Structural and Geographic Influences on Residential Parking Availability in Toronto*

Parking Type Strongly Predicts Multi-space Properties; Geographic Differences Are Modest.

Jing Liang

December 11, 2024

We analyzed residential parking data from Toronto to understand factors influencing the availability of more than one parking space. Our findings show that parking type, such as boulevard or front yard parking, strongly predicts the likelihood of multiple spaces, while geographic differences across wards are less significant. This highlights how structural factors play a critical role in shaping parking availability. These insights can inform urban planning policies to better address parking demand in residential areas.

1 Introduction

Residential parking availability plays a pivotal role in shaping urban mobility and land use policies, particularly in dense cities like Toronto. Previous studies have highlighted the importance of residential parking permits and supply in influencing local traffic conditions and land allocation (Van Ommeren, Groote, and Mingardo 2014). For instance, the convenience of home parking has been shown to significantly impact household car usage, underscoring the interplay between parking infrastructure and car dependency (Guo 2013). Moreover, research has demonstrated that vibrant urban districts often face unique challenges in balancing residential parking needs with limited space, requiring tailored policies to accommodate both residents and visitors (Molenda and Sieg 2013). Building on this body of work, this study investigates how structural and locational factors affect the likelihood of residential properties having more than one parking space, providing insights for more efficient and equitable urban planning. Residential parking availability is a critical aspect of urban planning, particularly in dense metropolitan areas like Toronto. The rise in vehicle ownership and the increasing

*Code and data are available at: https://github.com/JessLiang02/Residential-Front-Yard-Parking-Toronto.git.

demand for parking spaces have created challenges for both homeowners and city planners. This paper examines the factors influencing the availability of more than one parking space in residential properties, using a dataset from Toronto's Open Data Portal that documents front yard parking permits and related attributes. While parking availability has been explored in terms of policy and infrastructure, few studies have systematically analyzed how structural and geographic factors combine to affect the likelihood of multiple parking spaces at a property. This paper addresses this gap by applying a logistic regression framework to examine how location and parking type impact this outcome.

The primary estimand of interest is the conditional probability of a residential property having more than one parking space, given specific predictors such as geographic location (ward) and parking type. This estimand is operationalized through a binary outcome ("Yes" for more than one space, "No" otherwise) and modeled using predictors derived from administrative data on residential parking in Toronto. The goal is to understand how these factors jointly influence the likelihood of additional parking capacity while controlling for variability in urban layouts and zoning practices across the city.

The analysis reveals several key findings. First, parking type plays a significant role in determining whether a property has more than one parking space. Properties with boulevard parking show a higher likelihood of accommodating additional spaces compared to those with front yard or widened driveway parking, which are typically more constrained. Second, geographic location also emerges as a modest predictor, with some wards exhibiting slightly higher probabilities of multiple parking spaces, likely due to differences in zoning regulations and residential design. However, across all wards, the prevalence of properties with more than one parking space remains low, underscoring the challenges of expanding parking availability in a densely populated urban environment.

Understanding the factors influencing residential parking availability is crucial for informing urban planning and policy decisions. Insights from this analysis can guide city planners in identifying areas where parking demand is mismatched with supply and in revising zoning regulations to address these disparities. Moreover, this study highlights the importance of structural considerations, such as parking type, in determining parking capacity, offering practical implications for property design and development. By filling a critical gap in the literature on parking availability, this paper contributes to broader discussions on urban sustainability and equitable access to resources in growing metropolitan regions. The findings also provide a foundation for future research into the interplay between urban infrastructure, residential design, and municipal regulations.

The structure of this paper is organized as follows. Section 2 provides an overview of the data sources and key variables used in our analysis, offering a detailed description of the dataset on residential front yard parking in Toronto and the criteria used for variable selection. Section 3 describes the modeling approach, focusing on the Bayesian logistic regression framework. This section outlines the rationale for including predictors such as ward and parking type and discusses the assumptions underpinning the model. Section 4 presents the findings, highlighting the key factors influencing the likelihood of having more than one parking space and examining diagnostic checks such as model fit and residual behavior. Section 5 interprets the broader implications of the results, exploring their relevance to urban planning and parking policy in Toronto. This section also addresses potential limitations of the study and suggests areas for future research to improve the understanding of parking trends and availability in residential areas.

2 Data

2.1 Overview

For this analysis, we used the R programming language (R Core Team 2023) to explore data on residential front yard parking in Toronto. The dataset, sourced from Toronto's Open Data Portal (Gelfand 2022; Toronto 2024), provides detailed information on parking permissions and associated attributes across various residential areas in the city. Key aspects analyzed include the type of parking (e.g., front yard, boulevard, or widened driveway), the presence of more than one parking space, and geographic differences across city wards.

A range of `R` packages was employed for efficient data manipulation, modeling, and visualization. The `tidyverse` package suite formed the backbone of data processing, enabling seamless and reproducible workflows (Wickham et al. 2019). File path management was handled using `here`, ensuring reliable and consistent access to the dataset (Müller 2020). Data cleaning tasks were performed with `janitor`, which helped identify and rectify inconsistencies in the data (Firke 2023). Bayesian statistical modeling was conducted using `rstanarm`, allowing for robust inference and clear interpretation of results (Brilleman et al. 2018). The presentation of model outputs was streamlined using `modelsummary`, which facilitated the generation of concise and informative summaries (Arel-Bundock 2022). Additionally, `arrow` enabled fast and memory-efficient handling of the dataset, which was crucial for analyzing this spatially diverse data (Richardson et al. 2024). Lastly, `testthat` was used to test the validity and integrity of the simulated and analysis datasets (Wickham 2011).

