Marriage Data analysis*
STA304 WEEK 3

Maggie Zhang

September 27, 2024

This paper investigates the occupancy of shelters in Toronto from 2017 to 2020, focusing on capacity and usage across different shelter sectors. By analyzing the data, we explore how shelter demand {exceeded available resources in certain sectors. This research highlights significant challenges faced by the shelter system, providing insights into the strain on resources and potential gaps in support for the homeless population}. These findings are crucial for informing policies and improving resource allocation to better address homelessness in Toronto.

1 Introduction

Homelessness is a persistent and complex issue in many areas in Canada(Gaetz (2010)), and Toronto is no exception. As one of Canada's largest cities, Toronto faces significant challenges in helping homeless people, who are unable to spend on housing largely due to poverty, to meet the necessities of living(Jadidzadeh and Kneebone (2018)). The city operates a range of shelters that serve various groups, including men, women, youth, and families. The balance between availability and the demand for these services is crucial for both the well-being of individuals experiencing homelessness and the government's efforts to reduce homelessness population. Understanding how shelter occupancy changes over time is essential for identifying gaps in resources and shaping effective policy responses.

This paper focuses on analyzing shelter occupancy data from 2017 to 2020 in City of Toronto. The data which offers a detailed look at the city's shelter during this time of period is provided on Gelfand (2022) portal. By examining trends in occupancy through the data, we aim to assess how well Toronto's shelters met the demand for space and where shortages may have occurred. By focusing on specific sectors of the shelter system, pinpoint which groups may have been disproportionately affected by resource limitations.

---

*Code and data are available at:https://github.com/MaggieZ111119/CityofToronto_Daily_Shelter_Occupancy.Rproj.git

FINDING! This study addresses a clear gap in the current literature by providing an up-to-date analysis of Toronto's shelter capacity and usage during a time of unprecedented public health and social challenges. _____ NOTICE!!

The remiander of this paper are in presented in sections: Section 2 : Data, Section 3 :Results, and Section 4 :Discussion. The data section introduce the dataset used for analysis, expalinin its source and context that shaped its collection.This section thoroughly summarizes the data using visualizations created with `ggplot2` (Wickham et al. (2024)) and tables produced with `knitr` (Xie (2024)). Key variables chosen for the analysis, along with the reasoning behind their selection, are explained. The results section presents the findings drawn from the data, highlighting patterns in shelter occupancy and compare to it's availability. Following by the discussion section, which explores the implications of these findings, particularly in relation to resource allocation and policy development, offering insights into how Toronto's shelter system responded to increasing demand.

2 Data

City that has record of offering shleter servise incluede: Toronto, North York, Scarborough, and Etobicoke.

2.1 Overview

All data reviewed and analysis in this paper is the Daily Shelter Occupancy Toronto Shelter & Support Services (2022). The data is essential for evaluating how shelters serve different populations (e.g., men, women, youth, families) and for analyzing the current shelter servise system over time, such as capacity shortfalls or surpluses. It provides inforamtion on all the active shelters exist in the City of Toronto area, collected in forms of four separate datasets corresponding to the years 2017, 2018, 2019, and 2020. Each datasets include various characteristics of the shelters. The data was sourced from an Open Data Toronto Portal Gelfand (2022), with each row indicate a unique entry for a specific sector of specific shelter, on a specific date, and has Unique row identifier for Open Data database, "_id".

The collection process records information about corresponding name of the non-profit entity that is responsible for the shelter, as well as the name of the facility (e.g. hotel, residence building), along with and their capacity to accommodate homeless clients. The Capacity is being measured as number of bed or a mat/cot available. Something will happen is that for the family sector, it is possible to exceed available capacity depending on the bed number, because family could be accommodated in a room with number beds smaller than number of people in their family. The data was collected every day, 4:00 AM, to record number of homeless clients occupying the shelters. This way to gather information about occupancy provides a consistent snapshot of shelter use across all organizations.

Other similar datasets, especialy the ther dataset publish by Toronto Shelter & Support Services, were considered. However, they did not offer the cross-sector view of both occupancy and capacity needed for the focus of this analysis. The selected dataset allows for a broad comparison across time and sectors in different city, providing a fuller picture of trends within the shelter system.

