Understanding Factors Influencing Waterfowl Counts on Toronto Beaches[*]

A Bayesian Regression Approach Incorporating Environmental and Temporal Predictors

Yizhe Chen

November 27, 2024

This paper investigates the environmental and temporal factors influencing waterfowl counts on Toronto beaches using a Bayesian regression model. Key predictors include weather conditions, beach-specific characteristics, and time-related factors such as year and month. The analysis reveals significant associations between waterfowl counts and these variables, providing insights for ecological management and conservation efforts

Table of contents

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

1 Introduction

Urban environments, including Toronto's beaches, serve as critical habitats for waterfowl, supporting biodiversity while offering recreational spaces for humans. Understanding the factors influencing waterfowl populations is vital for ecological conservation, urban planning, and environmental management. Despite their importance, there is limited research on how environmental, temporal, and spatial factors interact to affect waterfowl abundance in urban settings, leaving a gap in the understanding necessary for effective conservation and management.

The estimand of this study is to find out the influence of environmental (e.g., wind speed, air temperature, water temperature), temporal (e.g., year, month), and spatial (e.g., beach name) factors on waterfowl counts at Toronto beaches. Specifically, the study seeks to determine the extent to which these predictors explain variations in waterfowl populations, contributing to a broader understanding of urban ecological dynamics.

This study employs a Bayesian regression model to analyze waterfowl count collected over several years at Toronto beaches by using data from Open Data Toronto (Parks, n.d.). The results reveal that factors such as year, month, beach location, and environmental conditions significantly impact waterfowl populations. For instance, temporal variables like year and month capture seasonal trends and long-term changes, while environmental predictors provide insights into how local conditions influence waterfowl activity.

This research is important because it bridges the gap between ecological monitoring and actionable insights, 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.

Telegraphing paragraph: The remainder of this paper is structured as follows. Section 2....

2 Data

2.1 Overview

This study utilizes observational data (Parks, n.d.) 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 comprehensive 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 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:

2.2.1 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.

2.2.2 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.

2.2.3 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.

2.2.4 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.

2.3 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.

2.4 Predictor variables

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

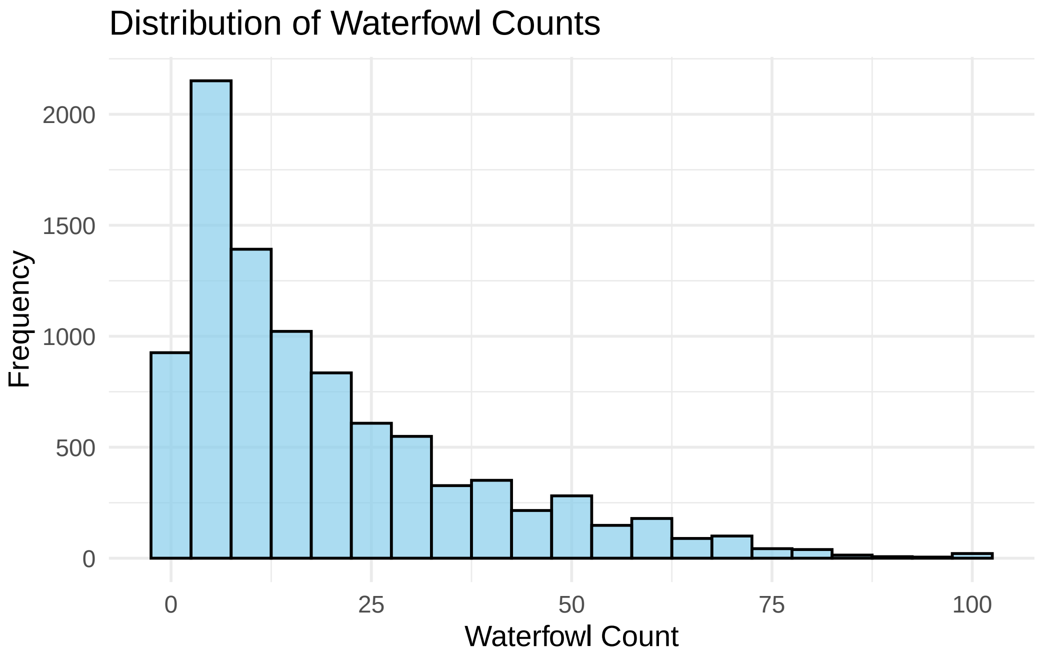

Figure 1: Distribution of waterfowl counts across beaches

2.4.1 Environmental Variables

2.4.1.1 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.

