Environmental, Temporal, and Spatial Influences on Waterfowl Counts at Toronto Beaches*

Suggestions from Bayesian Regression for Urban Ecological Management

Yizhe Chen

December 3, 2024

This paper studies the environmental, temporal, and spatial factors affecting waterfowl populations on Toronto beaches. Using a Bayesian regression model, key predictors such as wind speed, water clarity, and seasonal trends were analyzed to understand their impact on waterfowl counts. The findings show significant associations, including the influence of water clarity and wave action on habitat preferences, as well as long-term declines linked to environmental changes. These findings are vital for informing ecological conservation and urban planning efforts, supporting sustainable coexistence between human activity and biodiversity in urban waterfronts.

Table of contents

*Code and data are available at: https://github.com/YizheChenUT/Waterfowl-on-Toronto-Beaches.git.

1 Introduction

Urban environments, such as Toronto's beaches, serve as habitats for waterfowl, supporting biodiversity while coexisting with human recreational activities. These ecosystems are essential for ecological conservation, urban planning, and environmental management. However, there is a limited understanding of how environmental, temporal, and spatial factors influence waterfowl abundance in urban settings. This knowledge gap hinders effective conservation and management strategies, highlighting the need for research into the dynamics of waterfowl populations in urban ecosystems.

The estimand of this study is the relationship between waterfowl counts and key predictors, specifically environmental factors (e.g., wind speed, air temperature, water temperature), temporal factors (e.g., year, month), and spatial factors (e.g., beach name). The goal is to quantify the extent to which these predictors explain variations in waterfowl populations, offering understandings for the drivers of urban ecological dynamics.

To address this objective, the study employs a Bayesian regression model to analyze waterfowl counts collected from Toronto beaches over several years. This study uses the statistical programming language R (R Core Team 2023). The dataset, sourced from Open Data Toronto (Parks, Forestry & Recreation 2024), captures detailed environmental, temporal, and spatial variables. The analysis shows significant associations between these predictors and waterfowl counts, identifying key drivers of waterfowl activity. For example, temporal patterns highlight the role of seasonal trends, while environmental conditions such as water clarity and wave action reflect habitat suitability.

This research is important because it resolves the issues between ecological monitoring and actionable opinions, enabling urban planners and conservationists to make informed decisions that balance biodiversity conservation with human use of urban beaches. By identifying the key drivers of waterfowl populations, this study supports sustainable management of urban ecosystems.

The remainder of this paper is structured as follows. Section 2 details the data, including its sources, measurement processes, and key variables of interest, with a focus on environmental, temporal, and spatial predictors. Section 3 introduces the Bayesian regression models used in the analysis, providing explanations for their formulation, justification, and validation techniques. Section 4 presents the results, highlighting the significant factors influencing waterfowl counts, supported by visualizations and summary statistics. Section 5 discusses the broader implications of these findings for ecological management, identifies limitations of the current study, and suggests directions for future research. Finally, Section A provides additional methodological details, including sampling strategies, observational challenges, and linkages to relevant literature.

2 Data

2.1 Overview

This study utilizes observational data from Open Data Toronto (Parks, Forestry & Recreation 2024) collected from Toronto beaches to analyze the factors influencing waterfowl counts. The dataset consists of detailed environmental measurements, temporal records, and spatial identifiers, providing a view of the conditions at various beaches. The data was processed using the statistical programming language R (R Core Team 2023), and all analyses were conducted in a fully reproducible workflow. Following established guidelines (Alexander 2023), the data was carefully cleaned and structured to ensure accuracy and reliability.

The raw dataset contained measurements such as wind speed, air temperature, water temperature, water clarity, wave action, and more, recorded alongside waterfowl counts at multiple beaches. These records span multiple years and months, capturing both seasonal and long-term trends. By focusing on these key predictors, this paper aims to understand how environmental and temporal factors influence waterfowl populations.

2.2 Data Cleaning

The raw data from Toronto Beaches Observations underwent a cleaning process to ensure accuracy, consistency, and reliability. This included standardizing column names, handling missing values, and removing outliers specific to each beach's environmental conditions by using the janitor package (Firke 2023). Key variables were formatted into appropriate data types, with categorical variables like "wave action" and "water clarity" standardized into meaningful levels, and numeric variables adjusted for outliers using an interquartile range-based method. These steps transformed the raw dataset into a structured, high-quality dataset suitable for modeling, enabling precise analysis of the factors influencing waterfowl counts.

2.3 Measurement

The transformation of raw observational data into analyzable variables involved a series of well-documented steps. Each record in the dataset corresponds to an observation made on a specific date at a particular beach. The data cleaning process addressed inconsistencies, missing values, and outliers to ensure accuracy. Below is an overview of the key measurement considerations:

2.3.1 Date and Temporal Variables

- Year and Month: The `data_collection_date` column was transformed to extract `year` and `month` as separate variables. These temporal variables capture both seasonal patterns and long-term changes in waterfowl populations.
- Processing: Using R's `lubridate` package (Grolemund and Wickham 2011), the date format was standardized, and missing or invalid entries were removed.

2.3.2 Environmental Conditions

- Variables such as `wind_speed`, `air_temp`, and `water_temp` were converted to numeric types for consistency. Outliers in these variables were removed using an interquartile range (IQR)-based approach.
- `Rain` was converted into a binary variable (`1` for "Yes," `0` for "No") to simplify analysis.

2.3.3 Categorical Variables

- `Wave_action` was standardized to lowercase and categorized into levels: "none," "low," "mod," and "high."
- `Water_clarity` was grouped into categories such as "clear," "cloudy," and "unknown," based on textual patterns in the raw data.

2.3.4 Outlier Removal

- Outliers in numeric variables were handled using a custom function based on the IQR method, applied within groups defined by `beach_name`. This ensured outlier detection was sensitive to each beach's unique conditions.

By addressing these measurement challenges, the dataset was transformed into a reliable foundation for modeling.

2.4 Outcome variables

The primary outcome variable in this study is `water_fowl`, representing the count of waterfowl observed at a given beach on a specific date. This variable serves as the dependent variable in the Bayesian regression models.

The distribution of waterfowl counts across different beaches is shown in Figure 1 by using ggplot2 package (Wickham 2016).