To maintain clarity and reproducibility, we adhered to best practices in workflow organization and analysis, as outlined in Alexander (2023). These practices ensured that data manipulation, modeling, and visualization were seamlessly integrated into a coherent analytical pipeline. This approach supported an in-depth investigation of residential front yard parking trends and patterns across Toronto while ensuring transparency and replicability of results.

The dataset for this analysis, derived from Toronto's Open Data Portal, focuses on residential front yard parking permissions. To ensure the data was suitable for analysis, a thorough cleaning process was applied. Only relevant information, such as the type of parking and geographic details, was retained, while unrelated fields were removed. Entries representing properties without any licensed parking spaces were excluded, narrowing the focus to those with at least one space. A new classification was created to indicate whether a property had more than one parking space, simplifying the analysis of parking capacity. Geographic identifiers were transformed into meaningful names for better interpretability, and incomplete or invalid entries were removed to ensure consistency and accuracy. While alternative datasets on parking or zoning could have been used, this dataset was selected for its specificity to residential front yard parking and its relevance to understanding broader trends in urban parking infrastructure. The cleaning process ensured that the data was comprehensive, accurate, and ready for modeling and analysis.

2.2 Measurement

A thorough discussion of measurement, as it relates to the dataset on residential front yard parking, is provided in the data section. This section explains how real-world phenomena, such as parking infrastructure, city zoning regulations, and permissions granted by the municipal authorities, translate into structured data entries within the dataset. Specifically, it details the processes by which physical characteristics of parking spaces, such as front yard parking, boulevard parking, and widened driveways, are systematically documented and categorized. In addition, the dataset reflects administrative actions, including licensing and permits for additional parking spaces, capturing key decisions made to enforce zoning policies and maintain compliance across Toronto neighborhoods. This provides a snapshot of how parking infrastructure is managed and how residential properties interact with city policies.

The dataset reflects information meticulously collected and maintained by Toronto's municipal authorities, capturing instances of residential parking permissions and characteristics across the city's wards. These data entries stem from a series of systematic administrative processes, such as permit applications submitted by homeowners, inspections conducted by city officials, and approvals granted to ensure compliance with zoning and safety regulations. This formal documentation converts these processes into standardized data points, allowing for a consistent representation of real-world activities. For example, each property is classified by its parking type, geographic location (ward), and whether it has more than one licensed parking space. This standardization not only facilitates analysis but also ensures comparability across diverse neighborhoods with varying zoning and infrastructural constraints.

By focusing on these translated data points, we ensure that our analysis is grounded in how parking infrastructure and policies manifest in practice. This alignment enables us to investigate the relationships between geographical (ward-level) and structural (parking type) predictors and the likelihood of a property having more than one parking space. The dataset thus serves as a bridge between real-world urban phenomena and quantitative analysis, allowing for meaningful insights into residential parking dynamics. Additionally, the data section explores the implications of these measurement processes, highlighting potential limitations, such as the possibility of informal or unregistered parking arrangements that are not captured by the dataset. Such omissions could introduce biases or underreport the true extent of parking availability in some neighborhoods.

Furthermore, the data section underscores the importance of understanding the context in which the dataset was generated. For instance, while the dataset is robust in documenting licensed parking arrangements, it may not fully capture unlicensed modifications made by homeowners or discrepancies between zoning policies and actual usage. These gaps emphasize the need for cautious interpretation of the findings, as well as the importance of complementing administrative datasets with supplementary data sources, such as surveys or field observations, in future research. By addressing these nuances, the data section ensures a comprehensive understanding of the dataset's origins, its representational scope, and its role in analyzing Toronto's residential parking landscape. More details are provided in Appendix B.

2.3 Outcome variable

The outcome of interest in this analysis is whether a residential property has more than one parking space, categorized as "Yes" or "No." "Yes" indicates that the property offers more than one parking space, while "No" signifies a single parking space. The data shows that most properties fall under the "No" category, suggesting that having multiple parking spaces is relatively uncommon. This highlights the limited availability of properties equipped with additional parking capacity and underscores the need to explore factors that may influence this outcome.

2.4 Predictor variables

2.4.1 Ward

One of the key factors examined is the geographical location of the property, represented by different city wards. Each ward reflects unique characteristics, including zoning regulations, population density, and residential designs, which may affect parking availability. The analysis reveals that while the majority of properties in all wards have only one parking space, there are slight differences in the proportions of properties with more than one space across wards. These variations could be attributed to differing urban layouts or local policies influencing parking infrastructure.

2.4.2 Type of parking

Another important factor is the type of parking available on the property, such as street-side, front yard, or widened driveway parking. This structural characteristic appears to have a more pronounced impact on the availability of multiple parking spaces compared to location. For instance, properties with street-side parking are more likely to have multiple spaces than those with front yard or driveway parking, which are often constrained by property size or design. This finding suggests that the physical configuration of the parking area plays a critical role in determining whether a property can accommodate additional vehicles, providing valuable insights into the relationship between parking types and space availability.

2.5 Associations between variables

Figure 1 illustrates the proportion of responses for "more than 1 parking space" across the three parking types: Boulevard Parking, Front Yard, and Widened Driveway. The y-axis represents the percentage, scaled from 0% to 100%, while the x-axis shows the parking types. For Boulevard Parking, a larger proportion of "Yes" responses (blue) is observed compared to the other parking types, where "No" responses (red) dominate overwhelmingly. Front Yard and Widened Driveway parking types show minimal "Yes" responses, highlighting the disparity in the availability of additional parking spaces based on parking type. This visualization highlights how parking type correlates with the likelihood of having more than one parking space.

