

the data of Bodysafe*

qizhou xie

Invalid Date

1

Table of contents

1	Introduction	1
2	Data	2
2.1	Overview	2
2.2	Results	3
A	Appendix	5
A.1	Dataset and Graph Sketches	5
A.2	Data Cleaning	5
A.3	Attribution Statement	5
	References	6

1 Introduction

In March 2020, the World Health Organization proclaimed that the spread of COVID-19 constituted a pandemic (**whocovid?**). Shortly thereafter, in May 2020, the COVID-19 mortality rate in Canadian long-term care homes, where individuals live and receive healthcare on a continuous basis for reasons such as “advancing age, disability or declining health” (**ltchdefinition?**), reached approximately 35% and represented 81% of Canadian COVID-19 deaths (**cihi?**). In general, the risk of disease circulation is greater in healthcare settings

*A GitHub Repository containing all data, R code, and other files used in this investigation is located here:
<https://github.com/tx77777/marrige>

due to an increased presence of ill individuals and nearer physical contact between patients and staff (**cdc?**).

Consequently, with the recent advent of large volumes of respiratory diseases such as influenza and COVID-19 in Toronto (**covidflu?**), assessing the prevalence of disease outbreaks in healthcare facilities in Toronto is of paramount importance. According to (**tphdataset?**), an outbreak is “a localized increase (e.g. in an institution, or a specific ward or floor within an institution) in the rate of infection or illness, above that which is expected”. Thus, this investigation examines the number of outbreaks by location type (long-term care homes, retirement homes, hospitals, and transitional care facilities), outbreak type (respiratory and enteric/gastroenteric, the latter referring to intestinal and/or stomach illness (**enteric?**; **gastroenteric?**)), and first known cause (e.g., COVID-19, influenza A, etc.) to fill this knowledge gap.

To do this, 2023 Toronto Public Health data on outbreaks in healthcare facilities was first obtained as described in Section 2.1. Based on this dataset, it was discovered that long-term care homes comprised a majority of the outbreaks, that approximately 95% of the outbreaks were respiratory, and that almost two-thirds of the outbreaks had COVID-19 as their first known cause (Section 2.2). As discussed in **?@sec-discussion**, the high prevalence of outbreaks in long-term care homes has likely been exacerbated by the removal of a universal masking requirement, suggesting that the Government of Ontario should consider reinstating this policy to protect these vulnerable individuals. With respect to the structure of this paper, Section 2 includes an overview of the data and the results; **?@sec-discussion** includes a discussion of the results; and Section A includes supplementary information.

2 Data

2.1 Overview

The dataset used in this analysis is the 2023 installment of “Outbreaks in Toronto Healthcare Institutions” from Toronto Public Health (**tphdataset?**). In Ontario, the law mandates that respiratory and gastroenteric infections in healthcare facilities are followed and that any outbreaks, regardless of verification, are reported to public health entities such as Toronto Public Health (**tphdataset?**). Based on this information (or these measurements), Toronto Public Health updates this dataset on a weekly basis (**tphdataset?**). Moreover, this dataset is considered to be “open data” (**tphdataset?**) and can be utilized for a variety of purposes as long as an attribution statement is attached (Section A.3) and the link to the City of Toronto’s Open Data License is added (**tphlicense?**).

The variables or measurements included in this analysis are “Type of Location” (named “Outbreak Setting” in the original dataset), which refers to the nature of the healthcare offered at each facility; “Type of Outbreak”, which indicates whether each outbreak is respiratory, enteric, or other in nature; and “Outbreak First Known Cause” (named “Causative Agent-1”

in the original dataset), which denotes the first discovered pathogen based on at least one of the outbreak’s cases (**tphdataset?**).

Only two other datasets involving outbreaks, titled “COVID-19 Cases in Toronto” and “Daily Shelter & Overnight Service Occupancy & Capacity”, could be identified following a search of the City of Toronto’s Open Data Catalogue (**tphsearch?**). However, none of these datasets pertain specifically to healthcare settings in Toronto and were thus not used in this analysis.

Using the R programming language (R Core Team 2023), the **janitor** (**citejanitor?**) and **tidyverse** (**citetidyverse?**) packages were used to simulate the dataset and generate tests for it. The **opendatatoronto** (**citeopendatatoronto?**) and **tidyverse** (**citetidyverse?**) packages were then applied in order to download the raw Toronto Public Health dataset. Next, the **tidyverse** package (**citetidyverse?**) was used to clean the raw dataset and test the cleaned dataset.

2.2 Results

```
# A tibble: 1 x 11
  title      id      topics civic_issues publisher excerpt dataset_category
  <chr>    <chr>    <chr>  <chr>          <chr>    <chr>    <chr>
1 BodySafe c4052ff7-a09d~ City ~ <NA>          Toronto ~ This d~ Map
# i 4 more variables: num_resources <int>, formats <chr>, refresh_rate <chr>,
#   last_refreshed <date>
```

```
# A tibble: 9 x 4
  name      id      format last_modified
  <chr>    <chr>    <chr>    <date>
1 bodysafe 315f0f9f-cbf0-4b95-b8a5-a4afda0f~ GeoJS~ NA
2 bodysafe - 4326.geojson 9a9fbb34-308f-44d1-bc28-8db99eab~ GeoJS~ 2024-09-24
3 bodysafe - 2952.geojson 02ab8c29-be77-4296-bac5-fddc92fd~ GeoJS~ 2024-09-24
4 bodysafe - 4326.csv 05f7edf5-df38-4ca8-b46d-fcd40b10~ CSV 2024-09-24
5 bodysafe - 2952.csv 957ffc22-da39-4e69-be2a-20afab55~ CSV 2024-09-24
6 bodysafe - 4326.zip 33f4cf13-aab0-4dbe-9b6d-be871736~ SHP 2024-09-24
7 bodysafe - 2952.zip a17793b4-ccac-4096-972e-93371632~ SHP 2024-09-24
8 bodysafe - 4326.gpkg 6c6dbd9b-9550-4a17-aeaf-1cb5542d~ gpkg 2024-09-24
9 bodysafe - 2952.gpkg 77ee9801-eea2-410c-9ab6-ff844904~ gpkg 2024-09-24
```

After loading the dataset using the R programming language (R Core Team 2023) and the **here** package (**citehere?**), the **tidyverse** (**citetidyverse?**) package was used to generate graphs. In doing so, R code was adapted from (**tellingstorieswithdata?**).

