

Number of actively homeless people dramatically decreased may not mainly caused by moving into permanent housing

Ziyu Jiang

06/02/2022

Abstract

The dataset of information of homeless people who is or was using the night shelter service operated by the City of Toronto was collected through the Shelter Management Information System (SMIS). This report using the data to mainly discover the relationships between the change of number of actively homeless and the change of number of homeless people who moved to permanent housing, and in addition, it will investigate the relationships between the change of number of actively homeless and the change of number of other types of homeless. The data show that there are no significant relationship between the change of number of actively homeless and the change of number of homeless people who moved to permanent housing.

1 Introduction

On December 16, 2020, the City Council have considered the Emergency Housing Action which shows a plan for the City to accelerate the creation of 1,248 permanent affordable homes with support services within 12 months. (Toronto 2021) During the Covid-19 pandemic, homeless people had become one of the high-risk groups when they lived in a more crowded environment in the shelter. The government was committed to get homeless people into affordable permanent housing to reduce the number of the actively homeless people also to reduce the health risks they are facing to. (Toronto 2020) At the beginning of the pandemic in 2020, there was a rapidly decreasing of number of actively homeless people. It is interested that if this decreasing is caused by the change of the number of homeless who moved to permanent housing. The results would possibly be helpful for the government's purpose.¹

After analyzing the data in the certain period of time, we can see that the decreasing of the number of actively homeless people has no big deal with the change of the number of homeless who moved to permanent housing. In this report, I will use the Toronto Shelter System Flow data from the Toronto shelter system funded and operated by the City of Toronto to compare the changes of numbers of different type pf homeless people during the same period of time. I will discuss the limitation of the data and the potential bias from it. Then I will analyze the changes of number of homeless people groups by age. Finally I will discuss the relationships between the decreasing of the number of actively homeless people and the changes of number of other types of homeless people, also I will discuss the relationship between the decreasing and the changes of numbers of homeless people groups by age.

2 Data

R statistical programming language (R Core Team 2021) is used for analyzing, and the package `tidyverse` (Wickham et al. 2019) is used for data visualizing and data manipulating in this project. The `opendatatoronto` (Gelfand 2020) package is used for importing the data set directly from the City of Toronto Open Data Portal. The package `kableExtra` (Zhu 2021) is also used in this project to build tables, the package `janitor` (Firke 2021) is used to clean the dataset, the package `lubridate` (Grolemund and Wickham 2011) is used to

¹Github: <https://github.com/Owlll11/STA304>

modify the format of dates and the package `knitr` (Xie 2021) is used to knit the R markdown file to pdf form. Finally, the package `ggpubr` (Kassambara 2020) is used to combine several plots in one plot.

The Toronto Shelter System Flow data is collected by the Shelter Management Information System (SMIS), and is published by Shelter, Support & Housing Administration on the City of Toronto Open Data Portal. The SMIS is developed by the City of Toronto to provide a unified information management system for the Toronto's emergency shelter system; it provides the data about the shelter occupancy in a real time scale. (citeSIMS?) The open data for the shelter system flow is updated monthly, and for every new updated dataset, the data for the previous month is also included. The dataset is last updated on Jan 7th, 2022.

Since this dataset contains all of the information of the homeless people who are or were using the shelter service funded and operated by the City of Toronto using the SMIS, it is not able to represent the whole information of people experiencing homelessness in Toronto. The data only reflects people who have used the overnight service operated by the City of Toronto and via the SMIS, it does not record the status of the homeless people who slept outdoors or used other homeless services or used other overnight service which are funded by other levels of government which not using the SMIS. (Support and Administration 2021) The Shelter, Support & Housing Administration expects that there are around 18 percent of total number of people who are experiencing homelessness in Toronto are not included in current data. (Support and Administration 2021) This would probably cause some bias since the data we got is not really covering the whole population which is the all people who are experiencing homeless in Toronto. However, these bias created by this situation would not impact the project a lot, since we believe the main part of population which is recorded via the SMIS is a good representative of the population in this report. In additional, the data of the variables about the ages are all collected from self-reported date of birth, which might contain data which is not real since some people might fill in fake ages. This would cause the data to be inaccurate, but since we have no way to deal with it, we can not do anything about this and it might cause the wrong conclusion at the end.

