Factors Influencing 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

This paper investigates the allocation patterns of grants for pollinator gardens funded by the PollinateTO initiative in Toronto, focusing on how garden characteristics, neighborhood type, and community engagement influence funding decisions. A Bayesian logistic regression model reveals that gardens located in Neighbourhood Improvement Areas (NIA) or Emerging Neighborhoods (EN) are more likely to receive higher funding, especially for specific garden types like Multi-Residential Gardens and Multiple Property Gardens. Additionally, Indigenous-led gardens and those located in certain wards (e.g., Scarborough Centre) tend to receive larger grants. This research contributes to optimizing the allocation of urban biodiversity funding, ensuring equitable support for community-driven ecological projects.

Table of contents

---

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

1

1 Introduction

Urban biodiversity is recognized as vital to maintaining healthy ecosystems, particularly in the context of rapid urbanization. A key goal is the protection of pollinators, whose decline presents 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 areas. 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 under-explored area. Questions about which types of gardens or neighborhoods are more likely to receive funding, and the factors influencing these decisions, are important 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.

This paper addresses the lack of understanding how urban biodiversity funding decisions are made, with a focus on the 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 larger funding allocations from the PollinateTO program. By identifying these factors, this research provides important findings related to 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 larger funding allocations. These results offer actionable recommendations 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 analysis. Section 6 concludes the paper. The {-} provides supplementary materials, including additional model details related to cleaning and simulation, a section about survey, sampling, and observational data in the context of this paper, and technical model details such as posterior predictive checks and MCMC convergence diagnostics.

2 Data

2.1 Software Packages

This analysis utilizes the R programming language (R Core Team 2023), with the `rstanarm` (Goodrich et al. 2022), `tidyverse` (H. Wickham et al. 2019), `opendatatoronto` (Robinson, Denis, et al. 2021), `ggplot2` (H. Wickham 2016), `dplyr` (H. Wickham et al. 2023), `knitr` (Xie 2022), `arrow` (R. D. Wickham et al. 2020), `patchwork` (Pedersen 2020), `kableExtra` (Zhu 2020), and `broom.mixed` (Bolker, Dorie, and James 2020) 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 use 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 (Environment & Climate 2024), 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 important 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 important 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 a search 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, the variable is_large_garden was constructed. This binary variable was created 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 important information.

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 useful 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
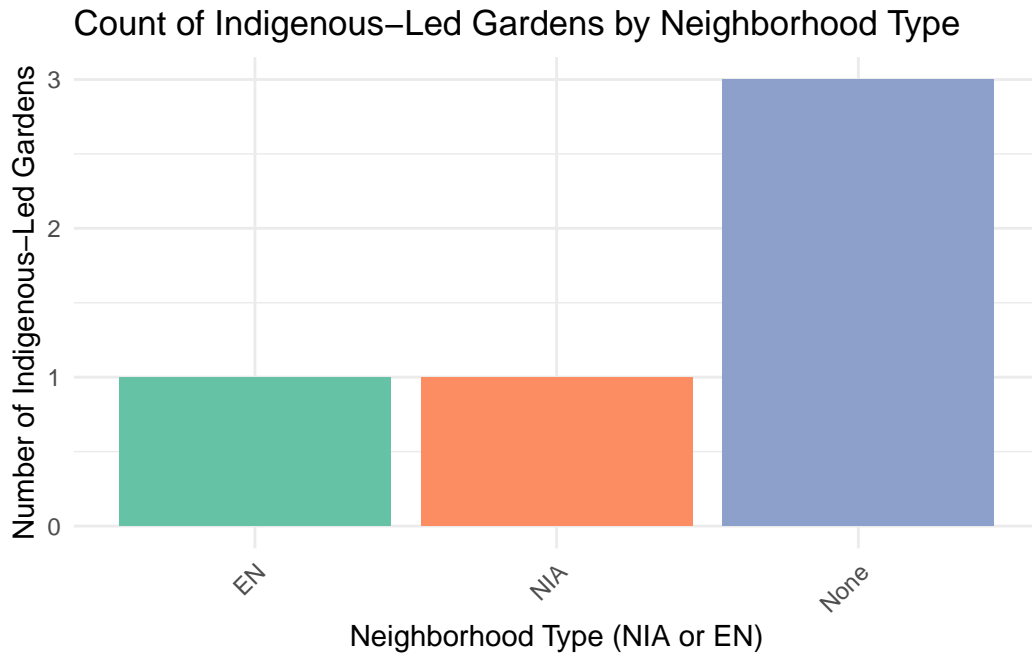
Figure 1: Number of Indigenous Gardens Funded Across Neighborhoods

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.
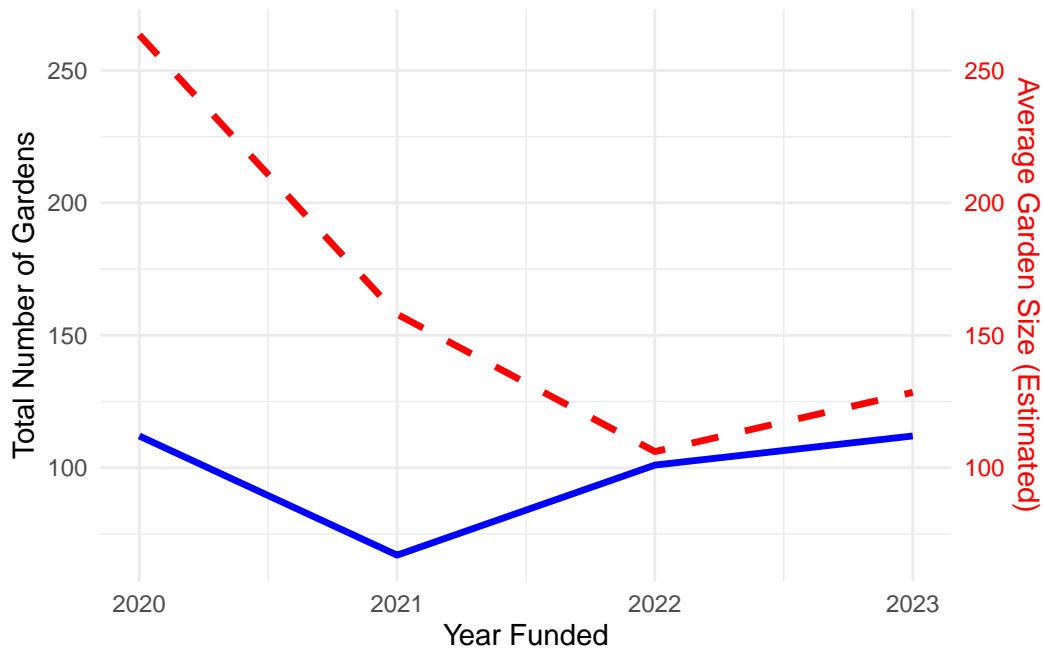
Figure 2: Trend of Garden Size and Number Over Time

