Rising Trends in Shelter Resident Deaths in Toronto Over the Past 17 Years*

Haowei Fan

September 25, 2024

Shelter resident deaths are an important indicator of social stability, reflecting systemic issues such as welfare systems, housing crises, and healthcare access. This project analyzed monthly shelter resident deaths in the Toronto region from March 2007 to August 2024, focusing on trends over time and seasonal variations. The study found that deaths have been increasing, with a significant rise between 2020 and 2023. While the number of deaths remained similar in spring, summer, and fall, winter months saw a significantly higher number of fatalities. These findings provide valuable data to support targeted interventions by governments and welfare organizations to address changing needs throughout the year.

Table of contents

*Code and data are available at: https://github.com/HaoweiFan0912/Deaths-of-Shelter-Residents-in-Toronto.git

1 Introduction

In 2023, as the COVID-19 pandemic receded, the Toronto government announced plans to phase out the shelter hotels that had been providing temporary accommodations for homeless individuals (Omstead (2023)). By the end of 2024, 24 shelter hotels will no longer be available due to expiring contracts, exacerbating an already overcrowded shelter system (Omstead (2023)). The loss of these temporary accommodations poses significant risks not only to the safety and stability of the community but also to the well-being of those who rely on shelters for survival. The influx of new shelter residents may potentially lead to a decline in living conditions, including sanitation issues and territorial conflicts. This highlights the urgent need to understand the trends and underlying factors affecting shelter residents, especially in terms of their mortality numbers.

While many sociological studies have explored the growing number of shelter residents. For example, Maureen Crane and Louise Joly have studied the growing number of shelter residents and the issue of aging in Western countries (Crane and Joly (2014)). Similarly, in their article, John M. Quigley, Steven Raphael, and Eugene Smolensky pointed out that the number of shelter residents in the United States is increasing, along with the underlying causes contributing to this rise (Quigley, Raphael, and Smolensky (2001)). Fewer have examined shelter resident mortality rates in detail. Given the challenges of accurately tracking the transient shelter population, studying actual death counts—concrete and verifiable data—offers a more reliable and feasible method for understanding changes in the needs of shelter residents, providing better insights for policy decisions than difficult-to-obtain population estimates (Rossi et al. (1987)).

This paper analyzes the number of deaths among shelter residents in Toronto from March 2007 to August 2024, aiming to identify long-term trends and seasonal variations. Using simple linear regression, the study estimates the rate of increase in deaths over time, while also examining anomalies by comparing yearly death numbers and assessing seasonal differences through average death rates across the four seasons. The findings reveal three key trends: 1) a steady rise in the number of shelter resident deaths over time; 2) a noticeable spike in deaths between 2020 and 2023; and 3) significantly higher mortality rates during winter compared to the relatively consistent rates in spring, summer, and fall. These insights provide critical data for governments and welfare organizations to develop more targeted interventions based on the needs of shelter residents throughout times.

The remainder of this paper is structured as follows: Section 2 discusses the data sources used in this project and details the data cleaning process, followed by a visualization of the dataset's key distribution and statistics, as well as a discussion of the measurement of the original data. Section 3 outlines the statistical methods used and presents the findings in detail. Section 4 addresses the limitations of the study and offers suggestions for future research.

2 Data

2.1 Raw Data

The dataset used in this study is titled "Deaths of Shelter Residents" and is sourced from OpenDataToronto (Gelfand (2022)). It was published by Toronto Public Health (TPH) and records the number of deaths among shelter residents in the Toronto area on a monthly basis, starting from January 2007. Additionally, the dataset includes the gender distribution of the deceased each month, with gender categorized into three groups: male, female, and other gender, which encompasses Transgender, Non-binary, and Two-Spirit individuals (Gelfand (2022)). The dataset is continuously updated. A detailed description of the variables can be found in Table 1.

Table 1: Column descriptions of the original data

Variables	Descriptions
_id	Unique row identifier for Open Data database.
Year	The calendar year being reported on.
Month	The month being reported on.
Total decedents	The total number of shelter residents who died in the reported month/year.
Male	The total number of male shelter residents who died in the reported month/year.
Female	The total number of female shelter residents who died in the reported month/year.
Transgender/Non-binary/Two-Spirit	The total number of transgender, non-binary, and Two-Spirit shelter residents who died in the reported month/year.

