Key Drivers of Community Garden Growth: Analyzing Trends in PollinateTO Funding Across Toronto's Neighborhoods*

A Bayesian Approach to Understanding the Impact of Time, Type, and Location on Garden Size and Number.

Aliza Abbas Mithwani

November 28, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

Table of contents

*Code and data are available at: [https://github.com/alizamithwani/PollinateTO.git]

1 Introduction

1.1 Overview

Urban biodiversity is increasingly recognized as vital to maintaining healthy ecosystems, especially in the face of rapid urbanization. Among the most important urban biodiversity goals is the protection of pollinators, whose decline poses significant risks to food security, agricultural productivity, and ecological stability. In response to this, many cities have launched initiatives to create pollinator-friendly habitats within urban spaces. Toronto's PollinateTO program is one such initiative, aiming to enhance biodiversity by funding community-led pollinator gardens. Since its launch in 2019, the program has supported over 190 projects, resulting in nearly 500 gardens and approximately 25,500 square meters of pollinator habitat.

Despite the program's success in expanding green spaces, the allocation of funding across these projects remains an understudied area. Questions about which types of gardens or neighborhoods are more likely to receive funding, and the factors influencing these decisions, are crucial for optimizing the program's impact. The objective of this paper is to examine these allocation patterns, specifically whether garden characteristics such as type (e.g., rain gardens, boulevard gardens) or location (e.g., Neighbourhood Improvement Areas) influence the likelihood of receiving funding and the size of the funding allocation.

In particular, this paper addresses the gap in understanding how urban biodiversity funding decisions are made, with a focus on the specific factors that contribute to higher funding likelihoods and amounts. The estimand in this analysis is whether particular garden features, community engagement strategies, or geographic locations are associated with receiving more substantial funding allocations from the PollinateTO program. By identifying these key factors, this research provides important insights into the effectiveness and equity of funding distributions for urban biodiversity projects.

Our findings indicate that garden type, community engagement, and the location of the project all play significant roles in funding decisions. Specifically, we find that gardens in underserved neighborhoods and those emphasizing community involvement are more likely to receive higher funding allocations. These results offer actionable insights for optimizing resource distribution in urban biodiversity programs and aligning ecological goals with social equity considerations.

The remainder of the paper is structured as follows: Section 2 describes the data used in the analysis, including its sources, key features, and considerations regarding measurement and visualization. Section 3 outlines the methodology, detailing the data visualization techniques employed and the justification for using a Bayesian logistic regression model to analyze the relationship between garden characteristics and funding outcomes. Section 4 presents the results, including model validation and key findings. Section 5 offers a discussion of the results, focusing on their implications through key points, followed by an examination of the study's limitations and suggested future steps for further exploration. Section 6 concludes the paper. The Appendix provides supplementary materials, including raw data previews, data feature definitions, data visualizations, and technical model details such as posterior predictive checks, MCMC convergence diagnostics, and additional figures.

2 Data

2.1 Software Packages

This analysis utilizes the R programming language (R Core Team 2023), with the 'rstanarm' (Goodrich et al. 2022) and 'tidyverse' (Wickham et al. 2019) packages for statistical modeling and data manipulation. Our data comes from the PollinateTO Primary Project Garden Locations dataset (Environment & Climate 2024), which is hosted on Toronto's Open Data Portal. Following the approach outlined in Alexander (2023), we leverage these tools to explore patterns and trends in the allocation of PollinateTO funding.

2.2 Data Source

The primary source of data for this study is the PollinateTO Primary Project Garden Locations dataset, which contains detailed information on community-led garden projects funded by PollinateTO. The dataset includes key features such as project locations, garden types, funding years, and other relevant information that enables an in-depth analysis of funding patterns and their spatial and socio-economic implications. This dataset is publicly accessible through

Toronto's Open Data Portal and is a vital resource for studying the impact and effectiveness of PollinateTO in promoting urban biodiversity.

