Some Correlations were Found between the efforts made by the Staff of Toronto's Central Intake Line and Toronto's Homeless Death Counts*

Sandy Yu

January 23, 2024

For the City of Toronto, the physical well-being of its local homeless population has always been one of its primary focuses. Numerous social efforts, funded by the City, were made in attempts to reduce the death rate of the local homeless population. In order to visualize the magnitude of social impact enforced by the City of Toronto, analyses were performed to investigate the correlation between the efforts made by the staff of Toronto's Central Intake Line and the monthly homeless death counts between November 2020 to June 2023 using data from OpenData-Toronto. Staff efforts such as the number of calls coded, number of referral to shelter, and number of information for homelessness prevention, were documented and compared against the monthly homeless death counts. An inverse correlation is found between the number of calls coded and the homeless death count. No discernible correlations were found between coded call types of interest and the monthly homeless death counts.

Table of contents

*Code and data are available at: https://github.com/Jingying-yu/central_intake_calls_and_homeless_death_count.git

1 Introduction

The City of Toronto has been consistently providing funding and staff in an effort to reduce the active homeless population within its governing district. One of the key aspects of reducing homelessness is the assignment of sheltering space. Amongst numerous city efforts, the 24/7 helpline — Toronto Central Intake Line — was created to offer referrals to emergency shelters, sleeping accommodations, and provide general information about homelessness and prevention. (Shelter 2020). Starting in 2020, detailed data about the Central Intake Line became available on the OpenDataToronto portal (Sharla Gelfand 2022).

In this paper, we want to measure the magnitude of positive impact created by the Central Intake Line in relation to the monthly cumulative number of deaths in the homeless community. Using Central Intake Call Wrap-up Codes data (Shelter 2020) from OpenDataToronto (Sharla Gelfand 2022), three different aspects of the Central Intake Line efforts were isolated: total calls coded, referral to shelter, and homelessness prevention information provided. These three key aspects were measured against the monthly cumulative Deaths of People Experiencing Homelessness Data (Health 2017) documented by the Toronto Public Health (Toronto 2024). The analyses yielded an inverse correlation between the number of calls coded and the homeless death count, and no discernible correlations between coded call types of interest and the monthly homeless death counts.

Analyses and findings in this paper are structured into several sections: Section 2 – Data, Section 3 – Result, Section 4 – Discussion, and Section 5 – Conclusion. The Data section examines all datasets and variables kept for analysis, followed by an explanation of their data cleaning processes. The Result section focuses on visualizing and presenting the correlation between variables of interest. The Discussion sections further evaluate the meaning behind the correlation presented in the previous section. Lastly, the arguments presented in all previous

sections are wrapped up by the Conclusion section which summarizes the main findings in this paper.

2 Data

All data used in this paper are obtained through OpenDataToronto Portal (Sharla Gelfand 2022). Two different datasets, Central Intake Call Wrap-Up Codes Data (Shelter 2020) and Deaths of People Experiencing Homelessness(Health 2017), are retrieved to analyze the correlation between different factors of Toronto's Central Call Line efforts and the monthly cumulative death counts of homeless individuals in Toronto. Data was cleaned and analyzed using the open source statistical programming language R (R Core Team 2022) and supporting packages tidyverse (Wickham et al. 2019), janitor (Firke 2023), dplyr (Wickham et al. 2023), lubridate (Grolemund and Wickham 2011), zoo (Zeileis and Grothendieck 2005), ggplot2 (Wickham 2016), and knitr (Xie 2023). Detailed process of data extracting and cleaning can be found in the subsections below.

2.1 Central Intake Call Wrap-Up Codes

On the OpenDataToronto portal, there are several datasets that reflect the City's effort to shelter the local homeless population. The Central Intake Call Wrap-Up Codes Dataset (Shelter 2020), stored in the Central Intake Calls Catalogue (Shelter 2020), is one of the freshest and most detailed. Data was stored and published by the Shelter, Support & Housing Administration since November of 2020 and refreshes on a monthly basis. The latest refresh occurred on January 15th, 2024.