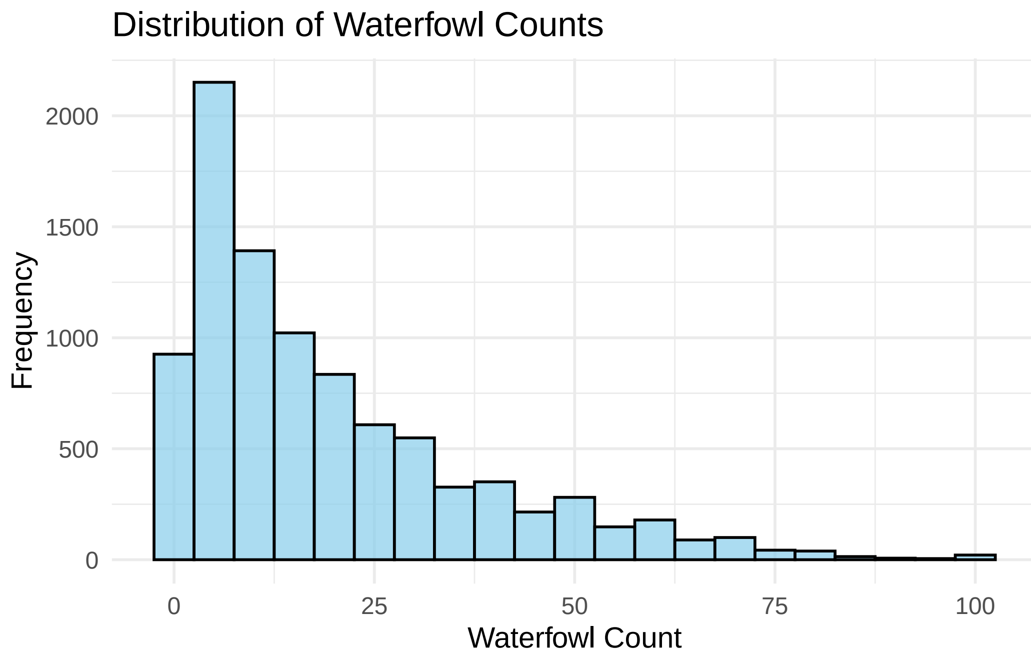

Figure 1: Distribution of Total Waterfowl Counts

2.5 Predictor variables

The study examines several predictor variables, categorized into environmental, temporal, and spatial factors. Each of these variables plays an important role in understanding the patterns and dynamics of waterfowl counts. Below, detailed explanations are provided alongside corresponding visualizations for these variables.

2.5.1 Environmental Variables

2.5.1.1 Wind Speed

Measures the speed of wind at the observation site. Wind conditions may influence waterfowl behavior, particularly their foraging and resting patterns.

Figure 2 illustrates that wind speeds are generally concentrated between 5 and 15 m/s, with a peak near 10 m/s. This distribution reflects typical wind conditions observed during data collection, with few extreme values.

2.5.1.2 Air Temperature

Represents the atmospheric temperature at the time of observation. Air temperature is a key ecological factor, influencing waterfowl activity levels and habitat use.

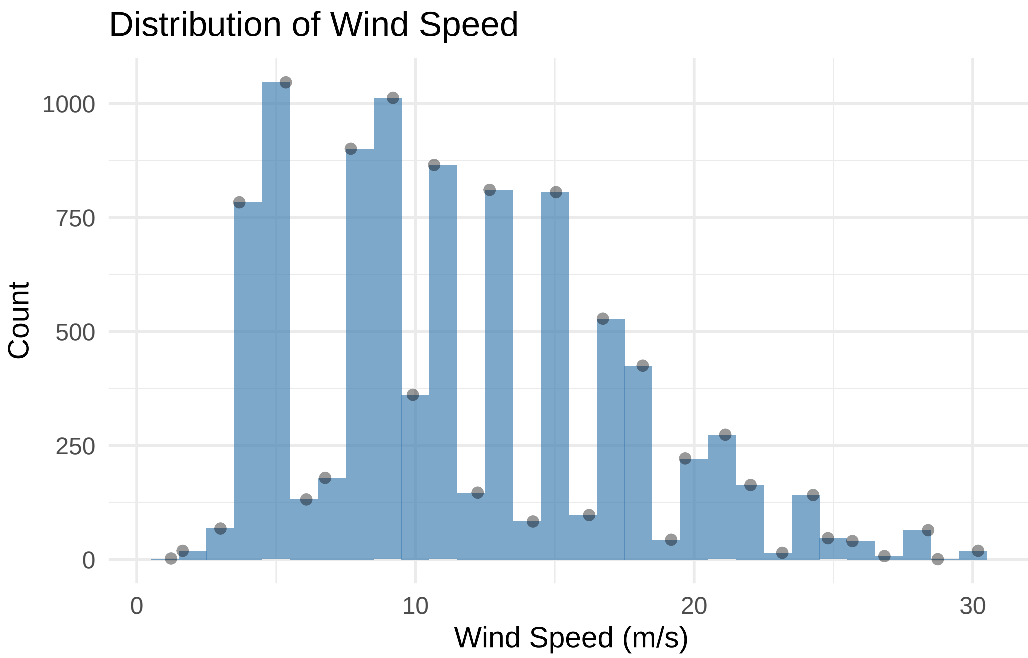

Figure 2: Frequency of Wind Speeds across All Observations

Figure 3 shows air temperatures primarily range between 15°C and 25°C, peaking around 20°C. This suggests the data predominantly represents warmer months, aligning with periods of heightened waterfowl activity.

2.5.1.3 Water Temperature

Indicates the temperature of the water, which directly impacts the suitability of the habitat for waterfowl.

Figure 4 indicates that water temperatures are mostly distributed between 10°C and 20°C, with a notable peak around 18°C. This distribution suggests moderate water conditions conducive to waterfowl activity during the observation periods.

2.5.2 Temporal Variables

2.5.2.1 Year

Captures long-term trends and shifts in waterfowl populations. These trends may reflect broader environmental changes, such as climate change or habitat alterations.