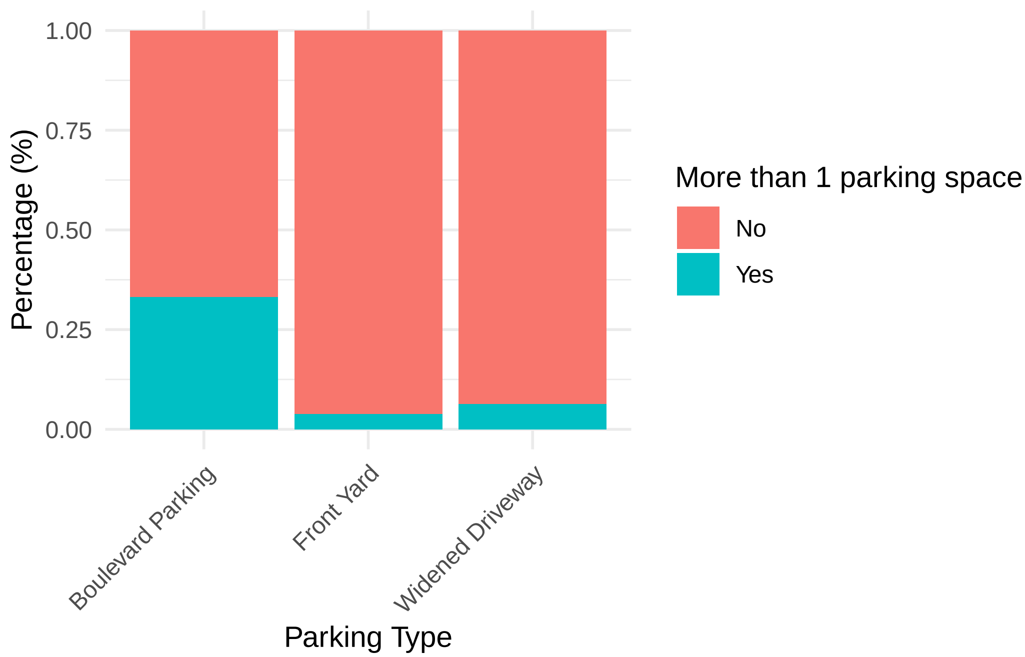


Figure 1: The figure illustrates the distribution of "More than 1 parking space" (Yes/No) as a percentage across various wards. Each bar represents a ward, and the height of the stacked segments within each bar corresponds to the proportion of responses for "Morespace." The responses "No" (red) and "Yes" (blue) sum to 100% for each ward.

Figure 2 provides a visual comparison of the proportion of residential addresses with more than one parking space across wards in the city of Toronto. The y-axis represents the percentage, scaled from 0% to 100%, while the x-axis shows the wards under study. The red portion of each bar indicates the percentage of “No” responses, whereas the blue portion indicates the percentage of “Yes” responses. From the chart, it is evident that “No” dominates the responses in all wards, with “Yes” contributing only a small proportion across the board. The distribution highlights the consistency in the prevalence of “No” responses irrespective of ward, with slight variations in the proportion of “Yes” responses between different wards. This visualization underscores potential patterns or uniformity in the proportion of more than 1 parking space within the study area.

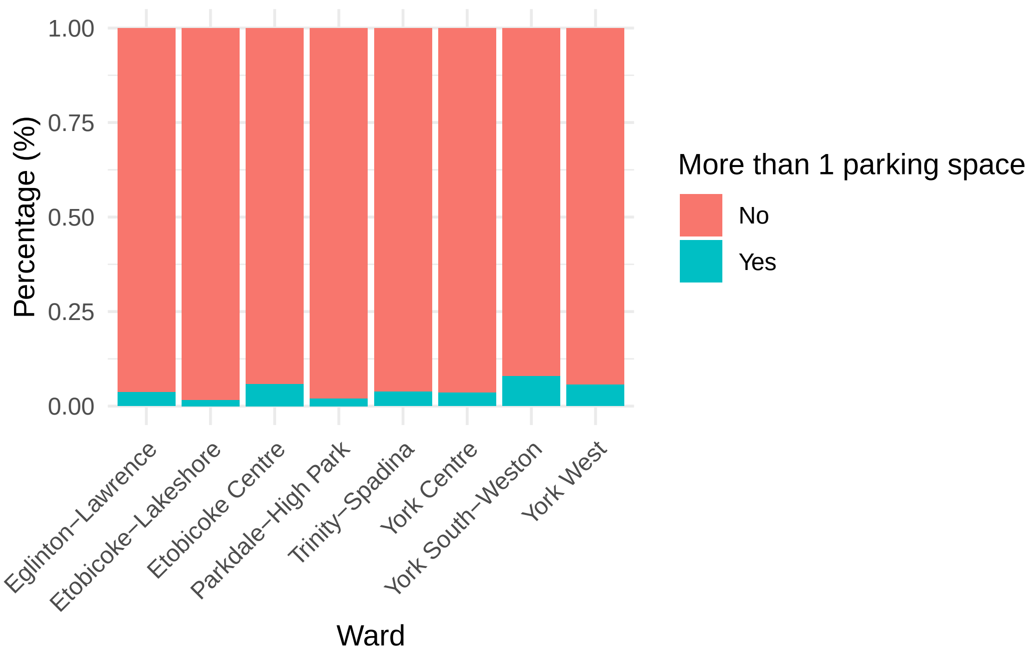


Figure 2: The figure illustrates the distribution of “More than 1 parking space” (Yes/No) as a percentage across various wards. Each bar represents a ward, and the height of the stacked segments within each bar corresponds to the proportion of responses for “Morespace.” The responses “No” (red) and “Yes” (blue) sum to 100% for each ward.

3 Model

The objective of our modeling strategy is twofold. First, we aim to understand the factors that influence whether a residential property has more than one parking space. Second, we seek to quantify the relative importance of structural factors, such as parking type, and geographic factors, such as ward, in predicting the likelihood of multiple parking spaces. To achieve these objectives, we employ a Bayesian logistic regression model that allows for probabilistic interpretation of the predictors' effects while incorporating prior information in a transparent manner. Background details and additional diagnostics are provided in Appendix A.