The data set provides a daily summary of the number of calls received, the number of calls classified into distinct wrap-up codes by the nature of the issue, and a count of calls under each wrap-up code. One of the example wrap-up codes in the original data set was: Code 1A - Referral to a Sleeping/Resting Space. The original data set includes 13 distinct wrap-up codes; only two codes, Code 1A — Referral to a Sleeping/Resting Space and Code 2C — Information - Homelessness & Prevention Services, were chosen for our analysis.

The process of data cleaning for this dataset started from renaming and selecting. Variables of relevance: Date, Total calls coded, Code 1A, and Code 2C, were renamed into appropriate column names and selected (Table 1).

Table 1: Sample of Renamed Central Intake Call Wrap-up Codes Variables

Date	Calls Coded	Referral to Shelter	Homelessness Prevention Info
2020-11-03	301	44	66
2020-11-04	337	56	77

Date	Calls Coded	Referral to Shelter	Homelessness Prevention Info
2020-11-05	315	66	92
2020-11-06	283	49	71
2020-11-07	168	52	41
2020-11-08	213	48	48

Cleaned data is then filtered to include only data before July of 2023. Remaining Data is then summarized into monthly cumulative counts using the group_by() function and summarise() function under the tidyverse package (Wickham et al. 2019) (See Table 2).The rationalization behind the filtering and summarizing actions will be explained in Section 2.3.

Table 2: Sample of Cleaned Central Intake Call Wrap-up Codes Dataset

Month	Calls Coded	Referral to Shelter	Homelessness Prevention Info
2020-11-30	8367	1433	2029
2020-12-31	10232	2027	2427
2021-01-31	12091	1726	2453
2021-02-28	10525	1662	2094
2021-03-31	12287	1771	2202
2021-04-30	12668	1232	1870

2.2 Deaths of People Experiencing Homelessness

The Deaths of People Experiencing Homelessness Dataset (Health 2017) contains monthly cumulative records of homeless deaths. The dataset is published by Toronto Public Health (Toronto 2024). The earliest data record started in January of 2017, and the latest record ends in June of 2023.

The original dataset contains three columns: Year of death, Month of death, and Count. The first step of data cleaning, after downloading the data, was to exclude any rows that contained an "unknown" value. Although this action created a source of error in future analyses, this is still a necessary step because we are plotting death counts against a timeline. A date column is then created to combine the information on the year and month of death. The format of this date-class column is "yyyy-mm-dd" where the "dd" is the last day of the month when the death count of the month is finalized. This design allows easier graphing of monthly cumulative counts as discrete variables against time. The date-class column is added to the original dataset by the mutate() function under the tidyverse package (Wickham et al. 2019).

Cleaned data is then filtered to include only data from November 2020 to June 2023 and a new column that contains the name of the "Month Year" is added to the dataset (Table 3). The rationalization behind the filtering will be explained in Section 2.3.

Table 3: Sample of Cleaned Deaths of People Experiencing Homelessness Dataset

Last Day of Month	Death Count
2020-11-30	15
2020-12-31	20
2021-01-31	20

2.3 Using Data from Different Datasets

During the data cleaning process, code chunks are written to ensure that both the Central Intake Call Wrap-Up Codes Dataset (Shelter 2020) and the Deaths of People Experiencing Homelessness Dataset (Health 2017) are filtered to include only data between November 1st, 2020 to June 30st, 2023.

The decision to only include this time period is reached by taking the common time period between the two datasets. In the latest refresh, the Central Intake Call Wrap-Up Codes Dataset (Shelter 2020) begins on November 3rd, 2020 and end on December 31st, 2023; whereas the Deaths of People Experiencing Homelessness Dataset (Health 2017) begins on January of 2017 and ends on June 2023.