In addition to this dataset, OpenDataToronto also provides a similar dataset titled "Deaths of People Experiencing Homelessness," which records monthly deaths of homeless individuals from January 2017 to March 2023 (Gelfand (2022)). However, I chose not to use this dataset due to its limited coverage, spanning only from January 2017 to December 2023. In comparison, the "Deaths of Shelter Residents" dataset offers a longer time frame, allowing for more comprehensive analysis and greater accuracy in the results (Gelfand (2022)).

2.2 Data Cleaning

During the data processing phase, tasks such as downloading, cleaning, simulating, analyzing, and code testing were all performed using R code (R Core Team (2023)). Additionally, the project utilized packages such as Tidyverse (Wickham et al. (2019)), Styler (Müller and Walthert (2024)), Dplyr (Wickham et al. (2023)), Lubridate (Grolemund and Wickham

(2011)), and Knitr (Xie (2024)) for data handling. These tools were instrumental in performing linear regression, creating visualizations, generating tables, and formatting data for analysis.

The data cleaning process involved several key steps, each designed to improve the dataset's suitability for analysis. First, the year and month variables were combined, with the month converted to numerical format to facilitate time series analysis. A new variable was then created to classify each month into one of the four seasons: March, April, and May as spring; June, July, and August as summer; September, October, and November as autumn; and December, January, and February as winter, enabling seasonal analysis. To streamline the dataset, unnecessary variables, such as ID and the three gender categories, were removed. Additionally, to ensure that the seasonal analysis of average shelter resident deaths covered equal time intervals for each season, certain samples from the dataset's beginning and end were removed, resulting in a total number of samples divisible by four. Finally, after reviewing the dataset, it was confirmed that there were no missing values, so no further treatment was necessary. Upon completion, the cleaned dataset was reduced to three variables: Time, Season, and Total Deaths, as few samples are outlined in Table 2.

Table 2: Samples of the cleaned data

Time	Season	Number of decedents
2007-03	Spring	3
2007-04	Spring	1
2007-05	Spring	2
2007-06	Summer	3
2007-07	Summer	2
2007-08	Summer	2

2.3 Statistical Summary of the Cleaned Data

Table 3 provides some basic statistics about the dataset. The project spans a total of 204 months. During this period, the highest number of deaths recorded in a single month was 19, while the lowest was 0. The average number of deaths per month is 3.848. Additionally, we can infer that the data is somewhat skewed, as the highest number of deaths in a single month is 19, while the average and median number of deaths are 3 and 3.83, respectively.

Table 3: Summary of the cleaned data

Time	Number of decedents
Length:204	Min. : 0.000
Class :character	1st Qu.: 1.000
Mode :character	Median : 3.000
NA	Mean : 3.848
NA	3rd Qu.: 5.000
NA	Max. :19.000

Figure 1 illustrates the distribution of monthly shelter resident deaths, showing a left-skewed normal distribution. This indicates that the majority of months have a death count in the range of 0-5. Additionally, there are two outliers, where the death count exceeds 15 in two months, which is an uncommon occurrence.