3.1 Model set-up

Define y_i as an indicator of whether a property has more than one parking space, where $y_i = 1$ for properties with more than one space and $y_i = 0$ otherwise. The predictors include x_{1i}, the type of parking available on the property, and x_{2i}, the geographic location (ward). Parking type is included as categorical variables for "Front Yard" and "Widened Driveway," with "Boulevard Parking" serving as the reference category. Wards are included as a set of dummy variables, with Etobicoke-Lakeshore as the reference group.

The model is specified as:

$$y_i \mid \pi_i \sim \text{Bernoulli}\left(\pi_i\right)$$

$$\text{logit}\left(\pi_i\right) = \alpha + \beta_1 \text{ FrontYard}_i + \beta_2 \text{ WidenedDriveway}_i + \gamma_1 \text{ EtobicokeCentre}_i + \gamma_2 \text{ ParkdaleHighPark}_i + \gamma_3 \text{ TrinityS}$$

Here, π_i represents the probability that property i has more than one parking space. The intercept α captures the baseline log-odds for properties in Etobicoke-Lakeshore with Boulevard Parking.

Coefficients β_1 and β_2 measure the effect of "Front Yard" and "Widened Driveway" parking types relative to the reference category. The coefficients $\gamma_1, \ldots, \gamma_6$ capture the effects of specific wards compared to the baseline (Etobicoke-Lakeshore), explicitly:

- γ_1 : Etobicoke Centre
- γ_2 : Parkdale-High Park
- γ_3 : Trinity-Spadina
- γ_4 : York Centre
- γ_5 :York South-Weston
- γ_6 : York West

The priors for the intercept and coefficients are weakly informative and specified as follows:

$$\begin{aligned} \alpha &\sim \text{Normal}(0, 10) \\ \beta_1, \beta_2 &\sim \text{Normal}(0, 10) \\ \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6 &\sim \text{Normal}(0, 10) \end{aligned}$$

These priors reflect minimal prior knowledge, allowing the data to influence the posterior estimates while discouraging extreme values. The model estimates how parking type and geographic location jointly predict the likelihood of a property having more than one parking space, with the logistic link function ensuring predictions are appropriately bounded between 0 and 1.

We fit the model in R (R Core Team 2023) using the `rstanarm` package from Brilleman et al. (2018). These priors ensure flexibility in the modeling process while allowing the data to drive the posterior estimates.

3.1.1 Model justification

This model is well-suited to the research question because the outcome variable, whether a property has more than one parking space, is binary. A logistic regression framework naturally models the relationship between this binary outcome and the predictors, capturing the log-odds of the outcome as a linear combination of the predictors. The Bayesian implementation of logistic regression adds further advantages, allowing for the inclusion of prior distributions to reflect uncertainty or prior knowledge about the relationships between predictors and the outcome. Weakly informative priors, as used in this model, ensure that the analysis remains data-driven while providing regularization to prevent overfitting. Additionally, the Bayesian framework yields posterior distributions for each parameter, offering richer insights into the uncertainty and variability of the estimates, which is particularly valuable in contexts like urban planning where decisions often involve uncertainty.

The inclusion of parking type and geographic location as predictors is highly appropriate given their theoretical and practical relevance to the research question. Parking type reflects structural constraints, as different designs (e.g., boulevard parking versus front yard parking) inherently allow for varying capacities. Geographic location, represented by ward, captures locational influences, including differences in zoning regulations, housing density, and property sizes across Toronto. Together, these predictors account for both the physical design of properties and the broader spatial context in which they exist, ensuring a comprehensive analysis of parking availability. Furthermore, these variables align with established findings in urban planning literature, which emphasize the role of structural and geographic factors in shaping parking dynamics. By incorporating these predictors, the model not only provides actionable insights but also positions the findings within a robust theoretical framework, enhancing their relevance for policy and planning.

3.1.2 Comparison with alternative models

The process involved splitting the dataset into training and testing sets, with 80% of the data used for training and 20% for testing. Both Bayesian and Frequentist logistic regression models were trained using the training set, with the Bayesian model incorporating priors and the Frequentist model relying solely on observed data. To evaluate prediction accuracy, the models were applied to the testing set to predict whether a property had more than one parking space. Probabilities were converted into binary predictions using a threshold of 0.5, and accuracy was calculated as the proportion of correct predictions. The results showed identical accuracy for both models at approximately 95.52%, suggesting that both approaches performed equally well in classifying the test data. This parity indicates that the Bayesian model's priors and the Frequentist model's reliance on maximum likelihood produced comparable predictive capabilities for this dataset.

While the Bayesian and Frequentist logistic regression models demonstrated identical accuracy in this analysis, there are compelling reasons to prefer the Bayesian approach in this context. The Bayesian model allows for the incorporation of prior information, which can be particularly advantageous when sample sizes are smaller or when there is prior domain knowledge about the effects of variables, such as parking type or ward characteristics. Even when priors are set to be non-informative, as in this case, the Bayesian framework naturally accounts for uncertainty in parameter estimates, offering a richer interpretation through posterior distributions rather than point estimates alone. This additional layer of insight is particularly valuable in urban planning contexts, where decisions often involve uncertainty and the need to weigh competing priorities.

Furthermore, Bayesian models provide posterior predictive distributions, allowing for a nuanced understanding of the probability of outcomes, rather than relying solely on a single predicted value. This can be especially useful for stakeholders seeking to quantify risks or probabilities associated with certain scenarios, such as the likelihood of a property having more than one parking space in specific wards. The flexibility of Bayesian modeling to incorporate future data or adjust priors also ensures that it remains robust and adaptive as more information becomes available, making it an ideal choice for iterative and long-term urban studies. While the identical accuracy suggests comparable predictive performance, the Bayesian model's ability to quantify uncertainty and provide a deeper probabilistic understanding justifies its use as the preferred approach.

