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 remains a persistent issue across Canada, including in Toronto(Gaetz (2010)). s one of the country's largest cities, Toronto faces ongoing challenges in assisting individuals who, primarily due to poverty, cannot afford housing and struggle to meet basic needs(Jadidzadeh and Kneebone (2018)). The City of Toronto operates a network of shelters catering to men, women, youth, and families, making it essential to balance the availability of these services with demand. Monitoring shelter occupancy trends with capacity is crucial for identifying resource gaps and developing effective policies to address the needs of homeless populations.

This paper analyzes shelter occupancy data from 2017 to 2020 in the City of Toronto, sourced from the Open Data Toronto portal ((**OepnDataToronto?**)). By examining occupancy trends during this period, particularly during the COVID-19 pandemic, we aim to evaluate how effectively the city's shelters met the increased demand. Our focus on specific sectors within the shelter system will help determine which groups were disproportionately affected and shed light on the system's adaptability in response to the crisis, particularly in addressing issues of overcrowding in family shelters.

By studying the shelter occupancy data from 2017 to 2020 across four cities, key findings emerged. There was an increase in shelter occupancy from late 2019 to early 2020, which

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

coincided with the onset of the COVID-19 pandemic, particularly in Scarborough, which was observed to have higher than usual capacities of sheltering service during that period. Despite the rise in shelter use, the utilization rates remained high, indicating that the shelters were effectively meeting demand without being overburdened. The government's timely increase in shelter capacity during the pandemic was especially evident. Over Capacity sheltering experience exists, in Toronto's shelters for families, but it likely resulted from flexible accommodation arrangements rather than systemic flaws. Overall, the shelter system demonstrated well adaptability and support for homeless populations during this period.

The remainder of this paper is presented in sections: Section 2 :Data, Section 3 :Results, Section 4 :Discussion, and Section 4.3 :Weaknesses and Next Steps. The data section introduces the dataset used for analysis, explaining its source and the 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)), also with `gridExtra` (Auguie (2017)) to help formatting. 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 its availability. Followed 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

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.

2.1 Overview

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 Figure 1.

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

Figure 1: Sample of Cleaned Shelter Data

2.3 Data Breakdown

As shown in Figure 2, 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. Across all cities, record for Youth age group being presents. In fact, Youth is the only population in Etobicoke and North York. In Toronto, which has the most number of observation recorded, equal number of sectors are being observed, but how these people in distributed inthe total popluation are not clear. This will be further addressed in the following sections.

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

Figure 2: Number of Observations by City and Sector

The following table Figure 3 provides a summary of the average, maximum, and minimum occupancy, as well as capacity metrics for each city. Toronto has most amount of shelter avaliable and also has most people occuupaying in. The Maximum occupany of Toronto in a day, combined all sectors, from 2017 to 2020,is 6803. Wheras Etobicoke only has 53 people occupanying. The fact that Etotobike has only hoemless people in the group of Youth recorded could explain alittle bit. But, the de=ifference still strong enough to see that the homeless population might be significant larger in Toronto compared to aother cities.

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

Figure 3: Statistics for Homeless Shelter Occupancy by City

The relastion ship of two key feature, occupancy and capacity, are essential for understaidng

current shelter servise system. The relaship is reflected in Figure 4. Occupany and capacity are greatly follwing a positive correlation, indicating that most there aren't much over-investment to the shelter system, nor heavily over-filled shelter. But there are noticible deviation from the perfect correlation to above, these are suggecting there are cases where occupany is higher than capacity, suggesting occupantee may living in a overly paked places. More will be dicussed in Section 3, including other learning on the shelter servise like thier trend, and how these shelter information possiblly reflect the homeless.

