Trends in Death License: Analyzing the Distribution of Death Licenses in Great Toronto Area*

Potentional Relation to Extreme Weather and Covid19

Shanjie Jiao

September 27, 2024

During the global pandemic, heightened sensitivity to mortality became evident, while urbanization is boosting disparities between urban and rural areas in terms of population density, economic growth, infrastructure, and public health services. While the number of deaths and death licenses are all conceptual data of death, but with two distinct concepts, COVID-19 showed a positive correlation with the number of recorded deaths. However, death licenses followed a different pattern, the peak of death license happens in January, whcih is recorded as the month with most sever weather, it may contribute to a higher death license number. And the data does not suggest it has a clear impact of COVID-19 on the issuance of death licenses. A proper understanding of the death license dataset will enable experts to conduct a more comprehensive analysis of the reasons for death and to improve health policy.

Table of contents

*Code and data are available at: https://github.com/Jie-jiao05/Paper-1.git

1 Introduction

While much of the existing research has focused on disparities in healthcare access, income, and community resources, few studies have examined the changes in death license issuance between urban and rural areas in response to various external factors during or after the COVID-19 pandemic. This study aims to fill that gap by analyzing death license data from Toronto and regions Outside City Limits, as recorded by civic centers, to observe how these areas have responded to different external factors.

Death licenses, as vital documents that provide detailed death information, offer a richer source of data than mere mortality numbers, allowing for a more nuanced approach to health-policy-making. The findings of this study may offer valuable insights for urban planners and public health experts, helping them understand the differences between urban and rural environments when extract data for further study that could used to investigate public health.

The paper is structured as follows: In Section 2 it will presents the general data trends from 2011 to 2024, including measurement techniques. Since the data for 2024 is incomplete, previous records are used to predict death license trends from September to December 2024. In Section 3 it critically examines the data, comparing the number of death licenses issued in Toronto and Outside City Limits on a monthly basis, with a focus on periods of high extreme weather frequency. The comparison also covers the years before and during the COVID-19 pandemic to investigate whether the increase in mortality caused by COVID-19 led to a higher issuance of death licenses.The results indicate that the peak in death licenses, particularly in January, may be linked to extreme weather events causing increased mortality. However, COVID-19 does not appear to have caused a significant change in the number of death licenses issued. In Section 4 The limitation of this research will be further elaborates .

2 Data

2.1 Data Source and Measurement

This report utilizes data collected by the City Clerk's Office, published in Open Data Toronto, titled "Death Registry Statistics" (Gelfand 2022). This specific dataset, used for the graphs and analyses in this paper, includes information about the number of death licenses issued, the civic centers (Scarborough, North York, Toronto, and Etobicoke) where they were issued, places of death, and time periods, along with unique identifiers for each record. Additionally, the dataset may not capture the complete number of deaths due to generate death license is totaly optional so not all death record will generate a death license. As this dataset only count the number of death license, so it do not include other personal information or detail in death. Importantly, no personal information is included in this dataset, as it is solely used to count the number of death licenses in Toronto and outside city limits, preserving the confidentiality of individuals' associated information (Gelfand 2022).

2.2 Data Tools

The dataset was analyzed using R (R Core Team 2023) and downloaded using the R package opendatatoronto (Gelfand 2022). Additionally, the packages tidyverse (Wickham et al. 2019), lubridate (Grolemund and Wickham 2011), readr (Wickham, Hester, and Bryan 2024), forecast (Hyndman and Khandakar 2008), ggplot (Wickham 2016), dplyr (Wickham et al. 2023), here(Müller 2020), forecast (Hyndman and Khandakar 2008), styler(Müller and Walthert 2024) have been used in data cleaning and visualization.

2.3 Data Characteristics

The raw dataset contained 948 unique observations and 5 variable columns. All data are recorded as either constants or strings, and each entry is filled with valid information. In the first cleaning step, this paper separated the "period" column into "year" and "month." Subsequently, I created a new "date" column by combining the year and month into a proper date format to facilitate easier data retrieval. A glimpse of the final cleaned dataset can be seen in Table 1.