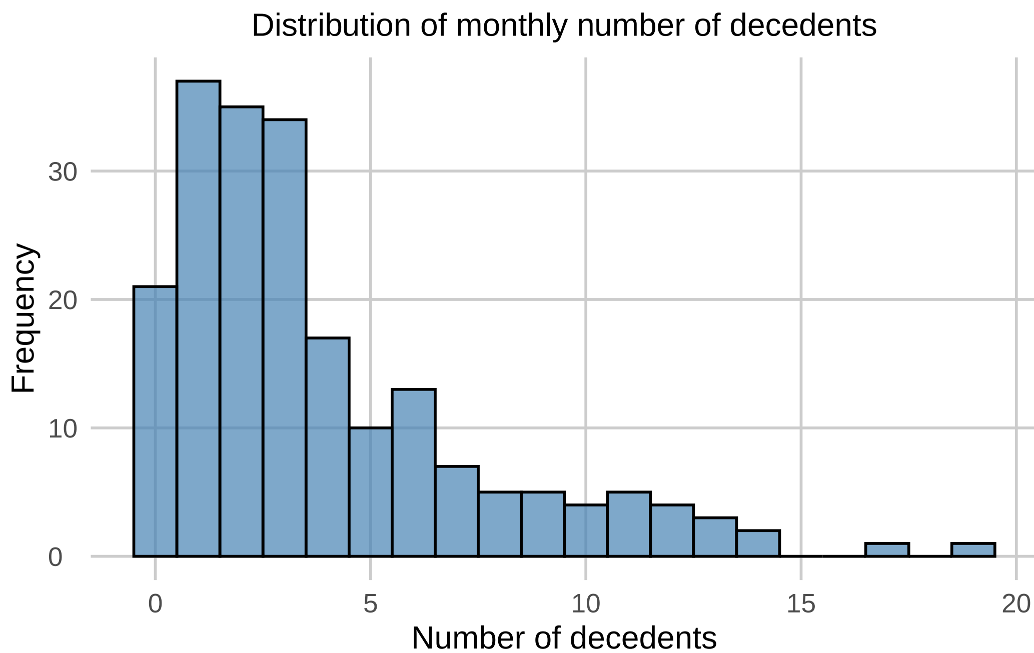

Figure 1: Distribution of monthly number of decedents

Figure 2 depicts the monthly variation in the number of deaths over time. The overall trend reveals a significant increase in the total number of shelter resident deaths in Toronto, particularly after 2018. Prior to 2018, the number of deaths remained relatively stable, rarely exceeding 5 per month. However, from 2018 onwards, there is not only a noticeable upward trend in deaths but also increased volatility. Spikes in the data become more frequent, with several months recording over 10 deaths, and some months reaching as high as 15 or more.

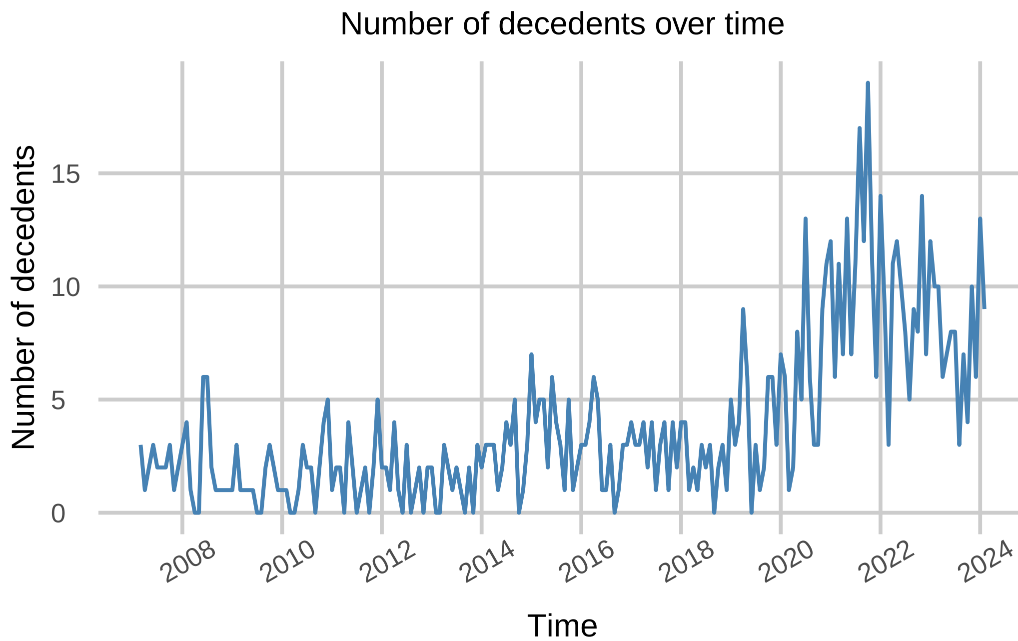

Figure 2: Number of decedents over time

2.4 Discussion of Data Measurement

The "Deaths of Shelter Residents" dataset is collected by Toronto Public Health (TPH), a governmental body (Gelfand (2022)). Information is sourced from several avenues, including a secure online form where agencies supporting homeless and under-housed populations report deaths (Gelfand (2022)). Other sources include the Toronto Shelter & Support Services (TSSS) division and monthly reports from the Toronto Homeless Memorial. Any duplicate entries are removed, and data are cross-checked by the Office of the Chief Coroner of Ontario (OCCO) (Gelfand (2022)).

