

Visualization Tool for NYC Open Data - A Time Lapse Analysis

Alan Yuan

Dept. of Computer Science
New York Institute of Technology
New York City, NY, USA
ayuan01@nyit.edu

Guang Wei Too

Dept. of Computer Science
New York Institute of Technology
New York City, NY, USA
gtoo@nyit.edu

Yihan Xu

Dept. of Electrical and Computer
Engineering
New York Institute of Technology
New York City, NY, USA
yxu66@nyit.edu

Ziqian Dong

Dept. of Electrical and Computer
Engineering
New York Institute of Technology
New York City, NY, USA
ziqian.dong@nyit.edu

Huanying Gu

Dept. of Computer Science
New York Institute of Technology
New York City, NY, USA
hgu03@nyit.edu

Roberto Rojas-Cessa

Dept. Electrical and Computer
Engineering
New Jersey Institute of Technology
Newark, NJ, USA
rcessa@ieee.org

Abstract—In this paper, we present a web-based tool that maps and visualizes datasets from New York City (NYC) Open Data, specifically on building energy efficiency and fallen trees reported to NYC311, as heatmap or pin map to show building energy efficiency and weather event impact at a neighborhood level over time. To evaluate its effectiveness in addressing climate-related urban challenges, we applied the tool to two case studies from NYC Open Data: the ENERGY STAR Score and Fallen Trees datasets. We demonstrate the heatmap function for the ENERGY STAR Score dataset and the use of a combination of heatmap and pin map functions of the Fallen Trees dataset to highlight spatial and temporal patterns. The interactive visualization tool effectively provides data distribution and trend analysis based on postal codes while also allowing for precise, location-specific insights using longitude and latitude. Beyond its applications in data visualization, the tool can facilitate decision-making in the design of future urban environments.

Index Terms—Heat map, pin map, NYC Open Data, data visualization, trend analysis

I. INTRODUCTION

New York City (NYC), one of the most densely populated cities in the United States, is home to countless buildings that run lights around the clock and provide heating and cooling to thousands of residents. Despite the massive consumption of energy, efforts to increase energy efficiency can be overshadowed by the lack of presentation of buildings' energy expenditure. Disclosure of the annual energy and water used is directed by the NYC Benchmarking Law [1], which helps to appreciate the efforts made on making energy consumption more sustainable. In addition to these disclosures, the U.S. government has expanded the ENERGY STAR program [2], an initiative designed to help consumers identify and

select energy-efficient products, by developing an energy performance rating system. This system not only provides labels for residential heating and cooling systems and new homes but also for existing buildings.

In NYC, buildings are responsible for approximately 80% of the city's greenhouse gas (GHG) emissions [3]. Understanding the ENERGY STAR Score is crucial for gauging building energy performance to reduce household-related energy and water bills. Additionally, it can help to develop effective strategies to achieve the city's 80 by 50 goal, which aims to reduce GHG emissions from energy use (heating, cooling, and power) by 30 % over the next decade [4]. NYC's aging, underutilized buildings offer a possibility for revitalization while cutting energy use and emissions. Many were built before modern efficiency standards, leading to higher energy costs, inefficient heating, and cooling [5].

These issues raise important questions: do higher ENERGY STAR scores genuinely translate into measurable environmental benefits, and can these scores guide policy or motivate building owners to invest in sustainability? To address these challenges, we need a visualization tool capable of displaying energy performance across different regions, economic conditions, and time frames. This tool helps us track evolving trends, expose key disparities, and provide supporting evidence needed to ensure equitable and data-driven improvements in building efficiency.

Some visualization tools have been developed to track data distribution and trends [6]–[8]. Zuk et al. [6] presented an interactive tool that effectively visualizes neighborhood-level changes related to displacement and gentrification. The tool draws on various open datasets

(such as NYC Open Data [7] and census data) and allows users to switch between layers that highlight different indicators, such as population changes, rising housing costs, and income levels. This functionality helps community groups, policymakers, and researchers identify which neighborhoods are experiencing (or at risk of) large-scale displacement. Although this approach produces valuable visualizations, it lacks the flexibility to accommodate additional datasets to provide further insights. A visualization platform that can not only overlay building-level performance on ZIP code layers, but also incorporate data with high spatial and temporal resolution is needed.