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.
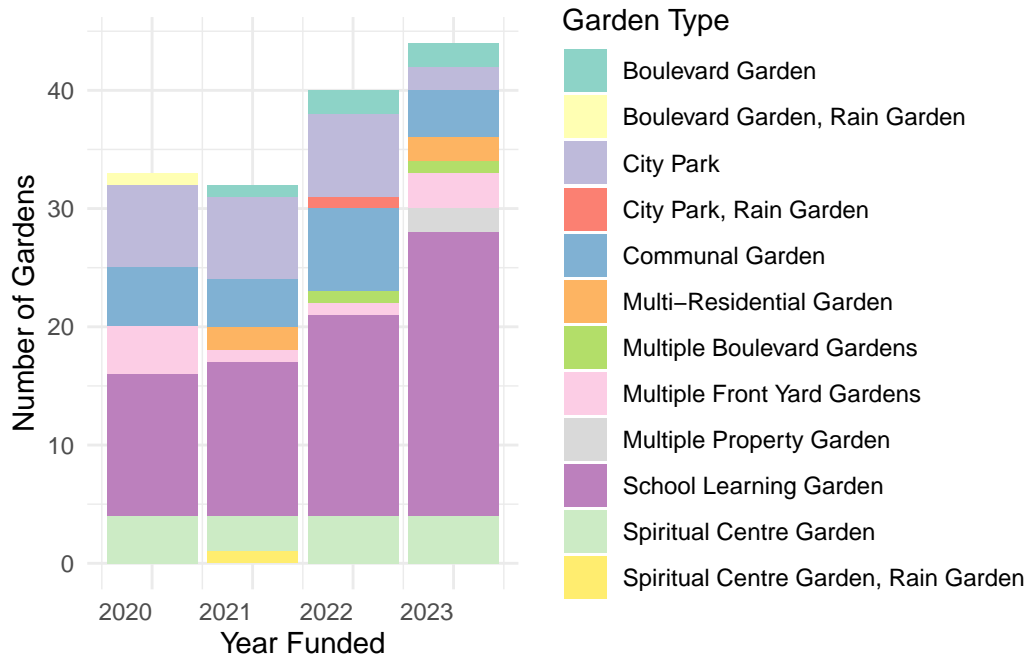
Figure 3: Distribution of Garden Types Funded Over Time

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).
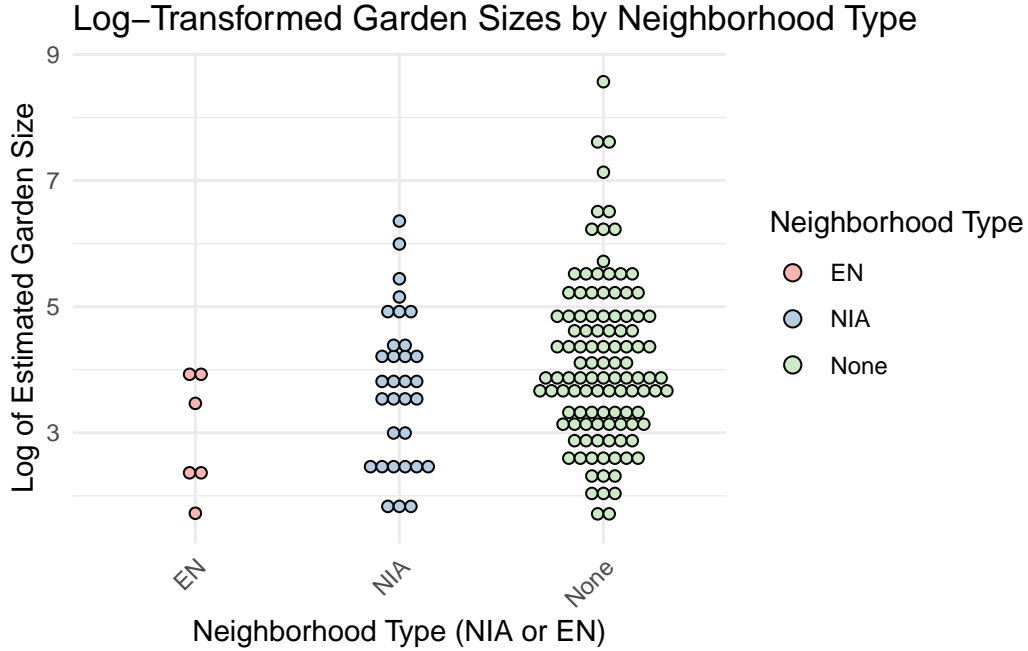
Figure 4: Log-Transformed Garden Sizes by Neighborhood Type

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 - C.

3.1 Model set-up

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

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

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

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

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

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

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

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

$$\beta_6 \sim \mathrm{Normal}(0, 2.5) \tag{9}$$

This Bayesian Logistic Regression model predicts whether a garden is classified as "large" (y_i), a binary outcome modeled as $y_i \sim \mathrm{Bern}(\pi_i)$, 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: $\mathrm{logit}(\pi_i) = \alpha + \beta_1 \times \text{garden type}_i + \beta_2 \times \text{year funded}_i + \beta_3 \times \text{nia or en}_i + \beta_4 \times \text{is indigenous garden}_i + \beta_5 \times \text{ward name}_i + \beta_6 \times (\text{garden type}_i \times \text{nia or en}_i)$, 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).

3.2 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 important 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, 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 shows 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.

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 understanding of the factors driving outcomes in the program.