Table 1: Sample of Cleaned Data of Death License in Toronto and Outside City Limits Area

id	civic_centre	death_licenses	place_of_death	year	month	date
27767	ET	69	Outside City Limits	2011	01	2011-01-01
27768	ET	341	Toronto	2011	01	2011-01-01
27769	NY	141	Outside City Limits	2011	01	2011-01-01
27770	NY	540	Toronto	2011	01	2011-01-01

Table 1: Sample of Cleaned Data of Death License in Toronto and Outside City Limits Area

id	civic_centre	death_licenses	place_of_death	year	month	date
27771	SC	129	Outside City Limits	2011	01	2011-01-01
27772	SC	545	Toronto	2011	01	2011-01-01
27773	TO	297	Toronto	2011	01	2011-01-01
27774	ET	83	Outside City Limits	2011	02	2011-02-01
27775	ET	224	Toronto	2011	02	2011-02-01
27776	NY	81	Outside City Limits	2011	02	2011-02-01

2.4 Data Record

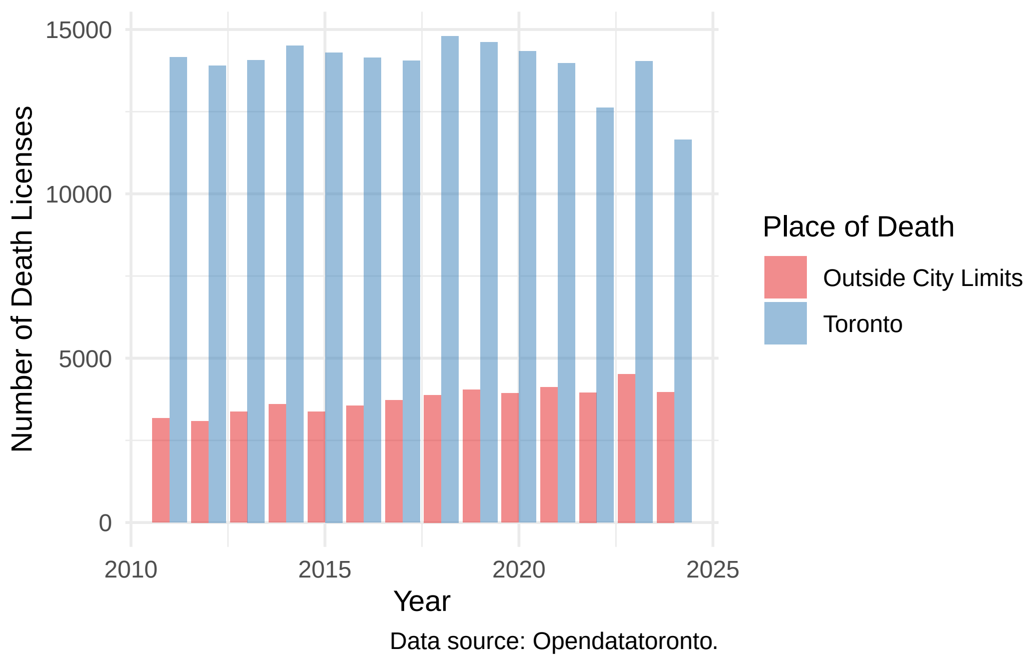

Figure 1: Histogram of Death Lincese From 2011 to 2024

Table Figure 1 illustrates a consistent difference in the number of death licenses issued in Toronto compared to the Outside City Limit area from 2011 to August 2024. While the general disparity between the two regions remains relatively unchanged, it is noteworthy that the number of death licenses in Toronto declined in 2022 but returned to the average levels in 2023. Since the dataset provided by OpenData Toronto, City Clerk's Office (Office, n.d.) is updated monthly, data for September to December 2024 cannot be obtained until the completion of this paper, which may introduce some bias for this specific period. In Section 2.5, this paper will attempt to predict this segment of data by building a predictive model using the historical data collected in this dataset.