To address this need, we developed a web-based visualization tool that can display a postal code (ZIP) code-based color-coded map of NYC, as well as pin markers based on the longitude and latitude location to pinpoint a high-resolution location of an event or point of interest. It will not only support the ENERGY STAR Score but also support a wide variety of available datasets for New York City. Thus, we can compare and analyze potential relationships among various data, enabling us to understand the existing intertwining among the presented urban data, the location, and possibly the communities. The tool is intended for rapid, general-purpose use, offering quick or sample visualizations that can be readily adapted to data associated with ZIP code or GPS location.

For more complex geospatial data visualization and analysis, established tools such as ArcGIS [8], QGIS [9], and Jupyter Notebook [10] offer extensive functions for global mapping and in-depth study. However, the visualization tool proposed herein is developed as an accessible lightweight solution for greater accessibility in analyzing NYC Open Data. Our primary goal is to provide users with a quick and interactive tool to identify urban trends in NYC Open Data. Users only need to enter the necessary information to achieve the distribution of data on the map. By simplifying the data visualization process, this tool lowers the entry barrier for those unfamiliar with Geographic information system (GIS) software and facilitates exploratory research that could lead to innovative sustainability studies. The presented visualization tool can be easily adapted for application on other cities' data.

The remainder of the paper is organized as follows. Section II describes the details of the visualization tool, including the data sources and processing and visualization technologies utilized. Section III demonstrates the visualization of heat maps and pinpoint capabilities using two sample datasets: ENERGY STAR Score for heat map visualization and Fallen Trees for pin map. Section IV discusses future work. Section V concludes the paper.



Fig. 1. Overview of the visualization tool.

II. METHODOLOGY

Figure 1 illustrates the components and interface of the proposed web-based mapping and visualization tool. The application is implemented in Python. It takes in a JSON file from NYC Open Data and utilizes Flask as the web server framework. The interface was designed using React.js with CSS and JavaScript for interactive map functionalities. Firebase was used for website hosting. The output is an interactive map that users can explore and analyze. The process begins with taking datasets from NYC Open Data, a free public platform hosting information collected by the city government [11], as input. In the following section, we outline the guidelines for using the tool, including details on data input requirements and data processing steps.

The system follows a structured pre-processing and visualization workflow, as shown in Figure 2. The users provide inputs, including the time range, the column of interest, ZIP code, and visualization type. The system then pre-processes, formats the data, and generates a descriptive statistical summary, an occurrence-based heatmap, or a pin map for geospatial representation.

A. Data Preparation

Figure 3 shows an example of the input for the application. There are four distinct input rows, each serving a specific purpose in generating the map. Users can specify the time range, chosen variable, ZIP code, and dataset

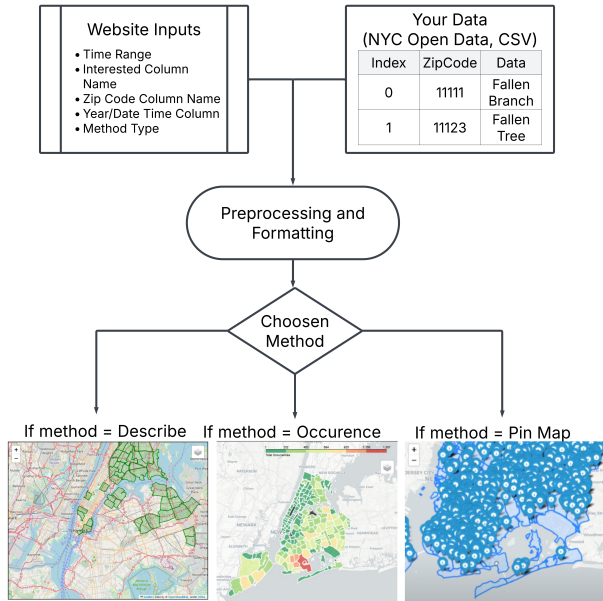


Fig. 2. Data processing and visualization workflow

Fig. 3. User interface for generating a time-lapse map based on NYC Open Data inputs

source in the application. In this case, `energy_star_score` data from 2017-2021 is mapped on a ZIP code basis.

