Key Drivers of Community Garden Growth: Analyzing Trends in PollinateTO Funding*

A Bayesian Approach to Understanding the Impact of Time, Type, and Location on Garden Size and Number in Toronto's Neighborhoods from 2020 to 2023.

Aliza Abbas Mithwani

December 3, 2024

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

Table of contents

---

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 the distribution of Indigenous gardens in each type of neighborhood (NIA, EN or None) in Table 1 and Figure 1. As seen in Table 1, Indigenous gardens take up a higher proportion of all gardens located in Emerging Neighborhoods; however, there are significantly less Emerging Neighborhood gardens than Neighborhood Improvement Area gardens. Figure 1 helps us realize that overall, there aren't a lot of Indigenous(-led) gardens in this initiative. There is only 1 in an Emerging Neighborhood, 1 in a Neighborhood Improvement Area and 3 in other neighborhoods (not EN or NIA).

Table 1: Proportion of Indigenous Gardens in Each Neighborhood Type

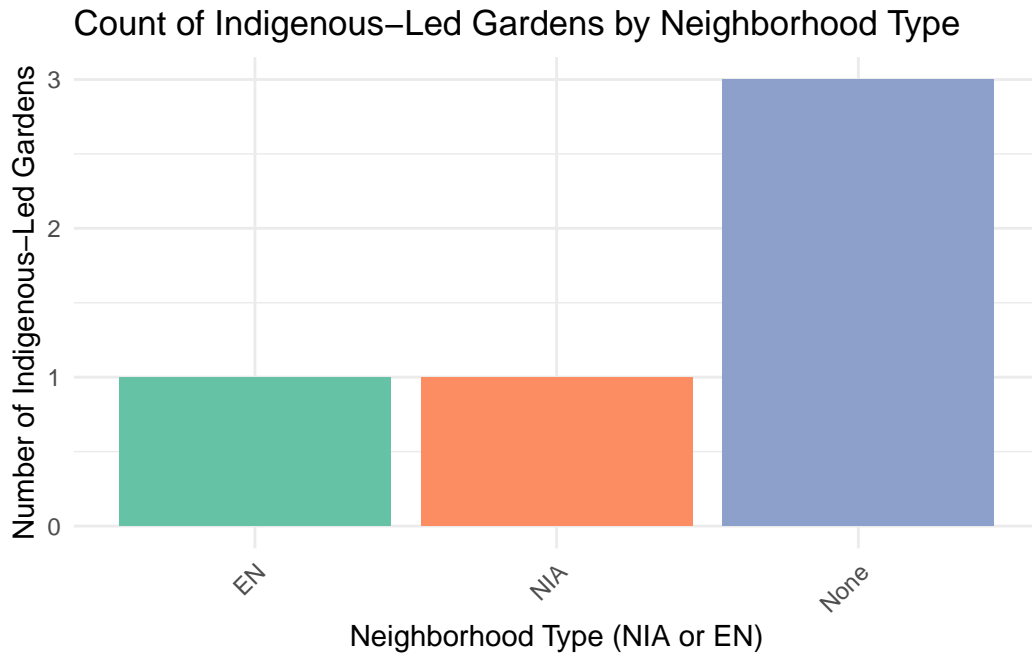| Neighborhood Type (NIA or EN)	Total Gardens	Indigenous-Led Gardens	Non-Indigenous Gardens	Proportion Indigenous (%)
EN	6	1	5	16.67
NIA	32	1	31	3.12
None	111	3	108	2.70

Figure 1: Number of Indigenous Gardens Funded Across Neighborhoods

Next, we visualize the trend of garden sizes and numbers across time to examine the impact of the initiative over time in Figure 2. The total number of new gardens funded by PollinateTO decreased significantly from 2020 to 2021, but then increased over 2021 to 2023 by a number greater than the initial drop. Estimated Average Garden Size has been decreasing significantly from 2020 to 2022, but recently increase by a small amount from 2022 to 2023.
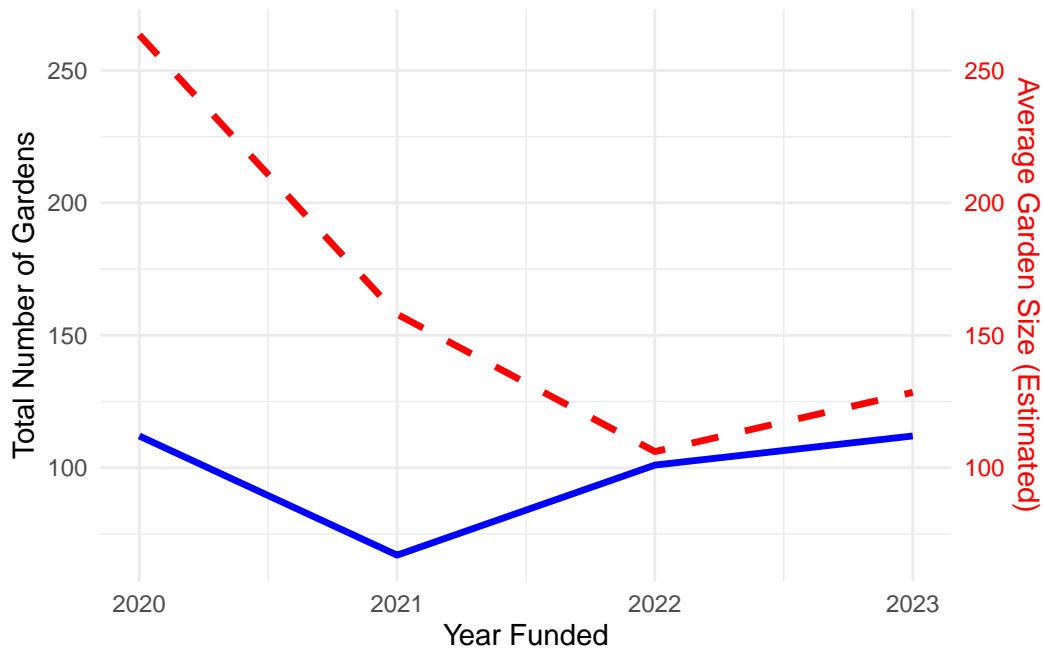
Figure 2: Trend of Garden Size and Number Over Time