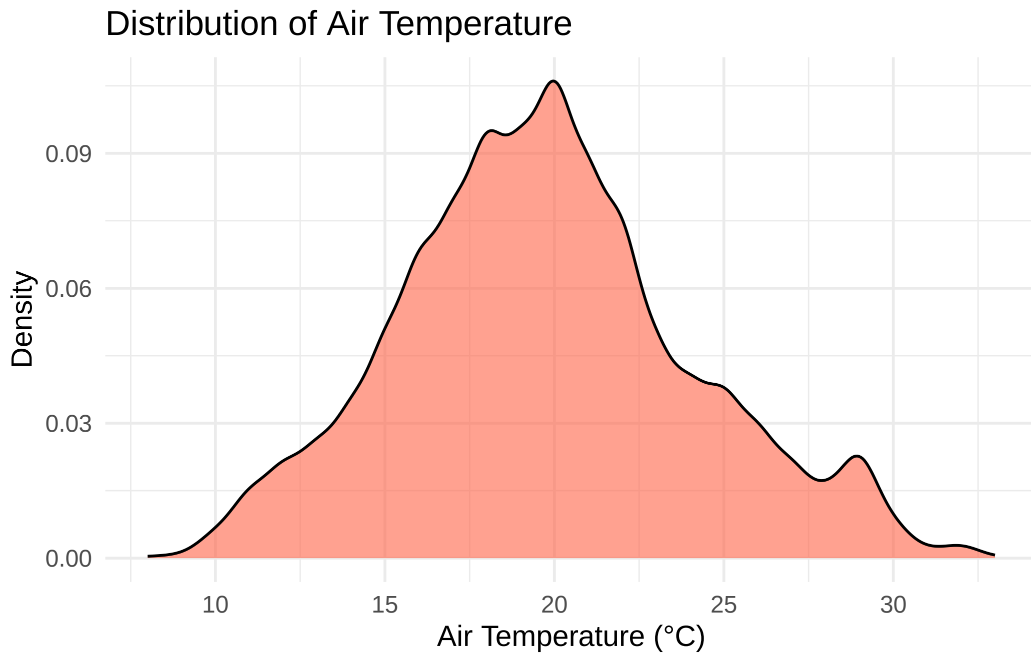

Figure 3: Distribution of Air Temperature across All Observations

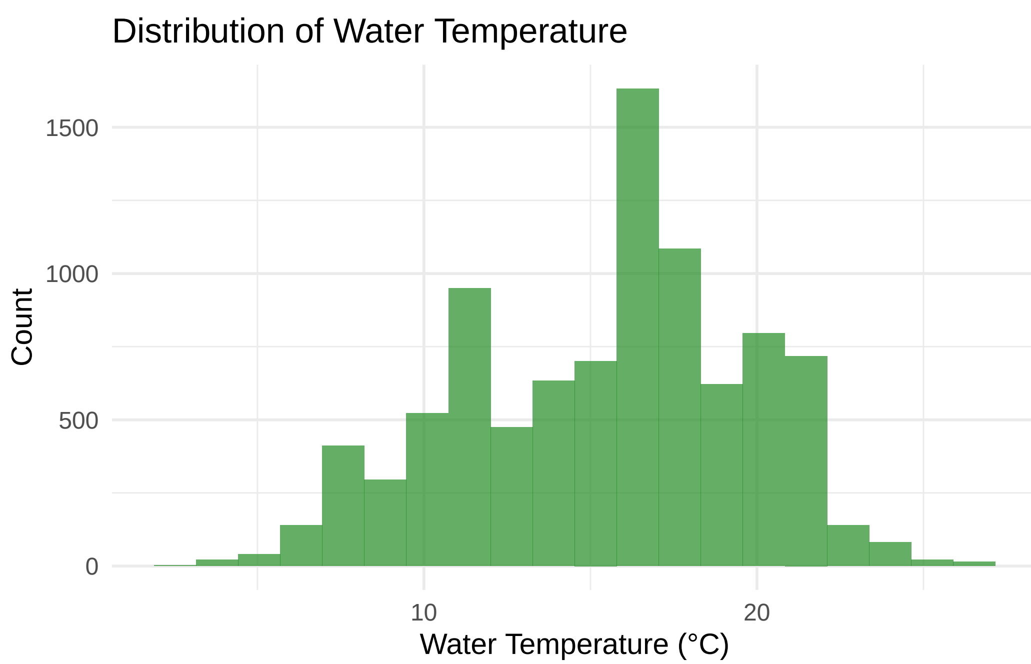

Figure 4: Frequency of Water Temperatures across All Observations

Figure 5 depicts the mean waterfowl counts per year. The data shows an initial peak around 2010, followed by a steady decline with some fluctuations. This suggests potential long-term changes in the waterfowl population or their habitat preferences.

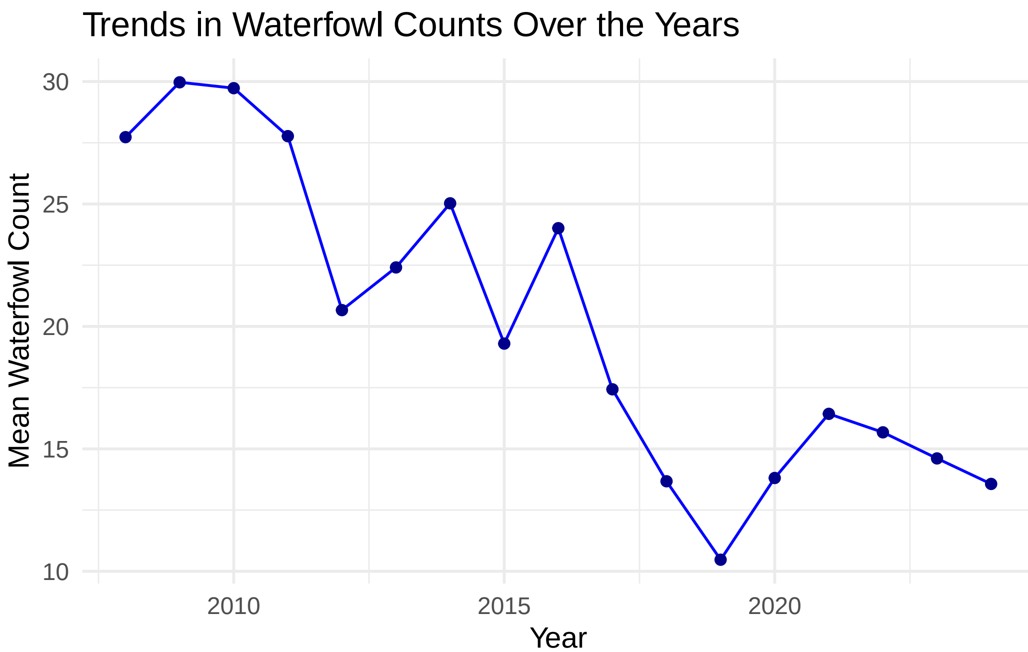

Figure 5: Waterfowl Counts over the Years

2.5.2.2 Month

Accounts for seasonal variations, providing views into the effects of migration patterns and breeding cycles on waterfowl activity.

Figure 6 illustrates monthly variations in waterfowl counts. While the median counts are relatively stable across months, the data exhibits significant variability, particularly in the summer months of June, July, and August, where larger counts are observed. These patterns highlight the importance of seasonality in understanding waterfowl dynamics.

2.5.3 Spatial Variables

2.5.3.1 Beach Name

Identifies the observation location, reflecting site-specific environmental or human influences on waterfowl counts.