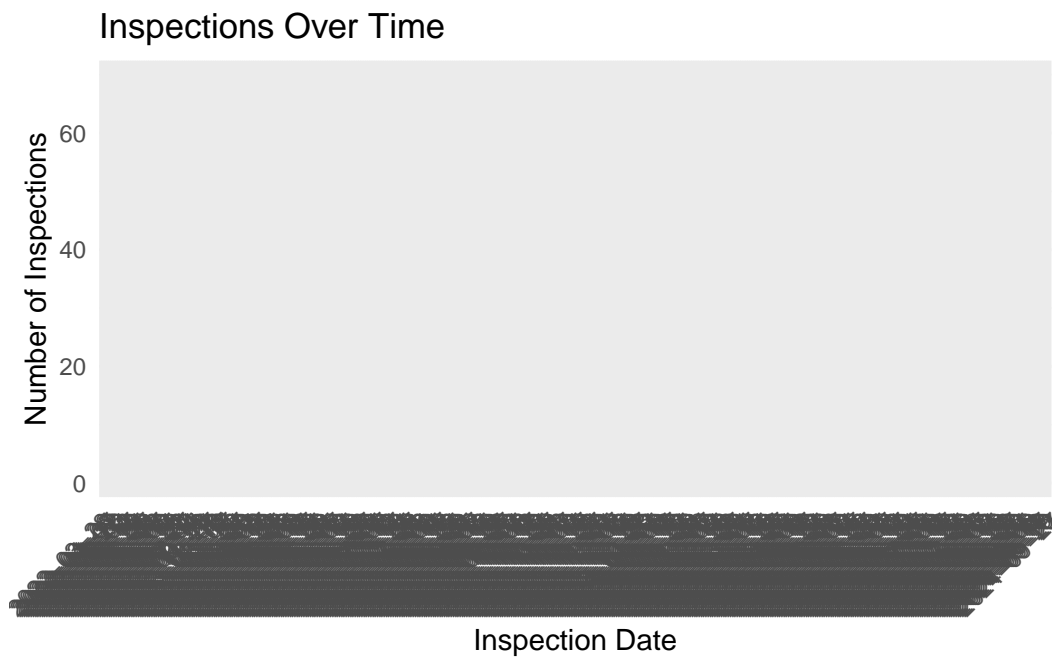
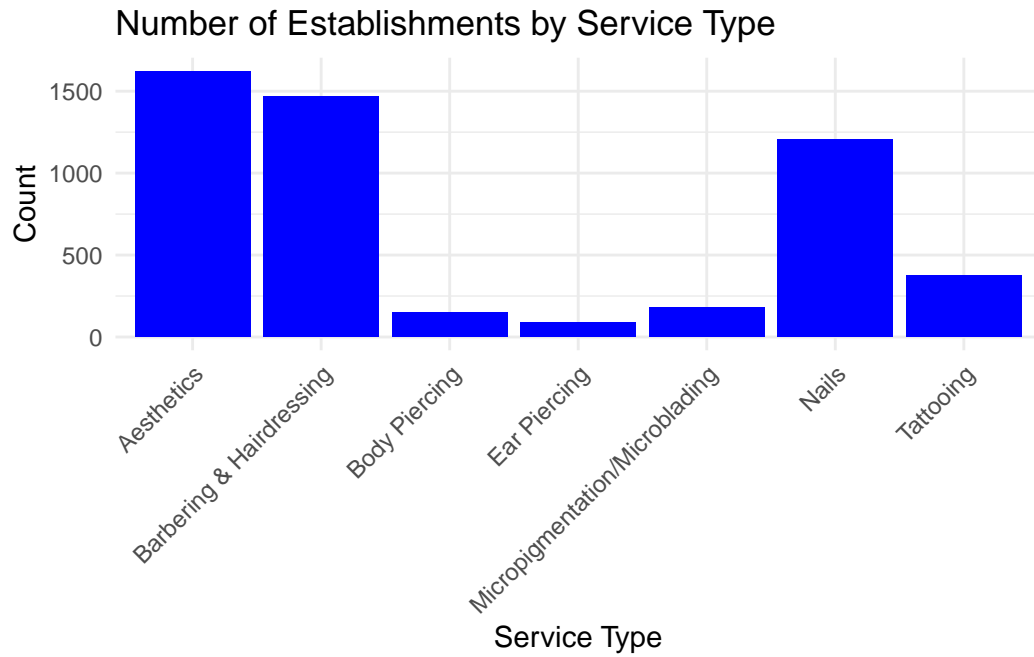


Figure 1: Number of outbreaks by type of healthcare location in Toronto in 2023

A Appendix

A.1 Dataset and Graph Sketches

Sketches depicting both the desired dataset and the graphs generated in this analysis are available in the GitHub Repository.

A.2 Data Cleaning

The data cleaning process involved filtering out some of the columns from the raw dataset and renaming some of the data entries for clarity and simplicity.

A.3 Attribution Statement

“Contains information licensed under the Open Government Licence – Toronto” (**tphlicense?**).

References

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