified here.

D. Query Management Panel

A query includes data of

(1) Spatial region;
(2) Time period;
(3) Keywords and category based sementic filters;
(4) GSE data inside the region, within the time period, and satisfy the semantic filters.

Figure 7: Query Management Panel

All the generated queries are managed in the Query Management Panel, as illustrated in Figure 7. The list shows the queries, how many active datasets are involved, and how many data records (events) are included.

Users can delete a query on the panel by clicking the garbage icon.

Users can click any query to make it active, so that the map view and visual analysis view show the corresponding query data for analysis.

A query may be applied over multiple datasets. Users can also select one specific dataset inside the active query to highlight it on the map view. In default, all datasets are assumed to be used in visualization.

E. Map view

Map view allows users to select regions for queries, display query regions, display GSE data on the map. Figure 8 shows one snapshot of the map view, where one active query (the rectangular region) is selected. The internal markers show the GSE events data of this query. They come from three different datasets (i.e., the checked data objects in the datasets

management panel). Users can click to select one marker to display its details. The legend on the lower left corner displays the colors of markers representing these datasets, and the number shows the number of markers (events).

On the map view ,multiple queries can be displayed simultaneously, while one active query is highlighted which is user selected on the query management panel.
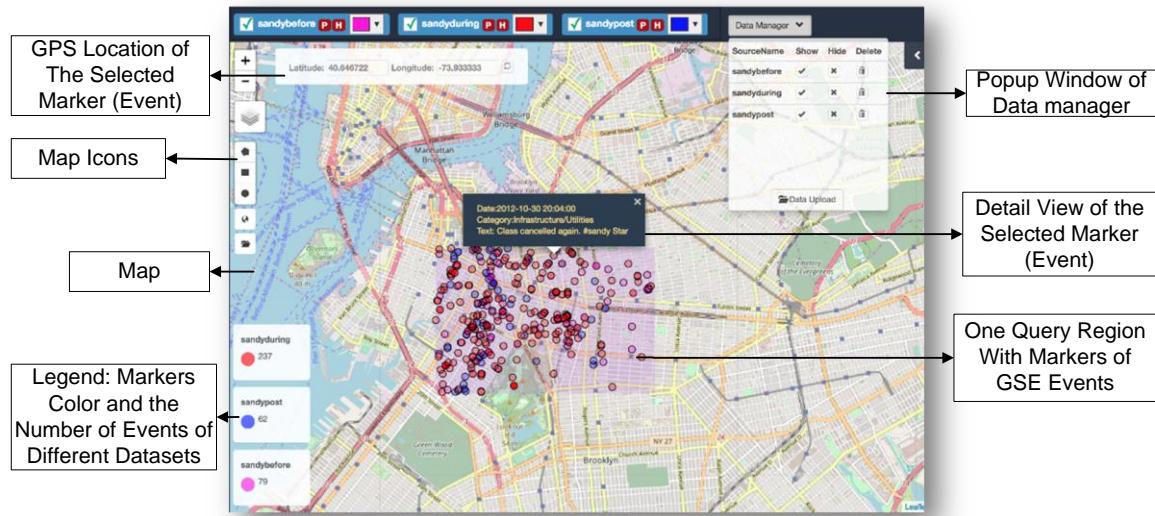

Figure 8: Map view overview

1. **Map Icons and Functions**
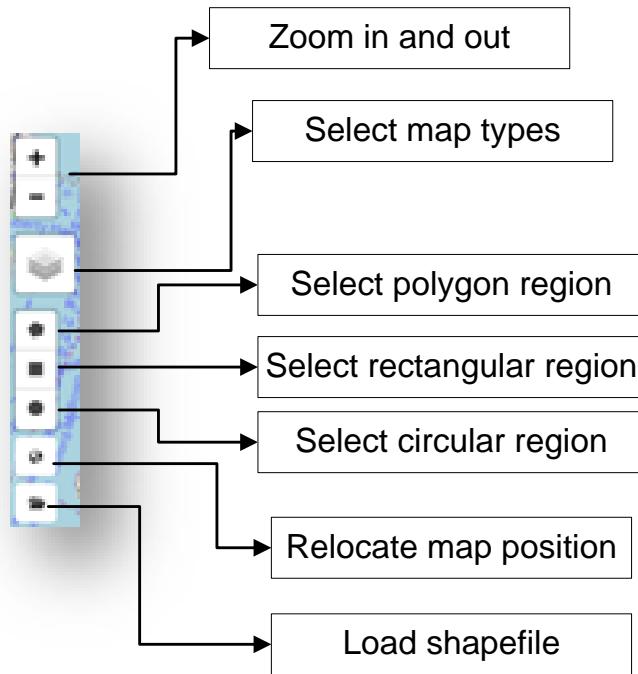
Figure 9: Map icons and their functions

Figure 8 shows the icons on the map view which also indicates the functions each icon represents including

(1) Query region selection types: allow users to select regions of interest to define a new query;
(2) Load pre-defined shapefiles to to define one or multiple queries;
(3) Map operations of zooming, map styles, and relocate the map position (if necessary when you change a different GSE dataset).