The visualization tool processes the data based on the selected map function. For the heat map function, the tool groups all data by ZIP code. Alternatively, for the pin map function, users specify the longitude and latitude fields of the dataset to pinpoint exact locations. The heat map handles missing data according to the method users choose. For example, if the user selects descriptive statistics for each ZIP code in a specific data column, the system creates a copy of the dataset and removes outliers before performing a mean calculation. The means value is then used to fill in missing numerical data.

For categorical data, the system removes rows with missing values before performing frequency analysis, ensuring a complete dataset and preserving data in-

tegrity. For numeric data, the system allows duplicates and accepts null values upon input, assuming that they represent meaningful data that users intend to visualize. The key steps in this process are the following:

- Validating data format: Ensures that the uploaded file adheres to JSON or CSV standards.
- Standardizing variable names: Maps the user-specified ZIP code, longitude/latitude, and variable fields to a uniform structure for processing.
- Filtering and aggregation: Filters the data based on the selected time range and aggregates values by location identifiers to simplify geographic representation.

B. Visualization Process

After parsing the data, the tool generates an interactive, color-coded map that displays the specified variable by ZIP code or by pins located by longitude and latitude. The map provides users with the following capabilities:

- Explore data trends over time: Use an interactive slider to observe changes along the selected time range. This feature allows users to visualize temporal patterns and trends within the dataset, making it easier to identify shifts and anomalies.
- Location-specific mapping: Zoom in on particular ZIP codes or pins to gain deeper insights of localized patterns. This approach allows users to isolate and study specific areas, uncovering trends or outliers that might be otherwise overlooked in a broader analysis.

III. CASE STUDIES

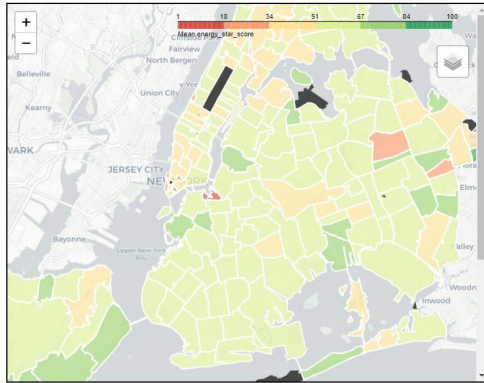
In this section, we demonstrate the interactive visualization tool using two case studies: the ENERGY STAR Score and Fallen Tree Datasets.

A. Energy STAR Score Dataset

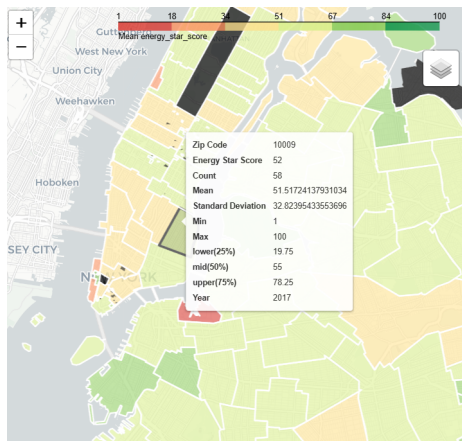
ENERGY STAR is a government-backed symbol for energy efficiency, and it includes a benchmarking system that evaluates a building's energy performance through the ENERGY STAR Score [12]. This score provides a comprehensive assessment of the energy efficiency of a building considering various factors, such as the physical characteristics of the building, operational practices, and the behavior of the occupants. The score is expressed on a scale from 1 to 100, where a higher score indicates superior energy efficiency and performance, making it an essential metric for evaluating sustainability initiatives in urban environments [13]. These scores are instrumental in evaluating a building's performance and identifying those with the greatest potential for improvement or recognition. For this case study, we reference the dataset titled "NYC Building Energy and Water Data Disclosure for Local Law 84 (2022-Present)" from NYC Open Data