The ultimate purpose of taking the common time period is to ensure that we can plot variables of interest in the two datasets onto the same graph. Thus a combined dataset that kept the variables of interest in both datasets is created (Table 4).

Table 4: Sample of Combined dataset

Month	Calls Coded	Referral to Shelter	Homelessness Prevention Info	Death Count
2020-11-30	8367	1433	2029	15
2020-12-31	10232	2027	2427	20
2021-01-31	12091	1726	2453	20

3 Results

Between November 2020 and June 2023, there were a total of 522 cases of homeless death recorded in Toronto. The mean monthly death count is 16.3125, with a minimum of 8 recorded

cases and a maximum of 26 recorded cases per month. The overall trend of homeless death cases is downward-sloping.

3.1 Number of Calls Coded & Homeless Death Count

The first graph (Figure 1) maps the overall trend of homeless death over time, represented by the blue line, against the adjusted cumulative number of calls coded per month. The value of calls coded per month, represented by the red line, is divided by 1000 for easier comparison against the death count. Original values are too big for any meaningful comparison.

The red line in the graph visualizes the number of calls coded per month obtained from the Central Intake Call Wrap-Up Codes Dataset (Shelter 2020) between November 2020 and June 2023. Throughout the time period, there were a total of 474515 calls coded, the mean monthly number of calls coded is 14828.59. The minimum number of calls coded per month was 8367, and the maximum number of calls coded was 30520. We can see a positive growth in the number of calls being coded per month over time (Figure 1).

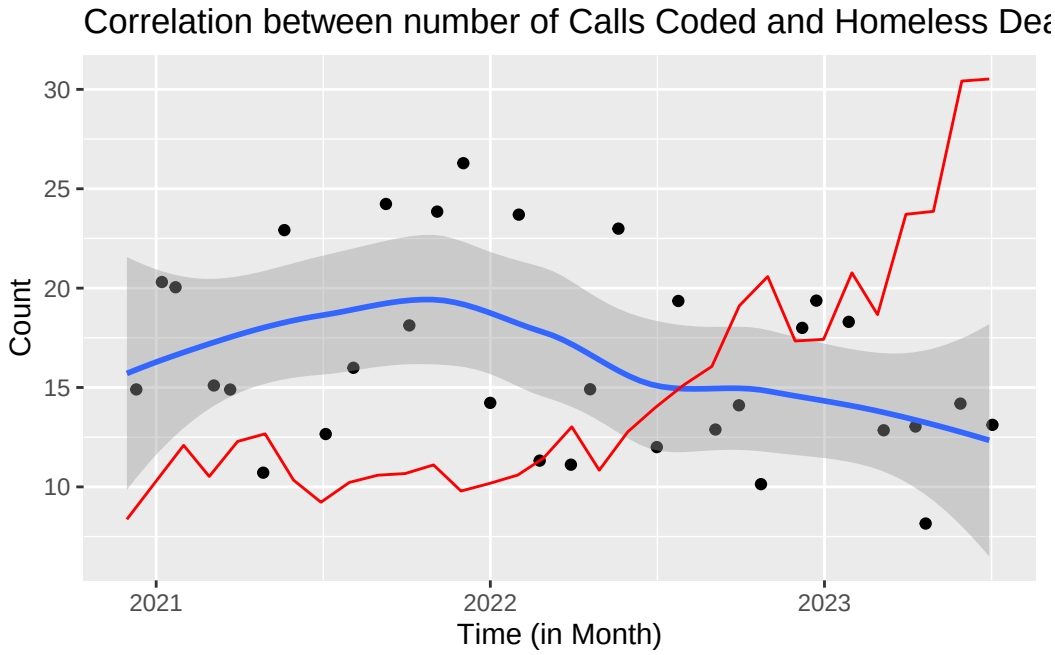

Figure 1: Visualization to show correlation between number of Calls Coded and Homeless Death Counts

3.2 Number of Callers Referred to Shelter & Homeless Death Count

The second graph (Figure 2) maps the overall trend of homeless death over time, represented by the blue line, against the adjusted cumulative number of callers referred to shelters per month. The number of callers referred to shelters per month, represented by the purple line, is divided by 100 for easier comparison against the death count. Original values are too big for any meaningful comparison.