2.3 Data Features

The dataset contains several key variables that are critical for understanding the distribution and characteristics of funded garden projects. Below, we describe each variable and its relevance to the analysis:

- YEAR_FUNDED: The year each project received funding from PollinateTO. This variable is crucial for understanding temporal trends in the program's funding allocation. It also allows us to analyze shifts in funding priorities over time.
- GARDEN_TYPE: This variable categorizes the gardens by their ecological functions (e.g., rain gardens, food gardens, pollinator gardens). These types reflect the intended ecological role of each garden and are important for understanding the program's impact on biodiversity and the environment.
- NIA_OR_EN: This classification indicates whether the project is located in a Neighbourhood Improvement Area (NIA), Emerging Neighbourhood (EN), or another area. These labels are derived from Toronto's planning data and allow for an exploration of whether funding is disproportionately allocated to areas with different socio-economic characteristics, enabling us to analyze equity in the distribution of funds.
- WARD_NAME: The specific ward in Toronto where the garden project is located. This geographic indicator helps us assess whether certain neighborhoods receive more funding based on local government priorities or demographic characteristics.
- IS_INDIGENOUS_GARDEN: A binary variable that identifies whether the project is Indigenous-led, based on self-identification in the funding application. This is a key variable for exploring whether the program is aligned with Toronto's commitments to Indigenous communities and ecological knowledge. PROJECT_NAME: The name assigned to each garden project. While this variable is useful for identification, it is not directly involved in the analysis but helps in linking data points to specific projects.
- LATITUDE and LONGITUDE: These geographic coordinates provide spatial context for each garden project. While these data points help in mapping and visualizing the geographic distribution of funded gardens, their primary role is in understanding spatial patterns and localizing project locations.

These features collectively allow for a broad analysis of trends, equity, and the ecological goals of the PollinateTO initiative.

2.4 Constructed Variables

In this study, variable is_large_garden was constructed. This binary variable was constructed to classify gardens as "large" or "small" based on their estimated garden size. Specifically, gardens with an estimated garden size greater than the mean size of all gardens in the dataset were coded as 1 (large garden), while those with a size less than or equal to the mean were coded as 0 (small garden). The mean garden size was calculated using the estimated_garden_size variable, excluding any missing values. This classification was chosen to capture differences in garden size relative to the overall distribution and to investigate whether larger gardens are more likely to receive funding.

2.5 Data Measurement

The dataset translates real-world community garden projects funded by PollinateTO into structured entries, enabling the analysis of patterns and relationships. It includes several key variables, with the following interpretations based on assumptions about the dataset's construction:

YEAR_FUNDED likely reflects administrative records of the year projects received funding. GARDEN_TYPE seems derived from categorizations in project applications, capturing intended ecological roles. NIA_OR_EN appears to use Toronto city planning designations to classify locations. IS_INDIGENOUS_GARDEN assumes self-identification during the funding application process. WARD_NAME is based on project location within Toronto's municipal wards.

These assumptions are informed by typical practices in grant-based datasets. Confirmation would require consultation with PollinateTO's documentation or administrators.

2.6 Data Consideration

Several considerations affect the dataset's completeness and reliability:

Self-Reported Data: Variables such as GARDEN_TYPE and IS_INDIGENOUS_GARDEN are based on self-reporting by project applicants, which may introduce some subjectivity and inconsistency.

Selection Bias: The dataset reflects funding decisions that may prioritize certain geographic areas, such as NIAs and ENs, which could lead to an overrepresentation of these areas relative to others. This potential bias should be considered when analyzing equity and resource allocation.

Temporal Incompleteness: The dataset covers the period from 2019 to the present, which limits our ability to analyze trends over a longer historical period.

Granularity: Some ecological measures, such as specific plant species or detailed environmental outcomes, are not captured in the dataset. This restricts the scope of ecological impact analysis, though the dataset's focus on garden characteristics and funding allocation provides valuable insights.

Geographic Precision: While LATITUDE and LONGITUDE provide the coordinates for each garden, there may be some inaccuracies in the precise location or boundaries of gardens, particularly in cases where they span multiple properties or are irregularly shaped.

2.7 Methodology

The PollinateTO dataset collects detailed information on community-led garden projects funded by Toronto's PollinateTO initiative. Data is sourced from the city's Open Data portal and includes variables like garden type, location, funding year, and size, enabling analysis of trends in urban biodiversity projects. The dataset uses standardized attributes and is updated annually, making it a valuable tool for evaluating the initiative's impact across various Toronto neighborhoods, though it may be subject to self-reporting biases and geographical limitations in precision.