Given that this dataset originates from major institutions in Toronto, its credibility is very high. The additional verification by the Office of the Chief Coroner of Ontario (OCCO) further enhances the reliability of the data.

Additionally, collecting data on the deaths of shelter residents is more reliable compared to tracking the entire population. First, Toronto is a large metropolitan area, where the presence of witnesses makes data falsification or omission nearly impossible. Second, tracking deaths is more dependable than attempting to count the entire population of shelter residents, as this group tends to be less organized, more transient, and harder to communicate with. As a result, it is not feasible to accurately account for the entire population in detail (House (2023)).

3 Result

I used the total number of shelter resident deaths per month as the target variable to be predicted, with time and season as predictors, and performed a linear regression analysis. Since time and season are evidently highly correlated, I decided to retain only time as the final predictor for the model. Additionally, since Figure 1 showed that the total number of decedents follows a normal distribution, it satisfied the conditions for using simple linear regression. After completing the simple linear regression, I obtained a positive slope, represented by a straight line (Figure 3), indicating that the number of shelter resident deaths in Toronto has increased over time. A comparison with the actual data revealed that the real values closely align with the predictions from the SLR model, demonstrating the reliability of my model.

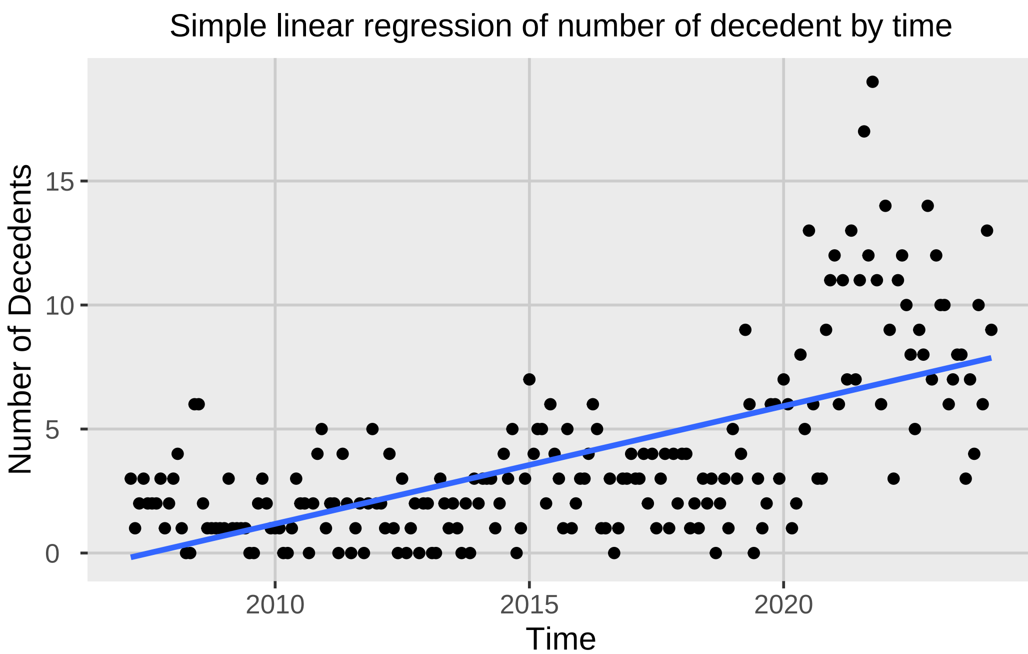

Figure 3: Simple linear regression of number of decedent by time

Figure 4 displays the total number of shelter resident deaths per year, along with their respective proportions. It is important to note that I excluded the data for 2007 and 2024, as these are incomplete years and including them could mislead the reader. From the chart, it is clear that the number of deaths between 2020 and 2023 was significantly higher, with each of these years accounting for over 10% of the total deaths during the 15-year period. The peak occurred in 2021, where the number of deaths exceeded 125, representing 17.8% of the total.