The raw dataset contains the information about the status of the homeless people entering and leaving the shelter system each month from January 2021 to December 2021. (Support and Administration 2021) More precisely, the dataset has the information about the population groups of the homeless people, the year and month of the information recorded, the number of homeless grouped by gender, the number of homeless grouped by age, the population group percentage of each population groups compare to the total population, and the number of different types of homeless people due to the time gap of visiting the shelters and the residence type before or after moving in the shelter. (Support and Administration 2021) It also contains the status of the homeless people, which is actively , moved to inactive, and returned to shelter. (Support and Administration 2021)

There were 180 observations and 18 variables in the raw dataset and there are no missing values, the variables are: unique identifier, date(month/year), population group, number of homeless returned from housing, number of homeless returned to shelter, number of newly identified homeless, number of homeless who has not use the shelter for last month, number of homeless whose age is under 16, number of homeless whose age is between 16 and 24, number of homeless whose age is between 25 and 44, number of homeless whose age is between 45 to 64, number of homeless whose age is over 65, number of male homeless, number of female homeless, number of other gender homeless, and the percentage of population group to the whole population. This dataset contains excessive different topics of variables, it would be simpler for the analysis if cleaning applied. The starting of the dataset cleaning process is applying the `clean_names` function (Firke 2021) to clean the column names of the dataset. After that, since the variable unique identifier(id) is not useful for this project, and the variables about the population groups and genders also the variables number of homeless returned from housing and number of homeless who has not use the shelter for last month are not helpful for the analysis of the main topic, all of these variables are removed for simplification. Then I modified the form of dates in the column "date_mmm_yy" using the function `my` (Grolemund and Wickham 2011) and saved the modified dates in a new column named "date_new." At last, I combined the rows which have same value of the variable "date_new"; that makes the variable "date_new" contains no repetitive dates values and added the values of other variables up for the same date. Finally, the information of homeless in Toronto contained in the cleaned dataset has the numbers of different age groups of homeless people and the total number of people who has been indicated as "actively homeless" groups by the date. Below is a brief view of

```
## Rows: 24
## Columns: 10
## $ date_new      <chr> "2020-01", "2020-02", "2020-03", "2020-04", "2020-~
## $ returned_to_shelter <int> 1508, 1564, 1463, 1068, 1250, 1008, 1619, 1341, 11~
## $ newly_identified  <int> 3115, 3275, 3157, 1835, 1622, 1554, 2070, 2388, 21~
## $ moved_to_housing  <int> 1926, 2090, 2303, 2107, 1747, 1570, 1838, 1739, 19~
## $ actively_homeless <int> 33219, 32914, 32507, 30272, 28170, 26619, 26254, 2~
## $ ageunder16        <int> 4747, 4707, 4563, 4078, 3616, 3393, 3134, 3070, 27~
## $ age16_24          <int> 4111, 3988, 3846, 3604, 3339, 3115, 3064, 3111, 30~
## $ age25_44          <int> 12775, 12756, 12619, 11643, 10596, 9978, 10021, 10~
## $ age45_64          <int> 9761, 9636, 9645, 9156, 8808, 8403, 8272, 8375, 85~
## $ age65over         <int> 1819, 1842, 1833, 1784, 1753, 1730, 1740, 1710, 18~
```



Table 1: Actively homeless summary table

min	Q1	median	Q3	max	IQR	mean	sd
26619	27782.25	29221	30830.75	32507	3048.5	29392	2559.976