2. Map View Types

    (1) Show each GSE data event as point markers
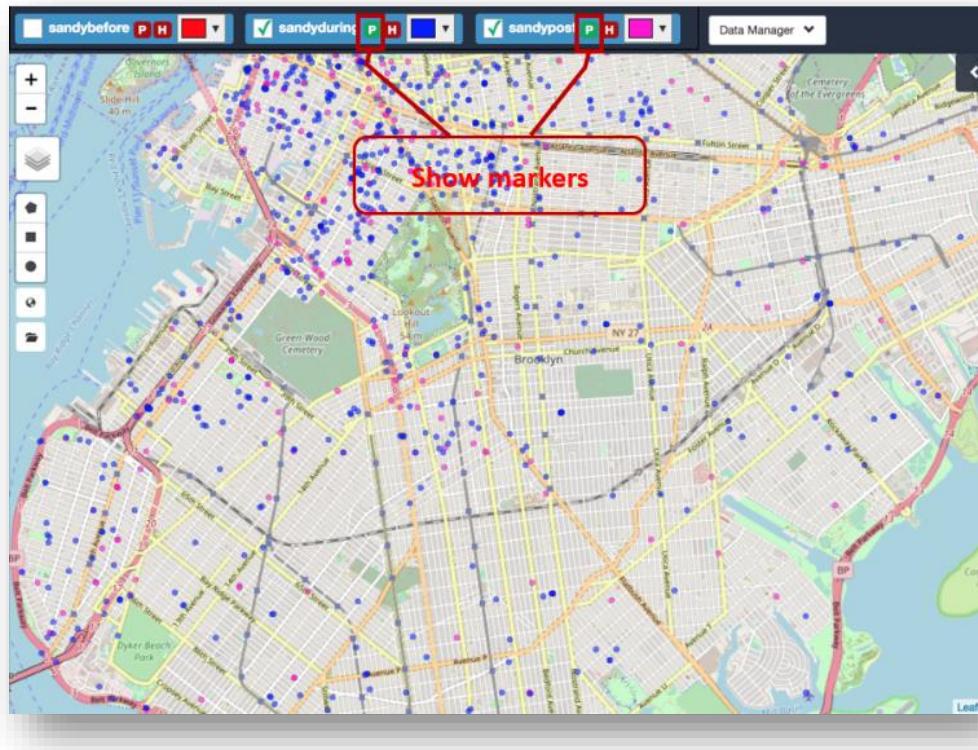
Users can click button ![P] for each dataset on Datasets Management Panel to globally show the events on the map.

    (2) Show GSE data distribution as heatmap
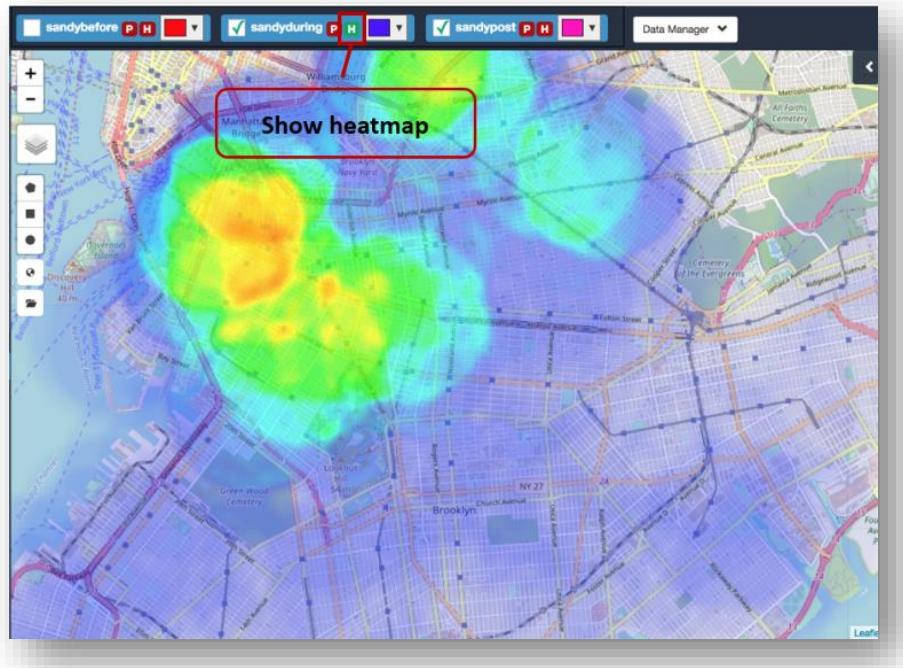
Users can click button [H] for each dataset on Datasets Management Panel to globally show the event distribution on the map.

3. **Show GSE Event Details**

(1) Users select one marker to show the event details in a popup window.

Date:2012-10-30 20:04:00
Category:Infrastructure/Utilities
Text: Class cancelled again. #sandy Star
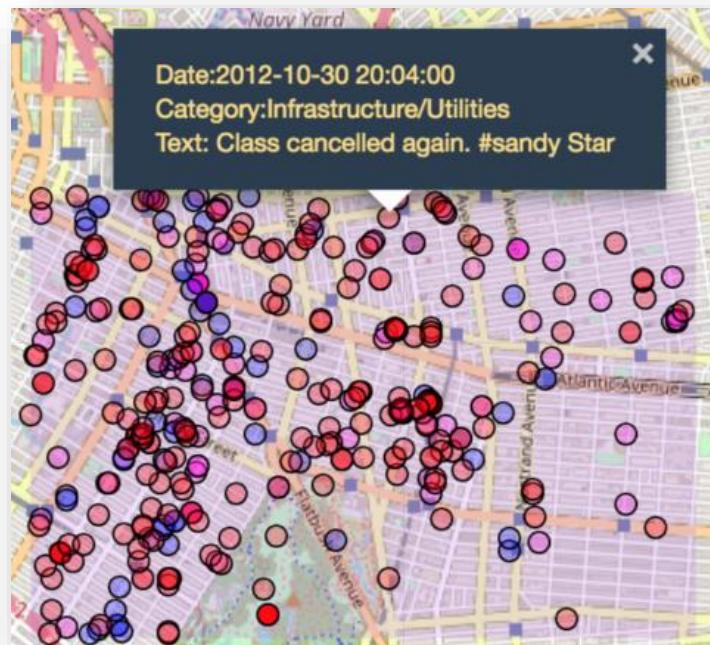
(2) Many GSE events happen at the same location. When users click the location, a time line view is popped up to show these events happening along a time axis. Users can interactively zoom in on the time line to see the time varied events. Users can click one event to see its details as well.