Now, let's analyze the proportion of each garden type across time in Figure 3 to see if certain garden types increased in proportion overtime, reflecting shifting funding priorities and initiatives. School Learning Gardens were consistently the most abundant type of garden compared to the rest in each year, and grew in proportion consistently from 2020 to 2023. City Park Gardens used to be the second most abundant type of garden in 2020, 2021, and 2022, but they recently fell significantly in proportion and were overtaken by Multiple Front Yard Gardens.

Figure 3: Distribution of Garden Types Funded Over Time

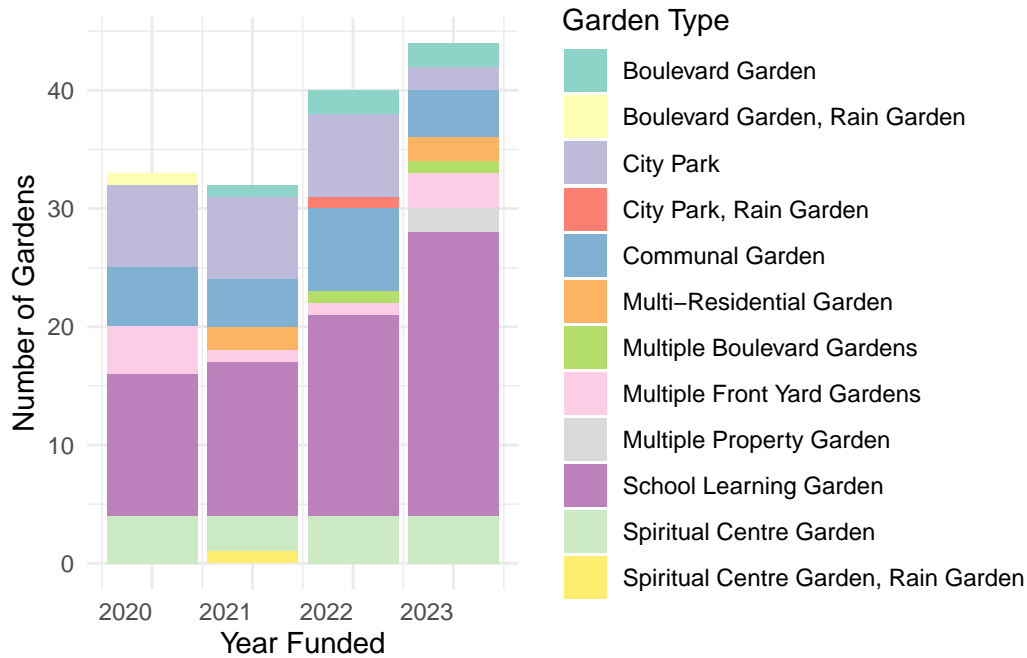To prevent outliars from making graphs unreadable, we use log-transformed garden sizes to analyze differences in garden sizes based on neighborhood type in Figure 4. A dot plot is used to do this where the 3 neighborhood categories are EN (Emerging Neighborhoods), NIA (Neighborhood Improvement Areas) and None (neither EN nor NIA).
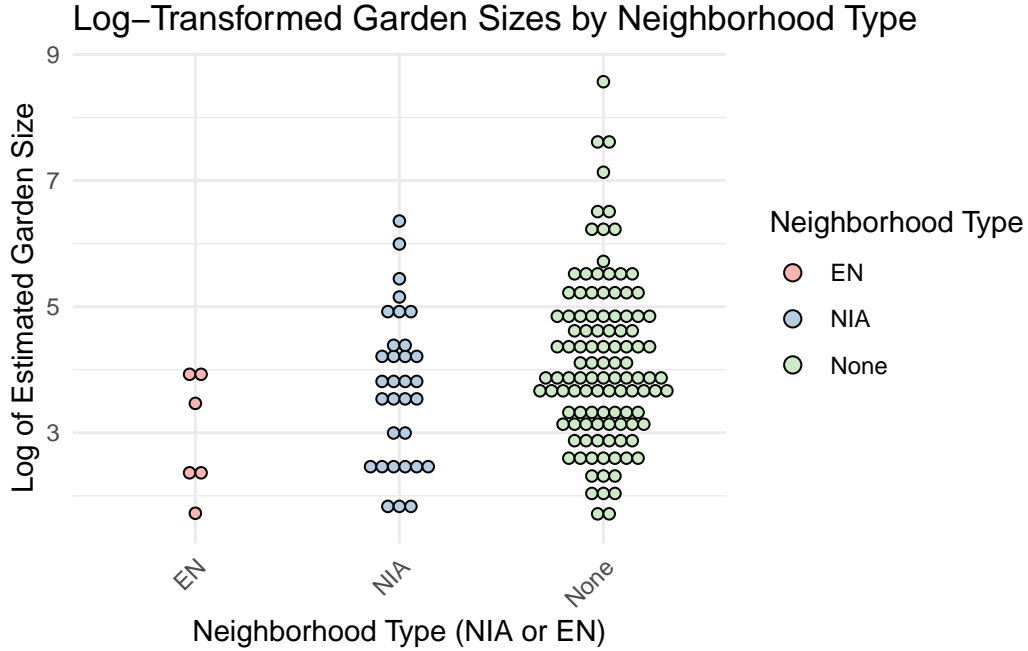
Figure 4: Log-Transformed Garden Sizes by Neighborhood Type