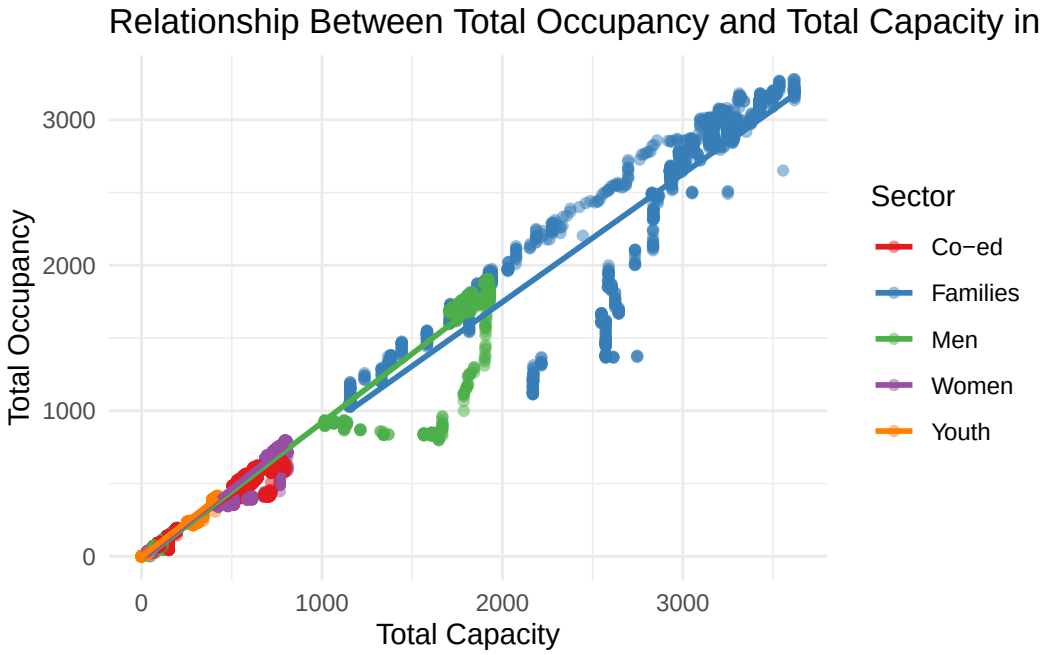

Figure 4: Relationship Between Total Occupancy and Total Capacity in Shelter of the City of Toronto

3 Results

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

3.1 Shelter Occupancy Trend over Time

The figure Figure 5 depicts the trend of homeless people occupying shelters in each city over time. A noticeable peak in occupancy is observed across all four cities between late 2019 and early 2020. Especially, there is an rising trend in Scarborough. This rise coincides with

the outbreak of COVID-19, as noted in the "About Daily Shelter Occupancy" section in the data website (**ShelterThe?**). significant decrease in occupancy from mid-2020 across all cities further supports this correlation, suggesting the pandemic had a direct impact on shelter use.

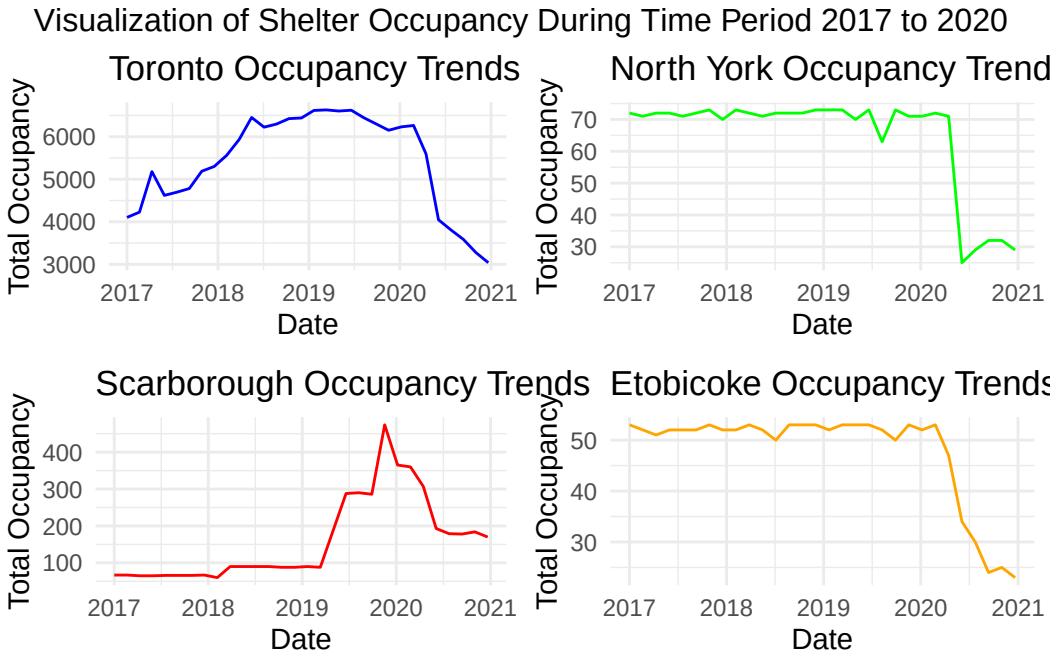

Figure 5: Visualization of Shelter Occupancy During Time Period 2017 to 2020

3.2 Shelter Utilization Analysis

The Utilization Rate as being calculated as:

$$\text{Utilization Rate} = \left(\frac{\text{Number of Occupancy}}{\text{Number of Capacity}}\right) \times 100\%$$