F. Query Analysis View

Query Analysis View is a panel on the right side of the map, which is initially hidden. Users can open it to visualize the details of the current active query. If users select one dataset of this query, only events related to this selected dataset is applied (see IV.D for details). Otherwise, all events in this query is visualized.

There are four sections in the Query Analysis View including

(1) Event records list.
(2) Time bar chart.
(3) Keyword cloud.
(4) Category bar chart.

All these components are interactively operated by users to filter by time, keywords and categories. The visualizations over these components, as well as on the map, are coordinated and synchronized with users' operations.

1. **Event Records List**

   It shows all events in the active query and the active dataset. Users can hover over each item to show it on the map view.

2. **Time Bar Chart**

   It shows the time distribution of all events in the given time period. Users can click to choose one particular bar so that only events inside this time range are used in map view and in the other views of this analysis panel. Users can choose to show the time bar charts in yearly, monthly or daily manner, according to the nature of the GSE events.
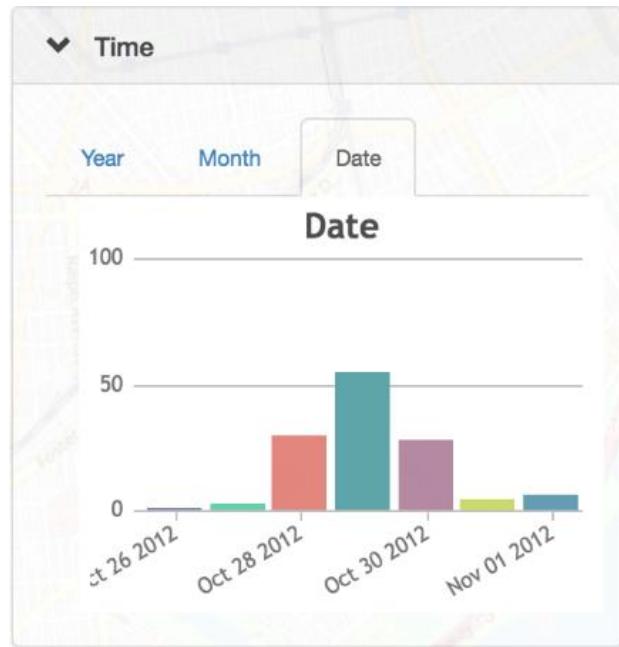
3. **Keyword Cloud**

It displays the keywords in the descriptions of all events of the query. The font of each keyword is used to visualize the importance of this keyword, currently measured by the term frequency. Users can click on a keyword so that the events on the map and on this panel will be filtered based on this selection.
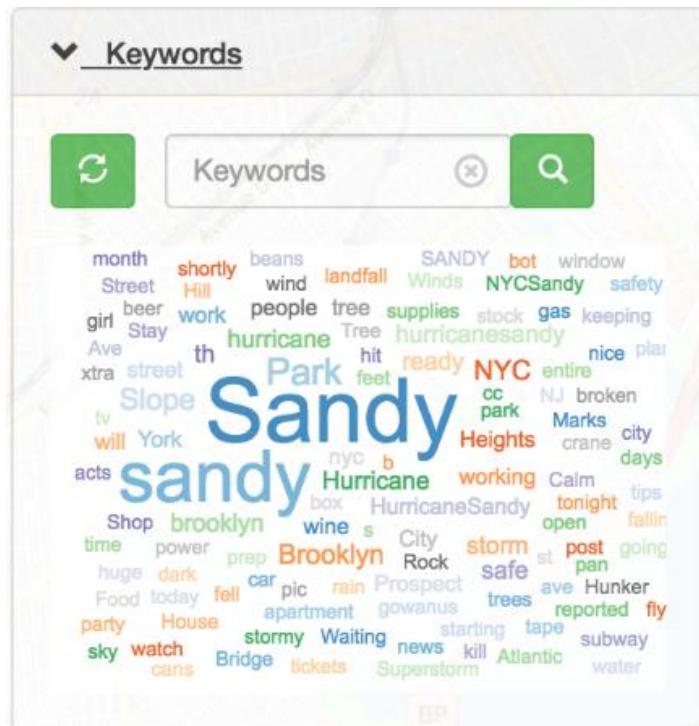
4. Category Bar Chart

It shows the histogram of events on the event categories. Users can click on a bar so that the events on the map and on this panel will be filtered based on this selection.
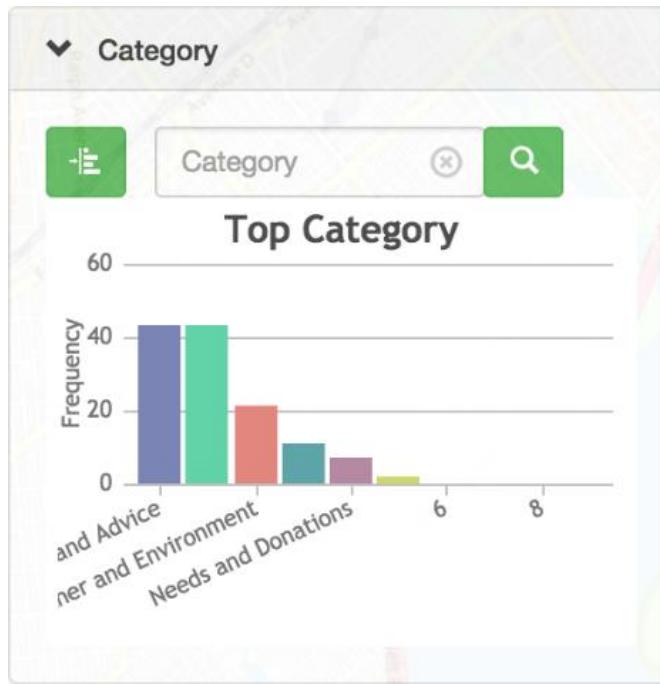
Figure 8: (3) Category bar chart shows top categories.