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

$$y_i | \pi_i \sim \text{Bern}(\pi_i) \tag{1}$$

$$\begin{aligned}
\text{logit}(\pi_i) = {} & \alpha + \beta_1 \times \text{garden type}_i + \beta_2 \times \text{year funded}_i \\
& + \beta_3 \times \text{nia or en}_i + \beta_4 \times \text{is indigenous garden}_i + \beta_5 \times \text{ward name}_i \\
& + \beta_6 \times (\text{garden type}_i \times \text{nia or en}_i) \tag{2}
\end{aligned}$$

$$\alpha \sim \text{Normal}(0, 2.5) \tag{3}$$

$$\beta_1 \sim \text{Normal}(0, 2.5) \tag{4}$$

$$\beta_2 \sim \text{Normal}(0, 2.5) \tag{5}$$

$$\beta_3 \sim \text{Normal}(0, 2.5) \tag{6}$$

$$\beta_4 \sim \text{Normal}(0, 2.5) \tag{7}$$

$$\beta_5 \sim \text{Normal}(0, 2.5) \tag{8}$$

This Bayesian Logistic Regression model predicts whether a garden is classified as "large" (y_i), a binary outcome modeled as $y_i \sim \text{Bern}(\pi_i)$, where π_i is the probability of the garden being large. The predictors include: garden type, a categorical variable indicating the type of garden (e.g., pollinator, vegetable); year funded, a continuous variable representing the year the garden received funding; nia or en, a categorical variable indicating whether the garden is in a Neighborhood Improvement Area (NIA) or elsewhere (EN); is indigenous garden, a binary variable identifying whether the garden is indigenous-led; and ward name, a categorical variable representing the geographic ward. Additionally, this model includes an interaction term between garden type and nia or en, which helps examine whether the effect of garden type on the likelihood of being large differs depending on whether the garden is in an NIA or EN neighborhood. These predictors influence π_i through the logit link function: $\text{logit}(\pi_i) = \alpha + \beta_1 \times \text{garden type}_i + \beta_2 \times \text{year funded}_i + \beta_3 \times \text{nia or en}_i + \beta_4 \times \text{is indigenous garden}_i + \beta_5 \times \text{ward name}_i + \beta_6 \times (\text{garden type}_i \times \text{nia or en}_i)$, where α is the intercept, β_k are the coefficients, and β_6 represents the interaction term between garden type and nia or en. Priors for α and β_k are defined as Normal(0,2.5), reflecting weakly informative assumptions. We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022).

3.1.1 Model justification

The Bayesian logistic regression model employed in this analysis is appropriate for investigating the factors influencing the likelihood of a garden being classified as large based on binary outcomes. Specifically, this model explores how various predictors—such as garden type, year funded, whether the garden is indigenous-led, the neighborhood classification (NIA or EN), and ward names—affect garden size. Given the binary nature of the outcome variable (is_large_garden), logistic regression is the suitable choice as it models probabilities between 0 and 1. The inclusion of an interaction term between garden type and nia or en further

enriches the analysis, allowing us to examine whether the effect of garden type on garden size differs by neighborhood type. This is especially relevant as different neighborhoods may have different conditions and priorities that could influence the type and size of gardens.

The Bayesian framework provides several advantages, including the ability to incorporate prior beliefs about the parameters, produce credible intervals, and assess model uncertainty. Furthermore, Bayesian methods enable a more robust assessment of convergence through posterior predictive checks and the examination of diagnostic tools such as trace and Rhat plots, ensuring reliable inferences.

The choice of the Bayesian approach in this context also allows for the flexible inclusion of multiple categorical predictors, which is crucial for this study, as it aims to assess the relationship between categorical factors like garden type, ward names, and the interaction between garden type and nia or en with a quantitative outcome. The model's use of informative priors helps regularize the estimates, particularly when dealing with categorical predictors that may have complex relationships with the outcome variable. Additionally, the Bayesian methodology provides the opportunity to assess model diagnostics comprehensively, ensuring that the model fits well with the data and that the parameters are estimated with high precision. Overall, this model is well-suited to capture the complexities of the data while offering interpretability and flexibility through its probabilistic nature.

4 Results

Our results are summarized in Table 2. The Bayesian logistic regression model reveals several important relationships between garden characteristics and the likelihood of a garden being classified as large. Among garden types, "Multiple Property Garden" stands out with a substantial positive coefficient (14.94), suggesting that such gardens are strongly associated with larger sizes. Conversely, "Boulevard Garden, Rain Garden" exhibits a notable negative coefficient (-13.77), indicating a reduced likelihood of being large. Other garden types, such as "Communal Garden" and "School Learning Garden," have smaller coefficients, implying a weaker or negligible relationship with garden size. This variability in the influence of garden type highlights the importance of understanding project-specific features when predicting garden outcomes.

The neighborhood context also plays a significant role. Gardens located in areas classified as "None" (neither EN nor NIA) have a relatively high positive coefficient (4.54), suggesting that these gardens are more likely to be larger compared to those in ENs or NIAs. Additionally, Indigenous-led gardens are associated with a negative coefficient (-8.86), suggesting they may be less likely to be large. The interaction terms, such as those between "Multi-Residential Garden" and neighborhood type, indicate that specific combinations of garden type and neighborhood characteristics can further modify the likelihood of a garden being large. These findings underscore the complexity of garden size determinants and suggest targeted strategies may be needed to ensure equitable garden funding and development.