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 (zoom in for a clearer view)

[!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)
estimates	ward_nameScarborough Centre	estimate	26.76
estimates	ward_nameScarborough Centre	mad	(17.94)
estimates	ward_nameScarborough Southwest	estimate	1.86
estimates	ward_nameScarborough Southwest	mad	(2.31)
estimates	ward_nameScarborough-Agincourt	estimate	-4.96
estimates	ward_nameScarborough-Agincourt	mad	(8.26)
estimates	ward_nameScarborough-Guildwood	estimate	3.08
estimates	ward_nameScarborough-Guildwood	mad	(2.50)
estimates	ward_nameScarborough-Rouge Park	estimate	4.27
estimates	ward_nameScarborough-Rouge Park	mad	(2.41)
estimates	ward_nameSpadina-Fort York	estimate	-5.25
estimates	ward_nameSpadina-Fort York	mad	(8.49)
estimates	ward_nameToronto Centre	estimate	4.85
estimates	ward_nameToronto Centre	mad	(2.12)
estimates	ward_nameToronto-Danforth	estimate	2.59
estimates	ward_nameToronto-Danforth	mad	(2.10)
estimates	ward_nameToronto-St. Paul's	estimate	-4.46
estimates	ward_nameToronto-St. Paul's	mad	(7.48)
estimates	ward_nameUniversity-Rosedale	estimate	3.64
estimates	ward_nameUniversity-Rosedale	mad	(2.30)
estimates	ward_nameWillowdale	estimate	5.50
estimates	ward_nameWillowdale	mad	(2.61)
estimates	ward_nameYork Centre	estimate	4.76
estimates	ward_nameYork Centre	mad	(2.35)
estimates	ward_nameYork South-Weston	estimate	-7.18
estimates	ward_nameYork South-Weston	mad	(9.50)
estimates	garden_typeCity Park × nia_or_enNIA	estimate	1.50
estimates	garden_typeCity Park × nia_or_enNIA	mad	(6.44)
estimates	garden_typeCity Park, Rain Garden × nia_or_enNIA	estimate	-9.01
estimates	garden_typeCity Park, Rain Garden × nia_or_enNIA	mad	(26.70)
estimates	garden_typeCommunal Garden × nia_or_enNIA	estimate	1.43
estimates	garden_typeCommunal Garden × nia_or_enNIA	mad	(6.29)
estimates	garden_typeMulti-Residential Garden × nia_or_enNIA	estimate	8.47
estimates	garden_typeMulti-Residential Garden × nia_or_enNIA	mad	(12.05)
estimates	garden_typeMultiple Front Yard Gardens × nia_or_enNIA	estimate	-11.30
estimates	garden_typeMultiple Front Yard Gardens × nia_or_enNIA	mad	(13.39)
estimates	garden_typeSchool Learning Garden × nia_or_enNIA	estimate	2.56
estimates	garden_typeSchool Learning Garden × nia_or_enNIA	mad	(5.22)
estimates	garden_typeBoulevard Garden, Rain Garden × nia_or_enNone	estimate	-13.65
estimates	garden_typeBoulevard Garden, Rain Garden × nia_or_enNone	mad	(25.14)
estimates	garden_typeCity Park × nia_or_enNone	estimate	-1.13
estimates	garden_typeCity Park × nia_or_enNone	mad	(4.80)
estimates	garden_typeCommunal Garden × nia_or_enNone	estimate	-0.07
estimates	garden_typeCommunal Garden × nia_or_enNone	mad	(5.38)
estimates	garden_typeMulti-Residential Garden × nia_or_enNone	estimate	-12.98
estimates	garden_typeMulti-Residential Garden × nia_or_enNone	mad	(14.72)
estimates	garden_typeMultiple Boulevard Gardens × nia_or_enNone	estimate	0.95
estimates	garden_typeMultiple Boulevard Gardens × nia_or_enNone	mad	(15.27)
estimates	garden_typeMultiple Front Yard Gardens × nia_or_enNone	estimate	3.11
estimates	garden_typeMultiple Front Yard Gardens × nia_or_enNone	mad	(7.67)
estimates	garden_typeMultiple Property Garden × nia_or_enNone	estimate	15.08
estimates	garden_typeMultiple Property Garden × nia_or_enNone	mad	(17.92)
estimates	garden_typeSchool Learning Garden × nia_or_enNone	estimate	-0.55
estimates	garden_typeSchool Learning Garden × nia_or_enNone	mad	(3.39)
estimates	garden_typeSpiritual Centre Garden × nia_or_enNone	estimate	6.53
estimates	garden_typeSpiritual Centre Garden × nia_or_enNone	mad	(4.73)
estimates	garden_typeSpiritual Centre Garden, Rain Garden × nia_or_enNone	estimate	-7.36
estimates	garden_typeSpiritual Centre Garden, Rain Garden × nia_or_enNone	mad	(26.17)
gof	Num.Obs.		149
gof	R2		0.488
gof	Log.Lik.		-42.825
gof	ELPD		-94.3
gof	ELPD s.e.		13.7
gof	LOOIC		188.6
gof	LOOIC s.e.		27.5
gof	WAIC		166.1
gof	RMSE		0.31

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 Polli-nateTO 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.
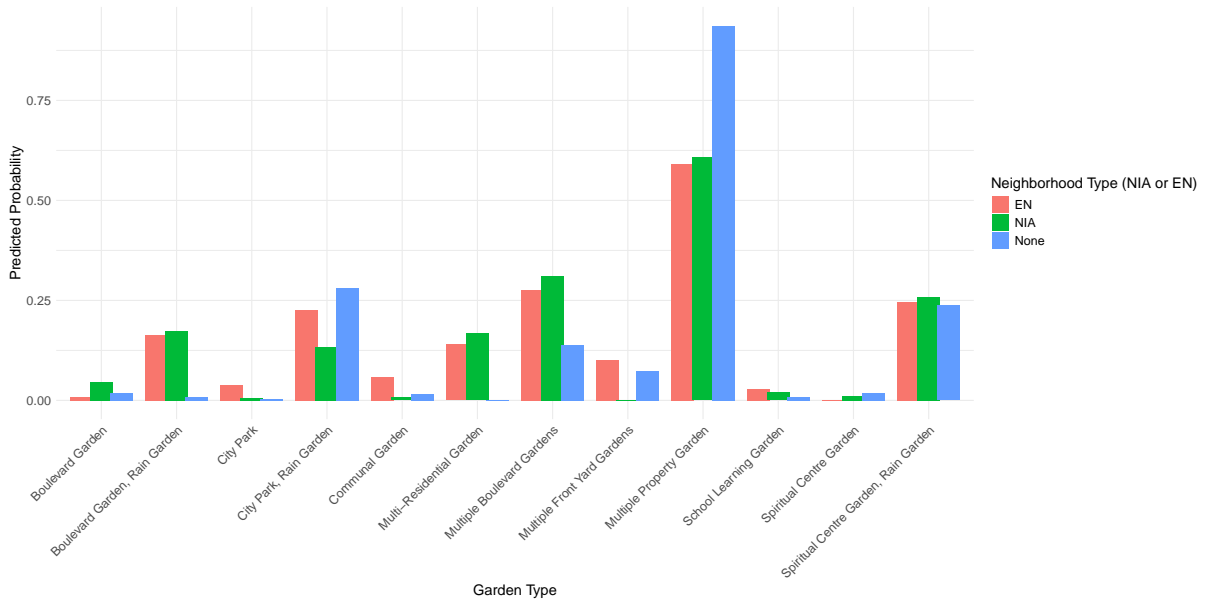

