Mapping the Spatial and Temporal Patterns of Bedbug Complaints in New York City

Final Project Info 628 Data Librarianship and Management

Willem Helf &Nene Villalobos

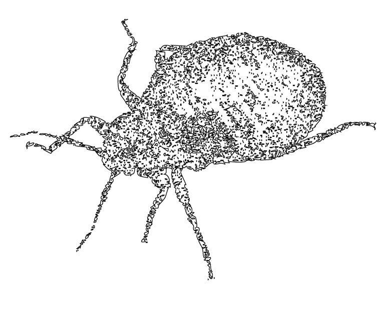


Table of Contents

Data Management Plan

Project Title: *Mapping the Spatial and Temporal Patterns of Bedbug Complaints in New York City*
Researchers: Nene Villalobos & Willem Helf
Date: October 2025

1. Overview and Research Objectives

This project investigates how the frequency and geography of bedbug complaints in New York City have evolved over the past decade (2010–2024). By examining Bedbug reports and 311 complaint data alongside demographic and housing indicators, this research aims to identify spatial and temporal trends in bedbug infestations, locate persistent or emerging hotspots, and explore correlations between complaint density and socioeconomic variables such as renter density and median income. The primary objective is to generate accessible visualizations that reveal how infestations map onto urban inequality and housing precarity. Findings will be shared through interactive data visualizations and interpretive summaries suitable for both public audiences and urban policy researchers.

2. Data Types and Sources

Primary Dataset

- **NYC 311 Bedbug Complaints Dataset** (available via NYC Open Data).
  Variables: complaint type, status ("Infestation," "Eradicated"), creation and closure dates, geographic coordinates, zip code, and borough.

- Time coverage: 2010–2024.

Secondary Datasets

- **NYC Open Data Demographic and Housing Indicators**:

  - NYC Open Data Bedbug and census tracts.
  - Variables: renter, household income, building age, and housing density.

Data Format and Volume
All data will be use API for retrieval and visualization. The datasets are public and free of personally identifiable information (PII).

3. Data Collection and Processing

Collection Methods

- Datasets using an NYC Open Data Bed Bug API to ensure data use with zero tiddying.

Cleaning and Transformation

- Clean attributes (latitude, longitude, zip, borough) using Python (GeoPandas).
- Standardize time fields to ISO 8601 format and remove duplicates.
- Join the complaint data with demographic indicators by geospatial location or ZIP code for analysis.

Documentation
A README file will accompany each dataset, including variable definitions, cleaning steps, and licensing information. Scripts used for cleaning and analysis will be version-controlled on GitHub with comments describing code functionality.

4. Data Analysis and Visualization

Analyses will include:

- **Temporal Analysis:** monthly and yearly trend plots showing complaint frequency over time.
- **Spatial Analysis:** heatmaps and choropleth maps displaying the density of complaints by borough and neighborhood.
- **Correlation Analysis:** comparative layers of complaint density and socioeconomic variables (% renters, median income).

An interactive web map ~~or dashboard~~ will enable users to explore changes in hotspots across time.

5. Storage, Backup, and Security

All working files will be stored in a shared Google Drive. Local copies will be version-controlled through GitHub and periodically archived to an external hard drive. Because the datasets are public and anonymized,confidential information will be handled.

6. Data Sharing and Reuse

Processed datasets and code will be deposited in an open repository (e.g., GitHub). The repository will include:

- Instruction on how to call the API
- README
- Jupyter Notebooks or Python scripts used for analysis
- Static visualizations and links to interactive maps
- Github pages to visulize the data

7. Preservation and Long-Term Access

Data and documentation will be retained for a minimum of five years post-completion (through 2030). Archival copies will be preserved in compressed formats (.zip)

8. Expected Outcomes and Broader Impacts

This project will produce a publicly accessible visual dataset and interactive map that traces the temporal and spatial distribution of bedbug complaints in NYC. By linking complaint density to housing conditions and income inequality, the study offers a model for using municipal open data to identify patterns of urban vulnerability~~.~~

Methodology Write-Up

1. Introduction and Research Focus

This project explores how the geography and frequency of bedbug complaints have shifted across New York City's boroughs from 2010 to 2024. Using publicly available municipal data, we aim to analyze and visualize the evolving spatial patterns of infestations in relation to neighborhood-level housing and socioeconomic characteristics. The broader intent is to understand how data can reveal urban inequities embedded within issues of pest control, sanitation, and tenant infrastructure.