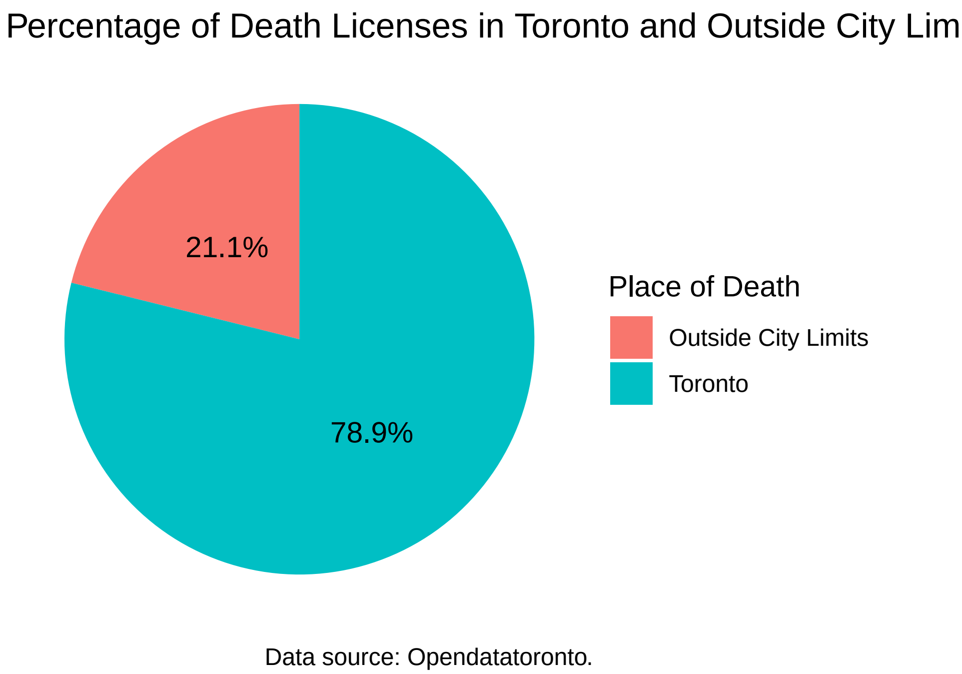

Figure 2

In Figure 2, it is evident that Toronto accounts for 78.9% of the death licenses issued during this time period, while Outside City Limits comprises the remaining 21.1%. This significant disparity in population density is a consequence of urbanization, which results in diminished resource inflow to Outside City Limits.

2.5 Data Prediction

To enhance the completeness of the data and improve year-over-year trend measurement, this paper utilizes time series analysis on the data from September to December for the years 2011 to 2023. This approach captures trends to forecast and generate future values. The results of this predictive model are illustrated in Figure 3. Additionally, we can simulate overall data by incorporating these predictions into the bar chart displayed in Figure 4, to maintain the rigor of the analysis, the generated prediction data will not be included in this study.

Based on the available data, a slow upward trend can be observed for Outside City Limits, while the Toronto area exhibits a more fluctuating trend. However, as indicated by predictions model shown in Figure 3, it is anticipated that the total number of death licenses in 2024 will exceed the mean line, resulting in a slight upward adjustment of the true mean line compared to the simulation presented in this paper.