2.2 Feature Selection and Aggregation

For the purpose of this analysis, several features are chose to be particularly focused on. variable of interest including the the city (named "SHELTER_CITY" in the dataset) in which shelters are located; the date of data recorded("OCCUPANCY_DATE"); the sector("SECTOR") of clientele the shelters serve, for example men, women, youth, families, and even co-ed; the shelter's capacities("CAPACITY") and occupancy("OCCUPANCY") levels. These variables were chosen for their relevance in which is to investigate shelter occupancy and capacity across different cities and sectors, focusing on the City of Toronto from 2017 to 2020.

There are also seven other detailed feature in the datasets: "ORGANIZATION_NAME", "SHELTER_NAME", "SHELTER_ADDRESS", "SHELTER_PROVINCE", "SHELTER_POSTAL_CODE", "FACILITY_NAME", and "PROGRAM_NAME". Using R programming lanaguage R Core Team (2023), the `janitor`Firke (2023), `tidyverse` Wickham (2023), and `dplyr` Wickham et al. (2023) pakages are used in data simulation, downloading, cleaning, and writting test. No new variables were created for this analysis, but the data was aggregated by grouping observations by city, date, and sector. This aggregation facilitated a more comprehensive comparison across cities and sector groups. A random sample of the cleaned data can be seen in (**sample-shelter?**).

Table 1: Sample of Cleaned Shelter Data

occupancy_date	shelter_city	sector	total_occupancy	total_capacity
2020-09-27	Toronto	Women	370	511
2017-07-11	Toronto	Men	1741	1801
2019-08-30	Toronto	Families	2949	3177
2017-04-30	Toronto	Women	680	697
2018-04-04	Toronto	Families	2535	2661

2.3 Data Breakdown

As shown in (**shelter-full-table?**), The four cities being observed in the dataset are Toronto, Etobicoke, North York, and Scarborough. The sectors recorded include Co-ed (mixture), Families, Men, Women, and Youth.

Table 2: Number of Observations by City and Sector

shelter_city	Co-ed	Families	Men	Women	Youth	Sum
Etobicoke	0	0	0	0	1461	1461
North York	0	0	0	0	1461	1461
Scarborough	1461	0	569	632	207	2869
Toronto	1461	1461	1461	1461	1461	7305

The following table provides a summary of the average, maximum, and minimum occupancy, as well as capacity metrics for each city.

Table 3: Statistics for Homeless Shelter Occupancy by City

City	Average Occupancy	Max Occupancy	Min Occupancy	Average Capacity	Max Capacity	Min Capacity
Toronto	5443.37577	6803	3001	6042.49281	7241	4460
North York	64.50650	73	20	67.36345	73	32
Scarborough	159.58727	564	60	178.18549	577	67
Etobicoke	48.06913	53	20	50.81383	53	33

The relastion ship of two key feature, occupancy and capacity, are essential for understaidng current shelter servise system. The relashipship is reflected in Figure 1. More learning on the shelter servise and thier trend, and how they reflect the fomeless population are done, and will be discuss in the next section **?@sec-result**.
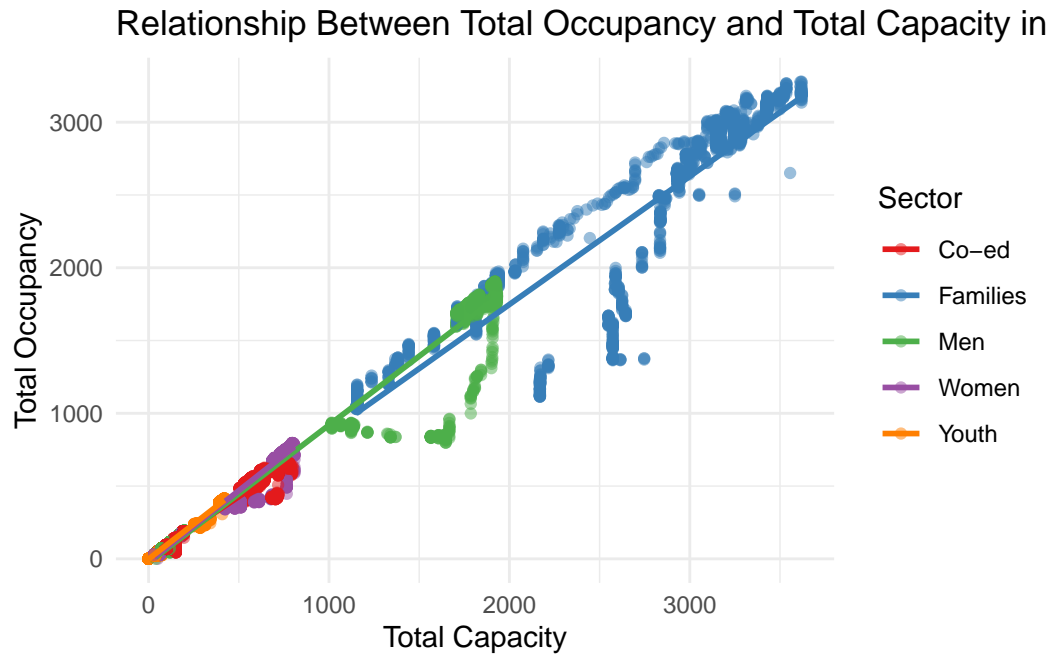
Figure 1: Relationship Between Total Occupancy and Total Capacity in Shelter of the City of Toronto