3.1.3 Model API

This model has been further enhanced by the development of an API that allows users to interact with the Bayesian logistic regression model in a practical and user-friendly manner. The API enables users to input their selected predictor values, such as specific parking types or ward locations, and obtain probabilistic predictions of whether a property is likely to have more than one parking space. This tool bridges the gap between advanced statistical modeling and real-world application, making the findings accessible and actionable for urban planners, policymakers, and other stakeholders.

4 Results

Our results are summarized in Table 1. The model results provide insights into the factors influencing the likelihood of a property having more than one parking space. The intercept represents the baseline log-odds for properties with boulevard parking in wards not explicitly listed in the model. Among parking types, properties with front yard parking and widened driveway parking show significantly lower likelihoods of having more than one parking space compared to boulevard parking, as indicated by their strongly negative coefficients (-2.57 and -2.12, respectively). This suggests that the physical and spatial constraints associated with these parking types limit the capacity for additional parking. These results highlight the critical role of parking type in determining parking space availability, with boulevard parking offering more flexibility for accommodating additional spaces.

The geographic location, as captured by the ward variable, also exhibits notable effects. Properties in York South-Weston have the highest positive association (0.90) with the likelihood of more than one parking space, followed by Etobicoke Centre (0.54) and York West (0.44). Conversely, properties in Etobicoke-Lakeshore (-0.64) and Parkdale-High Park (-0.57) show lower likelihoods of multiple parking spaces, while other wards exhibit smaller or non-significant effects. These geographic variations likely reflect differences in zoning regulations, property sizes, and urban layouts across Toronto. Although the model explains only a modest proportion of the variability in the outcome ($R^2 = 0.029$), the significant predictors offer valuable insights into structural and locational influences on residential parking availability. The model diagnostics, including log-likelihood, ELPD, and RMSE, indicate reasonable fit but suggest opportunities for further refinement, such as incorporating additional predictors, not included in the analysis dataset, to better capture complex relationships.

5 Discussion

5.1 The influence of structural and geographic factors on parking availability

This study undertakes a comprehensive examination of the factors influencing the availability of more than one parking space for residential properties in Toronto, offering new insights into the intricate interplay of structural design and geographic location. Employing a Bayesian logistic regression framework, the analysis highlights the pivotal role of parking type as a determinant of parking capacity. Properties featuring boulevard parking demonstrate a markedly higher likelihood of accommodating multiple spaces, a result that underscores the inherent spatial flexibility provided by this design. In contrast, front yard parking and widened driveways are constrained by their physical configurations, limiting their potential for expansion. These findings reflect broader implications for residential design, suggesting that spatial planning at the individual property level can have a profound impact on the ability of urban housing to meet growing parking demands. Such insights are particularly relevant for urban planners,

Table 1: Explanatory models of having more than parking space based on parking type and city ward.

	Bayesian logistic regression model
(Intercept)	−0.95
	(0.23)
parking_typeFront Yard	−2.57
	(0.19)
parking_typeWidened Driveway	−2.12
	(0.20)
wardEtobicoke-Lakeshore	−0.64
	(0.31)
wardEtobicoke Centre	0.54
	(0.17)
wardParkdale-High Park	−0.57
	(0.19)
wardTrinity-Spadina	0.15
	(0.17)
wardYork Centre	0.01
	(0.18)
wardYork South-Weston	0.90
	(0.15)
wardYork West	0.44
	(0.16)
Num.Obs.	17 527
R2	0.029
Log.Lik.	−3120.955
ELPD	−3130.9
ELPD s.e.	81.6
LOOIC	6261.8
LOOIC s.e.	163.3
WAIC	6261.8
RMSE	0.21

architects, and policymakers aiming to optimize land use while addressing the practicalities of residential parking.

Additionally, this analysis sheds light on the indirect implications of structural decisions, which often intersect with broader urban design principles. The strong influence of parking type highlights the need for a more deliberate approach to residential development that integrates parking considerations into zoning and building guidelines. For instance, prioritizing the adoption of boulevard parking in new residential developments could alleviate some of the pressures associated with parking shortages, particularly in high-demand areas. These results also call attention to the importance of aligning residential parking policies with urban growth strategies, ensuring that structural characteristics of properties are leveraged to address both current and future parking challenges.

5.2 Geographic variations in parking capacity

The geographic component of this analysis further reveals important variations in parking availability, demonstrating how broader spatial and regulatory contexts shape residential parking dynamics. By leveraging ward-level data, the study identifies nuanced differences in the probability of properties offering multiple parking spaces across Toronto's diverse neighborhoods. Notably, wards such as York South-Weston and Etobicoke Centre exhibit slightly elevated probabilities of additional parking capacity. These geographic disparities likely reflect a confluence of factors, including zoning regulations, housing density, and property size, as well as historical urban development patterns. For example, suburban wards may benefit from larger lot sizes and more lenient zoning requirements, enabling higher parking capacity, whereas more densely populated areas face physical and regulatory constraints that limit parking expansion.

This spatial variability underscores the importance of tailoring policy interventions to the unique characteristics of each neighborhood. Localized reforms, such as revisiting zoning codes or incentivizing parking capacity enhancements in areas with acute shortages, could mitigate the uneven distribution of parking resources. Additionally, this analysis highlights the broader urban planning challenge of reconciling the tension between accommodating private vehicles and promoting compact, sustainable urban design. By focusing on the geographic nuances of parking availability, policymakers can develop more targeted and contextually appropriate strategies that align with both local needs and broader sustainability goals.