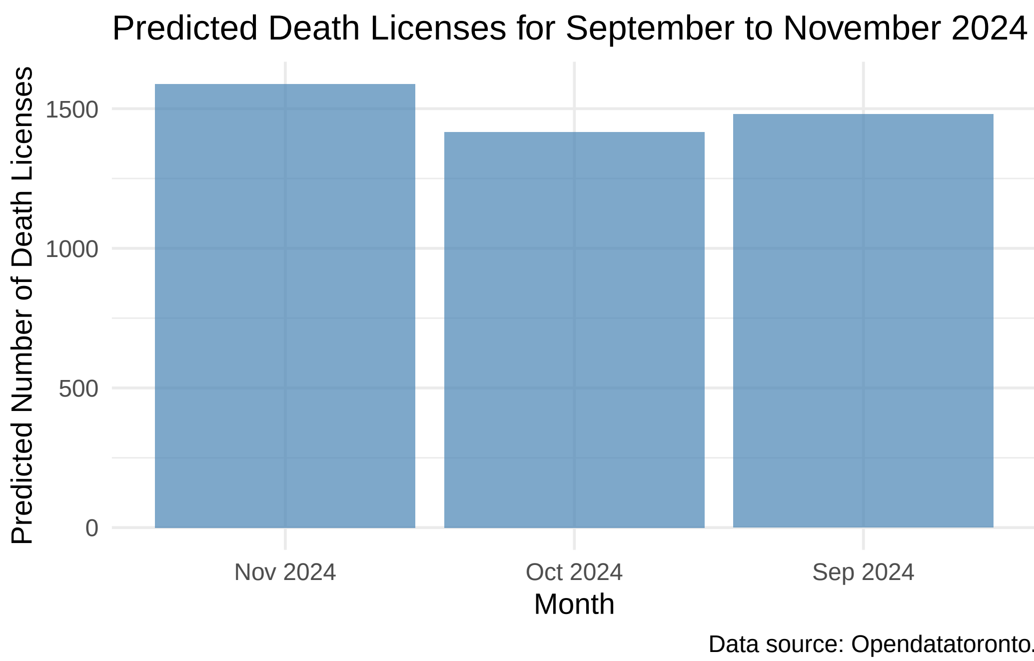

Figure 3

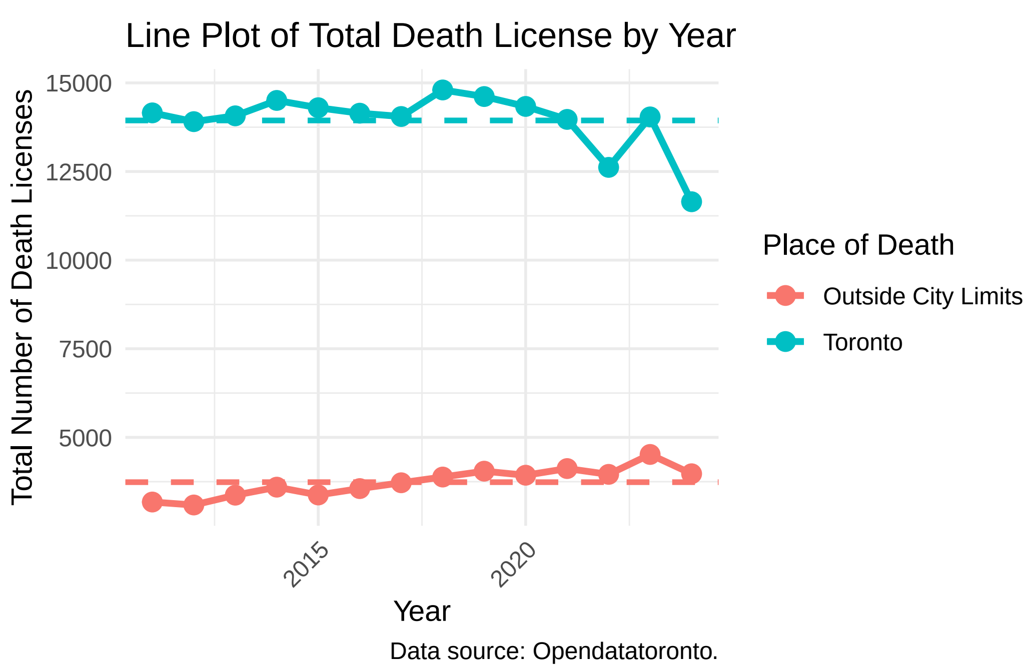

Figure 4: Line Plot of Total Death License by Year

3 Results

3.1 Death License with Month

The peak in death licenses consistently occurs in January each year. In recent years, extreme weather has been linked to an increase in mortality and may have an impact on the number of death licenses. As illustrated in Figure 5, the bar chart displays the total number of death licenses by month from 2011 to 2024 (excluding September to December 2024), revealing that January is the only month with a notable peak. In contrast, the remaining months generally fall within the mean line. According to data from the City of Toronto, there were a total of 262 extreme weather events from 2011 to 2022, with 115 occurring in January Toronto (2022) . Given that Toronto's summer climate is more favorable compared to other regions, it may have a lower impact from extreme heat on mortality in summer. However, severe cold