TABLE I
ENERGY STAR SCORE DATASET ATTRIBUTES

Dataset	NYC Building Energy and Water Data Disclosure for Local Law 84 (2022-Present)
Number of Rows	64.2K
Number of Columns	256
Agency	Department of Buildings (DOB)
Data Last Updated	November 25, 2024
Update Frequency	Annually



(a)



(b)

Fig. 4. Heat map of NYC, with colors representing ENERGY STAR scores of buildings across the city. The score ranges from 1 to 100, shown using a color gradient. The higher the score, the better the energy efficiency. (a) is the general visualization (b) shows the detailed information at the ZIP code level.

[14]. A brief overview of this dataset is provided in Table I.

In NYC, buildings exceeding 25,000 square feet are required to display a Building Energy Efficiency Rating Label, in compliance with Local Law 33 of 2018 and Local Law 95 of 2019 [1]. There are 50,000 buildings

larger than 25,000 square feet in NYC. According to the Urban Green Council Local Law, these buildings account for approximately 60% of all buildings and are responsible for two-thirds of the city's annual emissions, and significantly contributing to energy production impacts. Displaying data for nearly 50,000 buildings with a pin map would be challenging; therefore, we employ a heat map to analyze and visualize this large-scale dataset. By displaying ENERGY STAR Scores geographically, one can easily identify areas with good or poor scores and analyze significant trends. This visualization also enables users to observe how scores have changed over the years.

In addition, our tool allows users to zoom in on specific ZIP codes or neighborhoods. By integrating geographic, demographic, and contextual data, users can analyze patterns and correlations between scores and factors such as location, building age, proximity to industrial zones, or access to renewable energy sources. For example, clusters of low scores in specific neighborhoods may indicate issues such as outdated infrastructure, lack of energy retrofitting, or socioeconomic barriers preventing upgrades.

Figure 4 (a) shows the ENERGY STAR Score with seven (7) different colors each representing the ENERGY STAR score average in each of the ZIP codes. The higher the score, the better the energy performance for the average ENERGY STAR Score for each ZIP Code, along with their corresponding colors:

- **85-100** (Green)
- **67-84** (Light Green)
- **51-66** (Yellow-Green)
- **34-50** (Yellow)
- **18-33** (Orange)
- **1-17** (Red)
- **NULL** (Black): ENERGY STAR Scores are missing or unavailable data.

The ENERGY STAR Score average is calculated by aggregating all scores within a given ZIP code. The application removes any missing values from the dataset for calculating the average.

Figure 4 (b) illustrates that the tool can provide detailed ZIP-code level information. We picked ZIP code 10009 as an example, which includes the average ENERGY STAR Score, mean value, minimum and maximum values, year, etc. It is observed that the ENERGY STAR Score in recent years does not show any significant change. By providing a detailed, data-driven approach to visualizing ENERGY STAR Scores, our tool serves as a valuable resource for researchers, policymakers, and investors alike. Beyond research applications, policymakers can leverage this tool to assess the effectiveness of existing sustainability programs. From a decision-making perspective, the tool enables policymakers to pinpoint neighborhoods that would ben-

