Crowdsourced Ecology: A Visualization Platform for Deer Tick-Borne Disease Surveillance

Sanjana Prasad
Dept. of Computer Science
University of San Francisco
San Francisco, United States of America
sprasad10@dons.usfca.edu

***Abstract*—Data is a critical driver of scientific discovery, yet collecting it remains costly and time-consuming, particularly in the field of ecology. As ecosystems face increasing pressures from climate change and human activity, accessible and reliable data becomes even more essential. In this project, we explore the potential of crowdsourced ecological data by visualizing deer tick sightings reported across the state of New York. We create an interactive, web-based dashboard built with D3.js, and examine spatial and temporal patterns in tick populations to uncover trends and support public health awareness. This paper discusses our design process, challenges, and the implications of using volunteer-submitted data in ecological research.**

I. Introduction

Deer ticks (*Ixodes scapularis*), also known as blacklegged ticks, are small arachnids found commonly in wooded and grassy environments throughout the northeastern and midwestern United States. These ticks are of major concern to public health because they serve as vectors for several tick-borne illnesses. Among the most prevalent are Lyme disease, babesiosis, and anaplasmosis.

A. Key Terms

Lyme disease is caused by the bacterium *Borrelia burgdorferi*, through the bite of infected ticks. It often leads to flu-like symptoms, rashes, and, if untreated, neurological and joint complications. Babesiosis, caused by *Babesia microti*, is a malaria-like parasitic infection that infects red blood cells. Anaplasmosis is caused by the bacterium *Anaplasma phagocytophilum*. Borrelia miyamotoi disease, caused by the spirochete *Borrelia miyamotoi*, is a more recently recognized illness that shares some symptoms with Lyme disease.

B. Motivations

As climate change continues to affect ecological systems, warmer temperatures are believed to contribute to the expansion of tick habitats and prolonged tick activity seasons. This raises concerns about the increasing risk of exposure and disease transmission. Understanding and tracking tick populations and their correlation with disease outbreaks is thus critical for effective public health response and awareness.

However, collecting and analyzing such data is a lengthy and expensive process. In recent years, crowdsourced data collection and open-access databases have emerged as alternatives.

To address the need for accessible and applicable ecological data insights, we created an interactive dashboard designed to allow both the public and researchers to explore crowdsourced deer tick data through applicable visualizations. In this paper, we detail the methods used to build the visualizations and discuss the unique challenges associated with working with crowdsourced datasets.

C. Project Objectives

- Develop an interactive map showing density of deer tick properties by county across New York State.
- Implement line charts to visualize patterns across time and disease prevalence.
- Integrate histograms, scatterplot matrices (SPLOMs) to visualize relationships between tick properties
- Incorporate scroll-based storytelling features to guide users through key insights and observations.
- Allow comparison of disease-specific trends (Lyme disease, babesiosis, anaplasmosis) in relation to tick presence.
- Scrape, clean, and utilize mainly crowdsourced data

We also explore the following questions:

- How are Deer ticks distributed across New York State (NY)?
- How has the tick distribution in NY counties changed from 2008 - 2023?
- Which counties have higher or lower Deer tick populations?
- Do rural/ urban areas areas differ in terms of Deer tick distribution?
- Have tick populations expanded or retreated into new counties or regions from 2008 - 2023?
- What are the behavioral differences between the different species of Deer Ticks?
- Are there certain regions where one bacterium tends to dominate?
- Do the population dynamics of each bacterium species vary differently across different counties over time?
- Is there a relationship between a warming climate and Deer ticks?

- How does climate data correlate with the population dynamics of deer ticks across different regions?
- Are there emerging areas that may see a significant rise in tick populations due to climate change

II. Related Works

A number of prior works served as motivators for this project, spanning ecological data sources, disease research, climate change literature, and data visualization techniques.

The primary data on deer tick populations was sourced from New York State's open data portal [1], which provides comprehensive county-level records on tick presence and species identification. To analyze environmental factors, we incorporated temperature time series data from the National Centers for Environmental Information, alongside ecological data from the New York Flora Atlas [2].