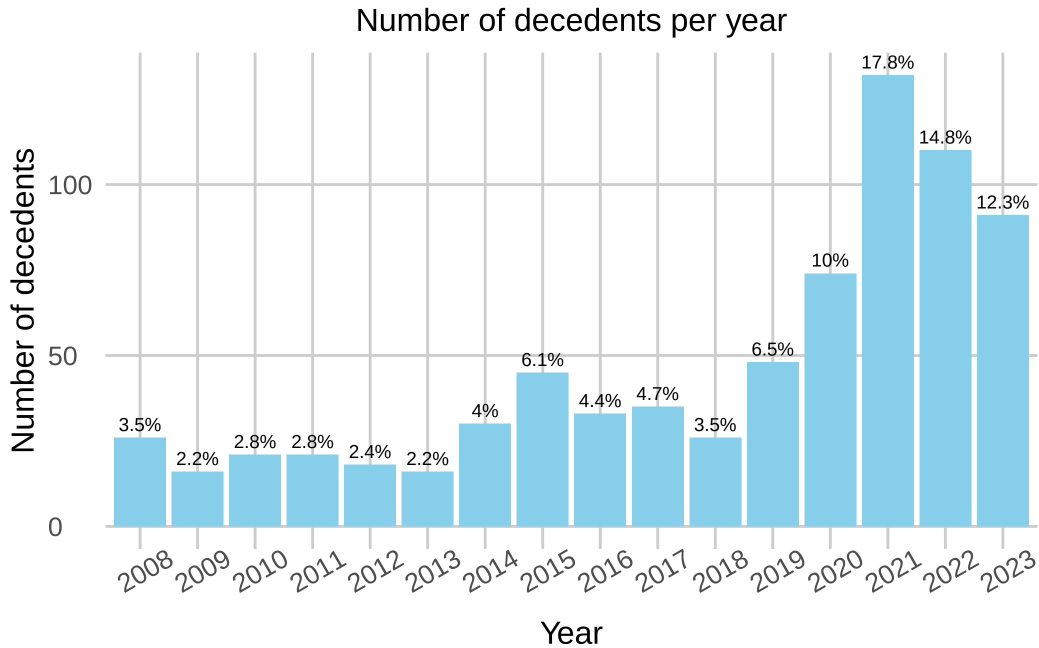

Figure 4: Number of decedents per Year

Additionally, Figure 5 shows the average number of deaths per season over the past few years. The proportions for spring, autumn, and summer are relatively similar, ranging between 23.6% and 23.8%. However, winter stands out with a significantly higher proportion, reaching 29%, notably surpassing the other three seasons.