Figure 6: Predictive Probability of Large Gardens

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, defined as a binary outcome (large vs. not large), is associated with several key predictors: garden type, year funded, neighborhood classification (Emerging Neighborhoods or Neighborhood Improvement Areas), Indigenous-led status, and ward-level geographic differences. The analysis shows several noteworthy patterns and disparities in the allocation of larger gardens. By leveraging this model, the study identifies which factors have the most significant influence on whether a garden is classified as large. This approach is enhanced by incorporating interaction effects, which provide a deeper understanding of the relationships between garden types and neighborhood characteristics.

The Bayesian logistic regression framework offers a probabilistic interpretation of these relationships, allowing for a more complex view of uncertainty in parameter estimates. By incorporating credible intervals, the model quantifies the uncertainty around the estimated coefficients, providing a more robust and transparent analysis compared to traditional frequentist methods. In addition to providing understanding of the predictive factors of garden size, this paper also assesses model fit through key measures such as R^2 and WAIC. These measures indicate that the model explains a significant portion of the variation in garden size, with nearly half of the variation accounted for by the chosen predictors.

The implications of these findings are significant for urban sustainability programs like PollinateTO. By highlighting the key drivers of garden size, this research offers an analytical lens through which policymakers can refine the allocation process for funding community gardens. If the goal of such initiatives is to foster environmental sustainability and equity, then understanding which factors contribute to larger gardens is important. For instance, urban greening efforts that are situated in more resource-rich neighborhoods or that prioritize certain types of gardens may need to be adjusted to address inequities and ensure that all communities have access to the same opportunities for greening and sustainability.

Additionally, the interaction terms between garden type and neighborhood classification are particularly important. They show that certain garden types (e.g., vegetable or pollinator gardens) tend to be more successful in certain types of neighborhoods, which may point to localized resources or specific neighborhood priorities. This finding suggests that urban greening programs should consider tailoring their strategies based on local needs and resources. Programs that adopt a more context-sensitive approach, considering both neighborhood characteristics and garden types, are likely to see more equitable and effective outcomes.

17

5.2 Neighborhood Inequities in Urban Gardening

A major finding from this analysis is the evident disparities in garden size across different neighborhoods. Specifically, gardens in areas classified as "None" (i.e., neighborhoods neither in the Emerging Neighborhoods nor in the Neighborhood Improvement Areas) are significantly more likely to be classified as large, while gardens located in Neighborhood Improvement Areas (NIAs) are less likely to achieve the same size. This finding suggests that neighborhoods with fewer systemic challenges, such as better access to resources, land availability, and socioeconomic stability, are more likely to develop large, successful urban gardens. Conversely, communities in NIAs, which often face challenges like poverty, limited access to green space, and lower levels of community investment, struggle to achieve similar outcomes.

This disparity underscores a broader issue in urban sustainability programs: systemic factors, such as socioeconomic inequalities, land availability, and access to resources, significantly affect the success of these initiatives. These inequalities are often compounded by historical disinvestment in marginalized neighborhoods, which can lead to lower levels of community engagement and reduced capacity for sustaining urban greening efforts. The difference in garden size between these neighborhoods points to the need for equity-focused policies in urban greening initiatives.

The implication is clear: sustainability programs must address neighborhood-level disparities in access to resources and opportunities. While it is important to continue fostering urban sustainability, it is equally important to ensure that these efforts benefit all communities, particularly those that have historically been underserved. Programs like PollinateTO should consider revising their funding allocation strategies to prioritize neighborhoods with fewer resources or greater systemic disadvantages. This could involve providing additional technical assistance, offering capacity-building programs for community members, or targeting more funding toward NIAs to ensure these communities can achieve similar environmental and social benefits. By doing so, urban sustainability programs can more effectively address environmental justice concerns, ensuring that all communities have the opportunity to benefit from urban greening.

Moreover, this finding points to a need for broader systemic change. To truly address inequities in urban gardening, policymakers must also focus on improving the underlying conditions in marginalized neighborhoods. This could include initiatives to improve access to affordable housing, better public transportation, and community infrastructure. Without such systemic changes, urban greening programs may fail to achieve their full potential in creating equitable and sustainable urban environments.

5.3 Redefining Success for Indigenous-Led Projects

Indigenous-led projects in urban greening initiatives, including those funded by PollinateTO, are less likely to result in large gardens compared to other projects. This finding raises im-

portant questions about the criteria used to evaluate the success of these projects. Indigenous communities may prioritize different values in their urban greening efforts, focusing on goals such as the preservation of native plant species, the restoration of traditional ecological practices, and the fostering of community cohesion, rather than on maximizing garden size. As a result, the existing framework for evaluating the success of urban gardens, which focuses primarily on the quantitative measure of size, may not fully capture the diverse goals and impacts of Indigenous-led initiatives.

This finding calls for a reevaluation of how success is defined and measured in urban sustainability programs. While garden size is a useful indicator of certain types of success, it may not adequately capture the full range of outcomes that Indigenous-led projects aim to achieve. For example, many Indigenous communities prioritize cultural, social, and ecological sustainability over the mere expansion of garden size. Gardens that are smaller but culturally significant or that provide vital ecosystem services may be just as important, if not more so, than larger gardens that lack these dimensions.

