Predicting the Water Temperature of Toronto Beaches*

Richard Guo

December 2, 2024

This paper analyzes the efficacy of using easily accessible weather information to predict the water temperature at a beach. Data recorded by City of Toroto staff stationed at various beachfronts between mid May and mid September is used. A linear regression model was used with air temperature, recent rainfall, and the location of the beach as the causal predictors.

Table of contents

*Code and data are available at: https://github.com/Richard-Guo1/predicting-beach-water-temperature.

1 Introduction

My estimand is the efficacy of using easily accessible weather information (such as from watching a news channel) to predict the temperature of the water at a beach. By analyzing this dataset, I aim to replicate a portion of the decision making process for when people decide to swim at a beach.

The result of the analysis done shows that there is a strong correlation between local air and water temperatures, with an increase in one degree Celsius in the air predicting an increase of 0.359 in the water.

Telegraphing paragraph: The remainder of this paper is structured as follows. Section 2 discusses how the dataset was obtained and analyzed. Section 3 discusses the model chosen to analyze the data. **?@sec-results** examines the results of the observations. Section 5 contains thoughts on swimming location comparisons, as well as weaknesses and next steps. Section A is an appendix. Section B contains references.

2 Data

2.1 Overview

The data used in this paper is derived from Open Data Toronto and is read into this paper through the opendatatoronto library (Gelfand 2022). The particular data set used to analyze the observations made by city staff on the conditions of all guarded Toronto beaches between the months of May and September (Toronto Department of Parks, Forestry and Recreation 2024). All the data analysis was done through R (R Core Team 2023) with the aid of the following packages: tidyverse (Wickham et al. 2019), arrow (Richardson et al. 2024), fastDummies(Kaplan 2023), here (Müller 2020), dplyr (Wickham et al. 2023), tibble (Müller and Wickham 2023), janitor (Firke 2024), ggplot2 (Wickham 2016), and knitr (Xie 2024).

2.2 Measurement

The measurement of each variable in the beach observation dataset was conducted by city of Toronto staff. Thermometers were used to record the temperature of both the water and the air. Some omitted datatypes such as a water fowl count were estimated due to lack of proper equipment. All observations made were then recorded.

2.3 Outcome variable

The variable we are interested in is the water temperature. Figure 1 shows that temperatures observed between mid May and September vary from slightly over 0 to slightly over 30 degrees Celsius. The majority of observations are concentrated between 10 and 20 degrees Celsius.