2.8 Data Visualization

To understand the data provided better, we visualize it using tables and graphs. To begin with, let's see how many gardens each type of neighborhood (NIA, EN or None) has and separate them by whether or not they are indigenous-led in Figure 1 and Figure 2.

Neighborhood Type (NIA or EN)	Total Gardens	Indigenous-Led Gardens	Non-Indigenous Gardens	Proportion Indigenous (%)
EN	6	1	5	16.666667
NIA	32	1	31	3.125000
None	111	3	108	2.702703

Figure 1: Summary of Gardens by Neighborhood Type

Next, we visualize the trend of garden sizes and numbers across time to examine the impact of the initiative over time in Figure 3.

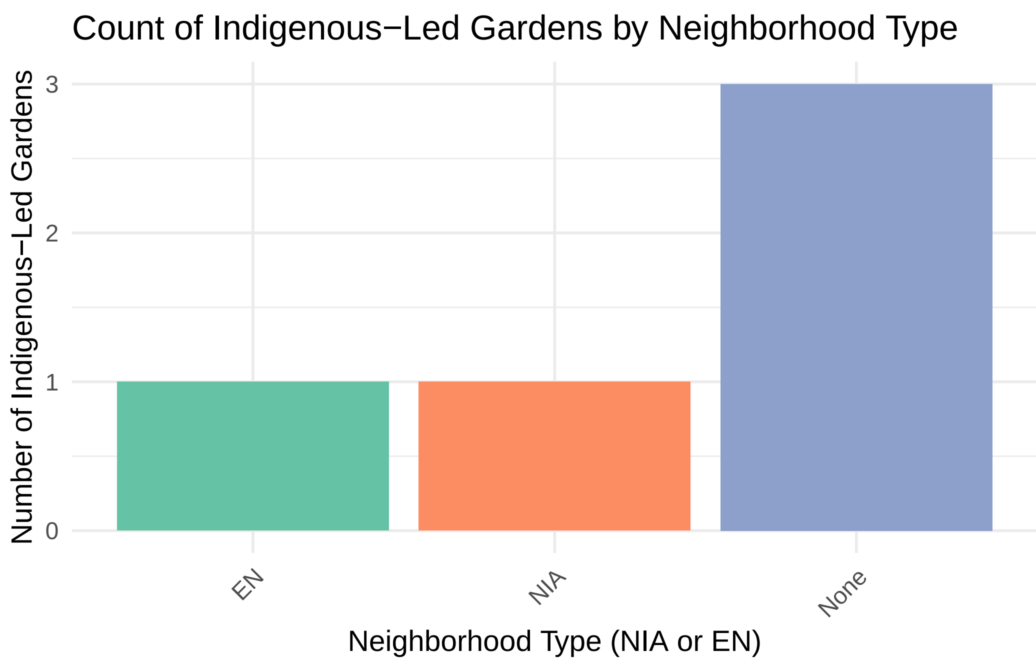

Figure 2: Distribution of Garden Types Funded Over Time