Table 2: Explanatory model of whether a garden is large based on certain factors: Indigenous-led indicator, Year Funded, Neighborhood, Garden type and Ward Names

[!h]

part	term	statistic	Model
estimates	(Intercept)	estimate	22.35
estimates	(Intercept)	mad	(591.06)
estimates	garden_typeBoulevard Garden, Rain Garden	estimate	-13.77
estimates	garden_typeBoulevard Garden, Rain Garden	mad	(25.17)
estimates	garden_typeCity Park	estimate	-0.81
estimates	garden_typeCity Park	mad	(4.69)
estimates	garden_typeCity Park, Rain Garden	estimate	-8.11
estimates	garden_typeCity Park, Rain Garden	mad	(25.56)
estimates	garden_typeCommunal Garden	estimate	0.10
estimates	garden_typeCommunal Garden	mad	(5.23)
estimates	garden_typeMulti-Residential Garden	estimate	-2.63
estimates	garden_typeMulti-Residential Garden	mad	(11.59)
estimates	garden_typeMultiple Boulevard Gardens	estimate	1.48
estimates	garden_typeMultiple Boulevard Gardens	mad	(15.19)
estimates	garden_typeMultiple Front Yard Gardens	estimate	-0.68
estimates	garden_typeMultiple Front Yard Gardens	mad	(7.33)
estimates	garden_typeMultiple Property Garden	estimate	14.94
estimates	garden_typeMultiple Property Garden	mad	(18.86)
estimates	garden_typeSchool Learning Garden	estimate	0.66
estimates	garden_typeSchool Learning Garden	mad	(3.35)
estimates	garden_typeSpiritual Centre Garden	estimate	-5.78
estimates	garden_typeSpiritual Centre Garden	mad	(4.79)
estimates	garden_typeSpiritual Centre Garden, Rain Garden	estimate	-7.32
estimates	garden_typeSpiritual Centre Garden, Rain Garden	mad	(25.90)
estimates	nia_or_enNIA	estimate	1.28
estimates	nia_or_enNIA	mad	(4.86)
estimates	nia_or_enNone	estimate	4.54
estimates	nia_or_enNone	mad	(4.09)
estimates	year_funded	estimate	-0.02
estimates	year_funded	mad	(0.29)
estimates	is_indigenous_garden1	estimate	-8.86
estimates	is_indigenous_garden1	mad	(7.25)
estimates	ward_nameDavenport	estimate	5.16
estimates	ward_nameDavenport	mad	(2.03)
estimates	ward_nameDon Valley East	estimate	4.11
estimates	ward_nameDon Valley East	mad	(2.31)
estimates	ward_nameDon Valley North	estimate	-14.77
estimates	ward_nameDon Valley North	mad	(18.12)
estimates	ward_nameDon Valley West	estimate	8.25
estimates	ward_nameDon Valley West	mad	(2.58)
estimates	ward_nameEglinton-Lawrence	estimate	12.12
estimates	ward_nameEglinton-Lawrence	mad	(3.87)
estimates	ward_nameEtobicoke Centre	estimate	-5.78
estimates	ward_nameEtobicoke Centre	mad	(8.61)
estimates	ward_nameEtobicoke North	estimate	-6.98
estimates	ward_nameEtobicoke North	mad	(10.12)
estimates	ward_nameEtobicoke-Lakeshore	estimate	3.87
estimates	ward_nameEtobicoke-Lakeshore	mad	(2.29)
estimates	ward_nameEtobicoke-North	estimate	-9.34
estimates	ward_nameEtobicoke-North	mad	(12.31)
estimates	ward_nameHumber River-Black Creek	estimate	-4.28
estimates	ward_nameHumber River-Black Creek	mad	(6.96)
estimates	ward_nameParkdale-High Park	estimate	4.44
estimates	ward_nameParkdale-High Park	mad	(1.97)

Figure 5 visualizes the estimated effects of various predictors on the likelihood of a garden being classified as "large" in the PollinateTO initiative. Each point represents the coefficient estimate for a predictor, indicating the direction and strength of its relationship with garden size. Predictors with positive coefficients increase the probability of a garden being large, while those with negative coefficients decrease it. The horizontal bars show the 95% credible intervals, capturing the uncertainty in the estimates. Predictors whose credible intervals do not cross the dashed red line at zero are likely to have significant effects. This plot highlights which variables, such as garden type, neighborhood classification, or year funded, are most influential in determining garden size, providing valuable insights into the factors driving outcomes in the program.