Background research on tick-borne illnesses—including Lyme disease, babesiosis, and anaplasmosis—was conducted through resources published by the Mayo Clinic [3] and Cornell University [4], both of which offer accessible medical and ecological context on tick biology and disease vectors.

Several studies exploring the effects of climate change on tick populations and disease spread helped guide the hypothesis development for our dashboard. In particular, articles from Medical News Today [5] and Time Magazine [6] detailed the growing geographic range of ticks due to warming temperatures and milder winters, highlighting the urgency of continued surveillance.

The structure and interactive design of this project drew direct inspiration from Eric Roston's What's Really Warming the World [7], a seminal example of animated, narrative-driven visualization (also known as "scrollytelling"). This model demonstrated the power of animated visual transitions in guiding readers through complex data relationships. To implement similar features, we referenced the Scrollama.js library [8], which enabled scroll-triggered visual updates within our dashboard.

For the creation of the main visual components—such as choropleth maps, scatterplot matrices, and coordinated views—we relied heavily on the official D3.js documentation and reusable examples from the community [9].

Together, these sources provided a technical and theoretical foundation for developing our interactive visualization platform to explore deer tick distribution and climate-linked trends.

III. Approach

To study the relationship between deer tick populations, environmental variables, and climatical changes in New York, we adopted a modular, data-driven approach that combines Python for data preprocessing with D3.js and React-like components for visual storytelling.

A. Data Collections and Cleaning

We downloaded multiple datasets including: Deer tick population data from New York State Open Data [1], Climate and temperature trends from NOAA/NCEI [11], and Flora information from the New York Flora Atlas [2].

Due to the nature of crowd sourced data, many entries contained missing values. We applied regression imputation, a statistical method that uses relationships between observed variables to estimate missing values, or 0 substituted the values

For data not dependent on time, we deleted values. The decision between deleting or regressing missing values was the hardest and most prevalent issue in this project. Thus is the nature of crowd-sourced data, and we also propose utilizing dashed lines to present uncertain, regressed data along a time-serialized visualization.

B. Backend Architecture

We implemented a component-based structure, borrowing from React's philosophy, to improve maintainability and reuse across visualizations. Most components deal with time-based data, so we abstracted key visualization logic (e.g., filtering, updating scales) to shared utility functions.

- Each visualization component reads from cleaned CSV files.
- A control panel (in control.js) manages user interactions such as scrolling, slider changes, and animations.
- All event listeners (click, scroll, play/pause) are centralized in one file for clean state management.

The choropleth map serves as the central filtering interface. The map:

- Is built using d3.geoAlbers with manual scaling and translation to fit New York State.
- Loads three datasets using Promise.all():

We define a filter() function that selects data based on the year and metric (e.g., Tick Population Density) and applies a color scale (d3.interpolateOranges) dynamically based on the selected metric. Unique filter functions updates all the visualiations colors and adds tooltips on hover.

TABLE I
CHALLENGES AND SOLUTIONS

Challenge	**Solution**
Missing values in datasets	Used regression imputation based on correlated variables; dashed lines show uncertain estimates.
Shared data access by visualizations	Separated filter logic per component to avoid conflicts and ensure consistency.
Tooltip flicker or misalignment	Positioned tooltips using cursor data and added transition smoothing.
Legend mismatch after filtering	Regenerated color scale legend dynamically using `d3.legendColor()`.
Conflicts during animation playback	Introduced an `animating` flag and delay loop to serialize updates.

Fig. 1. Issues with scrollytelling divs created huge whitespace.

C. Frontend Experience

To enhance the user experience, we incorporate easy to use filtering components such as sliders, play/ pause buttons, increment/ decrement buttons, and dynamic legend dropdown menus. Information is presented sequentially like a story to prevent overwhelming the user with the sheer amounts of data. Event listeners in the backend check for the viewport on the user's screen to present the information

D. Results

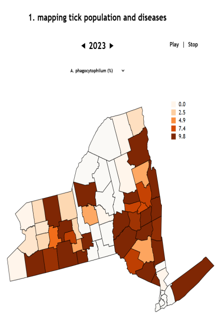

Fig. 2. Choropleth map with animations and tooltips

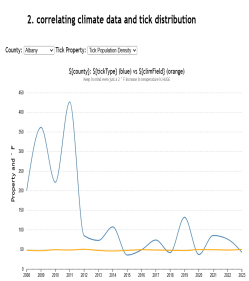

Fig. 3. Time Series Data with climate and tick property data

E. Success

Overall, the project met its objectives, though there were several opportunities for improvement, particularly in streamlining the user experience and handling the challenges of crowdsourced data. These considerations are explored further in the Discussion section.