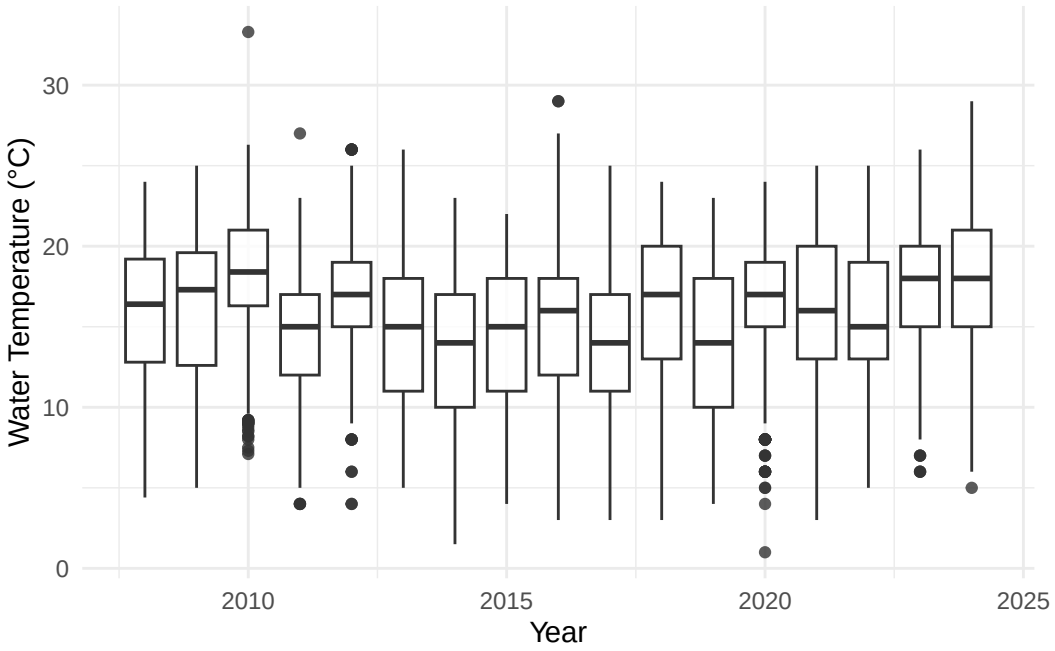

Figure 1: Water Temperature Quantiles by Year

2.4 Predictor variables

The chosen predictor variables from the dataset are the beach name, the air temperature, and the presence of rain. The rain amount category may have provided more accurate estimates, but was dropped due to lack of proper documentation, discussed more in (**additional?**).

The main predictor of water temperature is the surrounding air temperature. Figure 2 shows a strong positive correlation between the two variables. Figure 3 shows that each beach has a different median temperature. Figure 4 shows that there is a lower median air temperature when it is raining while Figure 5 shows a larger variance but similar median on comparison. This leads to the addition of an interaction predictor between rain and air temperature.