Figure 7 shows considerable variation. Beaches such as Sunnyside Beach and Kew Balmy Beach report higher median counts, while locations like Gibraltar Point Beach and Marie

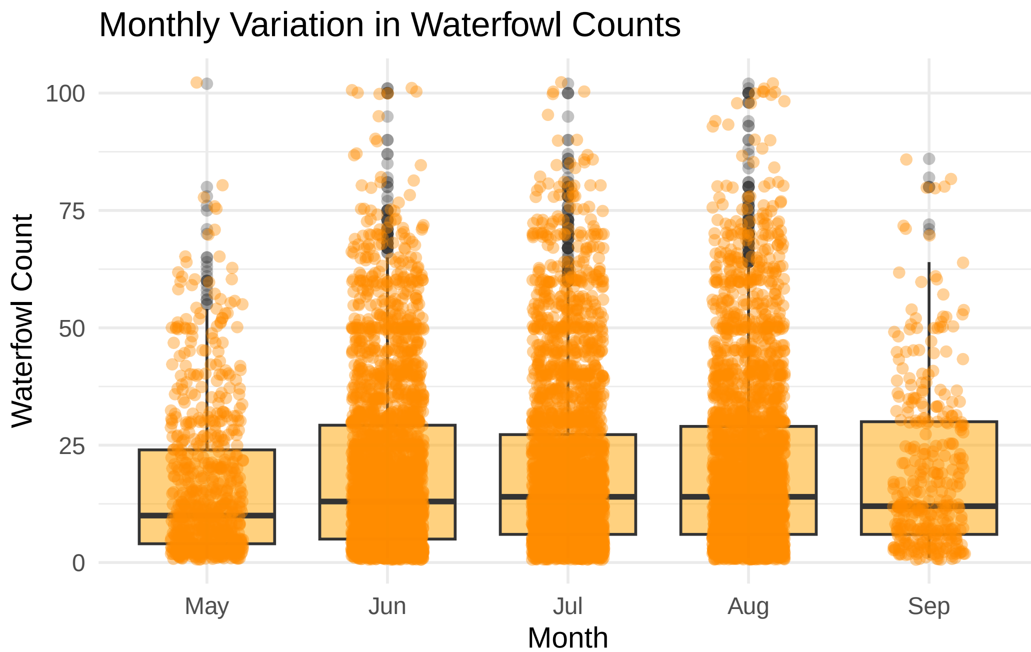

Figure 6: Waterfowl Counts by Month

Curtis Park East Beach have relatively lower median counts. The dispersion in counts is also notable, particularly at beaches with higher counts, indicating diverse environmental conditions and habitat preferences.

The relationships between these predictor variables and waterfowl counts are central to the analysis. The visualizations provided above illustrate their distributions and interactions, offering a overview of the dataset. These factors will be further explored in the modeling section to assess their significance and impact.

3 Model

The primary objective of this study is to model the factors influencing waterfowl counts on Toronto beaches. A Bayesian regression framework was chosen due to its ability to incorporate prior knowledge, handle uncertainty, and provide probabilistic estimates for both parameters and predictions. This approach is particularly suitable for ecological studies where uncertainty is inherent in both data collection and environmental processes.

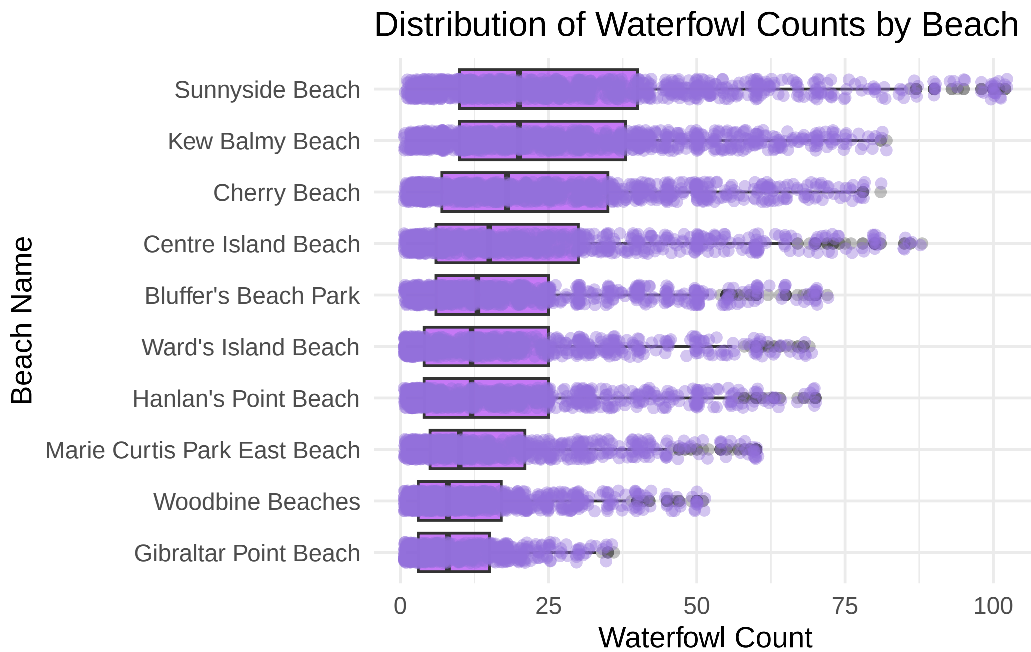

Figure 7: Waterfowl Counts by Beach

3.1 Model Setup

We focus on the second model, which incorporates both continuous and categorical predictors, as well as temporal variables, to provide a understanding of waterfowl counts. The mathematical representation is as follows:

$$y_i | \mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma)$$

Where:

$$\mu_i = \alpha + \beta_1(\text{wind speed}) + \beta_2(\text{air temperature}) + \beta_3(\text{water temperature}) + \gamma_1(\text{wave action})$$
$$+\gamma_2(\text{rain}) + \gamma_3(\text{water clarity}) + \delta(\text{beach name}) + \eta(\text{year}) + \theta(\text{month})$$

Key components:

- α: Intercept term representing the baseline waterfowl count.
- β: Coefficients for environmental predictors (continuous variables).
- γ: Coefficients for categorical variables such as wave action, rain, and water clarity.

- δ, η, θ: Coefficients capturing spatial and temporal effects.
- σ: Residual standard deviation, modeled as $\sigma \sim \text{Exponential}(1)$.

The analysis was conducted using the `stan_glm` function from the `rstanarm` package in R (Goodrich et al. 2024), with priors for α, β, γ drawn from $\text{Normal}(0, 2.5)$, reflecting a neutral prior belief centered around zero.