3 Results

Through phafical study of the datset, some usful information are found and the result will be dicussed in this section. These will _____

3.1 Shelter Occupancy Trend over Time
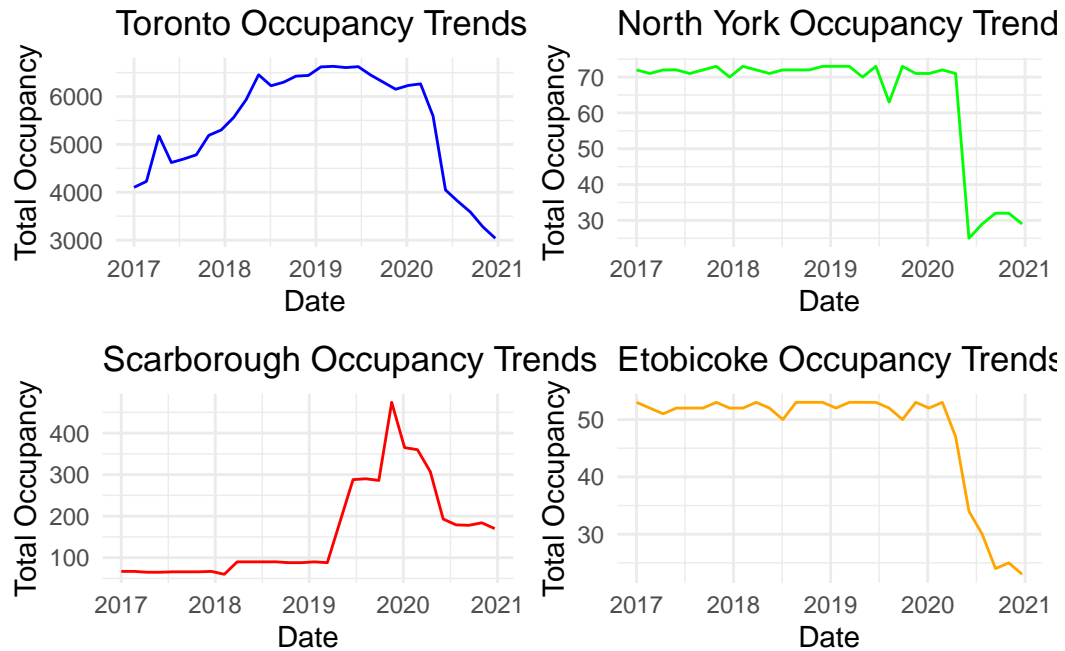
The figure

Figure 2: Visualization of Shelter _____ During Time Period 2017 to 2020

```
TableGrob (2 x 2) "arrange": 4 grobs
  z    cells     name           grob
1 1 (1-1,1-1) arrange gtable[layout]
2 2 (1-1,2-2) arrange gtable[layout]
3 3 (2-2,1-1) arrange gtable[layout]
4 4 (2-2,2-2) arrange gtable[layout]
```