The number of callers referred to shelters per month obtained from the Central Intake Call Wrap-Up Codes Dataset (Shelter 2020) between November 2020 and June 2023 is represented by the blue line in Figure 2. Throughout the time period, there were a total of 44760 callers referred, the mean monthly number of callers referred was 1398.75. The minimum number of callers referred per month was 456, and the maximum number of callers referred was 2192. We can observe a negative growth in the number of callers referred per month over time (Figure 2).

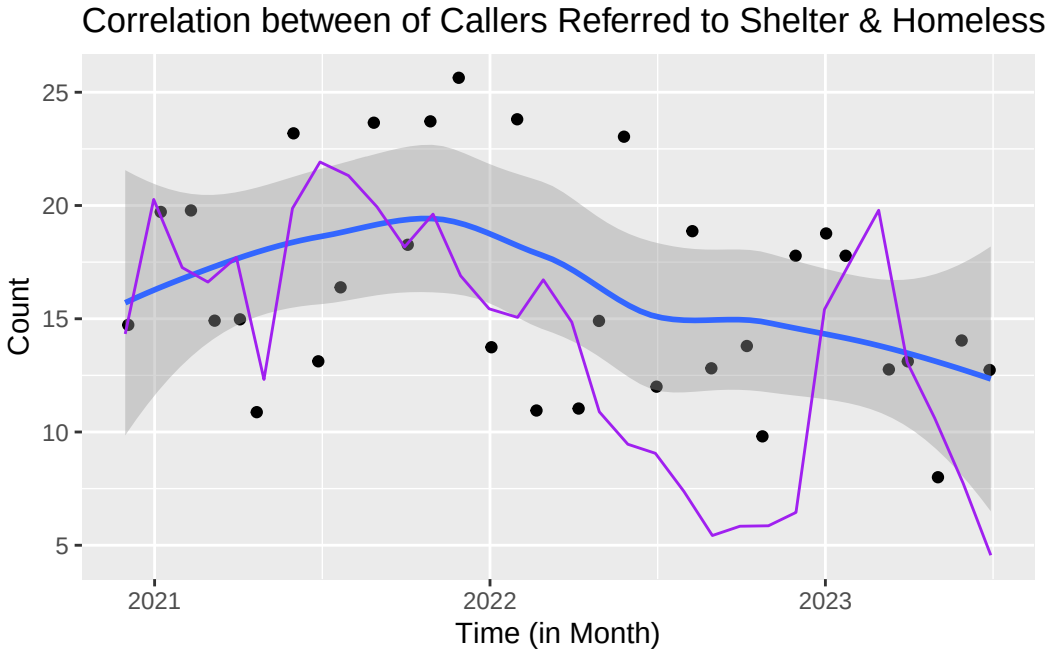

Figure 2: Visualization to show correlation between number of callers referred to shelter and Homeless Death Counts

3.3 Number of Callers given information on Homelessness Prevention & Homeless Death Count

The third graph (Figure 3) maps the overall trend of homeless death over time, represented by the blue line, against the adjusted monthly cumulative number of callers provided with

information about homelessness prevention. The number of callers provided with information, represented by the dark green line, is divided by 100 for easier comparison against the death count. Original values are too big for any meaningful comparison.

The number of callers provided with information on homelessness prevention per month obtained from the Central Intake Call Wrap-Up Codes Dataset (Shelter 2020) between November 2020 and June 2023 is represented by the dark green line in Figure 3. Throughout the time period, there were a total of 58333 callers provided with information, the mean monthly number of callers provided with information was 1822.906. The minimum number of callers provided with information per month was 1110, and the maximum number of callers provided with information was 2500. We can observe an initial negative growth followed by a positive growth in the number of callers provided with information per month over time (Figure 3).