We approach this study as both a **data-driven and interpretive project**, merging empirical analysis with a design-oriented mode of communication. Our methodological approach balances computational techniques such as geospatial mapping and correlation with narrative visualization, ensuring that the data is not only accurate but also accessible to non-technical audiences.

2. Data Collection and Preparation

2.1. Primary Data Source

The core dataset will be the **NYC 311 Service Requests for Bedbug Complaints**, accessible through the NYC Open Data portal. We will filter for records that explicitly list "Bedbugs" under complaint type and include fields for complaint status ("Infestation," "Eradicated"), creation date, and geographic coordinates (latitude, longitude, ZIP code, and borough).

This dataset will provide the foundation for our spatial-temporal analysis. The time span (2010–2024) ensures that our analysis captures both long-term and short-term variations, including potential policy shifts (e.g., NYC Housing Preservation and Development initiatives) and housing market changes that may influence reporting behaviors.

2.2. Secondary Data Sources

To contextualize complaint patterns, we will integrate secondary datasets:

- **Demographic and Housing Indicators:** NYC Neighborhood Tabulation Area (NTA) data, including renter-to-owner ratios, median household income, and population density.

Data cleaning and normalization will be performed using **Python (Pandas)** and **OpenRefine**, ensuring consistent coordinate systems and date formatting across datasets.

3. Analytical Methods

3.1. Temporal Analysis

To understand trends over time, we will aggregate complaint data by month and year. This temporal dimension will help reveal fluctuations in reporting frequencysuch as seasonal peaksand provide insight into long-term increases or decreases in infestations. Analyses will be conducted in **Python** using Pandas for aggregation and Matplotlib for plotting time series graphs. The resulting visualizations will highlight yearly changes per borough and offer a comparative view across neighborhoods.

3.2. Spatial Analysis

Spatial analysis underpins this project. Using **Python's Pandas** library, we will geocode complaint records and join them with demographic boundaries to generate:

- **Heatmaps** showing the concentration of bedbug complaints across boroughs.
- **Choropleth maps** illustrating normalized complaint rates per 1,000 residents.

These maps will be used to assess how infestations correlate with housing conditions, urban density, and income inequality.

3.3. Correlation and Contextual Analysis

To probe potential relationships between infestation density and social indicators, we will perform basic correlation and regression analyses in Python. For example, complaint rates will be compared to:

- **Percentage of renter-occupied housing** (as a proxy for housing mobility).
- **Median income levels** (as an indicator of economic vulnerability).

These correlations will be exploratory rather than causal, but they will help us identify where further qualitative research or city intervention might be warranted.

4. Visualization and Communication Methods

4.1. Interactive Mapping

The project's key deliverable will be an **interactive web-based map** that visualizes complaint patterns over time. We will use **Tableau Public** and **Python Folium** to design time-lapse layers, allowing users to explore the spread of infestations by year. The visualization will be designed to prioritize clarity and narrative engagement:

- Layer toggles for borough-level comparison.
- Animated temporal sliders showing how hotspots expand or contract.
- Tooltip pop-ups displaying contextual statistics (e.g., neighborhood median income).

4.2. Narrative and Design Integration

Our visualization strategy draws on principles of **information design and data storytelling**. Each map will be accompanied by short textual annotations highlighting notable changes such as emerging hotspots or sharp declines in particular neighborhoods.

We will also include a **comparative panel** overlaying complaint density with socioeconomic layers to make visible how specific urban populations bear disproportionate environmental and pest-related burdens. The aesthetic and structural design will reflect our shared background in communication design and data visualization, aiming for interpretive depth while maintaining analytic rigor.