5.3 Broader implications for urban planning

Beyond the specific structural and geographic insights, this study contributes to a broader understanding of the complex challenges surrounding residential parking in growing metropolitan areas like Toronto. The rapid increase in vehicle ownership, coupled with the finite availability of urban space, necessitates innovative solutions that balance the competing demands of accessibility and efficiency. This study's findings point to the critical need for integrated approaches that account for both structural and locational factors in addressing parking shortages. For example, incentivizing developers to adopt space-efficient designs, such as boulevard parking, could not only increase parking availability but also make more effective use of limited urban land. Similarly, revising zoning policies to allow for greater parking capacity in underserved neighborhoods could help alleviate geographic inequities in parking access.

The findings also underscore the value of data-driven decision-making in urban planning. By systematically analyzing how structural and geographic factors shape parking capacity, this study offers a replicable framework for evaluating residential parking challenges in other urban contexts. Moreover, the incorporation of advanced modeling techniques, such as Bayesian logistic regression, provides a robust foundation for exploring the multifaceted dynamics of parking availability. Policymakers and urban planners can draw on these insights to craft interventions that are not only effective in addressing immediate challenges but also aligned with long-term urban sustainability objectives.

5.4 Weaknesses and next steps

While the findings of this study offer valuable contributions to understanding residential parking dynamics, certain limitations warrant careful consideration. One key limitation is the reliance on administrative records, which may fail to capture unlicensed or informal parking arrangements. Such omissions could result in an underestimation of parking availability, particularly in areas where informal practices, such as shared driveways or unauthorized extensions, are prevalent. Future research could address this gap by incorporating alternative data sources, such as satellite imagery or field surveys, to provide a more comprehensive view of parking conditions.

Another limitation is the relatively narrow scope of predictors used in the analysis. While parking type and geographic location are critical factors, additional variables, such as property size, household income, or neighborhood-level car ownership rates, could offer deeper insights into the determinants of parking availability. Expanding the model to include these variables would not only enhance its explanatory power but also provide a more nuanced understanding of the socioeconomic and spatial dynamics of residential parking. Furthermore, the geographic resolution of this study, which focuses on ward-level analysis, may overlook important variations within smaller spatial units. Future research could benefit from a finer-grained approach, examining parking patterns at the neighborhood or even street level to uncover localized trends that are masked at broader scales.

Building on these limitations, future studies could also explore the interplay between residential parking and environmental factors, such as public transit accessibility or proximity to commercial hubs. Integrating qualitative data, such as surveys capturing residents' attitudes and preferences regarding parking, could complement quantitative findings and provide a richer understanding of the underlying drivers of parking behavior. These next steps would build on the insights of this study, advancing the development of targeted, equitable, and sustainable strategies for managing residential parking in rapidly urbanizing cities.

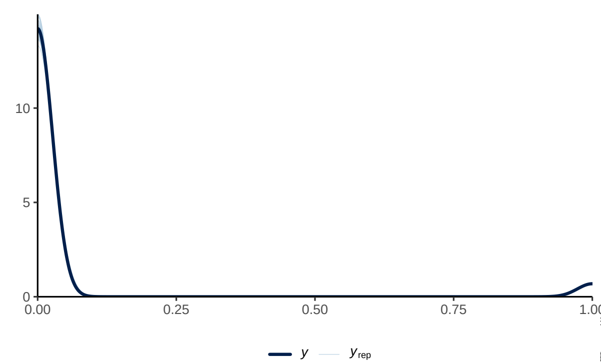
Appendix

A Model details

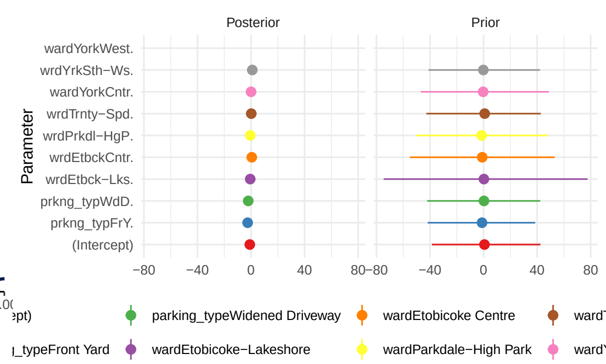
A.1 Posterior predictive check

In Figure 3a we implement a posterior predictive check. The comparison between the observed outcomes (y) and the replicated outcomes (y_{rep}) generated by the model demonstrates that the model performs reasonably well in capturing the general distribution of the observed data. However, deviations in the tail areas indicate that the model might have limitations in accounting for extreme values, suggesting potential areas for refinement.

In Figure 3b we compare the posterior with the prior. The shrinkage observed in the posterior distributions highlights the influence of the data in updating the initial beliefs (priors). For example, coefficients for parking type and ward display distinct shifts from the prior, reflecting the data's strong signal in identifying their effects. The comparison also shows that the priors were appropriately weakly informative, allowing the data to dominate the estimation process. This assessment ensures that the model outputs are data-driven and not overly influenced by prior assumptions.



(a) Posterior prediction check compares the observed data with the model's predictions, highlighting how well the model replicates the actual distribution and identifying areas where the fit could improve.



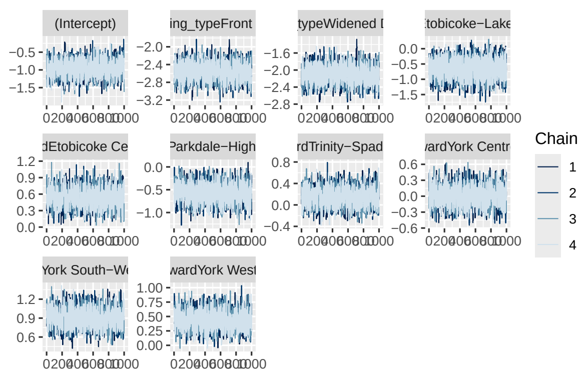
(b) Comparing the posterior with the prior shows the contrast between the prior and posterior distributions for each parameter, emphasizing how the data contribute to refining parameter estimates and reducing uncertainty.