5. Interaction

All these charts and keyword cloud are interactively operated by users to filter this query by different conditions. These conditions can be combined. For each component, the condition filter is shown on the top bar. Users can click on the check to remove this condition.

Meanwhile, all these conditions are shown in the top of the Query Analysis View as an information box. Map view and this Query Analysis View are updated automatically in real time based on the changes of these conditions.

V. Usage Example

We show one example of using NeighborVis to analyze a real-world data. This is a geo-tagged social media data which is collected and processed from Twitter API. The data is extracted for the area in Brooklyn, NY, during the time period of Hurricane Sandy 2012, which created particularly severe damage in New Jersey and New York of USA. All the events are divided into three GSE datasets. Each is a part of all the tweet events related to Sandy and has geo-tags (locations). Three of them refer to the time periods before, during, and post Hurricane Sandy, respectively. Next we show how NeighborVis is used in visual analysis.

1. **Upload a dataset and name it "sandypost". Other two datasets are uploaded in the same way (SandyBefore and SandyDuring).**
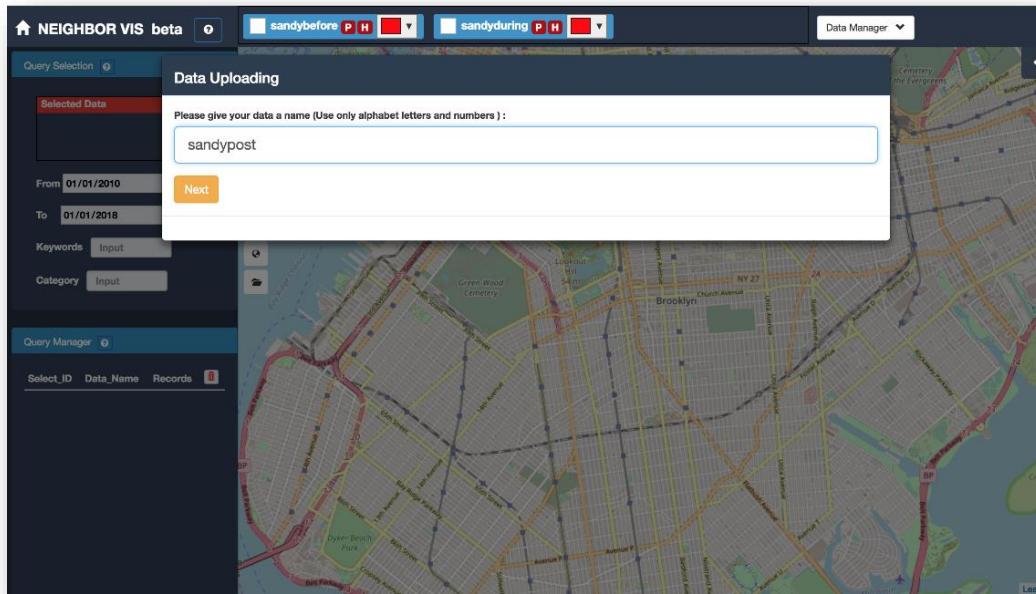
2. **All three datasets are stored in the database and selected as working datasets (SandyBefore, SandyPost and SandyDuring). They are all shown in the Datasets Management Panel.**
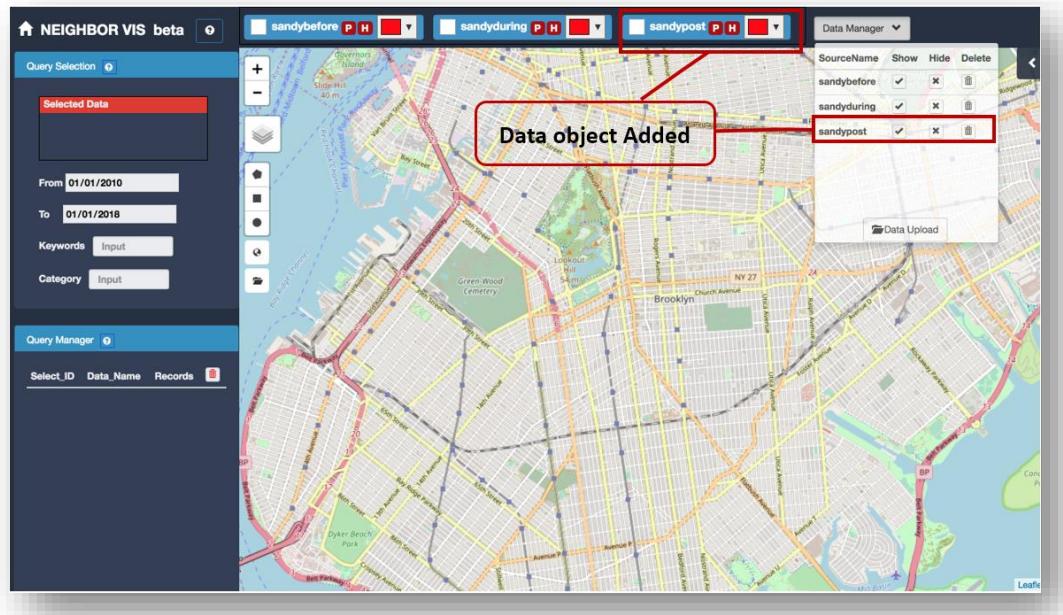