Figure 5: Coefficients of the Model

15

4.1 Predictive Probability

The bar chart in Figure 6 visualizes the predicted probability that a garden funded by PollinateTO will be classified as "large," based on the combination of garden type (garden_type) and neighborhood classification (nia_or_en). The height of each bar represents the average predicted probability for each garden type and neighborhood combination, with all other variables (such as year funded, Indigenous-led status, and ward) held constant. The chart helps identify which garden types and neighborhood categories are most likely to result in larger gardens. By examining the differences in predicted probabilities across the garden types and neighborhood types, we can discern patterns in how these factors influence garden size. This analysis can inform decisions about resource allocation, funding strategies, or identifying areas where specific types of gardens are more likely to grow larger, contributing to the broader goals of the PollinateTO program.
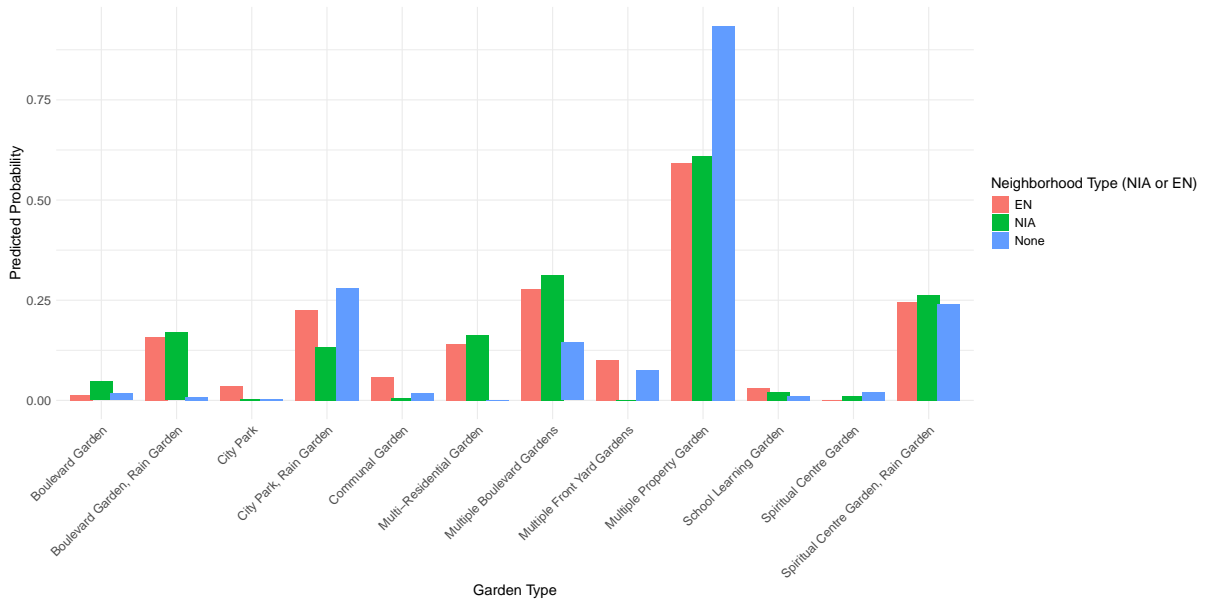

Figure 6: Predictive Probability of Large Gardens

5 Discussion

5.1 Unearthing Patterns in Garden Allocation

This paper investigates the factors influencing the size of gardens funded under Toronto's PollinateTO initiative using a Bayesian logistic regression model. Specifically, it explores how garden size (a binary outcome of "large" vs. "not large") is associated with variables such as garden type, year funded, neighborhood classification (Emerging Neighborhoods or Neighborhood Improvement Areas), Indigenous-led status, and ward-level geographic differences. By leveraging this model, the analysis identifies patterns and disparities in garden allocation, providing insights into which factors are most predictive of larger gardens. The approach incorporates interaction effects to capture nuanced relationships between garden types and neighborhood characteristics.

The inclusion of credible intervals from the Bayesian model ensures a robust interpretation of uncertainty in coefficient estimates. Additionally, key goodness-of-fit measures (e.g., R^2, WAIC) are evaluated, demonstrating the model's ability to explain nearly half of the variation in garden size. These findings offer a data-driven lens through which urban sustainability programs like PollinateTO can be assessed and improved.

5.2 Neighborhood Inequities in Urban Gardening

One of the central insights from this analysis is the existence of disparities in garden size across neighborhoods. Specifically, gardens in areas classified as "None" (neither EN nor NIA) are significantly more likely to be large, while gardens in NIAs face challenges in achieving similar sizes. This suggests that systemic factors, such as socioeconomic inequalities, land availability, and access to resources, influence the outcomes of urban greening initiatives.

This insight highlights a broader inequity in urban sustainability programs: neighborhoods with fewer resources or systemic disadvantages may struggle to achieve comparable environmental benefits. This finding emphasizes the importance of embedding equity-focused considerations into grant allocation processes. Addressing these disparities can improve not only the environmental impact of such initiatives but also their social relevance and inclusivity.

5.3 Redefining Success for Indigenous-Led Projects

Indigenous-led projects are less likely to result in large gardens, pointing to potential barriers that Indigenous communities face in participating fully in urban greening initiatives. This finding underscores the need to reevaluate the criteria for "success" in such programs. Indigenous-led projects may prioritize cultural and ecological goals—such as the preservation of native plant species or community education—over garden size.

This raises an important consideration for future urban sustainability programs: qualitative impacts, such as community engagement or cultural significance, may be equally (if not more) important than quantitative measures like garden size. By failing to account for these dimensions, existing frameworks risk marginalizing projects that offer critical social and cultural benefits. This insight encourages a shift in how success is defined and measured in urban greening efforts.

6 Weaknesses and next steps

I will address some of the weaknesses and limitations of the current analysis, focusing on areas where the methodology or data could be improved. Following this, I will explore potential next steps, outlining opportunities for future research and policy development that build on these findings to create more equitable and effective urban sustainability initiatives. This section aims to critically reflect on the study while charting a path forward for both academic inquiry and practical applications.

6.1 Challenges and Limitations in Modeling Garden Outcomes

While the model provides valuable insights, it is not without limitations. First, the binary categorization of garden size as "large" vs. "not large" may oversimplify the complexity of garden outcomes. A more nuanced approach, such as modeling size as a continuous variable or including qualitative measures of success, might provide richer insights.

Second, the dataset is limited to gardens funded by PollinateTO, which may not fully represent the broader urban gardening landscape in Toronto. Selection bias could influence the generalizability of these findings to other urban greening initiatives. Additionally, the potential for omitted variable bias exists, as factors like community engagement levels or prior gardening experience were not included in the analysis but may significantly impact garden outcomes.

Finally, while interaction terms capture some complexity, they add interpretive challenges, as not all combinations of garden types and neighborhoods had sufficient data to provide precise estimates. Future work could address this by gathering larger, more diverse datasets or employing hierarchical models to account for varying group sizes.