5. Tools and Workflow

Task	Tools/Software	Output
Data Cleaning	Python (Pandas)	API
Geospatial Mapping	GeoPandas	heatmaps

Interactive Visualization	Python	Web-based dashboard
Repository & Documentation	GitHub, Google Drive	Open data and reproducibility package

All scripts, intermediate datasets, and final outputs will be systematically labeled and stored in a structured folder hierarchy:

6. Ethical Considerations

All datasets originate from public sources and contain no personally identifiable information. Nevertheless, we will approach the visualization of neighborhood data with sensitivity, avoiding language or imagery that could stigmatize specific communities. Our interpretation will frame infestation trends as systemic outcomes tied to housing conditions and policy structures, not individual failings.

We also acknowledge that data like 311 complaints are proxies for visibility rather than absolute incidence; reporting may correlate with digital literacy, landlord responsiveness, or language access. This awareness will inform how we discuss data limitations in our final presentation.

7. Expected Outcomes

- A series of **maps and time-lapse visualizations** revealing how bedbug complaints have evolved spatially and temporally.
- A **comparative dataset** linking infestation rates with socioeconomic indicators.
- A public-facing **interactive dashboard** hosted on Tableau Public or GitHub Pages.
- A **short analytical summary** discussing observed trends, data caveats, and implications for urban policy and public health.

Beyond this specific case study, our methodology provides a reproducible framework for combining civic data with geospatial analysis to investigate other environmental or housing-related urban issues.

Mapping Bedbug Complaints in NYC: Data Analysis Write-Up

Bedbugs are often framed as an isolated public health nuisance, an unfortunate but minor domestic problem. However, when examined at scale, bedbug complaints reveal patterns in relationships among housing conditions, infrastructure, and inequality. This project approaches bedbug infestation not as an individual failure of cleanliness or behavior, but as a structural outcome shaped by housing precarity, density, and uneven access to remediation. Using New York City's 311 Bedbug Complaint dataset (2010–2024), this study analyzes spatial and temporal patterns of reported infestations and situates them within broader socioeconomic and housing contexts. The project asks: What does a dataset about pests reveal about people, power, and urban infrastructure? How do infestation reports cluster across neighborhoods, and what do those clusters suggest about who is most exposed to prolonged environmental discomfort?

By integrating complaint data with neighborhood-level demographic and housing indicators, this analysis demonstrates that bedbug reports consistently concentrate in areas marked by older housing stock, high renter occupancy, and lower median incomes. The resulting visualizations aim to shift the narrative around infestation from episodic inconvenience to systemic condition while highlighting how municipal data reflects both material realities and uneven capacities to report harm. The primary dataset for this study is the NYC 311 Bedbug Complaint dataset, accessed through NYC Open Data. Records include complaint dates, geographic coordinates, boroughs, ZIP codes, and complaint status. The fourteen-year temporal scope enables long-term analysis of trends, seasonality, and persistence.

Data were joined at the neighborhood level to support comparative analysis. While public datasets limit resolution and certainty, the combined sources provide sufficient structure to identify persistent spatial relationships between infestation reports and housing vulnerability. Records with missing or invalid geographic coordinates were removed, dates were standardized, and early reporting inconsistencies were corrected. Duplicate and ambiguous entries were reviewed to ensure consistent categorization across years. This approach allowed for parallel examination of where complaints cluster and how reporting changes over time.

Spatial patterns were examined using heatmaps. Complaint counts were normalized by population and housing density to avoid overrepresenting high-population areas. Spatial visualizations revealed persistent clusters of high complaint density in the Bronx, central Brooklyn, and northern Manhattan. These clusters remained stable across multiple years, suggesting that infestations are not episodic events but recurring conditions tied to long-term housing infrastructure. There is still room to investigate this notion.

Temporal aggregation revealed consistent patterns, with complaint peaks in late summer and early fall, but getting this developed into a map was still needing work. These peaks align with known bedbug life cycles and increased population mobility during warmer months. However, seasonal intensity varied significantly by neighborhood, indicating that environmental exposure and remediation capacity are unevenly distributed. Longitudinal trends also reflected shifts in reporting behavior influenced by public awareness campaigns and housing policy changes. Importantly, decreases in complaint volume did not consistently correspond to improved housing conditions, underscoring the distinction between infestation prevalence and reported data.