Figure 1 shows the number of actively homeless people, number of homeless who moved to housing, number of homeless newly identified, and number of homeless returned to shelter over time, each point represents the numbers of the four types of homeless people, and the lines were connecting the points to better show the trending of the data. We can see the overall trending of the change of the number of actively homeless through the timeline, the number is dramatically decreasing from around 32500 to around 26600 during the time between March, 2020 and June, 2020; however during the same period of time, the number of homeless people who moved to housing also sharp declined from around 2300 to around 1600. This is very abnormal since when we look at the other part of the two plots, we would notice that during the same period of time, the trends of the two plots should be exactly opposite. The homeless people are in a higher risk of transmitting and contracting COVID-19 since they are living in a more crowded environment in the shelter. (Ferreira 2020) The Toronto shelter death count does not want to make the detailed information about individual cases in public, but the incomplete data for 2020 provided by Toronto Public Health shows that there are 7 percent of homeless deaths occurred between January and June due to Covid-19; with other causes of homeless deaths during this time period totally added to 46 percent of total homeless deaths, there are still 47 percent of total homeless which causes is pending or is unknown. (Gibson 2021) In addition, we can see that the number of newly identified homeless decreased dramatically from March to April 2020 and gradually decreases from April to June 2020; and we can see that the number of returned to shelter homeless also dropped dramatically from March to April 2020, but increased a bit from April to May 2020 then finally dropped back from May to June 2020.

Table 1 would show the general information about the actively homeless people during the time period from March, 2020 to June, 2020. The standard deviation shows that there is a very large change of the numbers of actively homeless people during this time period.

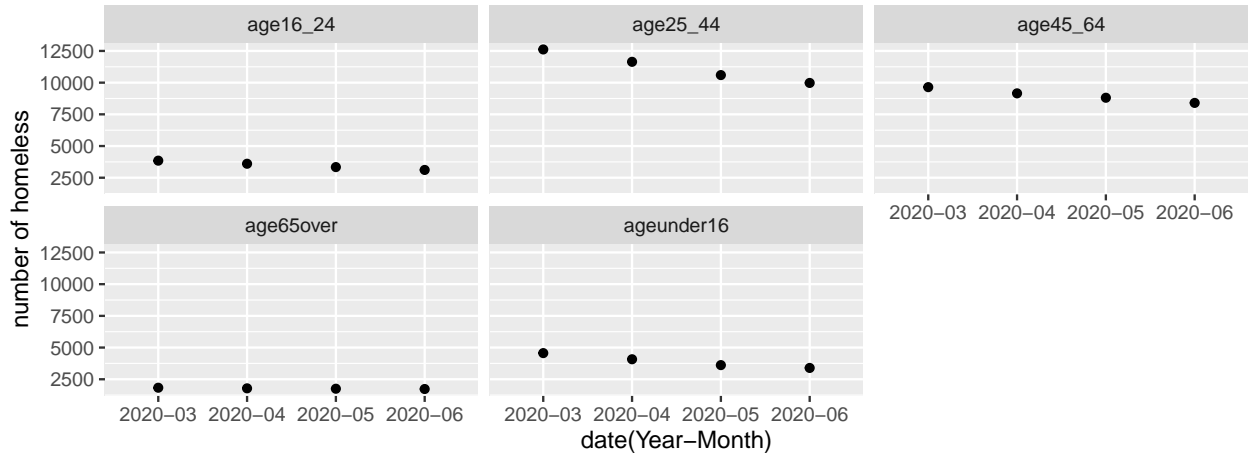


Figure 2: Number of homeless between March, 2020 and June, 2020, grouped by age

From figure 2, we can see the plots of number of homeless people using the night shelter during the time period between March, 2020 and June, 2020 grouped by age. Each point indicate the number of homeless people in the corresponding month for 5 different age groups. We can see from the plots that all plots more or less have decreasing trends over time. The fastest decreasing is occurring in the homeless whose age between 25 and 44, then the decreasing of number of homeless is a little bit slower for age group which is under 16 and age group which is between 45 and 64; it is even slower for the age group between 16 and 24 to decreasing

the number of homeless. And finally, the slowest decreasing is occurring in the homeless whose age is larger than 65. The data provided by the City of Toronto shows that the deaths of people experiencing homeless in 2020 has the largest proportion from age between 40-59 (36%), the second large proportion is from age over 60 (34%), the third largest proportion is from age between 20 and 39 (26%), only 1% of deaths from total Toronto 2020 homeless deaths is from age under 20. (City of Toronto and Epidemiology 2021)