To provide context, an initial model (referred to as the "first model") was also developed, incorporating only the numeric predictors: wind speed, air temperature, and water temperature. While this model served as a baseline for understanding the relationships among these variables, it lacked the richness to capture seasonal, spatial, and additional categorical effects, which were subsequently integrated into the second model.

3.2 Model Justification

The second model was chosen as the primary model due to its inclusion of additional predictors that align with known drivers of waterfowl activity:

3.2.1 Environmental Variables

Wind speed, air temperature, and water temperature reflect the physical conditions that directly impact waterfowl behavior and habitat use.

3.2.2 Temporal Variables

- The "year" variable captures long-term trends potentially linked to climatic or ecological shifts.
- "Month" accounts for seasonal variations in waterfowl presence.

3.2.3 Spatial Variables

Beach name stands for site-specific characteristics such as habitat type and human activity levels.

The richer specification of the second model allows for a more detailed research of factors influencing waterfowl counts, aligning closely with the study's objectives.

3.3 Model Validation and Diagnostics

The second model underwent rigorous validation and diagnostic checks, which demonstrated its robustness:

3.3.1 Posterior Predictive Checks

Simulated posterior predictions matched the observed waterfowl counts, indicating an acceptable fit.

3.3.2 Leave-One-Out Cross-Validation (LOO)

The second model significantly outperformed the first model in predictive accuracy, as indicated by LOO scores, justifying its enhanced complexity.

3.3.3 Convergence Diagnostics

Convergence was confirmed for all parameters ($\hat{R} \approx 1$), ensuring reliability in parameter estimates.

In conclusion, the second model serves as a robust analytical framework for understanding the factors influencing waterfowl counts, providing important views for ecological management and conservation efforts.

3.4 Limitations and Potential Improvements

While the Bayesian regression framework effectively captures key relationships, some limitations persist:

- Linear Assumptions: The model assumes linear relationships between predictors and waterfowl counts, which may oversimplify ecological interactions.
- Categorical Predictors: Variables such as wave action and water clarity rely on subjective categorizations, potentially introducing observer bias.
- Spatial Effects: While beach names capture site-specific factors, unmeasured variables (e.g., proximity to food sources) may contribute to unexplained variability.

Future iterations could explore hierarchical Bayesian models to account for nested data structures (e.g., beaches within geographic regions) and interactions between predictors.

4 Results

The results of the Bayesian regression analysis provide understandings of the factors influencing waterfowl counts on Toronto beaches. A summary of the estimates from the second model, which incorporates environmental, temporal, and spatial predictors, is presented in Table 1.

Table 1: Regression results for factors influencing waterfowl counts on Toronto beaches

term	estimate	std.error	conf.low	conf.high
(Intercept)	2606.45	115.39	2421.66	2805.06
wind_speed	-0.01	0.03	-0.06	0.05
air_temp	-0.06	0.05	-0.14	0.03
water_temp	0.05	0.06	-0.05	0.14
wave_actionlow	-1.58	0.53	-2.45	-0.77
wave_actionmod	-1.48	0.72	-2.68	-0.30
wave_actionhigh	-3.05	1.22	-5.06	-1.06
rain	0.23	0.38	-0.37	0.83
water_claritycloudy	1.70	0.51	0.84	2.52
water_clarityunknown	0.15	0.50	-0.68	0.94
beach_nameCentre Island Beach	2.30	0.79	1.01	3.63
beach_nameCherry Beach	3.17	0.79	1.86	4.48
beach_nameGibraltar Point Beach	-11.03	0.88	-12.51	-9.54
beach_nameHanlan's Point Beach	-2.35	0.81	-3.64	-0.95
beach_nameKew Balmy Beach	6.90	0.73	5.69	8.12
beach_nameMarie Curtis Park East Beach	-4.41	0.82	-5.69	-3.05
beach_nameSunnyside Beach	8.43	0.78	7.20	9.76
beach_nameWard's Island Beach	-3.56	0.83	-4.94	-2.16
beach_nameWoodbine Beaches	-6.58	0.77	-7.84	-5.32
year	-1.28	0.06	-1.38	-1.19
month.L	1.77	0.78	0.52	3.02
month.Q	-1.57	0.63	-2.63	-0.56
month.C	0.09	0.46	-0.67	0.84
month^4	-0.31	0.34	-0.84	0.28

4.1 Summary of Results

The Bayesian regression analysis highlights the factors influencing waterfowl counts on Toronto beaches. Environmental variables such as wave action, rain, and water clarity emerge as significant predictors, reflecting their direct impact on waterfowl habitat preferences and behaviors. Temporal trends, including a yearly decline and seasonal patterns, underscore the influence of long-term ecological shifts and migration cycles. Spatial variation across beaches indicates site-specific factors, with some beaches consistently supporting higher waterfowl counts, likely due to superior habitat quality or reduced human disturbance.

4.1.1 Environmental Variables

- Wind Speed: The coefficient for wind speed suggests a negligible effect on waterfowl counts, as its estimate is close to zero with wide confidence intervals.
- Air Temperature: A slight negative association is observed, indicating fewer waterfowl with increasing air temperatures.
- Water Temperature: A minimal positive relationship exists between water temperature and waterfowl counts with a low magnitude.
- Wave Action: The impact of wave action on waterfowl counts is significant. "Low" and "mod" wave actions show a positive influence, while "none" demonstrates the strongest positive association.
- Rain: The positive coefficient for rain suggests that rainy conditions might lead to a slight increase in waterfowl counts.
- Water Clarity: Compared to "unknown" water clarity, "cloudy" conditions result in higher waterfowl counts, suggesting potential habitat or food resource preferences.

4.1.2 Temporal Variables

- Year: A notable negative coefficient for the "year" variable indicates a decline in waterfowl counts over time, potentially reflecting environmental or ecological changes. Between 2010 and 2024, the mean annual waterfowl count decreased by approximately 50%, as shown in Figure 5.
- Month: Seasonal patterns emerge, with May displaying notably lower counts compared to other months, likely due to migratory or breeding behaviors.

4.1.3 Spatial Variables

- Beach Name: Waterfowl counts vary significantly by beach. For example, Cherry Beach has higher waterfowl counts compared to Gibraltar Point Beach, suggesting that localized factors such as habitat quality or human activity levels may influence these patterns.

4.2 Visualization of Results

To support the results and provide better interpretability, the following visualizations highlight key findings from the analysis.