Complaint density was examined in relation to key housing and socioeconomic indicators, including renter occupancy, building age, income, and housing density. The analysis revealed notably higher concentrations of complaints in renter-dominated neighborhoods and in areas with high densities of hotels and short-term lodging. This pattern is particularly striking given that infestations in hotel properties often receive rapid remediation through institutional resources and regulatory pressure, while surrounding residential areas especially renter communities continue to experience persistent infestations with limited support. These uneven responses suggest that bedbug complaints are not only unevenly distributed, but unevenly addressed. While certain property types benefit from expedited intervention and preventative resources, other neighborhoods struggle to secure sustained attention for what is ultimately a larger, systemic housing issue. The findings reinforce the central argument of this study: bedbug infestation is structurally mediated, shaped by housing systems, regulatory priorities, and access to resources rather than individual behavior. The most striking finding is the persistence of infestation clusters across time. Certain neighborhoods repeatedly appear as hotspots, suggesting chronic conditions rather than isolated outbreaks. These areas often overlap with neighborhoods historically affected by disinvestment, overcrowding, and limited housing enforcement. This persistence challenges narratives that portray bedbugs as randomly distributed or universally experienced. Instead, infestation emerges as a spatialized phenomenon, deeply entangled with urban inequality.

The analysis highlights the disproportionate burden borne by renters, particularly those in multi-unit buildings. Renters often lack authority to implement preventative measures, depend on landlord compliance for extermination, and may fear retaliation for reporting infestations. These dynamics help explain both high complaint density and prolonged infestation cycles in renter-dominated neighborhoods. Additionally, underreporting is likely significant. Residents facing language barriers, immigration precarity, or housing insecurity may avoid formal reporting systems altogether. As a result, the dataset likely underrepresents infestation in already vulnerable communities.

Areas with high concentrations of hotels and short-term lodging particularly in Manhattan show elevated complaint activity. High occupant turnover increases the likelihood of transmission and complicates eradication efforts. While hotels often have resources to address infestations quickly, surrounding residential communities may experience prolonged exposure. This pattern highlights how bedbugs move across boundaries between private homes, public accommodations, and neighborhoods while their impacts are unevenly absorbed. Filing a 311 complaint is itself a form of bureaucratic and emotional labor. It requires time, trust in municipal systems, and belief that reporting will result in intervention. Complaint density therefore reflects not only infestation, but who feels empowered to report harm. The dataset captures moments where discomfort becomes legible to the city. Absence of complaints should not be read as absence of infestation, but often as absence of institutional trust or protection.

This project treats visualization as an interpretive act rather than neutral display. Decisions around scale, normalization, and aggregation shape how infestation is understood. The visualizations emphasize continuity and correlation rather than sensationalized hotspots. By pairing infestation data with socioeconomic indicators, the project reframes bedbugs as an infrastructural issue. The design resists narratives of contamination and instead foregrounds patterns of neglect, inequality, and uneven care. The dataset reflects reported complaints rather than confirmed infestations and is subject to reporting bias. Neighborhood-level aggregation obscures building-specific dynamics, and correlations do not establish causation.

This project demonstrates how a dataset about pests becomes a dataset about people. Bedbug complaints operate as indicators of housing precarity, renter vulnerability, and infrastructural neglect. When analyzed over time and space, they reveal predictable and persistent patterns of inequality. By reading municipal data as a cultural and political artifact, this study argues for more structurally informed approaches to both housing policy and data design. Infestation is not an anomaly it is a symptom of uneven care embedded within the city’s housing systems.

All data, code, and documentation generated during the project.