2.4.1.2 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 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.

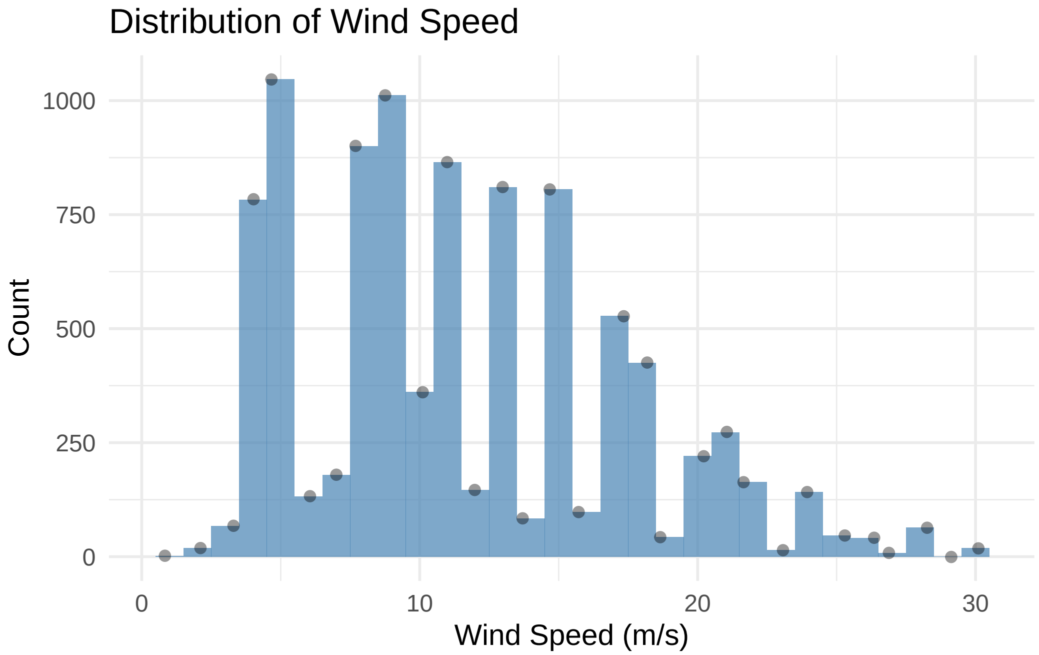

Figure 2: Distribution of wind speeds across all observations.