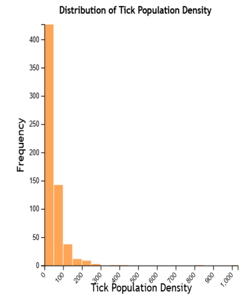

Fig. 4. Histogram of Bacterium Distributions with filtering

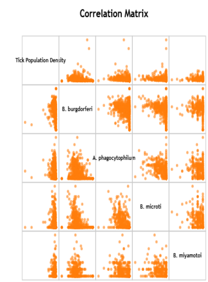

Fig. 5. SPLOM of Bacterium Distributions with filtering

The success of the project was measured by its ability to effectively educate users about the complexities of tick surveillance and the broader implications of climate change on public health. While the visualizations conveyed essential data-driven insights, it was the accompanying narrative brief but targeted text segments that truly reinforced the message. This integration of storytelling and visualization demonstrates the unique power of data journalism. Ultimately, the project's success lies in its ability to turn data into a compelling call to action, highlighting not just patterns, but the urgency behind them.

IV. Discussions

The approach we took combining data visualization with scrollytelling proved to be a promising way to engage users. By pairing journalistic explanation with interactive visuals, we were able to create a compelling narrative that encourages users. I would continue to use this method for similar projects with socially relevant themes.

However, working with crowdsourced data introduced a number of challenges. It often comes with inconsistencies and gaps. Based on this experience, I now better understand both the advantages and limitations of this type of data. With time and improved collection practices by organizations, its reliability will likely increase. That said, I would strongly advise others working with crowdsourced datasets to invest more effort in exploratory data analysis than I initially did. There are numerous statistical techniques for visualizing, understanding, and imputing missing data, and I recognize now that I could have used those methods more rigorously.

If I were to start the project again, I would dedicate more time upfront to data auditing and cleaning. I'd also consider building a more modular data pipeline to handle inconsistencies more efficiently. I find it difficult to imagine organizations using crowdsourced data for real time visualizations unless their backend automates the data analysis process well.

V. Future Work

In the future I would like to implement a better pipeline for visualizing uncertainty. I would also like to explore methods to automate the data cleaning and analysis process to use corwdsourced data in real time.

Acknowledgment

Special thanks to Professor Alark Joshi and my fellow classmates in the computer science department at the University of San Francisco.

References

[1] New York State Open Data, "Deer tick surveillance data," [Online]. Available: https://data.ny.gov

[2] New York Flora Atlas, "Flora distribution data for New York State," [Online]. Available: https://newyork.plantatlas.usf.edu

[3] Mayo Clinic, "Tick species that carry diseases," [Online]. Available: https://www.mayoclinic.org/diseases-conditions/alpha-gal-syndrome/in-depth/tick-species/art-20546861

[4] Cornell University, "New invasive ticks expanding in New York State: Things to watch," [Online]. Available: https://news.cornell.edu/media-relations/tip-sheets/new-invasive-ticks-expanding-new-york-state-things-watch

[5] Medical News Today, "What to know about deer ticks," [Online]. Available: https://www.medicalnewstoday.com/articles/deer-tick

[6] A. Ducharme, "How climate change is expanding tick territories," TIME Magazine, Mar. 2023. [Online]. Available: https://time.com/6262408/climate-change-lyme-disease-ticks/

[7] E. Roston, "What's really warming the world," Bloomberg Graphics, 2015. [Online]. Available: https://www.bloomberg.com/graphics/2015-whats-warming-the-world/

[8] R. Goldenberg, "Scrollama.js," GitHub repository, [Online]. Available: https://github.com/russellgoldenberg/scrollama

[9] D3.js, "D3 documentation and examples," [Online]. Available: https://d3js.org

[10] Stack Overflow contributors, Stack Overflow, [Online].