Date	Journal Entry
9/23/25	Today marked the real beginning of my technical learning curve. Will introduced me to the structure of the NYC Open Data API not just how to pull data, but how to read the endpoints, understand the parameters, and think about data as something actively queried rather than passively downloaded. I felt the gears shift in my mind: this project was no longer just about bedbugs, but about learning to speak the language of real-time data systems. Will walked me through the basics of constructing an API request, and while it felt abstract at first, I could feel the conceptual framework settling in.
9/30/25	I was still learning how to work with the API, so I focused most of my time on refining the Data Management Plan, tightening the structure, clarifying the data sources, and mapping out the workflow. Even while doing that, I kept checking in with Will to see where I could be helpful or contribute to the technical side. Will was deep in his own flow, experimenting with filters, limits, and different query structures, and I followed along as much as I could. We were working asynchronously but on parallel tracks, me grounding the project's framework, him pushing the technical boundaries. It felt productive, collaborative, and spacious: each of us locked into our tasks, but always looping back to compare notes and move the project forward together.
10/12/25	Together, we sketched out a rough wireframe on Neocities, playing with what the project could look like once code, API logic, and visual storytelling were woven together. Seeing the data architecture and the creative layer side by side made everything click: the API wasn't just a pipeline, it was part of the aesthetic and conceptual spine of the project. Will and I were bouncing ideas back and forth how interactive elements might respond to filters, how the visuals could carry the weight of the dataset, how the backend logic could shape the front-end experience. It was the first moment where technical learning and creative design felt inseparable. My analytical instincts and emerging technical skills finally met in a place that felt alive.
10/28/25	Today's work was shaped by the thoughtful feedback we received from classmates. Their comments helped us see that our original plan—to balance both qualitative interpretation and quantitative analysis was too diffuse for the direction our project was taking. We decided to pivot and center the quantitative foundation: the patterns, the clusters, the temporal shifts, and the socioeconomic correlations grounded in the dataset itself. The qualitative elements would remain present, but now as context rather than equal focus. It was a moment of realizing how much precision quantitative work requires and of feeling myself grow more capable of navigating it.
11/11/25	With the data in place, Will and I shifted toward interpretation and narrative. We reviewed the spatial clusters emerging from the merged dataset and talked through how socioeconomic patterns shape complaint concentrations. The technical skills I had gained made the interpretive work feel sharper and more grounded. I could now read a pattern and understand the mechanics behind its materialization. The

	API work wasn't just a skill; it became part of the project's analytical voice.
11/28/25	This stage felt like actual consolidation, where all the strands we had been developing finally began to weave into a cohesive whole. I spent the day refining our visualizations. Instead of juggling scattered downloads and manual steps, we began to see the project as a unified system. Will and I worked closely to align the technical pieces with the analytical structure we had been shaping for weeks. He helped me streamline sections of the workflow that had been clunky, and I could feel my understanding of the API its logic, its efficiency, and its constraints clicking more firmly into place. This consolidation phase made the project feel grounded and intentional: all the learning, experimentation, and revisions converging into a more straightforward narrative about how a dataset on pests reveals deeper stories about inequality, housing, and urban conditions.

The code generated during this project can be located on our GitHub:
https://github.com/willemhelf/bedbug_map