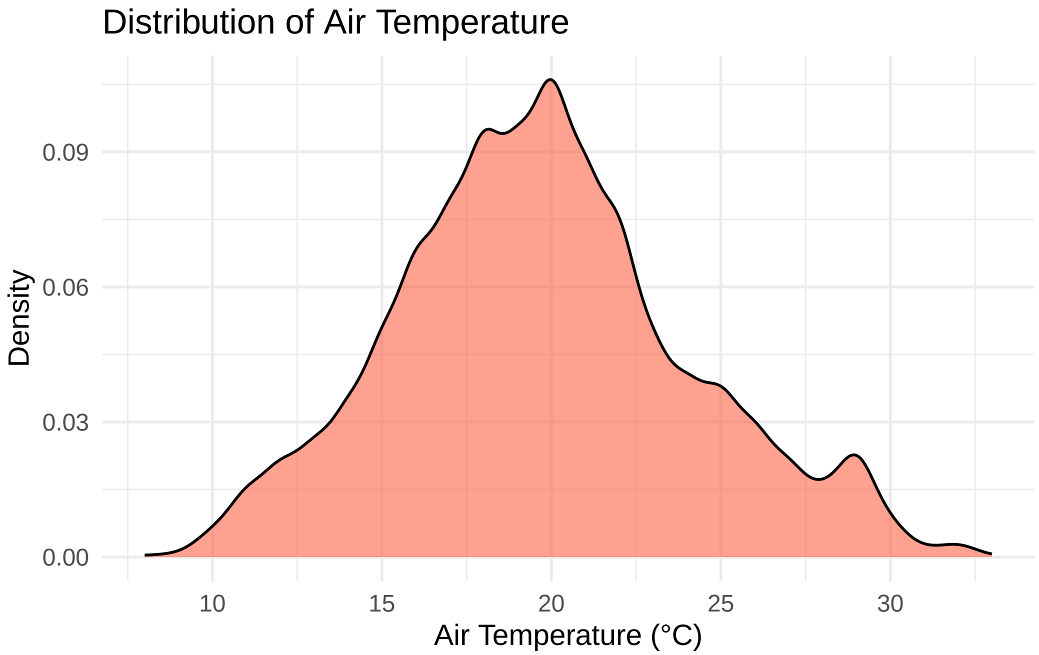

Figure 3: Distribution of air temperature across all observations.

2.4.1.3 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.

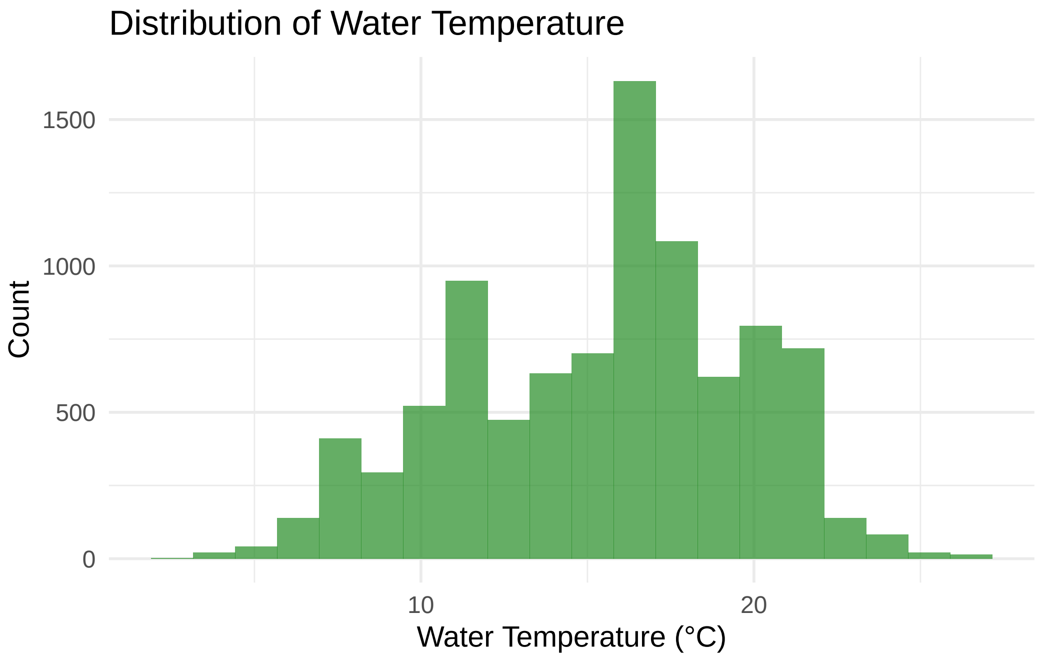

Figure 4: Distribution of water temperatures across all observations.

2.4.2 Temporal Variables

2.4.2.1 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 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.