The need to redefine success also speaks to a broader issue in urban sustainability programs: the tendency to prioritize Western, quantitative metrics of success, such as garden size, over qualitative outcomes that may be more meaningful to Indigenous communities. This discrepancy often leads to the marginalization of projects that offer important cultural and social benefits but do not conform to conventional measures of success. By failing to account for these dimensions, urban sustainability frameworks risk overlooking initiatives that are culturally relevant and community-driven.

This finding suggests a shift toward a more inclusive and holistic approach to evaluating urban greening projects. Success should not be measured solely by garden size, but rather by the broader social, cultural, and environmental impacts that these gardens have on the community. Future iterations of urban greening programs should consider the inclusion of qualitative measures, such as community engagement, cultural preservation, and ecological restoration, alongside traditional quantitative metrics. This would ensure that Indigenous-led projects are evaluated on their own terms and that they receive the recognition and support they deserve.

6 Weaknesses and next steps

While this analysis provides important findings, it is important to acknowledge the limitations of the current study and suggest areas for improvement in both the methodology and data collection. Addressing these weaknesses will help refine the findings and lead to more robust and equitable urban sustainability programs.

6.1 Challenges and Limitations in Modeling Garden Outcomes

The model employed in this study provides a useful understanding of the factors influencing garden size, but it has several limitations. One of the primary weaknesses is the binary categorization of garden size into "large" vs. "not large." This binary approach oversimplifies the complexity of garden outcomes, as garden size is a continuous variable that may better reflect the true diversity of urban gardens. Future research could model garden size as a continuous variable, allowing for a deeper understanding of how factors such as neighborhood characteristics and garden type influence garden size. Additionally, including qualitative measures of success, such as the impact of the garden on local biodiversity or community engagement, would enrich the analysis and provide a more thorough understanding of garden outcomes.

Another limitation is the focus on gardens funded by the PollinateTO initiative. While this dataset provides important findings, it may not fully represent the broader urban gardening landscape in Toronto. The sample is limited to one specific initiative, and the findings may not be generalizable to other urban greening programs or gardening projects outside of Toronto. Moreover, the potential for selection bias exists, as the gardens included in the dataset are those that were selected for funding, and these may not be representative of all gardens in the city. Future studies could broaden the scope to include data from other urban greening initiatives or include non-funded gardens to improve the generalizability of the findings.

Finally, while interaction terms in the model provide some understanding of how garden types and neighborhood characteristics interact, they add interpretive complexity. Not all combinations of garden types and neighborhoods had sufficient data to provide precise estimates. In particular, some combinations of garden type and neighborhood classification may have limited representation, leading to imprecise estimates. Future research could address this by gathering larger and more diverse datasets or employing hierarchical models that account for varying group sizes and complexities.

6.2 Opportunities for Future Research and Policy Improvements

Several open questions remain that could be explored in future research. One key area is the inclusion of community-level variables, such as demographic characteristics, access to volunteer networks, and proximity to green spaces. These factors could help explain why certain neighborhoods are more likely to host large gardens and may show underlying mechanisms driving these disparities. Incorporating demographic data could also help policymakers better understand the unique challenges and opportunities faced by different communities in relation to urban greening.

In addition to these community-level factors, further research should focus on exploring the qualitative outcomes of urban gardens, particularly for Indigenous-led projects. Incorporating interviews, case studies, or community feedback could provide a richer understanding of

how these gardens contribute to cultural preservation, social cohesion, and community well-being. By capturing the voices of community members, future studies can gain a more holistic understanding of the impacts of urban greening initiatives, particularly in Indigenous communities.

From a methodological lens, future research could explore alternative modeling techniques, such as hierarchical Bayesian models, to better account for ward-level variability and cross-neighborhood interactions. This would allow for deeper interpretations of how different neighborhoods and wards contribute to garden outcomes. Expanding the dataset to include data from other urban greening initiatives or longitudinal data that tracks changes over time would also provide important understanding of how garden outcomes evolve and whether the impact of different factors changes over time.

Finally, there is a need for future studies to test the effectiveness of policy interventions informed by these findings. For example, targeted funding for Neighborhood Improvement Areas (NIAs) or revised evaluation criteria for Indigenous-led projects could be piloted and evaluated to determine their impact on garden size, community engagement, and other outcomes. This iterative process of evidence-based policy design and evaluation can help create more equitable and effective urban sustainability programs. By testing different policy interventions, cities can identify the most effective strategies for promoting urban greening and achieving more equitable outcomes for all communities.

Appendix

A Additional data details

A.1 Data Cleaning

The cleaning process for the PollinateTO dataset involved several key steps to ensure the data was prepared for analysis. Initially, the raw dataset, downloaded from Open Data Toronto, was read into R. The janitor package was used to clean column names for consistency and readability. Redundant columns, such as 'x_id', which duplicated the 'id' column, were removed to streamline the dataset. Additional irrelevant columns, including 'description', 'image_name', 'postal_code', and others containing only missing values or unnecessary information, were also discarded. Columns such as 'year_funded', 'number_of_gardens', and 'estimated_garden_size' were explicitly converted to numeric formats to ensure proper data types, while 'nia_or_en', 'garden_type', and 'ward_name' were converted to character strings. The 'is_indigenous_garden' column was recoded as a binary variable, with 1 indicating Indigenous-led projects and 0 otherwise. Furthermore, any rows with missing values were removed to maintain the integrity of the dataset. Lastly, the cleaned data was saved in Parquet format for efficient storage and further analysis.

A.2 Data Simulation

To simulate the dataset for this study, a random sampling approach was employed to generate synthetic data that mirrors the characteristics of the PollinateTO initiative's garden locations. The simulation process involved assigning values to several key variables, including the year of funding, garden type, the number of gardens, estimated garden size, neighborhood classification (NIA, EN, or None), Indigenous-led status, and ward name. For each of these variables, values were randomly selected from predefined lists. For example, the "year_funded" variable was drawn from the years 2020 to 2023, while garden types were chosen from a list that includes a variety of garden types such as "Communal Garden" and "Spiritual Centre Garden." Additionally, random values were generated for the number of gardens and estimated garden sizes within realistic ranges. To ensure variability across different wards and neighborhood classifications, the "nia_or_en" and "ward_name" variables were sampled from their respective options. The Indigenous-led status was simulated as a binary variable with values of 0 or 1. The resulting dataset includes 149 entries, with each row representing a simulated garden project. The simulated data was then saved for further analysis, ensuring it captures the diversity and distribution of factors relevant to the PollinateTO initiative.