3.2 Shelter Ustillization Analysis

To analyze capacity utilization for the four cities—Toronto, North York, Scarborough, and Etobicoke— calculate the capacity utilization rate for each city over time and visualize this data.

```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
```
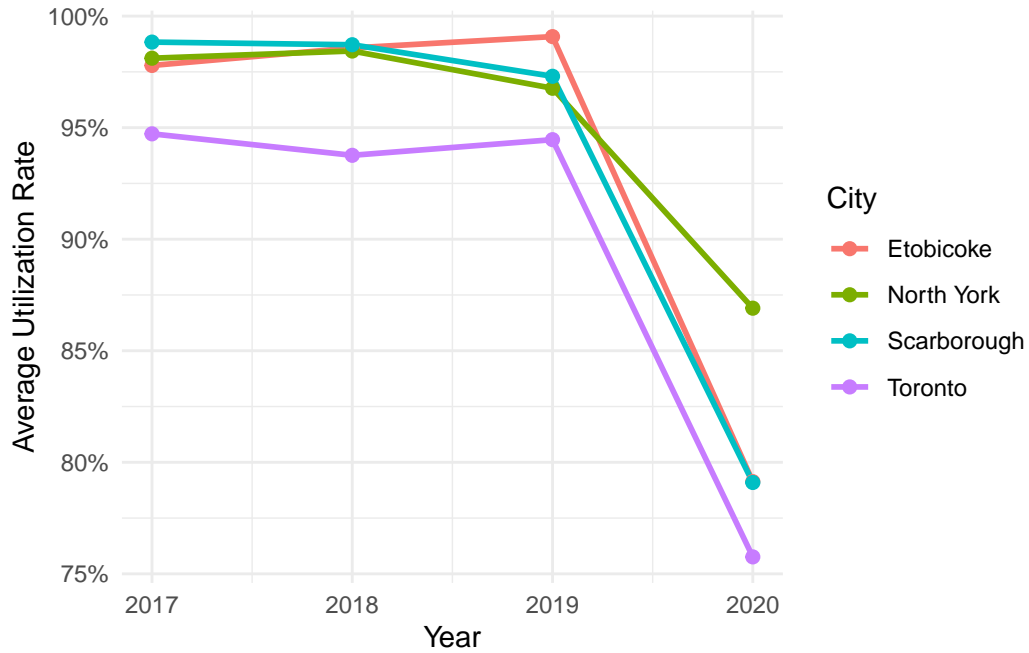
6

Figure 3: Change in Utilization Rate Average on All Shelters by City (2017-2020)

3.3 Over Capacity Cases

Back to the original dataset, identify cases where shelter capacity is exceeded (i.e., occupancy > capacity) and understand the frequency and distribution of these events.

Table 4: Overcrowded Cases by City and Sector

shelter_city	Families	Youth	Co-ed	Men	Women
Toronto	87	0	0	0	0
Etobicoke	0	0	0	0	0
North York	0	0	0	0	0
Scarborough	0	0	0	0	0

4 Discussion

4.1 First discussion point

from the study, the Marriage number across Toronto is …..

4.2 Second discussion point

[My second point]

4.3 Third discussion point

[My third point]

4.4 Weaknesses and next steps

The study is not finalized, more detailed and careful data examination is needed.

Appendix

References

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

Gaetz, Stephen. 2010. "The Struggle to End Homelessness in Canada: How We Created the Crisis, and How We Can End It." *The Open Health Services and Policy Journal* 3 (1).

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

Jadidzadeh, Ali, and Ron Kneebone. 2018. "Patterns and Intensity of Use of Homeless Shelters in Toronto." *Canadian Public Policy* 44 (4): 342–55.

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

Toronto Shelter & Support Services. 2022. "Daily Shelter Occupancy." https://open.toronto.ca/dataset/daily-shelter-occupancy/.

Wickham, Hadley. 2023. *Tidyverse: Easily Install and Load the Tidyverse.* https://tidyverse.tidyverse.org.

Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, Hiroaki Yutani, Dewey Dunnington, and Teun van den Brand. 2024. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics.* https://ggplot2.tidyverse.org.

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/.