6.2 Opportunities for Future Research and Policy Improvements

Several open questions remain. One key area for future research is the inclusion of community-level variables, such as demographic characteristics, access to volunteer networks, and proximity to green spaces. These factors may help explain why certain neighborhoods are more likely to host large gardens.

Additionally, further work is needed to explore the qualitative outcomes of urban gardens, particularly for Indigenous-led projects. Incorporating interviews, case studies, or community feedback could provide a deeper understanding of how these gardens contribute to cultural preservation and community well-being.

From a methodological perspective, future studies could explore alternative modeling techniques, such as hierarchical Bayesian models, to better account for ward-level variability and cross-neighborhood interactions. Expanding the dataset to include other urban greening initiatives or longitudinal data could also provide insights into how garden outcomes evolve over time.

Lastly, there is a need to test the effectiveness of policy interventions informed by these findings. For instance, targeted funding for NIAs or revised evaluation criteria for Indigenous-led projects could be piloted and evaluated to determine their impact on garden size and community engagement. This iterative process of evidence-based policy design and evaluation can help create more equitable and effective urban sustainability programs.

Appendix

A Additional data details

## A.1 Data Cleaning

A.2 Data Simulation

B Model details

B.1 Posterior predictive check

In Figure 7 we implement a posterior predictive check. The posterior predictive check (pp_check) helps assess how well the model fits the observed data in the context of your research on the factors influencing community garden funding by PollinateTO across different Toronto neighborhoods. By comparing the model's predictions (posterior distribution) with the observed data, this diagnostic reveals whether the model is accurately capturing the key influences, such as garden type, funding year, and neighborhood characteristics. A good fit suggests that these factors are well-represented in the model, while significant deviations between the predictions and the data may indicate that the model needs refinement. This check is essential for validating the model's effectiveness in answering your research question.
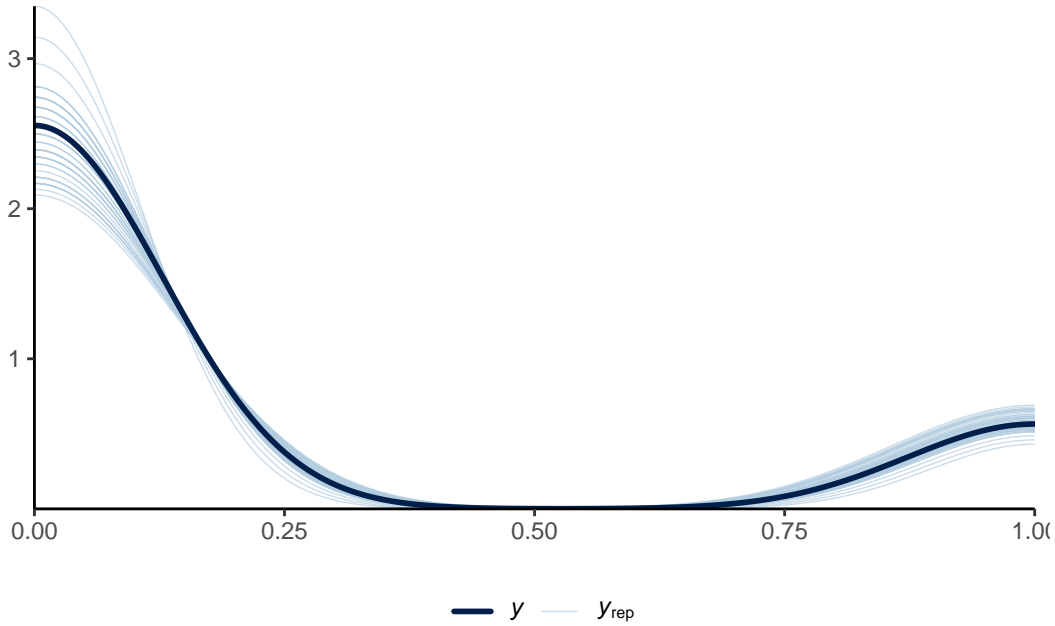
(a) Posterior prediction check

Figure 7: Examining how the model fits, and is affected by, the data

In Figure 8 we compare the posterior with the prior. The posterior vs. prior comparison visualizes how the model's posterior estimates for various parameters compare to the prior distributions. In the context of your research on the factors influencing the size and number of community gardens funded by PollinateTO, this comparison helps evaluate whether the data has significantly updated the prior beliefs embedded in the model. A well-calibrated model should show a substantial difference between the prior and posterior for key parameters, indicating that the observed data has influenced the model's estimates. This comparison is crucial for understanding which variables are most influential in determining garden size and funding across neighborhoods.

Given the large number of parameters in the model, a random sample of 5 garden types and 5 ward names was taken for this visualization. This reduces complexity and ensures the comparison remains interpretable, focusing on a manageable subset of the factors that may affect the size and number of community gardens across Toronto.