4.2.1 Residual Distribution

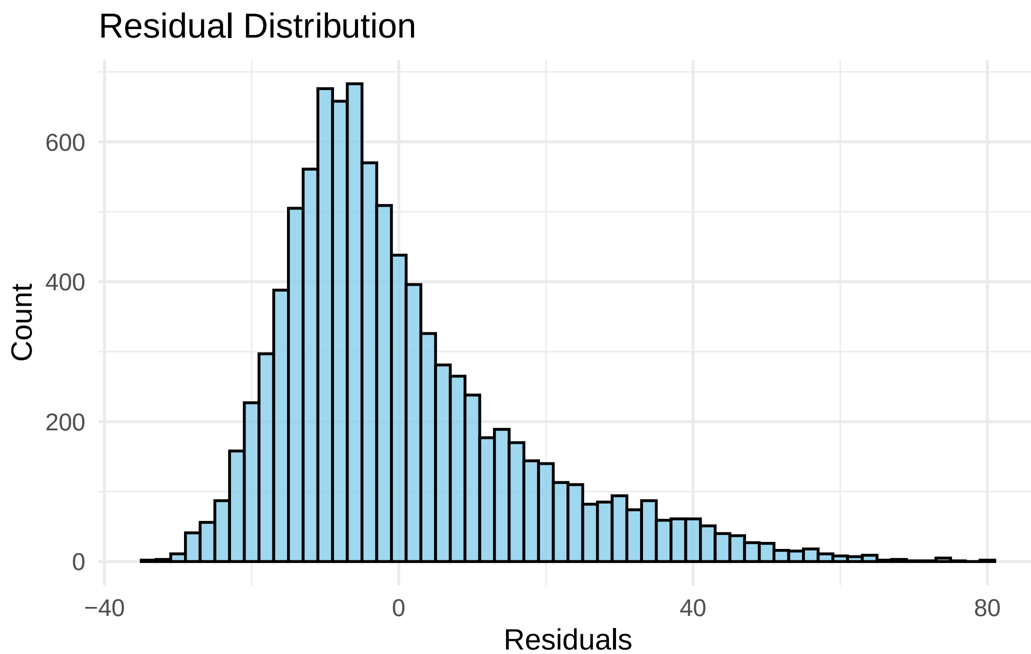

Figure 8: Residual Distribution of the Model

Figure 8 shows the histogram of residuals from the model. The residuals are concentrated around zero, indicating that the model predictions are unbiased. This suggests that the model effectively captures the relationships between the predictors and waterfowl counts without significant systematic errors.

4.2.2 Water Clarity vs. Waterfowl Counts

Figure 9 highlights the relationship between water clarity and waterfowl counts. Clear water conditions consistently support higher waterfowl counts, aligning with ecological preferences for visibility during foraging.

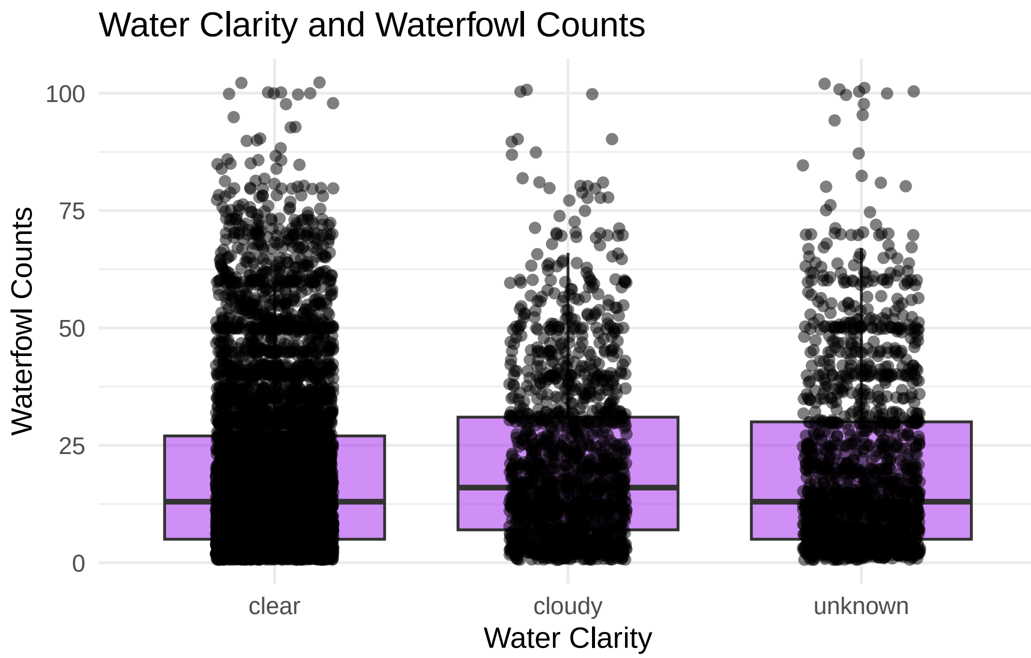

Figure 9: Waterfowl Counts by Water Clarity

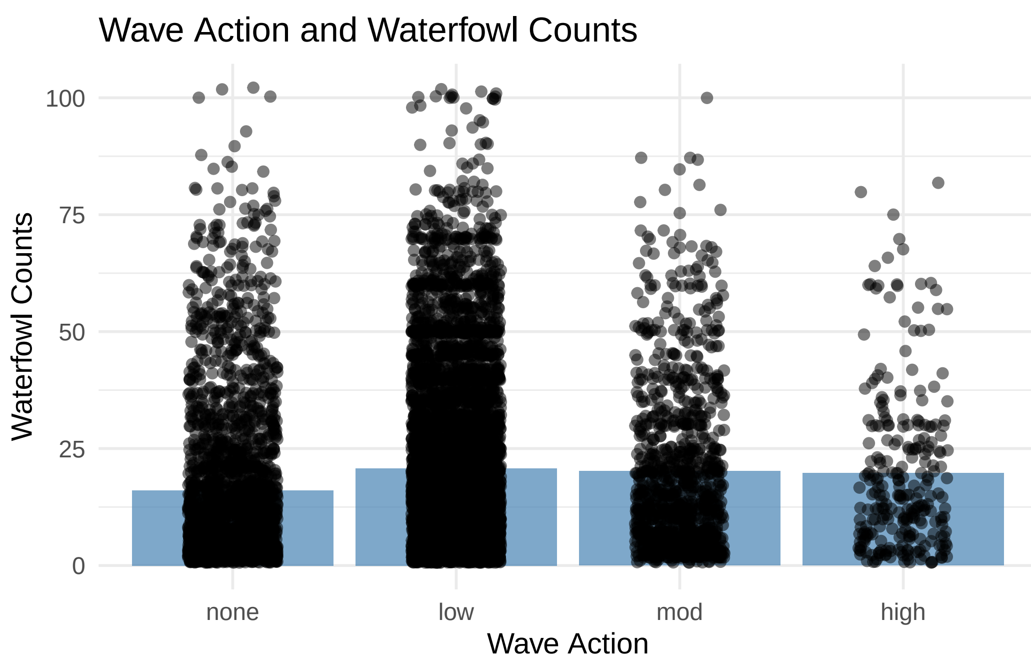

Figure 10: Waterfowl Counts by Wave Action

4.2.3 Wave Action vs. Waterfowl Counts

Figure 10 illustrates the impact of wave action on waterfowl counts. Different wave action levels—none, low, moderate, and high—show distinct patterns in waterfowl behavior. The results suggest that wave action plays a significant role in influencing waterfowl distribution, with calmer conditions (low or no wave action) generally corresponding to higher counts. Moderate wave actions create calm but dynamic water conditions that may reduce human disturbance and enhance foraging opportunities.