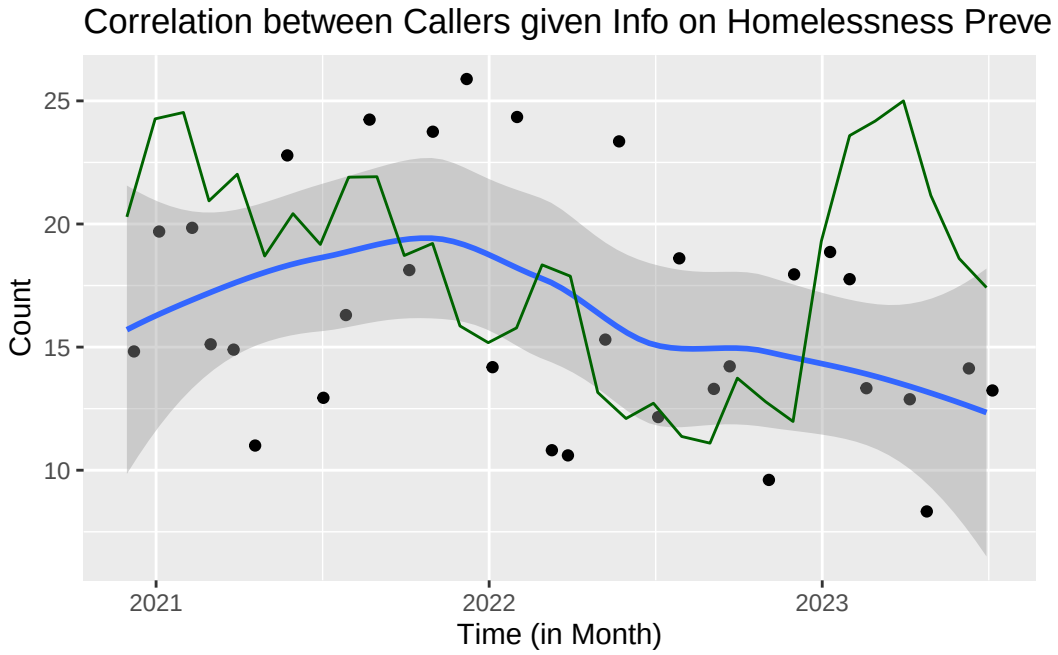

Figure 3: Visualization to show correlation between number of callers given information on homelessness prevention and Homeless Death Counts

4 Discussion

4.1 Correlation between number of Calls Coded & Homeless Death Counts

Based on Figure 1, there seems to be an inverse relationship between the directions of change for the homeless death count trend over time, the blue line, and the monthly cumulative calls coded represented by the red lines. When the red line moves upward, the blue line moves down.

This inverse relationship indicates a negative correlation between the two variables. A negative correlation between these two variables implies that the efforts the Central Intake Line placed into coding calls has created a positive impact in reducing homeless death in Toronto.

4.2 Correlation between number of Coded Calls of Interest & Homeless Death Counts

According to the Figure 2 and Figure 3, both the number of callers referred to shelters (represented by the purple line in Figure 2) and the number of callers provided with information on homelessness prevention (represented by the dark green line in Figure 3) change in the same direction as the monthly cumulative homeless death count. Through pure visual observation, a positive correlation between homeless death count to each of the variables would be found. This relationship would imply that the Central Intake Line (Shelter 2020) has actually created a negative impact in reducing the homeless death count in Toronto by referring callers to shelter and providing callers with information on homelessness prevention. However, the magnitude of fluctuation for both the purple and dark green line are very large throughout the entire time period. Especially from the beginning of 2022 to the end of 2023 for the purple line. Due to the magnitude of fluctuation for both lines, we cannot responsibly conclude any causal relationship between the two variables by pure observation of the trends. We can only say that the general trend of the purple and dark green line moves in the same direction as the homeless death count trend. Therefore, no meaningful correlation between the two variables were concludable using only the Central Intake Call Wrap-Up Codes Dataset (Shelter 2020) and Deaths of People Experiencing Homelessness Dataset (Health 2017) obtained from OpenDataToronto portal (Sharla Gelfand 2022). However, this result does not imply that no correlation between variables exists in the real world. It is completely plausible that correlation between the two variables can be found if more detailed data were provided, or more advanced statistical models were used for analysis.