From Figure 6, it's the utilization rate in Toronto remains the lowest over 4 years almost all the time compared to other cities, but still in a relativly high number. This suggest that shelter servise croass all four cities are not over-inveseting, their servise are actually helping the hoemelss population since alot of people are using them. There is a significant decrease in utilization rate from Fall 2019 to Summer 2020, despite the increase in occupation number as sugessted in Figure 5. This incicated that more shelter are timely provided, the goverment are providing larger amount of shelter than usual during pandemic. This actully can be visulized in Figure 7, espcially in the case for Scarborugh, there is a noticibke peak at very end of 2019 and start of 2020.

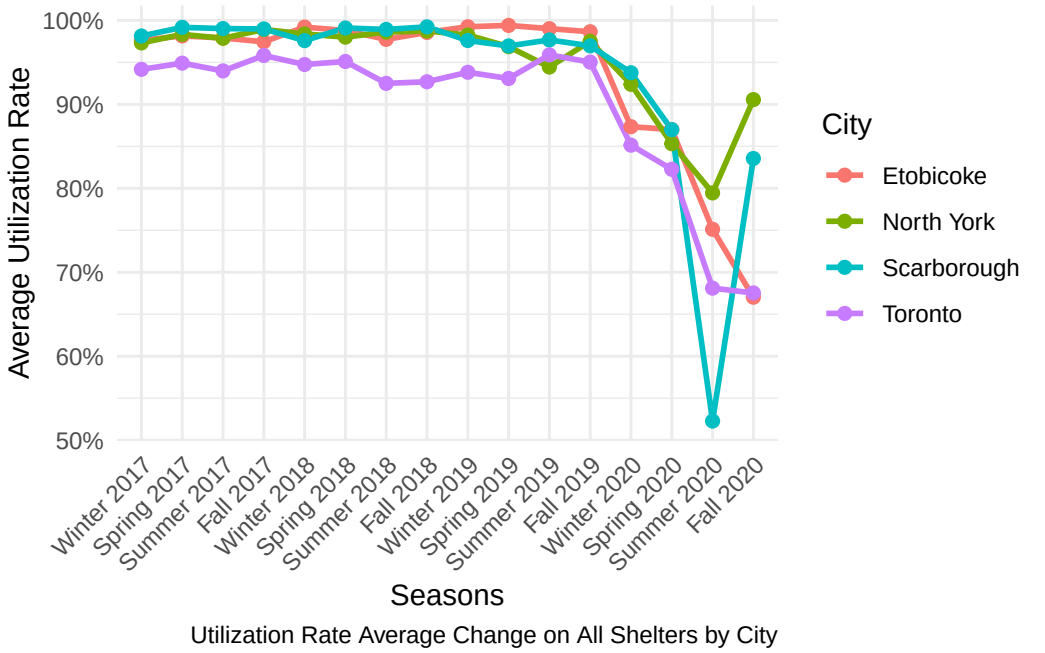

Figure 6: Yearly change in Utilization Rate Average on All Shelters by City (2017-2020)