B.2 Markov Chain Monte Carlo Convergence Check

Figure 9 displays trace plots of the model. Trace plots are essential diagnostics in Bayesian logistic regression models, like the one used in your research examining predictors of whether a garden is larger than average, including garden type, indigenous-led status, funding year, and location. These plots display the sampled parameter values across MCMC iterations, helping
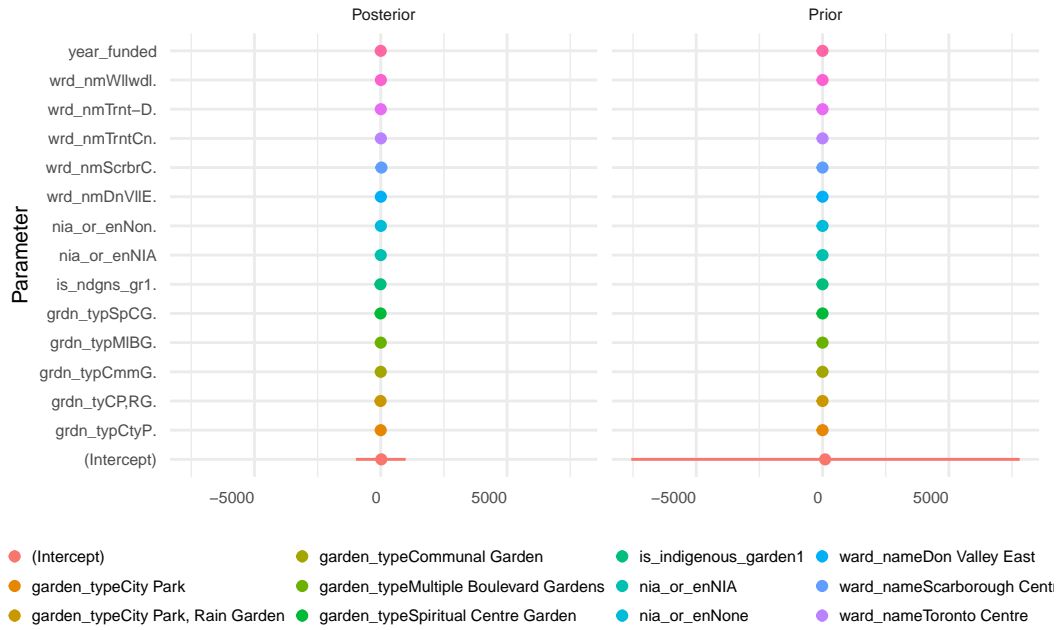
Figure 8: Compare the posterior with the prior

assess convergence and mixing. A well-behaved trace plot resembles a "fuzzy caterpillar," indicating that the chain has stabilized and is sampling consistently from the posterior distribution, ensuring reliable estimates. This matches the description of what we observe in our trace plots. Non-convergence, shown by trends or failure to stabilize, may result from poor model specification or insufficient warmup. Good mixing is characterized by rapid oscillations and variability, reflecting effective exploration of the parameter space, whereas slow mixing or long stretches of similar values may indicate inefficiency or high autocorrelation. Proper convergence and mixing are critical to draw robust inferences about how factors like indigenous-led status or funding year influence garden size, ensuring the validity of your findings. This suggests...

Figure 10 is a Rhat plot. Rhat plots are critical diagnostics for assessing convergence in Bayesian logistic regression models like this one. The Rhat statistic, or Gelman-Rubin diagnostic, compares within-chain and between-chain variance to determine if the Markov chains have converged to the posterior distribution. Ideally, Rhat values should be close to 1 (e.g., below 1.01 or 1.05), indicating convergence. This matches the description of what we observe in our Rhat plot. An Rhat plot visually summarizes these values, with deviations above the threshold signaling potential non-convergence and the need for more iterations, better model specification, or improved priors. By ensuring all parameters in the model have Rhat values near 1, the plot provides confidence in the stability and reliability of the posterior estimates. This is crucial for interpreting how predictors like indigenous-led status and funding year influence garden size accurately.
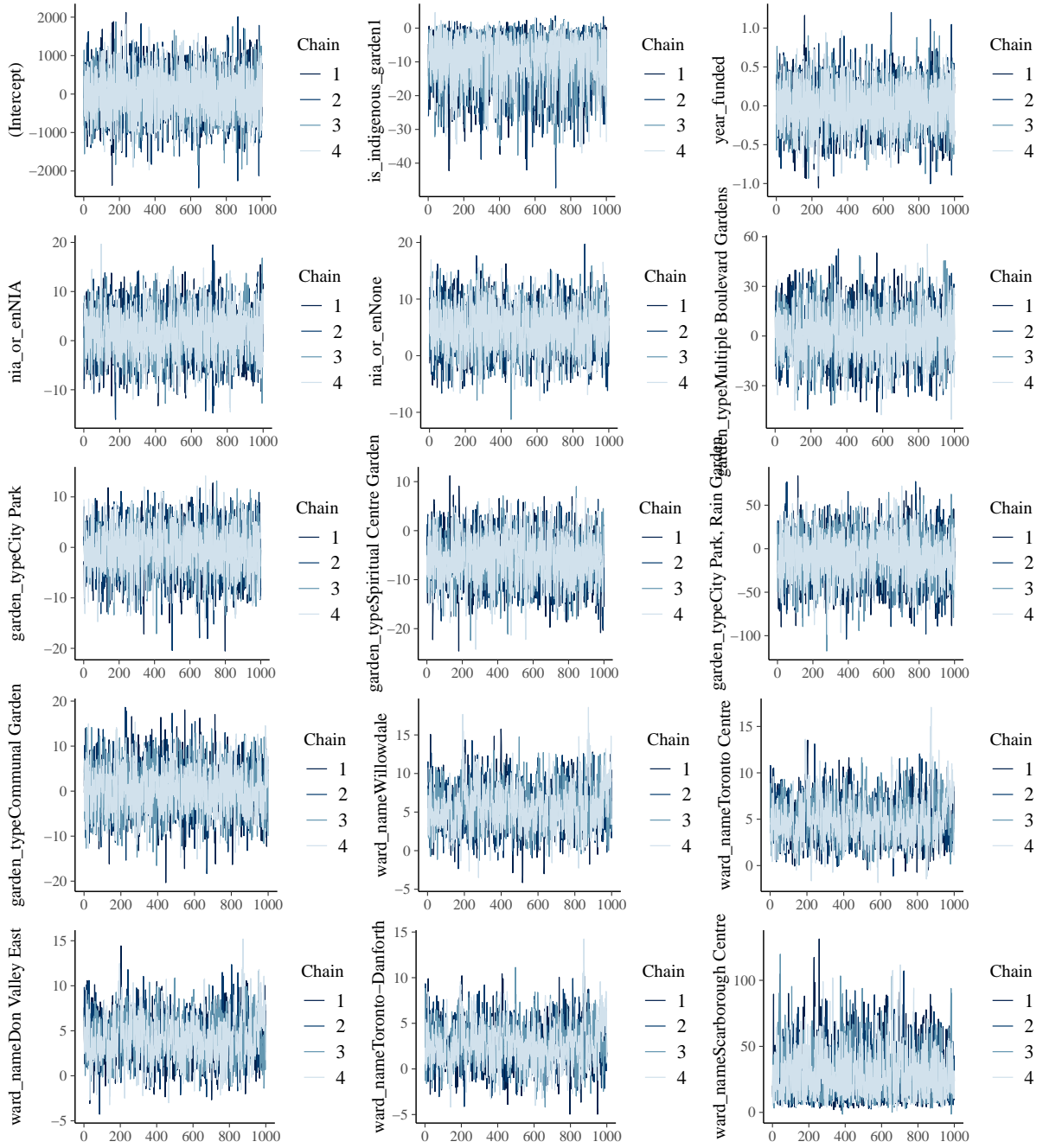
(a) Trace plot of Intercept, Indigenous-led indicator, Year Funded, Neighborhood, Garden type (Sample) and Ward Names (Sample)