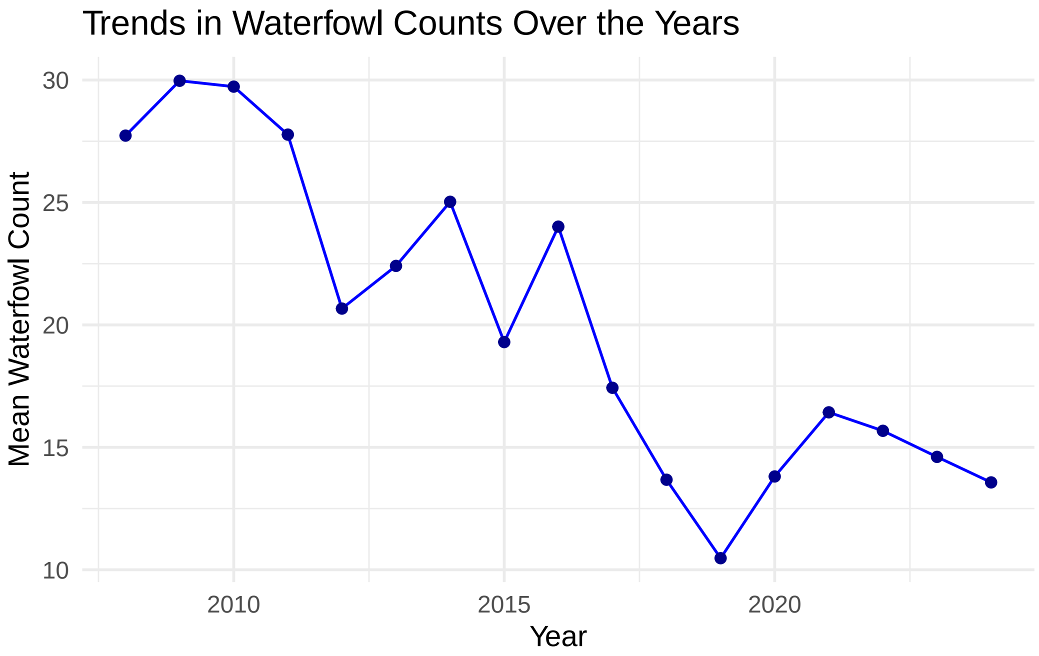

Figure 5: Waterfowl counts over the years.

2.4.2.2 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.

###Spatial Variables

2.4.2.3 Beach Name

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

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

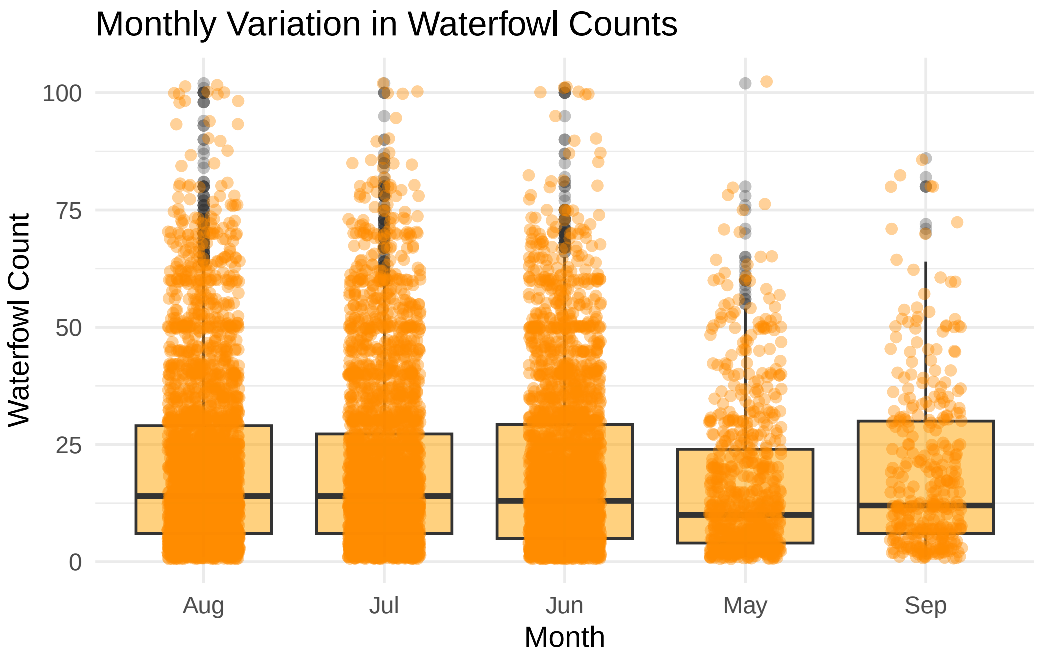

Figure 6: Monthly variation in waterfowl counts.