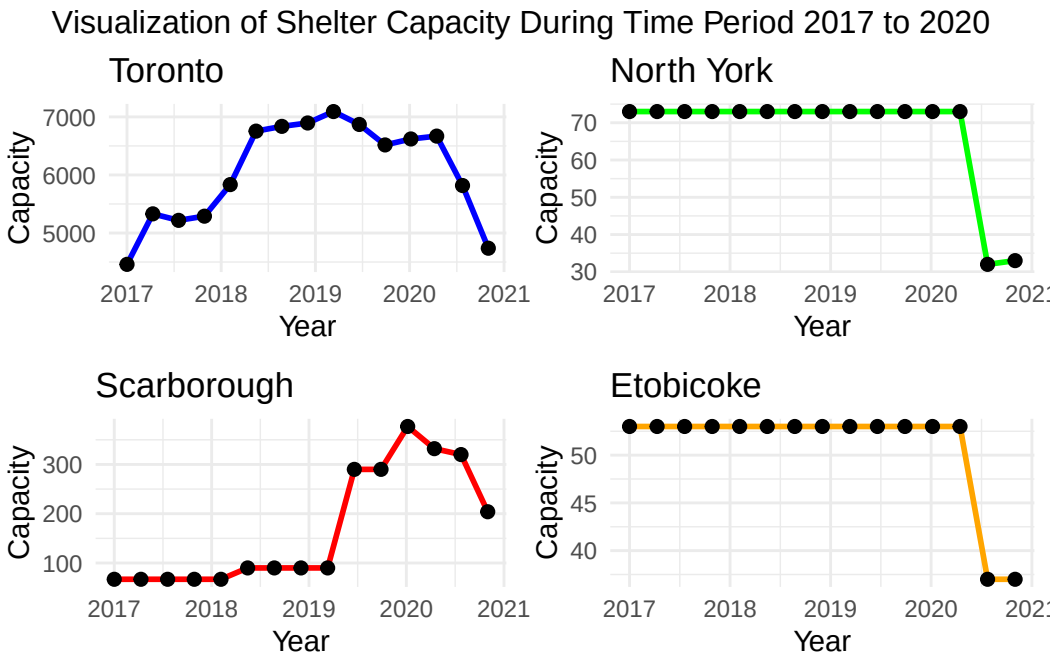

Figure 7: Shelter Capacity During Time Period 2017 to 2020 of CIties

3.3 Over Capacity Cases

The relashipship between Occupancy and Capcity over this four year (see Figure 4) are revealling cases where more tha expectted people are living in the shelter, as we dicusssued in Data section Section 2. These cases where capacity is exceeded (when $occupancy > capacity$) are importent when considering improvement shelter system improvement. By further invisted to the dataset, we are ablt to identify cases are actually happens to Families sector in Toronto (see Figure 8). And all other sector didn't expereince overcorwaaed shelter. Think incidate the quility of shleter servise is good at most case. Families sector are expereiceing overcraodded shelter does not implies the shelter system's fault, it is likely that families with children are assigned to a room with bed/mat that is smaller compared to thier family size. 2 children are able to sleep in one bed, this does not seems like a proalem. For example, a family of five with small children, can elect to be accommodated in a room with four beds, as stated in explanamtion of "CAPACITY" feature in the Portal Toronto Shelter & Support Services (2022).

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

Figure 8: Overcrowded Cases by City and Sector

4 Discussion

4.1 Shelter servise evaluation

Based on the information said in Section 3, particularly Figure 6, we can see the selter servise system in the CIty of Toronto are indded good structured. The untillization rate is usually in the 90% - 100% region. Suggesting that what being provided are well-used and supportivie to the homeless community.

4.2 COVID-19 Effect

Thorugh looking at the plot Figure 5, the ococupanny of shleters are at the highest around WInter 2019 to the begining of 2020. The pendenmic is indeed infecting the homeless population across four cities. Government actual in react to the outbreak of COVID-19, in intend

to support homeless population are represnted on Figure 7, with higher capacities of living nessecities provided. Thorugh higher capaciticies of existing shelter,s or adding in new shelters.

4.3 Weaknesses and next steps

COVID-19 is inded empacting on homeless populations, and the effor of mainting thier living needs is a continuse effort even 5 years after the outbreak. This data contains evidence for govemetn support during the pendemic, but is not inclusive enought to evluated on their performance on shelter servises. This defeinlty reqire more rigours research. And it's also not enought to say about what is the govement' support now, since this data only untill 2020. ALso, the increase untillization rate during covid outbreak could be the case that more people are facing more severe peoeblems so that not even make it to the shelter. Understanidng the wellvelopneness and imparotance of support networks among people experiencing homelessness is sessential. A more cirtical proalblem comes along it how should the supporte being delievered. A good way to to support while ensuring not to disrupt them, for instance by avoiding practices that displace or isolate people from their communities such as moving people to shelters located far from their usual supports Boucher et al. (2022). This is defienlt a long journey to go.

References

Auguie, B. 2017. "gridExtra: Miscellaneous Functions for "Grid" Graphics." https://CRAN.R-project.org/package=gridExtra.

Boucher, Lisa M, Zoë Dodd, Samantha Young, Abeera Shahid, Ahmed Bayoumi, Michelle Firestone, and Claire E Kendall. 2022. "'They Have Their Security, We Have Our Community': Mutual Support Among People Experiencing Homelessness in Encampments in Toronto During the COVID-19 Pandemic." *SSM-Qualitative Research in Health* 2: 100163.

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/.