3. **Relocate the map to the corresponding area (Brooklyn). Set color for each dataset.**
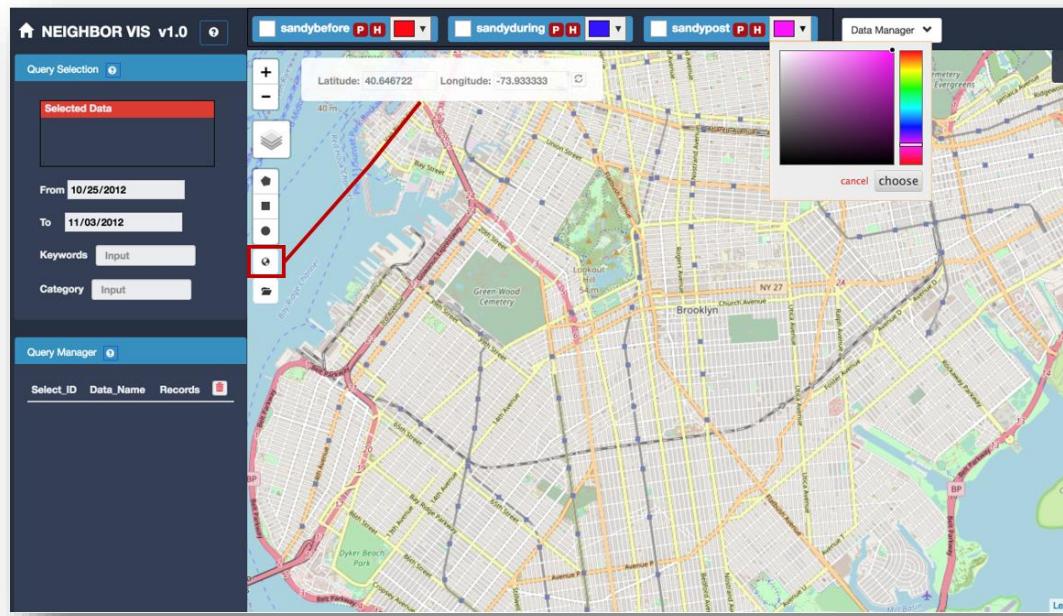
4. **Check to see heatmaps (or markers) of active datasets on the map. Two datasets (SandyPost and SandyDuring) are activated for the following query.**
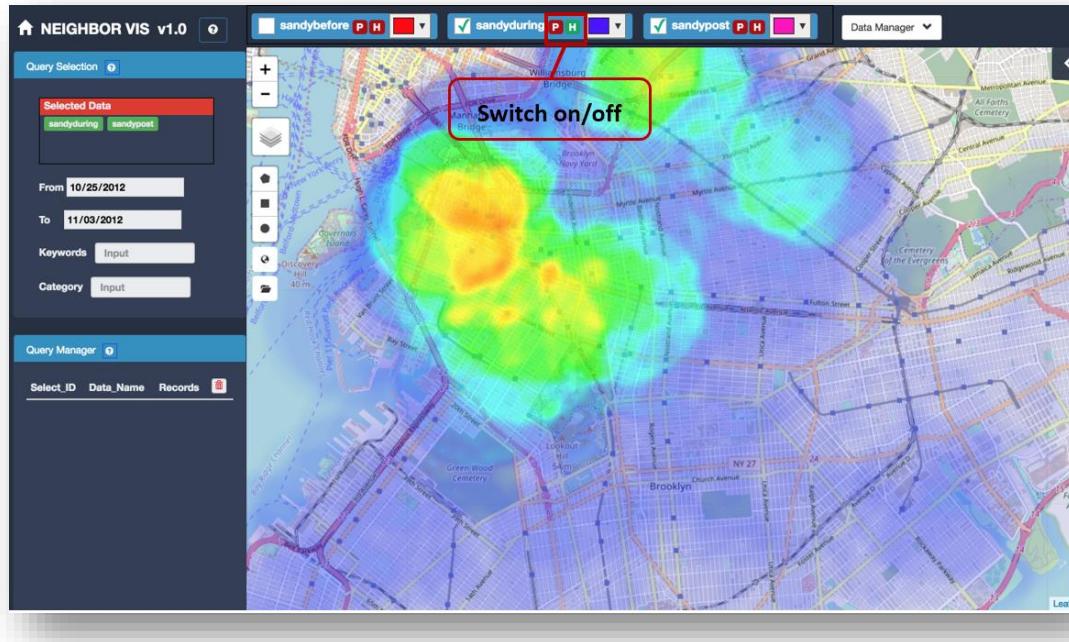
5. Set up query selection filters by giving a time period from 2012-10-25 to 2012-11-03, the keyword of interest "sandy".
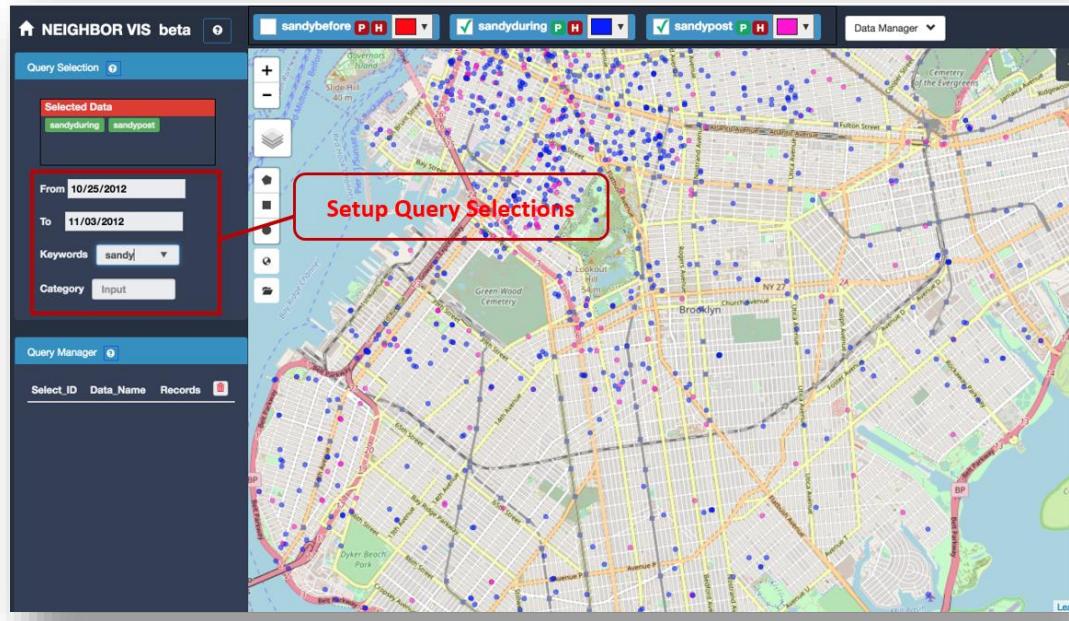