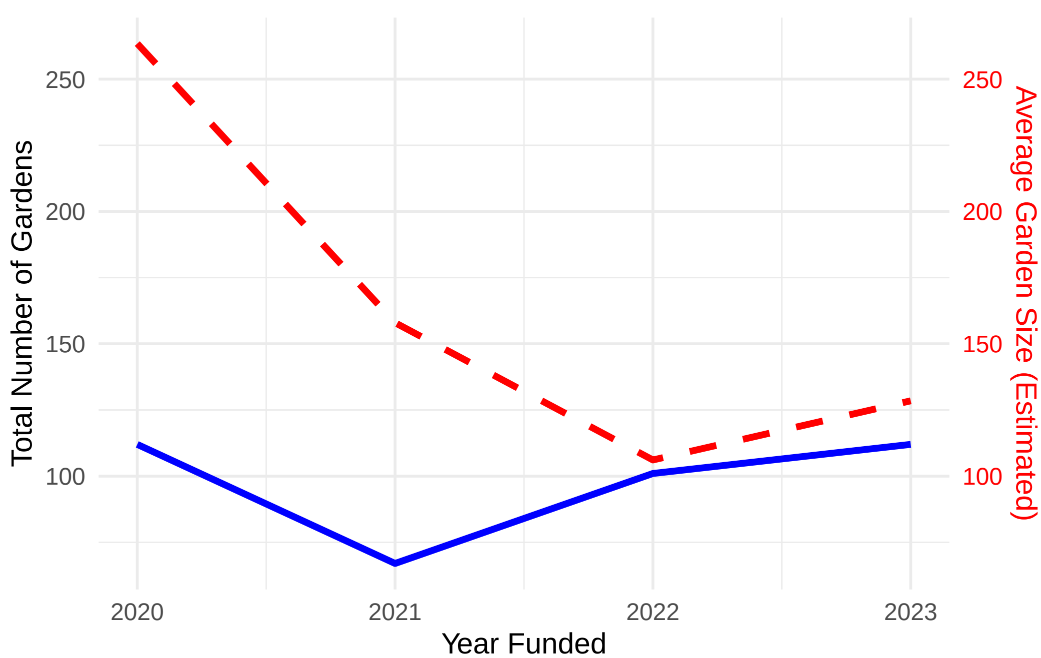

Figure 3: Trend of Garden Size and Number Over Time

Now, let's analyze the proportion of each garden type visually across time in Figure 4 to see if certain garden types increased in number and proportion overtime, reflecting shifting funding priorities and initiatives.

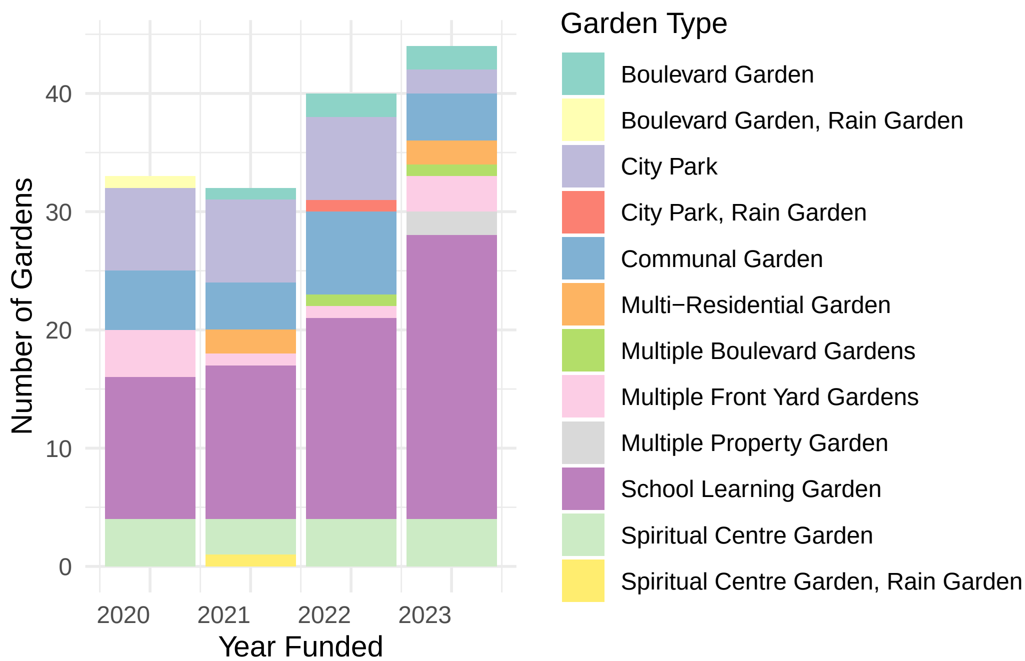

Figure 4: Distribution of Garden Types Funded Over Time

To prevent outliars from making graphs unreadable, we use log-transformed garden sizes to analyze differences in garden sizes based on neighborhood type in Figure 5. A boxplot is used to do this where the 3 neighborhood categories are EN (Emerging Neighborhoods), NIA (Neighborhood Improvement Areas) and None (neither EN nor NIA).