B Surveys, Sampling, and Observational Data

In this section, we go into the methodology underlying the dataset used for this study, which focuses on analyzing the factors influencing the size and allocation of gardens funded by the City of Toronto's PollinateTO initiative. The data, derived from the open-source dataset on primary project garden locations provided by the City of Toronto, offers a unique opportunity to examine the relationships between various community and garden-level characteristics. This dataset includes information on garden size, neighborhood classifications, garden types, and the presence of Indigenous-led gardens, which serves as the foundation for the analysis in this paper. However, the observational nature of the data raises important questions about sampling methods, biases, and generalizability, which we address below.

B.1 Survey and Data Collection Methodology

The PollinateTO dataset is primarily observational, as it includes data from gardens that were funded through the initiative. However, while the dataset provides a rich source of information, there are several aspects of the data collection process that must be scrutinized. First, the data is not drawn from a random sample of all urban gardens in Toronto, but specifically from those that received funding from the PollinateTO initiative. This introduces a potential selection bias in the dataset, as gardens that are funded by PollinateTO may differ from those that are not, in terms of size, location, and community characteristics. The lack of randomization in the selection of gardens funded by the initiative means that any conclusions drawn from the data may be limited to this specific subset of gardens, potentially reducing the external validity of the findings when applied to broader urban gardening contexts.

The dataset includes key variables such as garden type, neighborhood classification (Emerging Neighborhoods or Neighborhood Improvement Areas), the year funding was provided, and the ward in which each garden is located. These variables were likely chosen to capture factors that influence garden success and size, but it is important to note that other relevant variables, such as community engagement or access to resources, are not included. The absence of these factors may lead to omitted variable bias, as unmeasured factors could also influence garden size, and therefore the results of this study may only be partial.

B.2 Sampling Considerations and Biases

In terms of sampling, the PollinateTO initiative targets gardens that aim to support pollinator species and contribute to biodiversity in urban areas. The initiative is aligned with the broader goals of sustainability and environmental resilience in Toronto, and it emphasizes inclusivity by supporting gardens in various types of neighborhoods. However, the program's funding decisions are likely influenced by factors such as available funding, the perceived urgency of greening initiatives in specific neighborhoods, and the organizational capacity of communities

to implement and maintain gardens. As a result, neighborhood characteristics such as socioeconomic status, access to green spaces, and local community engagement in sustainability initiatives may systematically influence which areas receive funding, potentially skewing the data towards neighborhoods with higher resources or stronger community organizations.

Furthermore, because the dataset is observational rather than experimental, the model used in this study must account for potential confounding variables. For instance, neighborhood-level characteristics like income inequality or infrastructure quality may be correlated with both garden size and the likelihood of a neighborhood receiving PollinateTO funding, yet these factors are not explicitly included in the dataset. This type of confounding bias can make it challenging to draw causal inferences from the analysis, as the relationships between the predictors (e.g., garden type, neighborhood type, and ward location) and garden size may be influenced by other, unobserved factors.

B.3 Methodological Considerations and Simulation

To address some of these biases and limitations, future research could incorporate simulation techniques to generate counterfactual scenarios, such as estimating the outcomes for neighborhoods that did not receive PollinateTO funding. This approach would help assess the extent of selection bias and better understand how the distribution of gardens might have differed under alternative conditions. Moreover, propensity score matching could be employed to create a matched sample of funded and non-funded gardens, ensuring that both groups are comparable on observed characteristics, thereby mitigating potential confounding effects.

Additionally, longitudinal data on garden outcomes, including follow-up data on garden size and sustainability, would provide more understanding of how factors like neighborhood characteristics and garden types evolve over time. This would help disentangle short-term effects of funding from long-term impacts, and improve the understanding of how urban sustainability programs influence communities over extended periods.

In conclusion, while the PollinateTO dataset provides an important resource for examining urban gardening initiatives, careful consideration of the sampling and observational data structure is essential when interpreting the findings. The potential biases—such as selection bias, omitted variable bias, and confounding bias—must be accounted for when drawing conclusions about the factors influencing garden size. Future research could benefit from additional data sources, such as surveys of garden participants or longitudinal data, to further explore these relationships and improve the robustness of the findings.

C Model details

C.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 shows 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.
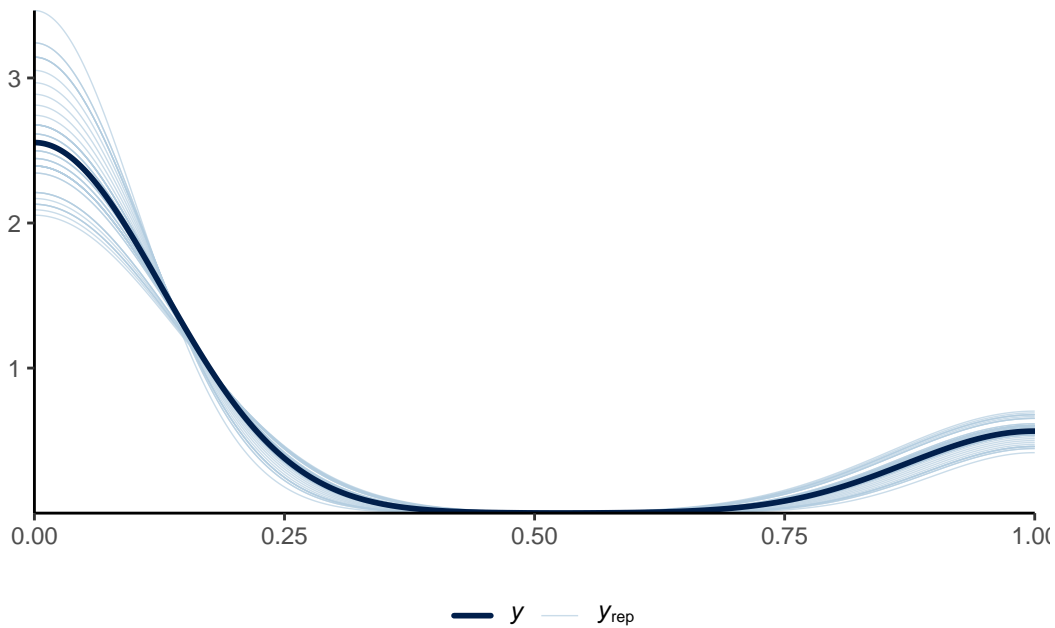
(a) Posterior prediction check