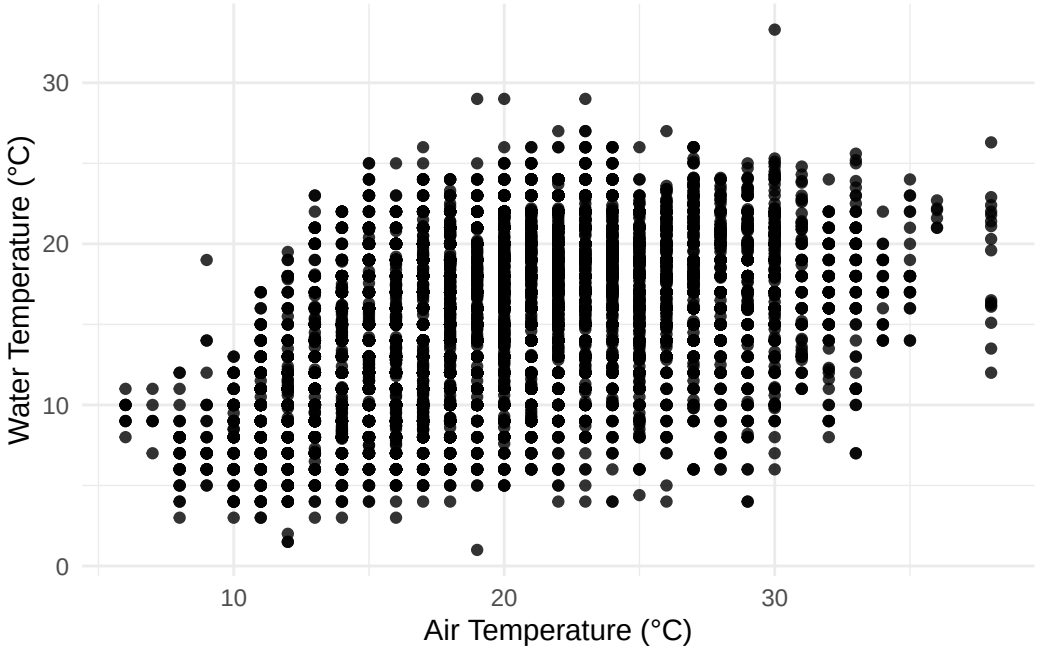

Figure 2: Relationship between Air Temperature and Water Temperature

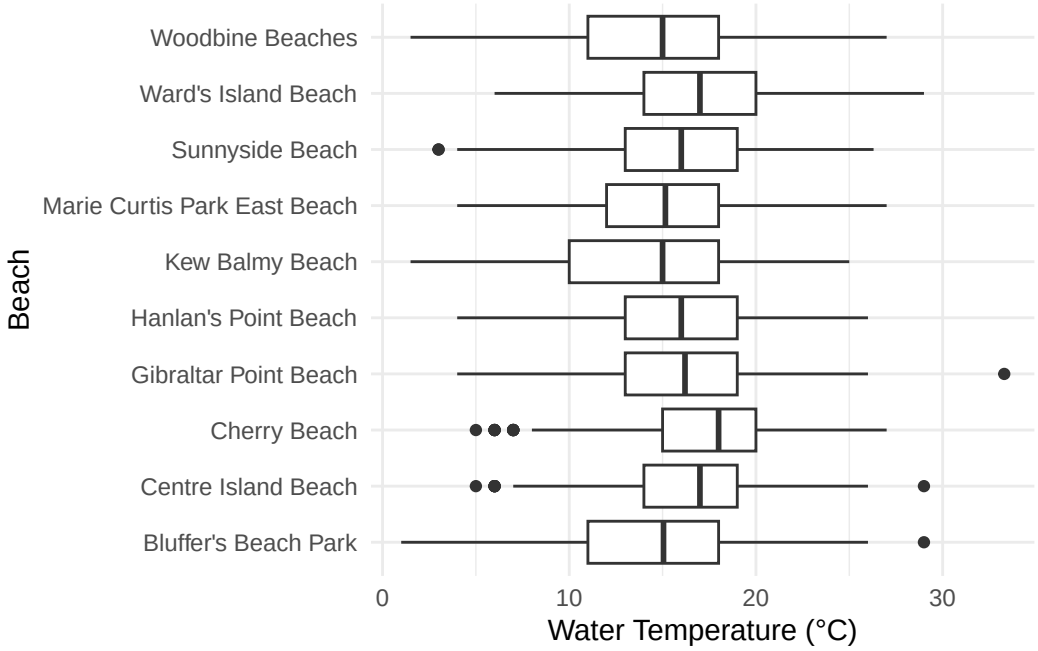

Figure 3: Water Temperature Quantiles by Beach

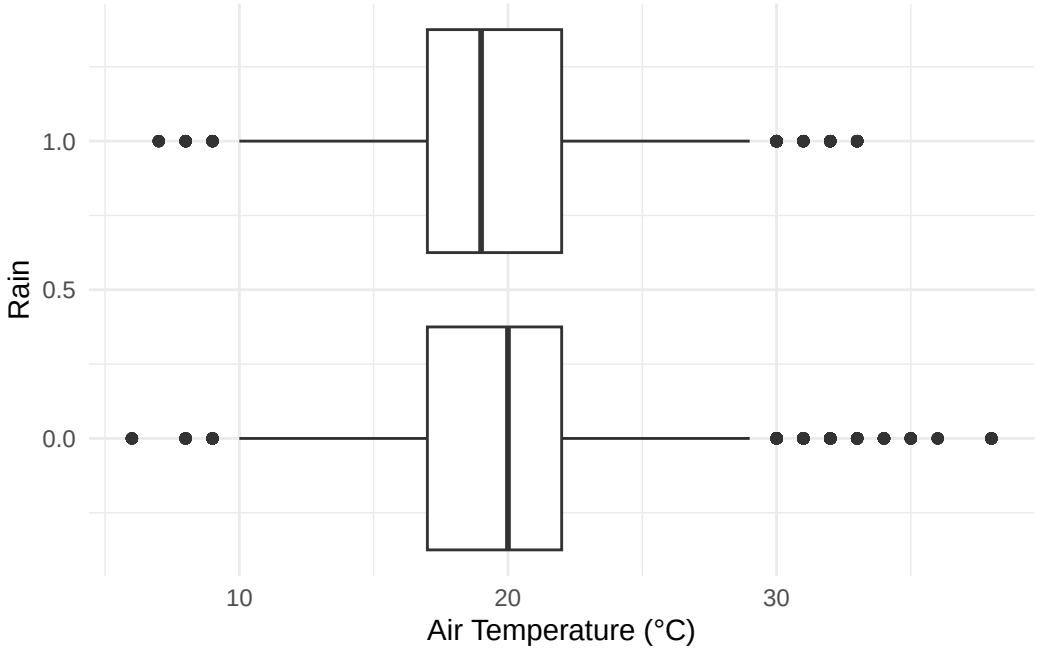

Figure 4: Air and Water Temperature with Rain

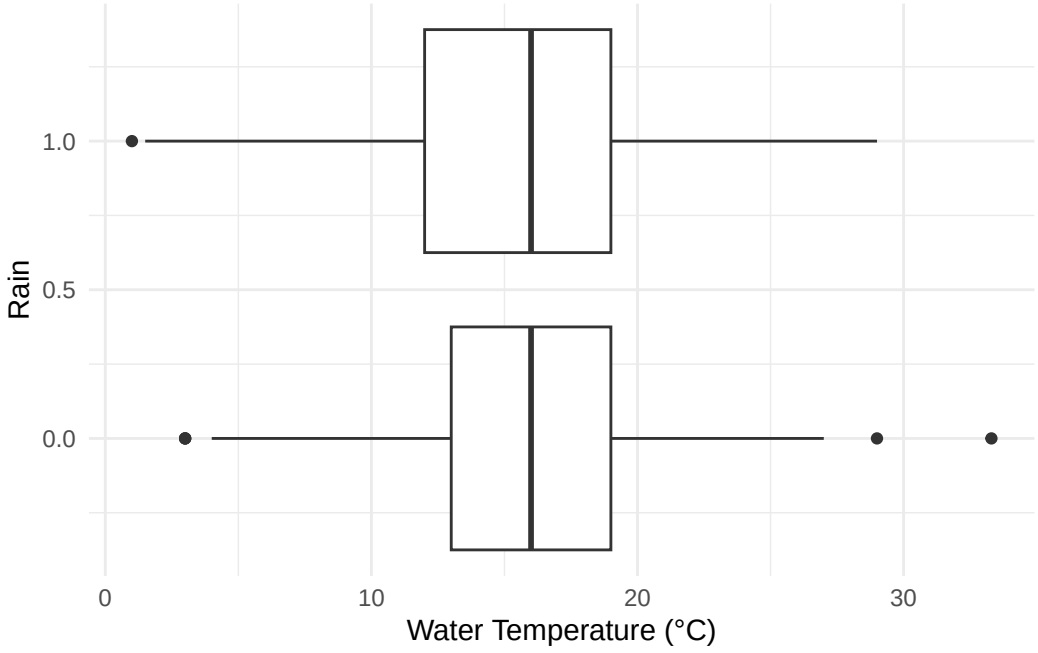

Figure 5: Air and Water Temperature with Rain

3 Model

3.1 Model set-up

3.1.1 Model justification

4 Results

Our results are summarized in Table 1.

5 Discussion

5.1 Beaches, Outdoor Pools, and Indoor Pools

5.2 Weaknesses and next steps

Table 1: Explanatory models of water temperature based on air temperature, rain, and location

	Water Temperature
(Intercept)	7.547
	(0.185)
rain	−1.737
	(0.280)
airTemp	0.359
	(0.008)
isCentreIsland	1.744
	(0.135)
isCherry	2.746
	(0.134)
isGibraltarPoint	1.325
	(0.135)
isHanlansPoint	0.986
	(0.135)
isKewBalmy	−0.382
	(0.134)
isMarieCurtis	0.403
	(0.135)
isSunnyside	0.985
	(0.133)
isWardsIsland	2.049
	(0.134)
isWoodbine	−0.313
	(0.134)
rain × airTemp	0.092
	(0.014)
Num.Obs.	15 911
R2	0.230
R2 Adj.	0.230
Log.Lik.	−43 748.536
RMSE	3.78

A Appendix

B Additional data details

References

Firke, Sam. 2024. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://github.com/sfirke/janitor.

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*.https://sharlagelfand.github.io/opendatatoronto/, https://github.com/sharlagelfand/opendatatoronto/.

Kaplan, Jacob. 2023. *fastDummies: Fast Creation of Dummy (Binary) Columns and Rows from Categorical Variables*.

Müller, Kirill. 2020. *Here*. https://here.r-lib.org/index.html.

Müller, Kirill, and Hadley Wickham. 2023. *Tibble: Simple Data Frames*. https://tibble.tidyverse.org/, https://github.com/tidyverse/tibble.

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://github.com/apache/arrow/.

Toronto Department of Parks, Forestry and Recreation. 2024. *Toronto Beach Observations*. https://open.toronto.ca/dataset/toronto-beaches-observations/.

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

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org.

Xie, Yihui. 2024. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.