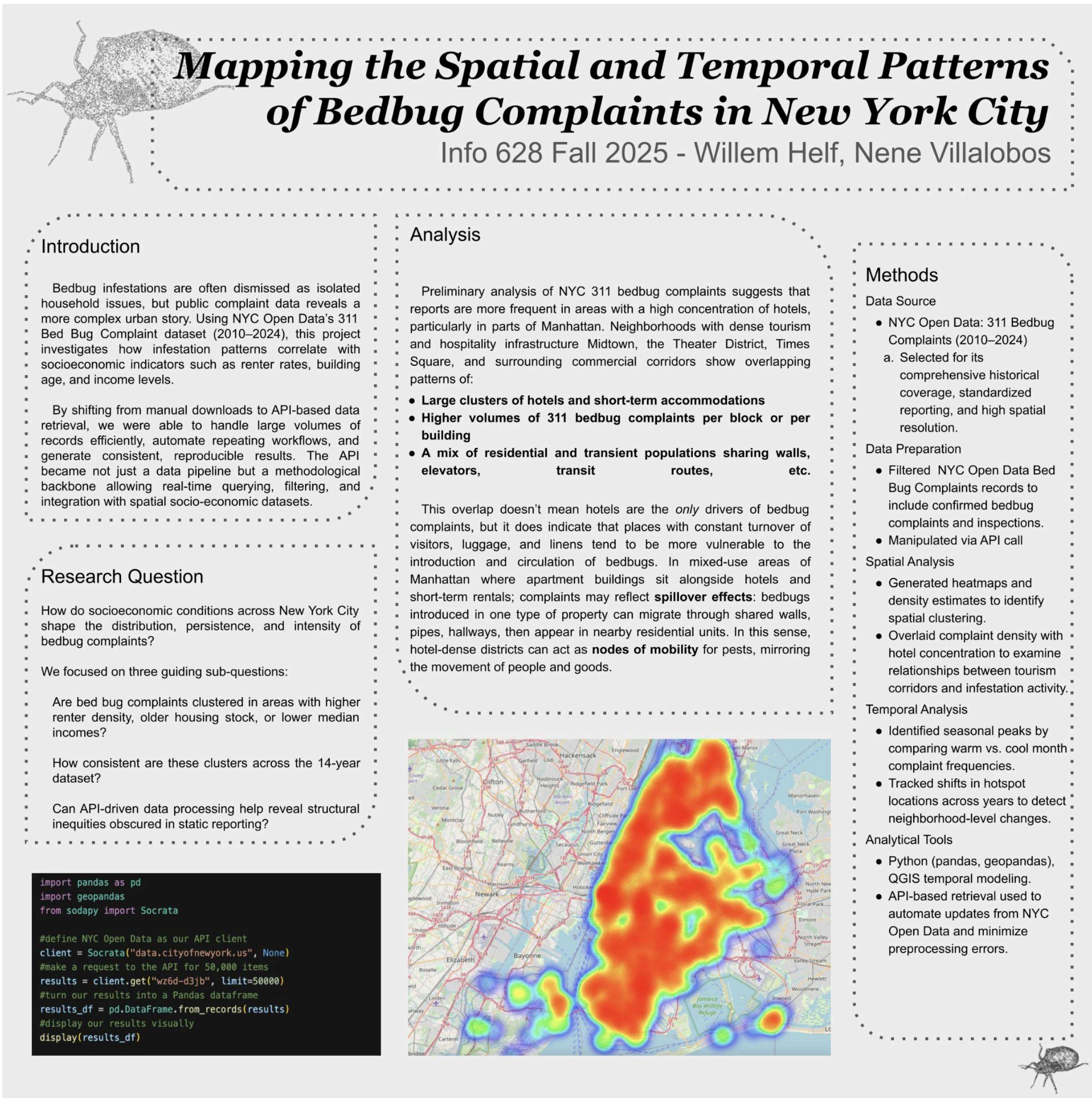
An example of this map and the prototype of the map can be found here:

Version One as a concept: https://nene-villalobos.neocities.org/BedBug/bedbug

Heatmap: https://nene-villalobos.neocities.org/BedBug/Data_Lib

Poster

https://docs.google.com/presentation/d/e/2PACX-1vQemW5QuTXhYLnWWTBV1kkkRhUcsa9b9ixqjVTyEsbz095iUxmpF2-7fcKhVG0q_rPz2l6BhCMhG-Dc/pub?start=false&loop=false&delayms=3000



References

and, Preservation. “Bed Bug Complaints.” *Cityofnewyork.us*, 15 Mar. 2015, data.cityofnewyork.us/Housing-Development/Bed-Bug-Complaints/gih3-4epm/about_data. Accessed 12 Dec. 2025.

Frenzel, Prof. “Dear Fellow Data Scientists,.” *Medium*, 5 Oct. 2024, prof-frenzel.medium.com/kb-api-data-retrieval-194315e0f3c0. Accessed 12 Dec. 2025.

Health, NYC. “Bedbugs - NYC Health.” *Nyc.gov*, 2025, www.nyc.gov/site/doh/health/health-topics/bedbugs.page.

US EPA, OCSPP. “Bed Bugs: A Public Health Issue.” *US EPA*, 10 Apr. 2013, www.epa.gov/bedbugs/bed-bugs-public-health-issue.

Villalobos, Nene. “Neocities.” *Neocities.org*, 2025, neocities.org/site_files/text_editor?filename=BedBug%2FData_Lib.html. Accessed 12 Dec. 2025.

willemhelf. “GitHub - Willemhelf/Bedbug_map.” *GitHub*, 2025, github.com/willemhelf/bedbug_map. Accessed 12 Dec. 2025.