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 important 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.
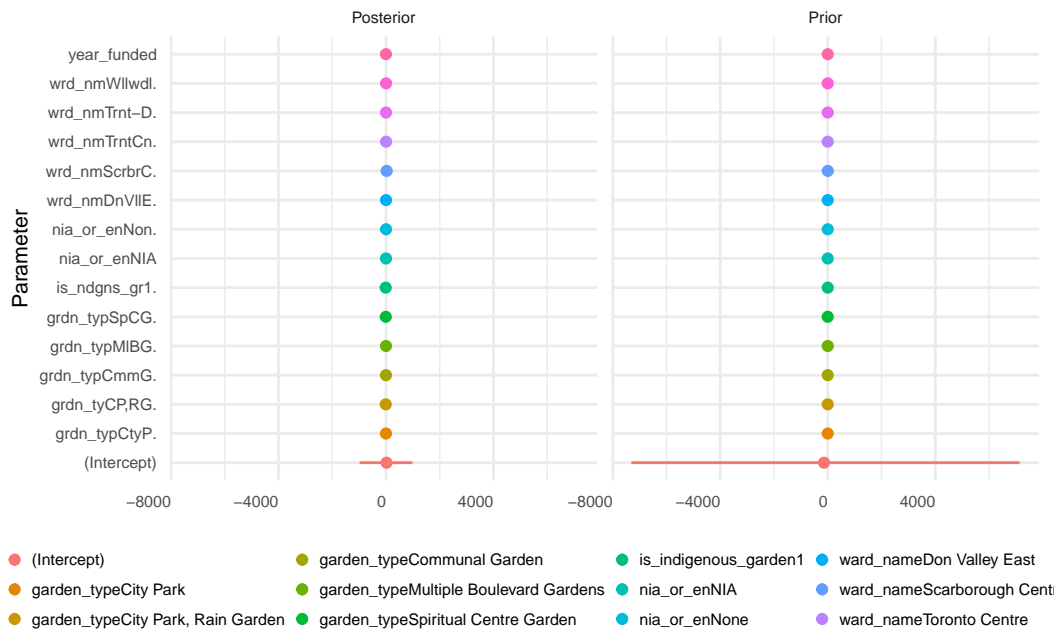

Figure 8: Compare the posterior with the prior

C.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 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 understanding of the parameter space, whereas slow mixing or long stretches of similar values may indicate inefficiency or high autocorrelation. Proper convergence and mixing are important 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 important 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 important for interpreting how predictors like indigenous-led status and funding year influence garden size accurately.
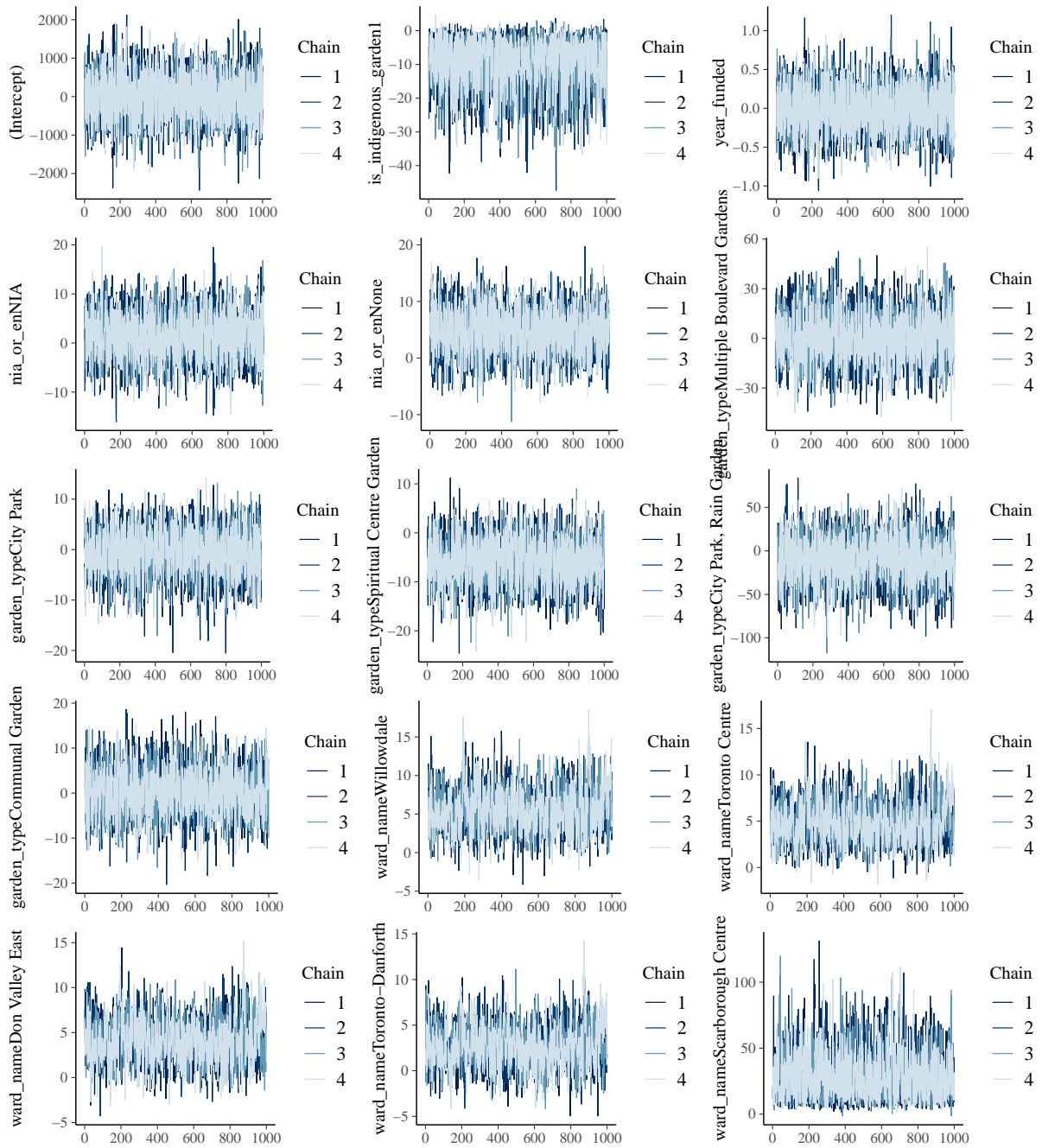
(a) Trace plot of Intercept, Indigenous-led indicator, Year Funded, Neighborhood, Garden type (Sample) and Ward Names (Sample)