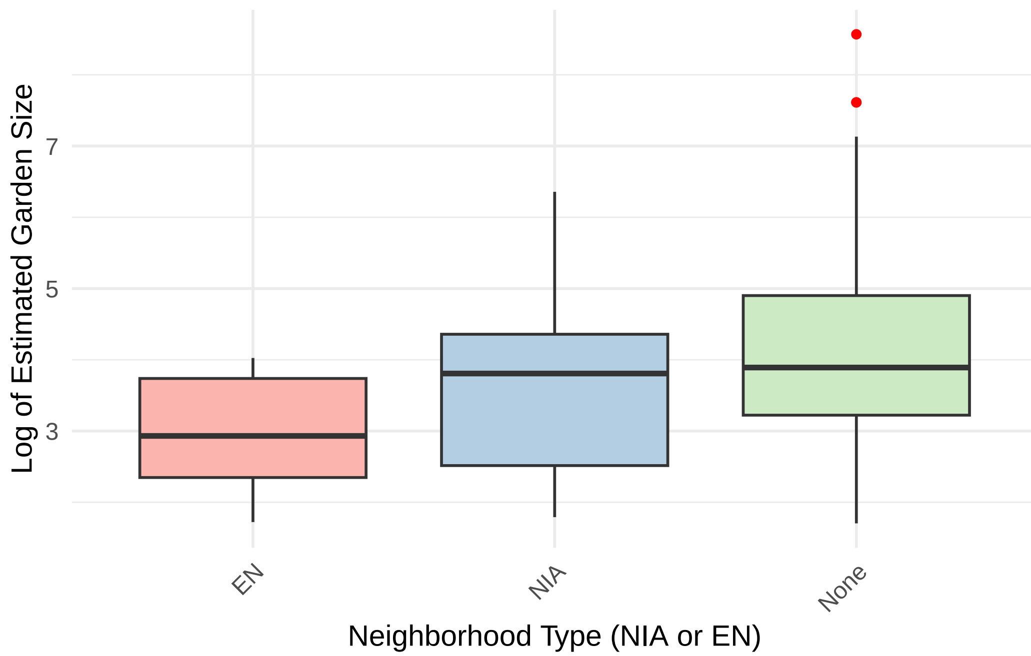

Figure 5: Log-Transformed Garden Sizes by Neighborhood Type

3 Model

In the analysis, I used a Bayesian Logistic Regression Model to examine the relationship between garden size and other factors like the type of garden, the year the project was funded, the neighborhood type (emerging neighborhood, EN vs neighborhood improvement area, NIA), and if it is an indigenous-led project. We are interested in modeling whether a garden is large (binary outcome) as a function of various factors such as the garden type, year funded, neighborhood type, indigenous garden status, and ward name. Background details and diagnostics are included in Appendix B.

3.1 Model set-up

Table 2: Explanatory models of flight time based on wing width and wing length

```
\begin{align}
y_i|\pi_i &\sim \mbox{Bern}(\pi_i) \\
\mbox{logit}(\pi_i) &= \alpha + \beta_1 \times \mbox{garden_type}_i + \beta_2 \times \mbox{y
\alpha &\sim \mbox{Normal}(0, 2.5) \\
\beta_1 &\sim \mbox{Normal}(0, 2.5) \\
\beta_2 &\sim \mbox{Normal}(0, 2.5) \\
\beta_3 &\sim \mbox{Normal}(0, 2.5) \\
\beta_4 &\sim \mbox{Normal}(0, 2.5) \\
\beta_5 &\sim \mbox{Normal}(0, 2.5)
\end{align}
```

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.
We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

3.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular…

We can use maths by including latex between dollar signs, for instance θ.

4 Results

Our results are summarized in Table 2.

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

Please don't use these as sub-heading labels - change them to be what your point actually is.

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

B.1 Posterior predictive check

In **?@fig-ppcheckandposteriorvsprior-1** we implement a posterior predictive check. This shows…

In **?@fig-ppcheckandposteriorvsprior-2** we compare the posterior with the prior. This shows…

Examining how the model fits, and is affected by, the data

B.2 Diagnostics

?@fig-stanareyouokay-1 is a trace plot. It shows… This suggests…

?@fig-stanareyouokay-2 is a Rhat plot. It shows… This suggests…

Checking the convergence of the MCMC algorithm

References

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

Environment & Climate. 2024. *PollinateTO Primary Project Garden Locations*. https://open.toronto.ca/dataset/pollinateto-primary-project-garden-locations/.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "rstanarm: Bayesian applied regression modeling via Stan." https://mc-stan.org/rstanarm/.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.