Figure 3: Examining how well the model captures the observed data and how the data influence the parameter estimates.

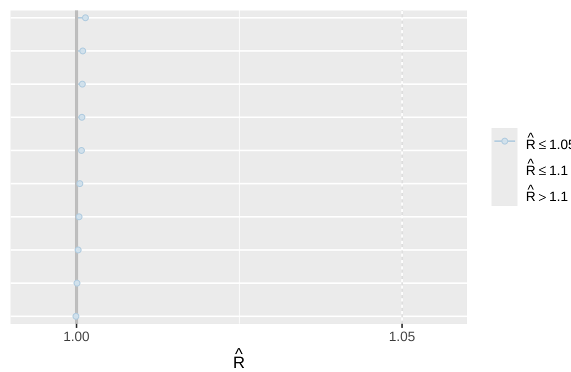
A.2 Diagnostics

Figure 4a is a trace plot. It shows the sampling paths of the four Markov chains for each parameter in the model. The plot demonstrates that all chains mix well and explore the posterior distribution without noticeable patterns or divergences. The oscillations within a stable range indicate that the chains have reached convergence, suggesting that the samples are representative of the posterior distribution. This provides confidence in the reliability of the parameter estimates derived from the model.

Figure 4b is an Rhat plot. It shows the Gelman-Rubin diagnostic (R) for each parameter in the model, which measures the convergence of the chains. All parameters have $\hat{R}$ values very close to 1, indicating that the chains have converged well and there is no significant between-chain variation. This suggests that the model's posterior samples are valid and that the results can be interpreted with confidence. Additionally, the lack of any outliers in the $\hat{R}$ values further reinforces the robustness of the sampling process.



(a) Trace plot shows the sampling trajectories of four MCMC chains for each parameter in the model. The consistent oscillations across the chains, within a stable range, indicate good mixing and convergence, ensuring the samples are representative of the posterior distribution.



(b) Rhat plot illustrates the Gelman-Rubin diagnostic (Rhat) for each parameter. All Rhat values are close to 1, signifying that the chains have converged and there is minimal between-chain variability, confirming the reliability of the posterior estimates.

Figure 4: Checking the convergence of the MCMC algorith and ssessing the convergence of the Markov Chain Monte Carlo (MCMC) algorithm to ensure reliable parameter estimates.

B Survey details

B.1 Overview of Survey Design

Surveys and sampling techniques are foundational to data collection in many of the datasets hosted on the City of Toronto's Open Data Portal. These datasets often include observational data derived from surveys conducted by city departments to inform public policies, track progress on civic issues, or manage resources effectively. Understanding the nuances of surveys and sampling methods is essential for interpreting such data accurately, ensuring its quality, and recognizing its limitations.

Types of Sampling Techniques

The representativeness and reliability of data often depend on the sampling techniques used during its collection. The following sampling techniques are commonly applied in observational studies and surveys:

- Simple Random Sampling (SRS): Ensures each individual in the population has an equal probability of selection. For example, a dataset tracking public transit usage might rely on SRS to obtain unbiased estimates of ridership patterns.
- Stratified Sampling: The population is divided into strata based on certain characteristics, and samples are drawn from each stratum. For instance, datasets on housing affordability may use stratified sampling to ensure representation across income groups or geographic areas.
- Cluster Sampling: Entire clusters, such as neighborhoods or wards, are sampled instead of individuals. This method is cost-effective for large-scale surveys but may introduce design effects that affect variance estimates.
- Convenience Sampling: Data is collected from readily available sources. While not ideal for generalizations, it is common in some City of Toronto datasets due to resource constraints.
- Systematic Sampling: Every nth individual is selected from a list, providing a more structured approach that is sometimes easier to implement than SRS.

Challenges in Observational Data

- Observational data - data collected without intervention—is prevalent in datasets like residential front yard parking, traffic counts, or community service usage. Such data comes with unique challenges:
- Bias: Selection bias and non-response bias can occur if survey participants are not representative of the population.

- Confounding Variables: Observational data often captures multiple variables simultaneously, making it difficult to establish causal relationships without accounting for confounders.
- Data Quality: Observational data depends on the accuracy of the instruments or methods used during collection. Issues like misreporting or gaps in data can lead to misleading conclusions.

Residential Front Yard Parking Dataset

The "Residential Front Yard Parking" dataset includes observational data about licensed parking spaces across Toronto. Key considerations for this dataset are:

- Sampling Frame: The dataset applies only to residential addresses within Toronto's city boundaries. This limitation highlights the importance of clearly defining the population of interest.
- Variable Observations: Variables such as the number of licensed spaces and parking type are inherently observational and recorded based on city licensing records.
- Non-transferability Note: The license for a parking space is not transferable to new property owners, introducing an additional layer of complexity for longitudinal analysis.

Data Quality Considerations in Surveys

The City of Toronto has implemented the Data Quality Score to enhance transparency about dataset usability. This score evaluates datasets based on usability, metadata, freshness, completeness, and accessibility. Survey datasets, in particular, benefit from:

- Metadata: Clear definitions for variables, like "Parking Type" or "Ward," ensure consistent interpretation.
- Completeness: Missing data points can lead to biased inferences. For example, a lack of data for certain wards might misrepresent city-wide patterns.
- Freshness: Regular updates are critical for dynamic datasets like those related to traffic or housing trends.

Implications for Analysis

When using open data for research or policy-making, understanding the underlying survey design and sampling methodology is critical:

- Generalizability: Can the dataset's findings be extrapolated to the larger population? This depends on how well the sampling method ensures representativeness.
- Causal Inference: While observational data is valuable for identifying correlations, caution must be exercised in drawing causal conclusions.