6. Select map icons to draw a region on the map. All events satisfy the filters and inside this region are used to form one active query. That is: all GSE events (geo-tagged tweets) from 2012-10-25 to 2012-11-03 and contain keyword "sandy". This query also

appears in the Query Manager Panel. Other events out of this query are shown in small dots as background. Users can make them disappear by operating on the above panel as well.
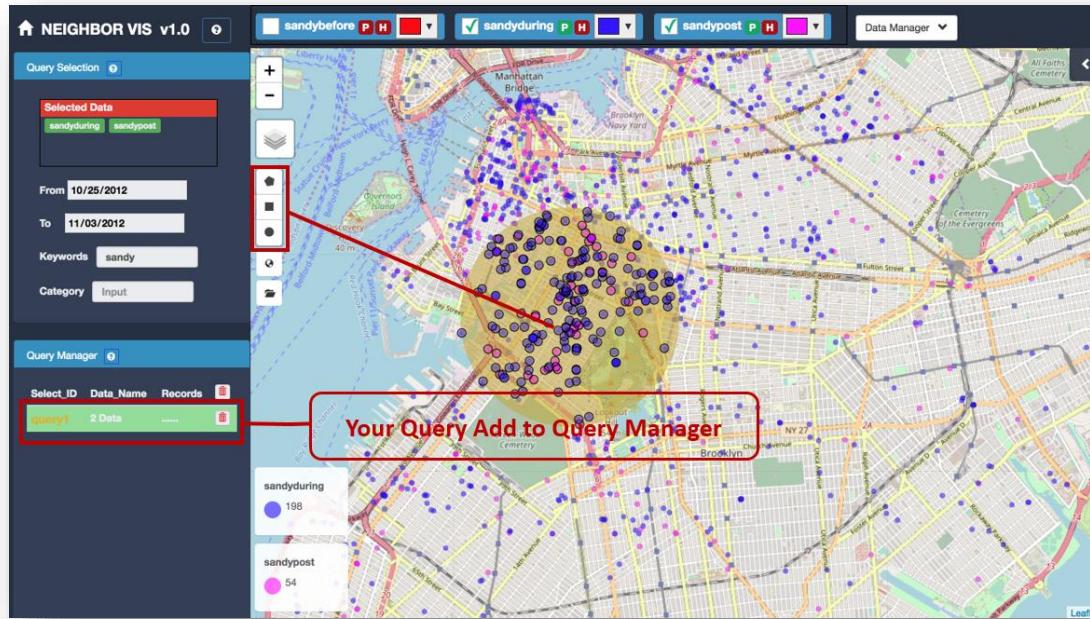

7. **Perform two more queries. These queries are shown together on the map with Query 1 being highlighted. Users can select different active queries on the Query Manager.**
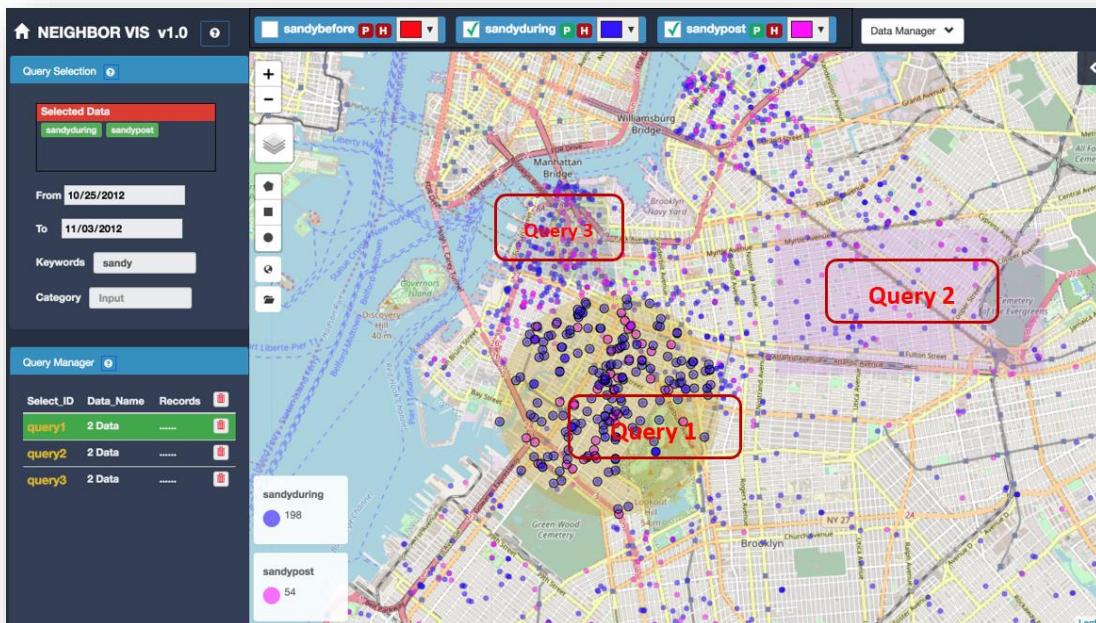
8. By clicking "Query 1" in the Query Manager, a sublist shows the two datasets (SandyPost SandyDuring) associated with this query. By selecting one of them (SandyDuring), the Map View and Query Analysis View are updated to show only the events in this specified dataset (SandyDuring).
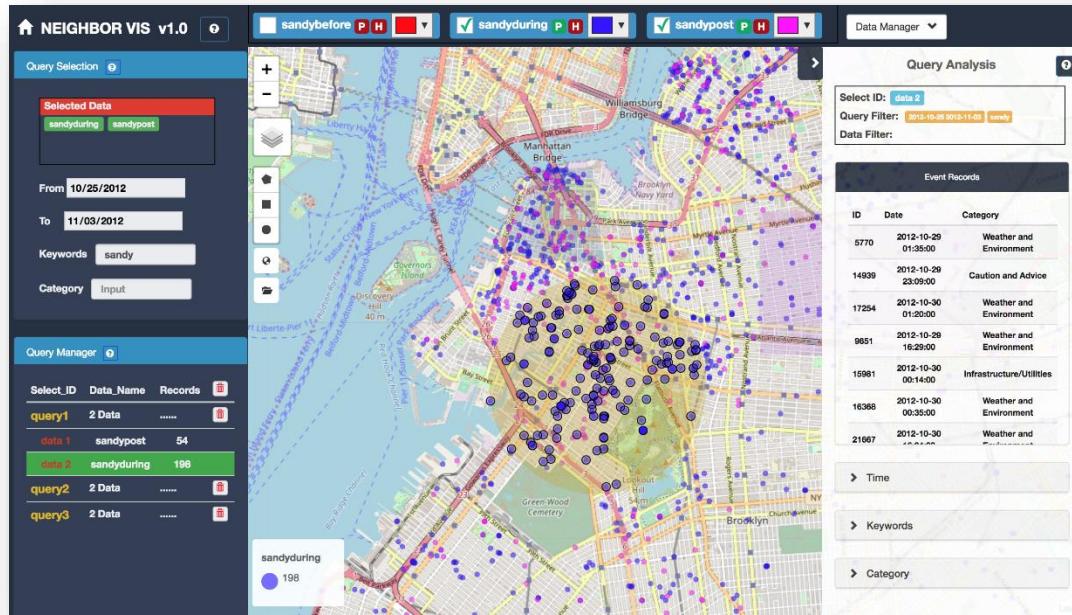