These results underline the significant influence of environmental, temporal, and spatial factors on waterfowl counts. The results from this study can inform ecological management practices aimed at preserving biodiversity and promoting sustainable use of urban waterfronts.

5 Discussion

5.1 Gains from the Study

This study identifies key factors influencing waterfowl counts on Toronto beaches, offering vital understandings into urban ecology. Among environmental variables, **water clarity** emerged as a significant predictor. Clear water was associated with higher waterfowl counts, likely due to improved visibility, which facilitates foraging. This finding aligns with ecological studies suggesting that visual clarity in aquatic habitats plays a vital role in supporting avian populations (Thompson and Price 2006). **Wave action**, another environmental factor, also demonstrated notable effects: low and moderate wave action correlated positively with waterfowl counts, while high wave action deterred their presence, possibly due to reduced safety and increased energy expenditure required for navigation. Moderate wave action could help waterfowl maintain stability in feeding areas while avoiding predation or competition often exacerbated by turbulent waters (Rubert et al. 2021).

Temporal variables, including **year** and **month**, illustrated essential trends. The consistent decline in waterfowl counts over the years might reflect broader ecological changes, such as habitat loss, pollution, or the impacts of climate change (Zedler and Kercher 2005). Seasonal variations, as captured by the month variable, highlighted the influence of migratory patterns, with peak activity during certain months aligning with known breeding and migratory behaviors. This temporal view underscores the importance of aligning conservation efforts with seasonal ecological cycles.

Spatial differences, captured through **beach name**, indicate that specific beaches offer unique habitat characteristics influencing waterfowl behavior. Beaches like Kew Balmy and Sunnyside, which showed higher waterfowl counts, may provide better foraging grounds, shelter, or other favorable conditions. Understanding these spatial dynamics can inform localized conservation strategies.

5.2 Implications for Ecological Management

The findings provide actionable opinions for policymakers and conservationists. For example, maintaining **water clarity** through stringent pollution control measures could significantly enhance habitat quality. Similarly, **managing wave conditions** near key habitats, potentially through engineering solutions like wave breaks, could create safer and more attractive environments for waterfowl. Conservation efforts could also benefit from applying temporal trends, such as increasing protective measures during migration peaks or breeding seasons. Specific beaches with declining waterfowl counts may require targeted interventions, such as establishing vegetation buffers, reducing urban runoff pollution, and implementing seasonal restrictions on human activities, to prevent further population declines (Murgui and Hedblom 2017).

Moreover, the observed long-term decline in waterfowl populations highlights the urgent need for continuous ecological monitoring. Coordinated efforts between local authorities and environmental organizations could establish regular surveys to assess trends and address emerging threats proactively.

5.3 Limitations of the Study

While this study provides understanding for the factors influencing waterfowl counts on Toronto beaches, several limitations should be noted to contextualize the findings:

- Unmeasured Ecological Variables: The dataset does not capture ecological factors such as food availability, predator-prey dynamics, or specific human disturbances like beach cleaning or construction. These unmeasured variables could play a significant role in influencing waterfowl behavior and populations.
- Assumptions of Linear Relationships: The study assumes linear relationships between predictors and waterfowl counts. However, ecological systems are inherently complex and often involve non-linear interactions. Important difference may have been overlooked due to this simplification.
- Categorical Variable Subjectivity: The use of categorical variables for factors like wave action and water clarity introduces observer subjectivity. Different observers may interpret these categories inconsistently, which could affect data reliability. Observer bias in categorizing variables such as water clarity and wave action could influence the consistency of the dataset.
- Spatial Variables and Site-Specific Factors: While beach names serve as proxies for spatial variation, they may not fully capture site-specific characteristics such as habitat type, pollution levels, or human activity intensity. These unmeasured factors could confound the observed relationships.

- Temporal Resolution: The temporal granularity of the dataset limits understandings for short-term ecological dynamics, such as diurnal activity patterns or responses to sudden weather changes. Incorporating finer temporal data could improve the robustness of the analysis.
- Data Exclusions Due to Missing Values: Variables with significant missing data, such as rain amount and turbidity, were excluded from the analysis. This exclusion may result in a partial understanding of environmental influences.
- Model Limitations: The Bayesian linear regression framework, while effective for many applications, may not capture all complexities of ecological data. For instance, hierarchical models or machine learning approaches might better account for nested structures or interactions among predictors.

5.4 Future Research Directions

Future research can build on these findings by addressing the limitations. Incorporating **additional predictors**, such as food availability, predator presence, and human disturbance metrics, would provide a more complete view of waterfowl ecology. For example, linking waterfowl counts with fish or invertebrate abundance data could illuminate trophic interactions.

The application of **non-linear modeling approaches**, such as hierarchical Bayesian models or machine learning techniques, could capture more complex relationships and interactions among predictors. These methods could also explore how the effects of certain predictors vary across beaches or seasons.

Longitudinal studies focused on specific **beaches with declining waterfowl trends** could offer opinions on localized threats and solutions. For example, identifying whether habitat degradation or pollution is a primary driver at a specific location could guide targeted restoration efforts. Additionally, **collaborative citizen science initiatives** could enrich the dataset by providing broader spatial and temporal coverage, enhancing accuracy and engagement in ecological monitoring (Storey et al. 2016).

Finally, the impact of climate change on waterfowl populations warrants further investigation. Studies incorporating **climate data**, such as precipitation, temperature extremes, and wind patterns, could identify how these broader factors interact with local conditions to shape waterfowl behaviors. Such research would be vital for developing adaptive conservation strategies in a changing climate.

This study underscores the delicate balance between human activity and ecological integrity in urban environments. By integrating environmental, temporal, and spatial factors, it contributes to a deeper understanding of urban waterfowl ecology and lays the foundation for future work aimed at sustainable urban planning and biodiversity conservation.