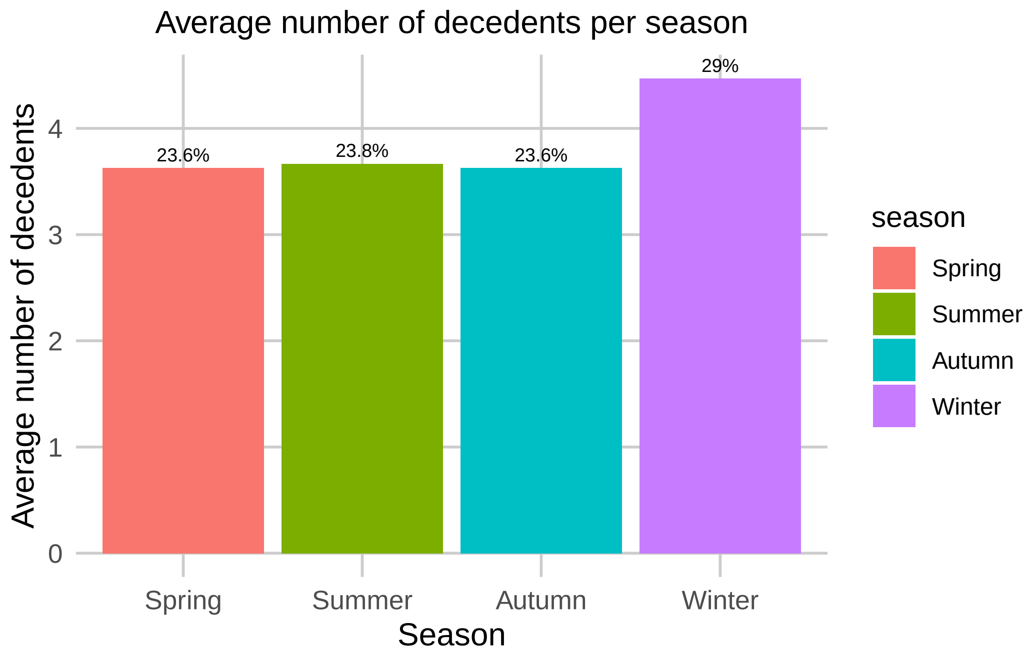

Figure 5: Average number of decedents per season

4 Discussion

4.1 Limitations

This study acknowledges several limitations. First, although the number of shelter resident deaths is more reliably recorded than the total shelter resident population, due to the unstructured nature of this group, the conclusions drawn about death trends may not fully reflect the challenges faced by the entire population. For instance, an increase in the total shelter population with a stable mortality rate would still result in a rise in deaths, which may not indicate worsening conditions. Second, the project utilizes a relatively small sample of data (204 months), and therefore, the results from the linear regression model may be subject to bias. However, this dataset represents the most comprehensive data available from government sources. Third, when analyzing seasonal variations, the sharp rise in death counts after 2020 introduces potential bias. It is possible that there were no seasonal differences before 2020, but new patterns emerged thereafter, which may have skewed the overall average and affected the seasonal analysis.

4.2 Suggestions for Future Research

There are several avenues for future research that could build upon the findings of this study. First, as noted in the limitations, this research focuses on the number of shelter resident deaths, which may not fully reflect the conditions of the entire shelter population. Future studies examining the relationship between the total shelter population and the number of deaths could provide a more comprehensive understanding of the current situation. Second, while seasonal analysis in this study may have been influenced by recent years' data, it is undeniable that shelter residents face heightened risks during the winter months at least in recent years. Research into the specific reasons behind increased winter threats, and an exploration of the actual needs of shelter residents during this period, would provide governments and welfare organizations with clearer insights to improve policies and intervention strategies. Finally, this research highlights the surge in death counts after 2020, but the underlying causes behind this spike should be further investigated by future scholars to prevent similar occurrences in the future.

References

Crane, Maureen, and Louise Joly. 2014. "Older Homeless People: Increasing Numbers and Changing Needs." *Reviews in Clinical Gerontology* 24 (4): 255–68. https://doi.org/10.1017/S095925981400015X.

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

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/.

House, Covenant. 2023. "Ending Homelessness for All: The Data Problem." https://www.covenanthouse.org/news/ending-homelessness-all-data-problem#:~:text='The%20transient%20nature%20of%20the,collection%20also%20leads%20to%20underreporting.

Müller, Kirill, and Lorenz Walthert. 2024. *Styler: Non-Invasive Pretty Printing of r Code*. https://CRAN.R-project.org/package=styler.

Omstead, Jordan. 2023. *Residents Demand Action from Toronto Mayor Olivia Chow as Another Shelter Hotel to Close*. globalnews.ca. https://globalnews.ca/news/9835385/toronto-shelter-hotel-to-close-residents-demand-action/.

Quigley, John M., Steven Raphael, and Eugene Smolensky. 2001. "Homeless in America, Homeless in California." *The Review of Economics and Statistics* 83 (1): 37–51. https://doi.org/10.1162/003465301750160027.

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

Rossi, Peter H., James D. Wright, Gene A. Fisher, and Georgianna Willis. 1987. "The Urban Homeless: Estimating Composition and Size." *Science* 235 (4794): 1336–41. https://doi.org/10.1126/science.2950592.

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://CRAN.R-project.org/package=dplyr.

Xie, Yihui. 2024. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.