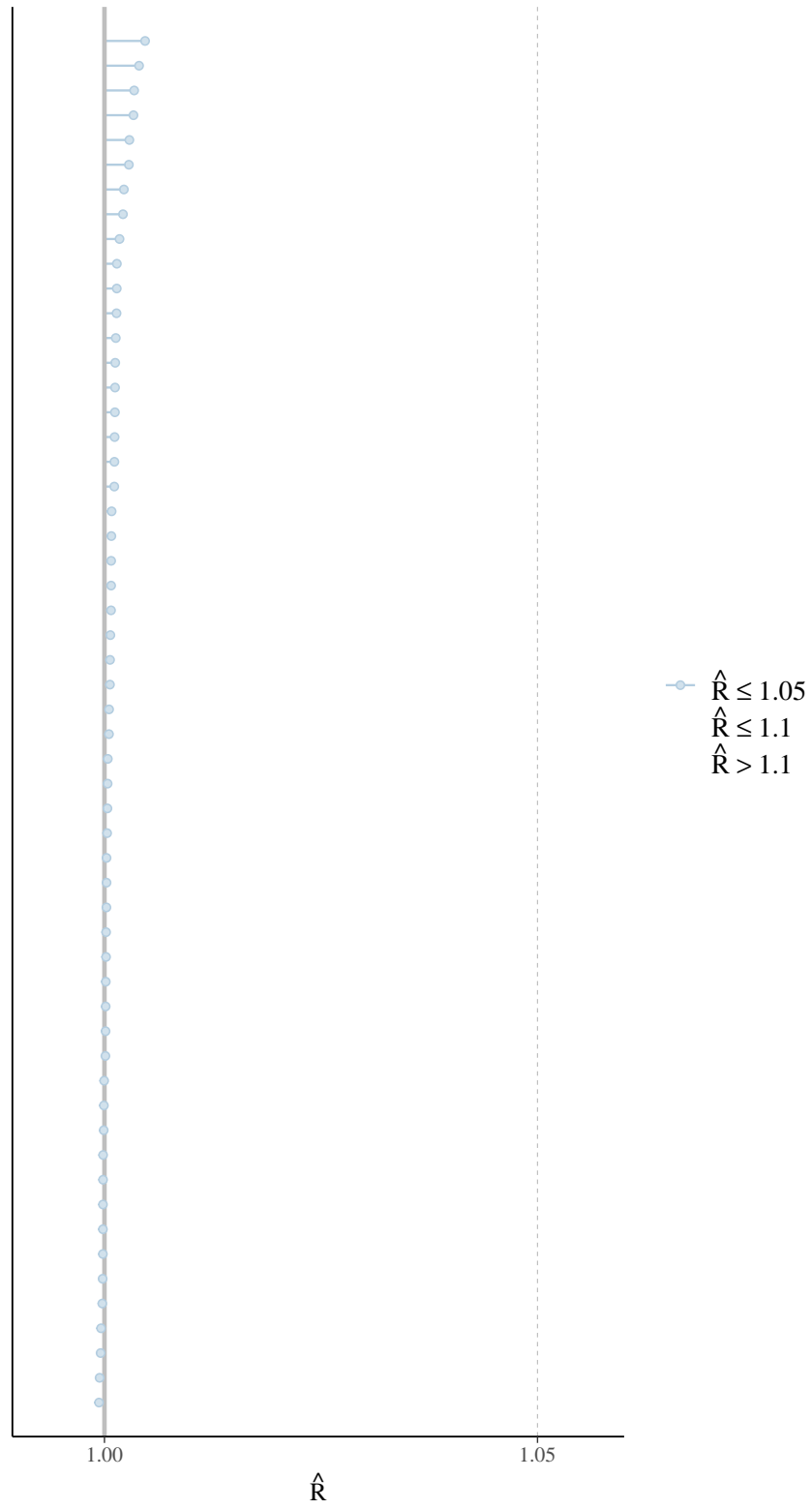Figure 9: Checking the convergence of the MCMC algorithm

Figure 10: Rhat plot

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.