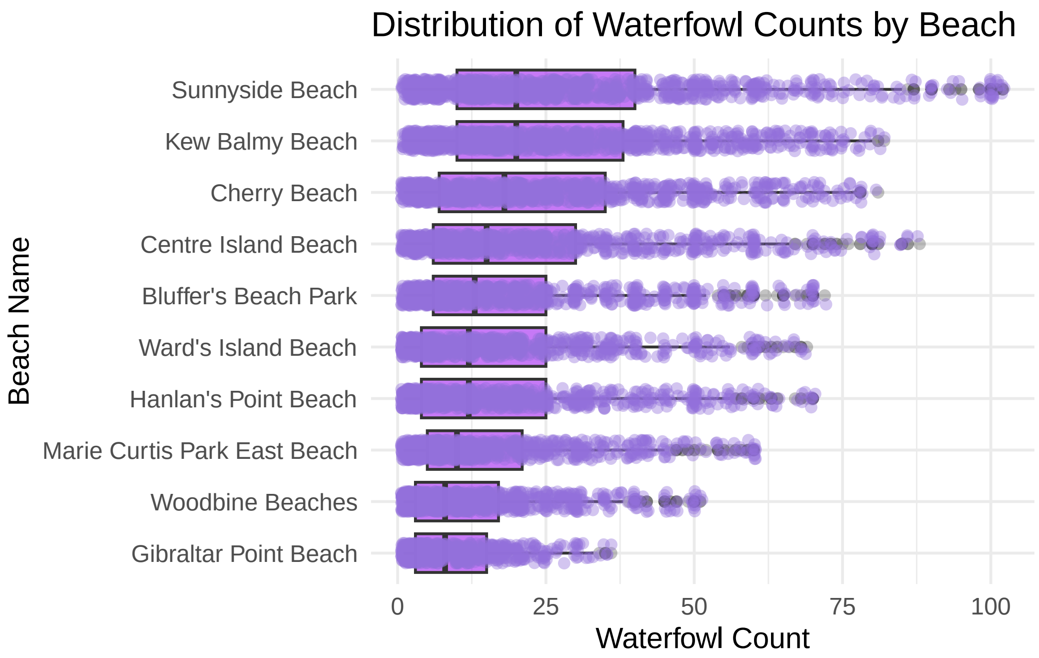

Figure 7: Distribution of waterfowl counts by beach.

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 comprehensive overview of the dataset. These factors will be further explored in the modeling section to assess their significance and impact.

3 Model

The goal of our modelling strategy is twofold. Firstly,...

Here we briefly describe the Bayesian analysis model used to investigate... Background details and diagnostics are included in Appendix B.

3.1 Model set-up

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

$$
\begin{aligned}
y_i|\mu_i, \sigma &\sim \text{Normal}(\mu_i, \sigma) && (1)\\
\mu_i &= \alpha + \beta_i + \gamma_i && (2)\\
\alpha &\sim \text{Normal}(0, 2.5) && (3)\\
\beta &\sim \text{Normal}(0, 2.5) && (4)\\
\gamma &\sim \text{Normal}(0, 2.5) && (5)\\
\sigma &\sim \text{Exponential}(1) && (6)
\end{aligned}
$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of (**rstanarm?**). We use the default priors from `rstanarm`.

3.1.1 Model justification

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

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

4 Results

Our results are summarized in **?@tbl-modelresults**.

5 Discussion

5.1 First discussion point

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

5.2 Second discussion point

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

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

B.1 Posterior predictive check

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

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

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

Figure 8: **?(caption)**

B.2 Diagnostics

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

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

Checking the convergence of the MCMC algorithm

Figure 9: **?(caption)**

References

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

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/.

Parks, Forestry & Recreation. n.d. "Open Data Dataset." *About Toronto Beaches Observations*. https://open.toronto.ca/dataset/.

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