9. Study Query Analysis View for the list of events, the time distribution, keyword clouds and category distributions. Users can check event details on the map as well.
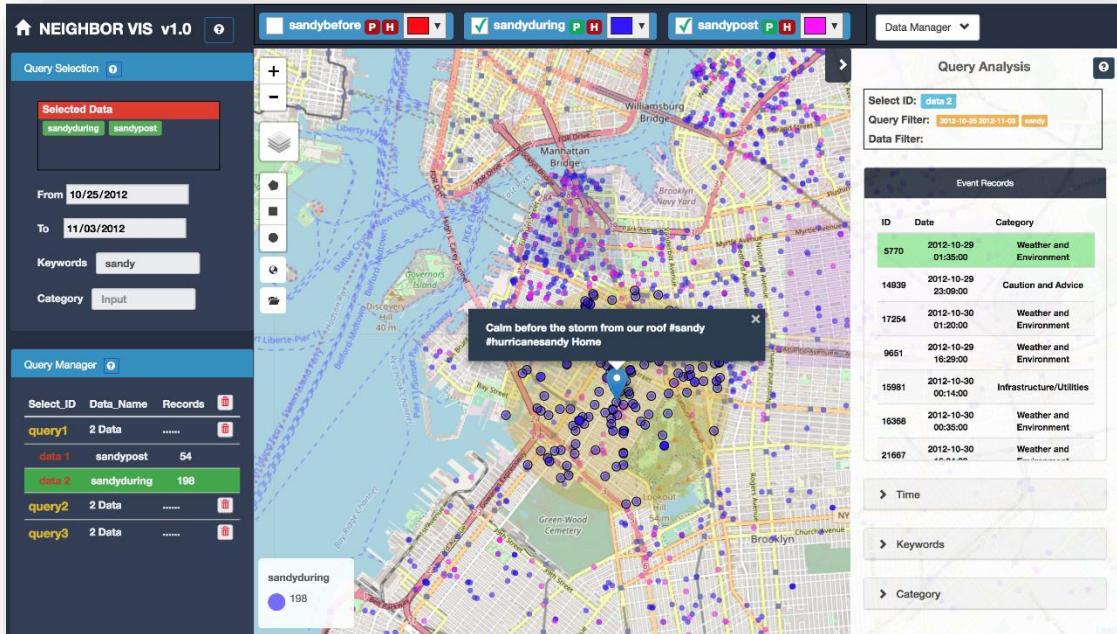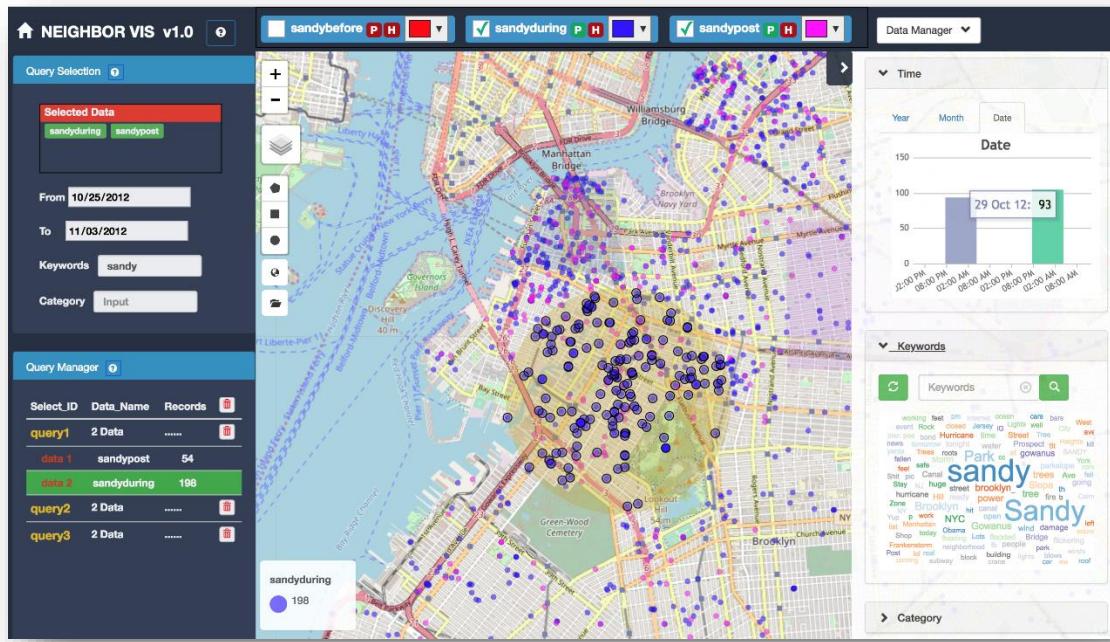
10. **Show time line details of events at a specific location, which may be the same person published multiple tweets.**
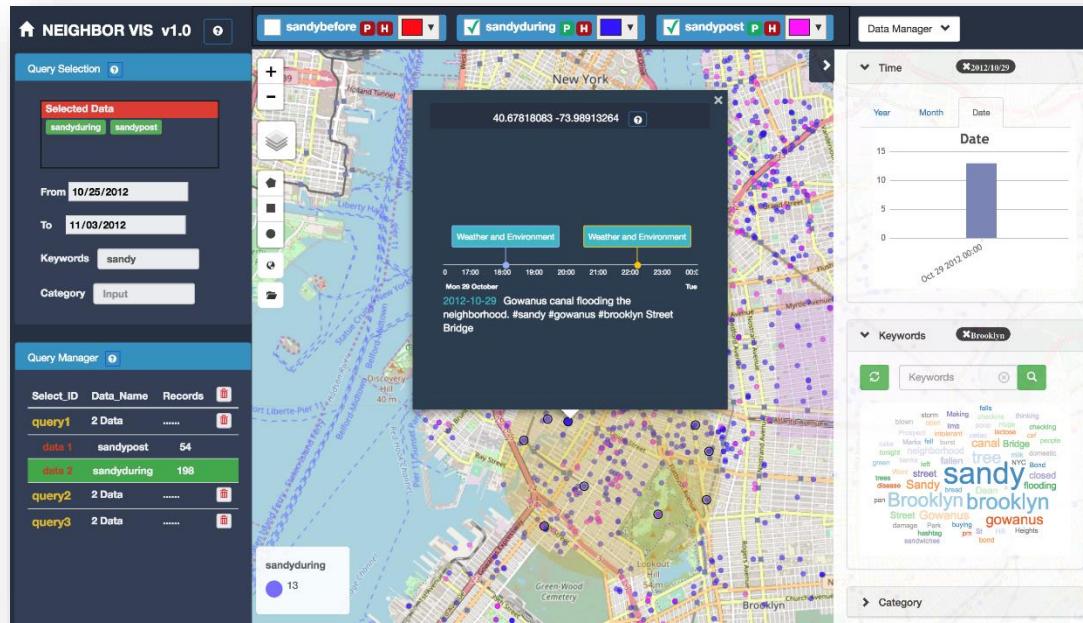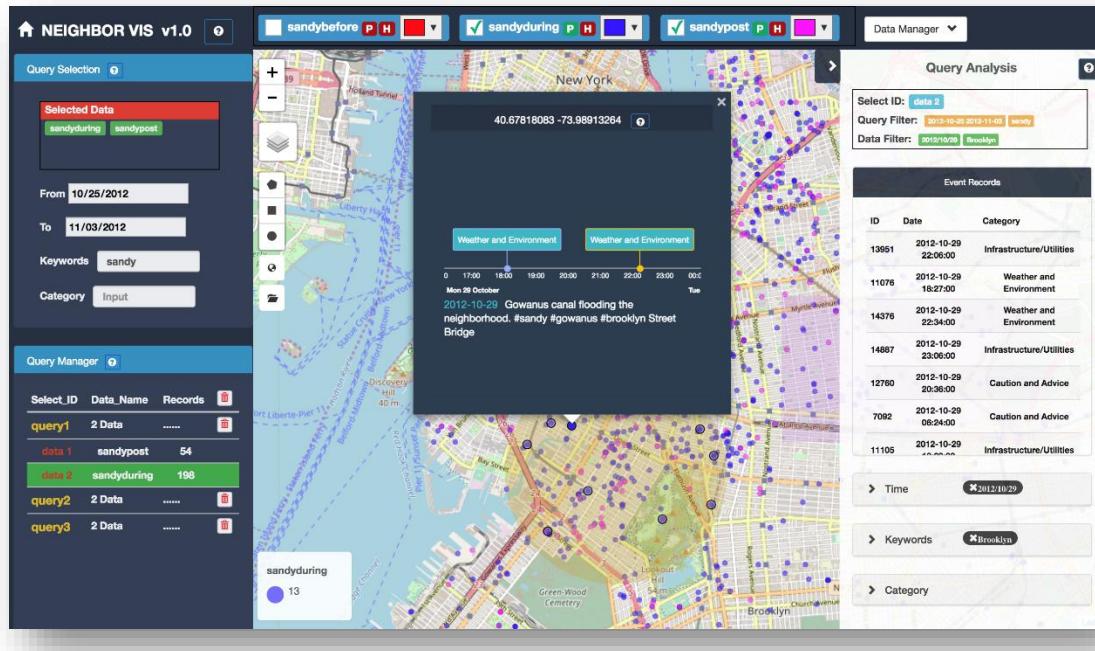
VI. Summary

NeighborVis software is used for analyzing geospatial temporal events with semantic contents. NeighborVis has two ways to access:

➢ Downloadable version to be used in desktop/laptop, please download it from our website and refer to the installation guide for technical details.

➢ Internet-based Visual analysis through Web browsers, please go to the website for details

NeighborVis will be updated with more functions in both data processing and visualization. Please help to provide your feedback, suggestions and comments on software website to help us improve the software.