3 Discussion

Due to the City of Toronto reported, in December 2020, the City Council is about to assign 1,248 permanent affordable homes with support services within 12 months.(Toronto 2021) Which means before that time, there is no way of suddenly increasing of number of homeless who moved to permanent housing, for example, from March, 2020 to June, 2020; the abnormal trending of the two plots confirmed with this point. By comparing the two plots in figure 1 and the additional researches at the beginning, we might think the main reason of the strange phenomenon occurred in the actively homeless and the moved to housing homeless would be the unexpected deaths caused by the Covid-19 pandemic during the time period between March, 2020 and June, 2020. When we start analyzing the relationship of the reasons of number of actively homeless decreasing and the age groups, we can notice that for the total actively homeless population, the main reason of decreasing in number is not the unexpected death caused by Covid-19; the reason for it is very clear: the proportion of deaths of homeless in Toronto, 2020 from the age group over 65 is much larger than the proportion of deaths from age group under 16, but the decreasing rate for the number of homeless in age group under 16 is significantly larger than the decreasing rate for the number of homeless in age group over 65. For the other age groups, the main reason of decreasing of number is not very clear, we are not able to get the related answers from the dataset; so we are only able to discuss the possibilities. From figure 1 we have noticed that the number of newly identified homeless decreased dramatically from March to April 2020 and gradually decreases from April to June 2020, also the number of returned to shelter homeless also dropped dramatically from March to April 2020, but increased a bit from April to May 2020 then finally dropped back from May to June 2020. These might apply that lack of newly identified homeless and lack of returned to the shelter homeless might be some of the reasons that the number of actively homeless dropped dramatically.

References

- City of Toronto, Surveillance, and Epidemiology. 2021. *Deaths of People Experiencing Homelessness*. *Free Data Visualization Software*. <https://public.tableau.com/app/profile/tphseu/viz/DeathsofPeopleExperiencingHomelessness/HomelessDeathsFinal?publish=yes>.
- Ferreira, Jennifer. 2020. *The Toll Covid-19 Is Taking on Canada's Homeless*. *Coronavirus*. CTV News. <https://www.ctvnews.ca/health/coronavirus/the-toll-covid-19-is-taking-on-canada-s-homeless-1.4950722>.
- Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://github.com/sfirke/janitor>.
- Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal*.
- Gibson, Victoria. 2021. *Toronto Shelter System Records Deadliest Year on Record, with 74 Deaths in 2020*. *Thestar.com*. <https://www.thestar.com/news/gta/2021/01/12/toronto-shelter-system-records-deadliest-year-on-record-with-74-deaths-in-2020.html>.
- 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/>.
- Kassambara, Alboukadel. 2020. *Ggpubr: 'Ggplot2' Based Publication Ready Plots*. <https://rpkgs.datanovia.com/ggpubr/>.
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Support, Shelter, and Housing Administration. 2021. *Shelter System Flow Data*. *City of Toronto*. <https://www.toronto.ca/city-government/data-research-maps/research-reports/housing-and-homelessness-research-and-reports/shelter-system-flow-data/>.
- Toronto, City of. 2020. *City of Toronto Continues to Take Extraordinary Steps to Help and Protect People Experiencing Homelessness During COVID-19*. *City of Toronto*. <https://www.toronto.ca/news/city-of-toronto-continues-to-take-extraordinary-steps-to-help-and-protect-people-experiencing-homelessness-during-covid-19/>.
- . 2021. *Housing and Homelessness Research and Reports*. *City of Toronto*. <https://www.toronto.ca/city-government/data-research-maps/research-reports/housing-and-homelessness-research-and-reports/>.
- 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>.
- Xie, Yihui. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://yihui.org/knitr/>.
- Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*.