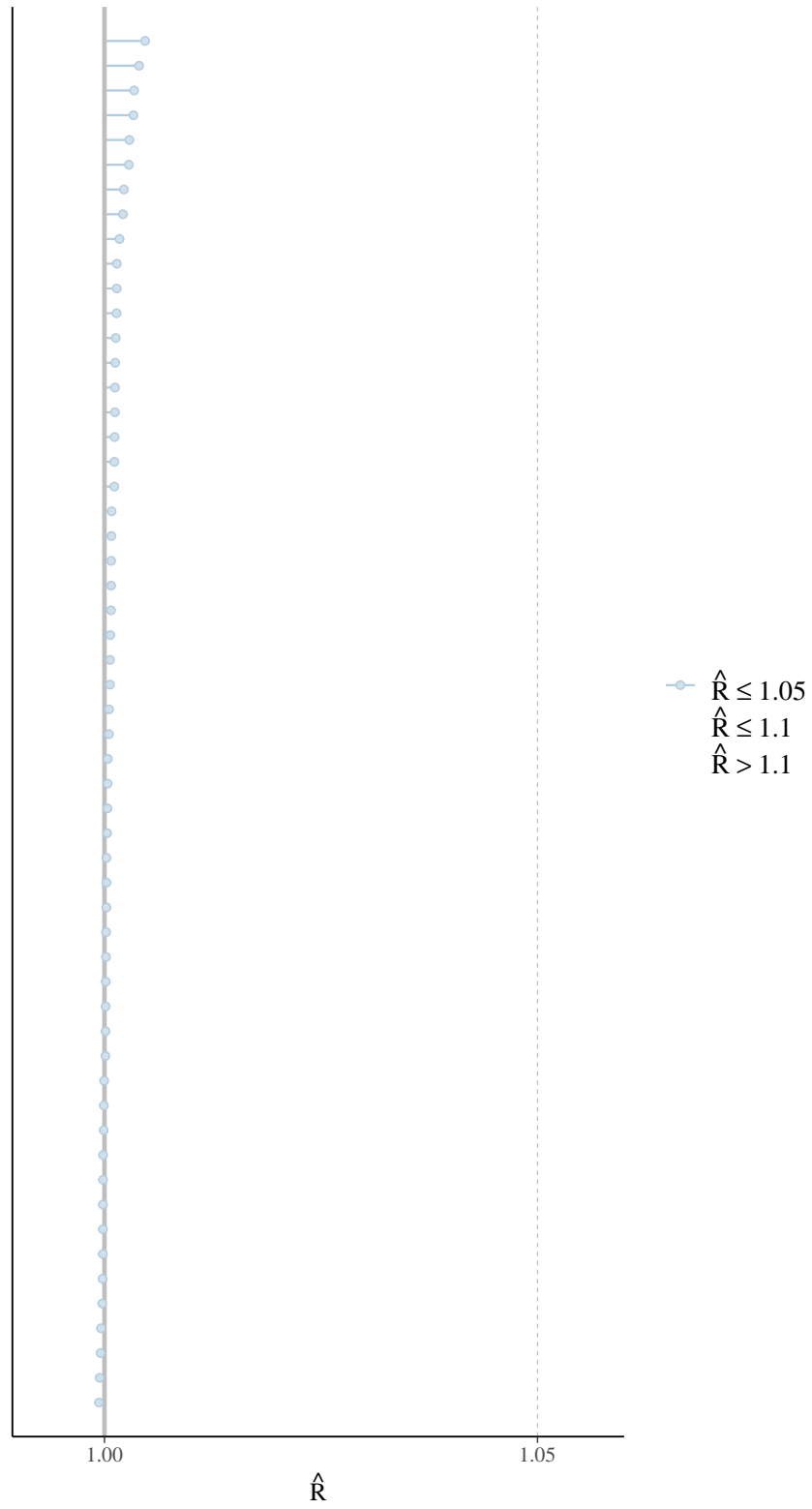Figure 9: Checking the convergence of the MCMC algorithm

Figure 10: Rhat plot

References

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

Bolker, B. L., M. H. Dorie, and M. R. James. 2020. "Broom.mixed: Tidying Methods for Mixed Models." *Journal of Open Source Software* 5 (58): 2331. https://doi.org/10.21105/joss.02331.

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/.

Pedersen, K. 2020. "Patchwork: The Composer for r." *Journal of Open Source Software* 5 (55): 2117. https://doi.org/10.21105/joss.02117.

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

Robinson, David A., Maria St. Denis, et al. 2021. "Opendatatoronto: Access Toronto Open Data from r." *Journal of Open Source Software* 6 (65): 2611. https://doi.org/10.21105/joss.02611.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis.* New York, USA: Springer-Verlag. https://ggplot2.tidyverse.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.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2023. "Dplyr: A Grammar of Data Manipulation." *Journal of Open Source Software* 4 (51): 2151. https://doi.org/10.21105/joss.02151.

Wickham, R. D., Hadley Wickham, J. M. Bryan, and W. Chang. 2020. "The arrow Package: High-Performance Data Processing in r." *Journal of Open Source Software* 5 (49): 2234. https://doi.org/10.21105/joss.02234.

Xie, Yihui. 2022. *Knitr: A General-Purpose Package for Dynamic Report Generation in r.* https://yihui.org/knitr/.

Zhu, Y. 2020. "kableExtra: Construct Complex Table Layouts for kable." *Journal of Open Source Software* 5 (56): 2482. https://doi.org/10.21105/joss.02482.