A Appendix

A.1 Data Source and Collection

The dataset utilized in this study, Toronto Beaches Observations, is an open-access resource provided by the City of Toronto through its open data portal (Parks, Forestry & Recreation 2024). This dataset is part of an initiative to promote transparency and accessibility, offering important environmental and ecological data collected across Toronto's beaches. It encompasses a wide range of variables, including environmental conditions (e.g., water clarity, wave action, water temperature), temporal factors (e.g., date, year, and month), and observations of waterfowl counts.

A.2 Sampling Strategy

The data collection process involves routine observations conducted by city staff or designated environmental agencies. Sampling is likely performed using systematic observation techniques, where inspectors visit specific beaches regularly to record environmental conditions and waterfowl counts. This approach ensures temporal coverage, capturing trends and seasonal variations. However, the systematic nature of the data collection may introduce sampling biases. For instance, observations might coincide with particular times of the day, potentially missing diurnal variations in waterfowl activity.

The dataset does not explicitly detail the exact methods used to count waterfowl. It is plausible that visual counts were conducted by trained personnel, but factors such as observer variability, weather conditions, and the availability of counting technologies (e.g., drones or automated sensors) could influence the accuracy and consistency of the data.

A.3 Observational Challenges and Limitations

- Observer Subjectivity: Some variables, such as "wave action" (categorized as none, low, moderate, or high) and "water clarity" (categorized as clear, cloudy, or unknown), are inherently subjective. Differences in interpretation across observers may lead to inconsistent categorizations.
- Data Gaps: The dataset contains missing or incomplete data for certain variables, which necessitated imputation or exclusion during data cleaning. For instance, variables such as "rain amount" and "turbidity" had significant gaps and were excluded from this analysis to maintain data integrity.
- Temporal and Spatial Limitations: While the dataset provides extensive spatial and temporal coverage, the observational schedule might not fully capture short-term ecological dynamics, such as sudden weather changes or predator disturbances. Additionally,

observations are limited to designated beaches, potentially overlooking other important waterfowl habitats in Toronto.

A.4 Simulation and Methodological Considerations

To address gaps and validate findings, simulation-based approaches were considered. For example, the study explored using bootstrapping techniques to estimate confidence intervals for waterfowl counts across various environmental conditions. Simulations also tested the robustness of predictors, such as water clarity and wave action, under hypothetical scenarios (e.g., assuming more frequent "clear water" conditions or higher "wave action").

A.5 Linkages to Literature

This study draws from various interdisciplinary opinions to provide a analysis of waterfowl counts in urban environments. McKinney's research on urbanization and biodiversity underpins the study's finding of environmental factors such as water clarity and wave action, offering vital context for understanding urban ecological dynamics (McKINNEY 2002). The Bayesian modeling framework employed aligns with the methodologies outlined by Gelman et al. (2013), ensuring accurate parameter estimation while incorporating uncertainty—a key consideration in ecological studies. Furthermore, Nudds and Bowlby (1984) emphasis on capturing seasonal and annual trends in waterfowl populations informed the inclusion of temporal variables like year and month in the model. Finally, the utilization of the Toronto Beaches Observations dataset exemplifies advocacy of Bonney et al. (2009) for citizen science as a means to enhance data accessibility and ecological research. These linkages highlight the study's grounding in established literature while advancing methodologies to address urban ecological challenges.

A.6 Future Recommendations for Data Collection

- Standardization of Variables: Implementing clear operational definitions and training for observers can reduce subjectivity in variables like wave action and water clarity.
- Technological Integration: Using drones, automated sensors, or image recognition software for waterfowl counting can improve accuracy and consistency.
- Expanded Temporal Sampling: Collecting data at varied times of the day and during extreme weather events can provide a more complete understanding of ecological dynamics.
- Collaborative Efforts: Encouraging citizen science participation can enhance spatial and temporal coverage while fostering community engagement in ecological monitoring.

By addressing these challenges and integrating techniques, future iterations of the Toronto Beaches Observations dataset can provide even more robust support for urban ecological research.

References

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

Bonney, Rick, Caren B. Cooper, Janis Dickinson, Steve Kelling, Tina Phillips, Kenneth V. Rosenberg, and Jennifer Shirk. 2009. "Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy." *BioScience* 59 (11): 977–84. https://doi.org/10.1525/bio.2009.59.11.9.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://github.com/sfirke/janitor.

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. 2013. *Bayesian Data Analysis*. 0th ed. Chapman; Hall/CRC. https://doi.org/10.1201/b16018.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2024. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. https://www.jstatsoft.org/v40/i03/.

McKINNEY, Michael L. 2002. "Urbanization, Biodiversity, and Conservation." *BioScience* 52 (10): 883. https://doi.org/10.1641/0006-3568(2002)052%5B0883:UBAC%5D2.0.CO;2.

Murgui, Enrique, and Marcus Hedblom. 2017. *Ecology and Conservation of Birds in Urban Environments*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-43314-1.

Nudds, Thomas D., and James N. Bowlby. 1984. "Predator–Prey Size Relationships in North American Dabbling Ducks." *Canadian Journal of Zoology* 62 (10): 2002–8. https://doi.org/10.1139/z84-293.

Parks, Forestry & Recreation. 2024. "Open Data Dataset." *About Toronto Beaches Observations*. https://open.toronto.ca/dataset/toronto-beaches-observations/.

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

Rubert, B., J. O. Branco, G. H. C. Barrilli, D. C. Melo, and A. P. Ferreira. 2021. "Behavioral Aspects of Waterbirds." *Brazilian Journal of Biology* 81 (1): 164–77. https://doi.org/10.1590/1519-6984.225048.

Storey, Richard G., Aslan Wright-Stow, Elsemieke Kin, Robert J. Davies-Colley, and Rebecca Stott. 2016. "Volunteer Stream Monitoring: Do the Data Quality and Monitoring Experience Support Increased Community Involvement in Freshwater Decision Making?" *Ecology and Society* 21 (4): art32. https://doi.org/10.5751/ES-08934-210432.

Thompson, Stephanie A., and J. Jordan Price. 2006. "Water Clarity and Diving Behavior in Wintering Common Loons." *Waterbirds* 29 (2): 169–75. https://doi.org/10.1675/1524-4695(2006)29%5B169:WCADBI%5D2.0.CO;2.

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

Zedler, Joy B., and Suzanne Kercher. 2005. "WETLAND RESOURCES: Status, Trends, Ecosystem Services, and Restorability." *Annual Review of Environment and Resources* 30

(1): 39–74. https://doi.org/10.1146/annurev.energy.30.050504.144248.