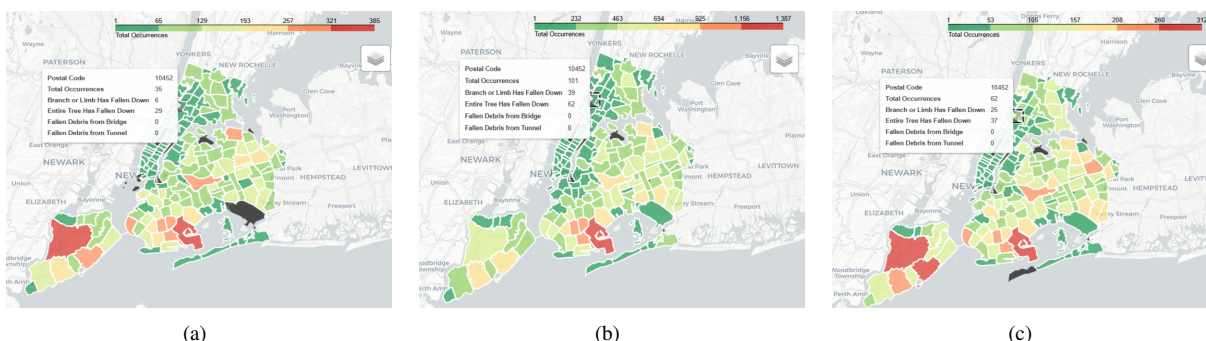


Fig. 5. Fallen tree visualization on heat and occurrence map. Greener areas mean fewer fallen trees and redder areas mean more fallen trees. Black means missing data. (a) data in 2019 (b) data in 2020 (c) data in 2021.

enefit most from targeted retrofitting programs, financial incentives, or stricter energy performance regulations. Similarly, real estate developers and property owners can use the tool to make informed investment decisions, identifying high-potential areas for green building improvements.

B. Fallen Trees Dataset

Severe weather events, such as storms and hurricanes, are primary causes of fallen trees and can lead to extensive damage [15]. By mapping clusters of fallen trees, our tool not only reveals the substantial impact of climate-related events on urban settings but also provides practical insights that can help NYC Parks allocate resources more effectively during post-storm cleanups. To showcase how the tool addresses localized data, we used an interactive pin and occurrence map to visualize each fallen tree incident by the exact longitude and latitude. The fallen tree dataset from the 311 data on NYC’s Open Data Portal [16]. As depicted in Figure 5, the data present information on fallen trees within the specified ZIP code for years 2019, 2020, and 2021. Users can explore the number of fallen trees, branches, limbs, and debris in each ZIP code. Additionally, the figure allows users to observe trends in tree falls over the three-year period. Figure 6 presents a case data on fallen trees in ZIP code 11697. This area is chosen because it has the fewest fallen trees, making it an ideal case for demonstrating the results. Figure 6 (a) provides detailed information about ZIP code 11697, while Figure 6 (b) illustrates a pin map of the fallen trees, highlighting their precise geographical locations.

In this case, we chose a pin map over a heat map to achieve the granularity needed for swift planning and response, enabling stakeholders to pinpoint high-risk areas accurately. Although they illustrate the overall density, heat maps often obscure individual data points, making it difficult to identify localized issues. In contrast, pin maps reveal the precise locations of each fallen tree, highlighting clusters along specific streets, near parks, or



Fig. 6. Visual map of the Fallen Tree dataset in the ZIP code 11697. (a) display the detailed information in this ZIP code area. (b) shows the fallen tree’s pins on the map.

within neighborhoods, these patterns that may indicate aging trees, inadequate maintenance, or the effects of wind corridors. This level of detail is invaluable for urban forestry management and infrastructure planning. Additionally, when combined with temporal analysis, the visualization allows users to track changes over time, assess the impact of interventions, and monitor trends re-

lated to climate events, ultimately helping urban planners and policymakers allocate resources more effectively and mitigate risks in vulnerable communities.

IV. FUTURE WORK

Although the current version of the interactive visualization tool for NYC Open Data provides significant benefits by visualizing neighborhood data, additional functions may be needed to improve data interpretation. One area of focus is adding functionality to compare multiple datasets, enabling cross-sectional and multi-metric analysis. Another limitation of our approach is the NYC Open Data platform's restriction of 1000 rows per request, which significantly impacts data retrieval. While our tool is capable of processing larger datasets, the primary challenges come from the data accessing rather than computational capabilities. Although the tool can handle extensive datasets, larger queries require more time to process. To address this limitation, users can upload external CSV files containing large datasets, bypassing the API constraint and allowing for more comprehensive analysis. However, additional challenges remain, partially in implementing automated data cleaning and handling missing value, which require further refinement.