4.3 Errors in analysis, Limitations and Next Steps

Potential errors exist in the final datasets used for analyses. During the data cleaning process, if rows in the Deaths of People Experiencing Homelessness Dataset (Health 2017) contain "Unknown" as a Month or Year value, then these rows are filtered out of the final dataset used for relationship analyses. This filtering action would cause an inaccuracy in the total cumulation of the homeless death count; however, incoherence of dates would occur if these rows were left inside the dataset. Another potential error could occur even before I download datasets from OpenDataToronto (Sharla Gelfand 2022). We cannot rule out the possibility of human error during the data inputting process before the data is submitted to the OpenDataToronto portal (Sharla Gelfand 2022).

Limitations to this analysis exist for a few reasons. Firstly, each Central Intake Line staff has their own understanding about which code they should assign their calls. We have no way of

verifying whether the staff have categories their calls correctly. Secondly, the staff do not have enough time to conduct repeated follow-ups with each caller they were assigned to. We have no way of verifying what percentage of callers referred to a shelter actually got there. Lastly, we do not have enough information to sort the coded calls by demography. A subset of data that sorts the homeless death count by demography exists in the Deaths of People Experiencing Homelessness Catalogue (Health 2017), if more information about the demography of the callers were provided, more information about the relationship between variables of interest might be found. Further analysis on homeless death counts by demography can be conducted based on shelter data that contains demographic values.

5 Conclusion

This paper investigated the magnitude of impact created by the Central Intake Line in relation to the monthly cumulative number of deaths in the homeless community. Variables of interest such as: calls coded, number of callers referred to shelter, number of callers provided with information on homelessness prevention, and homeless death count, were obtained from the Central Intake Call Wrap-Up Codes Dataset (Shelter 2020) and the Deaths of People Experiencing Homelessness Dataset (Health 2017) stored in the OpenDataToronto portal (Sharla Gelfand 2022). Statistical analyses yielded a negative correlation between the number of calls coded and the homeless death count, which implies a positive impact in reducing homeless death in Toronto. Other variables, such as number of callers referred to shelter and number of callers provided with information on homelessness prevention, yielded no meaningful correlation to the homeless death count. Therefore we cannot conclude the magnitude of their contribution to the reduction of homeless death count. Further analysis on homeless death counts by demography can be conducted based on shelter data that contains demographic values.

References

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

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/.

Health, Toronto Public. 2017. "Deaths of People Experiencing Homelessness." https://open.toronto.ca/dataset/deaths-of-people-experiencing-homelessness/.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Sharla Gelfand, City of Toronto. 2022. "Opendatatoronto: Access the City of Toronto Open Data Portal." https://open.toronto.ca.

Shelter, City of Toronto, Support & Housing Administration. 2020. "Central Intake Call Wrap-up Codes Data." https://open.toronto.ca/dataset/central-intake-calls/.

Toronto, City of. 2024. "Toronto Public Health." https://www.toronto.ca/city-government/accountability-operations-customer-service/city-administration/staff-directory-divisions-and-customer-service/toronto-public-health/.

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. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

Zeileis, Achim, and Gabor Grothendieck. 2005. "Zoo: S3 Infrastructure for Regular and Irregular Time Series." *Journal of Statistical Software* 14 (6): 1–27. https://doi.org/10.18637/jss.v014.i06.