- Error Margins: Sampling error and non-sampling errors (e.g., measurement errors) must be factored into the analysis to avoid overconfidence in results.

Recommendations for Users of Open Data

- Evaluate Metadata: Always review the "Data Features" and "Limitations" sections for information about sampling methods and data quality.
- Use Statistical Weighting: If the dataset includes survey weights, apply them to account for unequal probabilities of selection or non-response.
- Cross-validate with Other Sources: Compare findings from the dataset with external sources to check for consistency and validate conclusions.

This appendix underscores the significance of understanding surveys, sampling, and observational data when working with open datasets. These principles are crucial for ensuring that analyses are robust and aligned with the inherent characteristics of the data.

B.2 Idealized Survey Design

To enhance the quality and representativeness of the "Residential Front Yard Parking" dataset, an idealized survey design would incorporate a stratified random sampling approach with robust data collection protocols. The survey would focus on gathering comprehensive data on licensed parking spaces, parking type, and associated structural and locational characteristics across all Toronto wards. This survey design would ensure high data reliability and validity while being resource-efficient through careful budget allocation and planning.

Budget Allocation and Staff Recruitment

A well-planned survey requires a structured budget to cover personnel, equipment, and operational costs. For this survey, the following budget allocation is proposed:

- Survey staff: A team of 12 field surveyors, each earning $50,000 annually, would be tasked with data collection across different wards, resulting in a staff budget of $600,000. Additionally, three data analysts ($70,000 per year each) would process, clean, and analyze the data, adding another $210,000. A project manager ($90,000 annually) would oversee the entire operation, ensuring timelines and objectives are met.
- Training: initial training sessions for the survey team on data collection protocols, ethical guidelines, and technical tools would require a one-time investment of $15,000, with $5,000 allocated annually for refresher training.

- Data collection tools: Portable devices, such as tablets with GPS capabilities, would be provided to each surveyor for real-time data entry and geotagging. The estimated cost for 15 devices (12 for field surveyors and 3 as backups) is $30,000. Calibration tools for measuring parking dimensions and verifying vehicle counts would cost an additional $10,000.
- Operational costs: Monthly travel allowances for surveyors, estimated at $500 per person, would result in $72,000 annually. Miscellaneous expenses, such as printing survey forms, purchasing software licenses, and contingency funds, would require $50,000.
- Outreach and recruitment: To recruit staff, a $10,000 marketing budget would be allocated to advertise job postings and conduct interviews. For public engagement, an additional $15,000 would be allocated to run community information sessions, encouraging property owners to participate.

In total, the annual budget for this idealized survey design is approximately $1 million, ensuring high-quality data collection and analysis.

Data Collection Protocol and Costs

The survey would use stratified random sampling, dividing Toronto into strata based on wards and neighborhoods. Within each stratum, random samples of residential properties with front yard parking would be selected to ensure representativeness across geographic and demographic dimensions. A target sample size of 10,000 properties is proposed, offering a margin of error of less than 1% for city-wide estimates and allowing for meaningful analysis at the ward level.

- Survey timing: Data collection would be conducted over a 12-month period to capture seasonal variations in parking behavior. For instance, winter months might reveal the impact of snow clearance on parking availability, while summer months might highlight peak usage patterns.
- Protocol: Field surveyors would visit selected properties to collect observational data, including parking type, the number of licensed spaces, and dimensions of the parking area. Each observation would be geotagged using GPS-enabled tablets to ensure precise spatial mapping. Surveyors would also record contextual variables, such as proximity to public transit and adjacent land use.
- Digital integration: Data entry would occur in real time via a custom-built mobile application to minimize transcription errors. The app would include dropdown menus for categorical data, photo upload functionality for verification, and automated validation checks to flag inconsistencies.
- Quality assurance: Weekly audits would be conducted on 10% of the collected data to ensure accuracy and consistency. Analysts would cross-check recorded data with licensing records and flag discrepancies for follow-up by field staff.

Cost Summary:

- Staffing: $900,000 (including surveyors, analysts, and a project manager)
- Equipment and Tools: $40,000 (tablets, calibration tools)
- Operational Expenses: $72,000 (travel) + $50,000 (miscellaneous)
- Training and Outreach: $30,000
- Quality Assurance: Incorporated within the analyst budget.

Total Estimated Cost: $1.1 million annually

Broader Impacts

This idealized survey design ensures a high-quality dataset that is representative of Toronto's diverse neighborhoods. By addressing sampling biases, integrating digital tools, and implementing rigorous quality assurance protocols, this survey would significantly improve the reliability and usability of the residential parking dataset. The insights derived could inform zoning reforms, urban planning policies, and targeted interventions to optimize residential parking across the city.

References

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

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Brilleman, SL, MJ Crowther, M Moreno-Betancur, J Buros Novik, and R Wolfe. 2018. "Joint Longitudinal and Time-to-Event Models via Stan." https://github.com/stan-dev/stancon_talks/.

Firke, Sam. 2023. *janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

Guo, Zhan. 2013. "Home Parking Convenience, Household Car Usage, and Implications to Residential Parking Policies." *Transport Policy* 29: 97–106.

Molenda, Inga, and Gernot Sieg. 2013. "Residential Parking in Vibrant City Districts." *Economics of Transportation* 2 (4): 131–39.

Müller, Kirill. 2020. *here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

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

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

Toronto, City of. 2024. "Residential Front Yard Parking." https://open.toronto.ca/dataset/residential-front-yard-parking/.

Van Ommeren, Jos, Jesper de Groote, and Giuliano Mingardo. 2014. "Residential Parking Permits and Parking Supply." *Regional Science and Urban Economics* 45: 33–44.

Wickham, Hadley. 2011. "testthat: Get Started with Testing." *The R Journal* 3: 5–10. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf.

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.