V. CONCLUSIONS

This paper presents an interactive visualization tool designed to enhance the accessibility and usability of NYC Open Data or large data sets to provide insights into urban sustainability. The tool can show time lapse of data as heat maps and pin point locations of interest. We tested the visualization function using two NYC Open Data datasets, the ENERGY STAR Score and Fallen Trees, demonstrating its capacity to visualize time-sensitive and incident-specific data alongside city-wide trends. This visual tool can offer policymakers, researchers, and community stakeholders an intuitive interface to explore complex datasets, and help city planners prioritize sustainability initiatives, such as targeted retrofitting programs or incentive-based energy efficiency upgrades.

VI. ACKNOWLEDGMENT

This material is based upon work partially supported by the National Science Foundation under Grant Awards 1856032, 1834099, and 2402240.

REFERENCES

- [1] NYC Department of Buildings, "Benchmarking," 2025, accessed: February 16, 2025. [Online]. Available: <https://www.nyc.gov/site/buildings/codes/benchmarking.page>
- [2] S. ENERGY, "Energy star@," *History: ENERGY STAR*, 2011.
- [3] New York City Department of Buildings, "Energy efficiency," 2025, accessed: 2025-03-05. [Online]. Available: <https://www.nyc.gov/site/buildings/property-or-business-owner/energy-efficiency.page>
- [4] NYC Built to Last, "A roadmap for new york city's buildings," accessed: 2025-03-05. [Online]. Available: <https://www.nyc.gov/site/builttolast/goals/goals.page>
- [5] City Limits, "Opinion: The untapped potential of nyc's existing buildings in the climate fight," 2025, accessed: 2025-03-09. [Online]. Available: <https://citylimits.org/2025/02/21/opinion-the-untapped-potential-of-nycs-existing-buildings-in-the-climate-fight/>
- [6] M. Zuk and K. Chapple, "Urban displacement project," *URL* <https://www.urbandisplacement.org>, 2015.
- [7] K. Okamoto, "Introducing open government data," *The Reference Librarian*, vol. 58, no. 2, pp. 111–123, 2017.
- [8] B. Booth, A. Mitchell *et al.*, "Getting started with ArcGIS," 2001.
- [9] N. Moyroud and F. Portet, "Introduction to QGIS," *QGIS and generic tools*, vol. 1, pp. 1–17, 2018.
- [10] D. Yin, Y. Liu, A. Padmanabhan, J. Terstriep, J. Rush, and S. Wang, "A CyberGIS-Jupyter framework for geospatial analytics at scale," in *Practice and Experience in Advanced Research Computing 2017: Sustainability, Success and Impact*, 2017, pp. 1–8.
- [11] K. Okamoto, "What is being done with open government data? an exploratory analysis of public uses of new york city open data," *Webology*, vol. 13, no. 1, 2016.
- [12] U.S. Environmental Protection Agency, "Energy star score," accessed: 2025-02-17. [Online]. Available: <https://portfoliomanager.energystar.gov/pdf/reference/ENERGY%20STAR%20Score.pdf>
- [13] —, "Analyze benchmarking results - energy star," 2024, accessed: December 22, 2024. [Online]. Available: <https://www.energystar.gov/buildings/benchmark/analyze-benchmarking-results>
- [14] New York City Mayor's Office of Sustainability, "NYC Building Energy and Water Data Disclosure for Local Law 84 (LL84)," 2024, accessed: 2024-03-09. [Online]. Available: <https://data.cityofnewyork.us/Environment/NYC-Building-Energy-and-Water-Data-Disclosure-for-15zyy-y8am>
- [15] T. W. Schmidlin, "Human fatalities from wind-related tree failures in the united states, 1995–2007," *Natural Hazards*, vol. 50, pp. 13–25, 2009.
- [16] New York City Department of Information Technology and Telecommunications, "311 Damaged Tree Dataset," 2024, accessed: 2024-03-09. [Online]. Available: <https://data.cityofnewyork.us/Social-Services/311-Damaged-Tree/xb2u-ubdp>