weather in winter could contribute to an increase in deaths, leading to a higher number of death licenses issued in January. A population-based study published in CMAJ, titled "Assessment of the Effect of Cold and Hot Temperatures on Mortality in Ontario, Canada," reveals that each 5°C change in daily temperature is estimated to induce 7 excess deaths per day during cold seasons and 4 excess deaths during warm seasons Chen et al. (2016) . This suggests that extreme winter weather may be a contributing factor to the increase in mortality which may have a positive relation with death license.

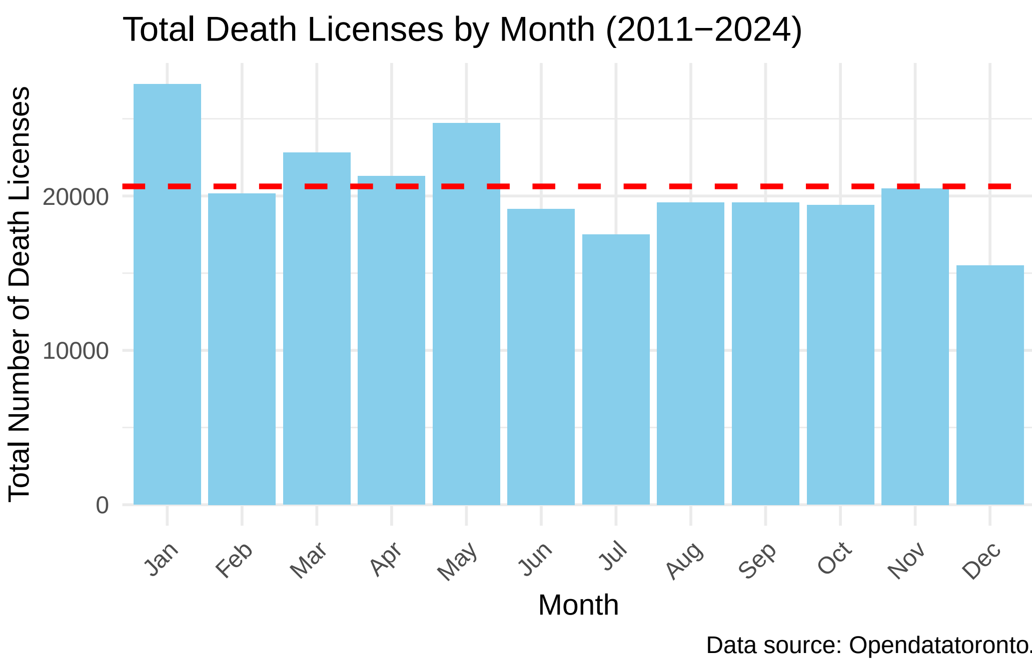

Figure 5: Total Death License by Month

3.2 Death License with Covid

Figure 6 presents the dataset displayed beneath the dot plot, which has been narrowed down to the period from 2017 to 2024 to facilitate a visual observation of changes in the number of death licenses during the COVID-19 pandemic. The data for each month is presented separately. There was not a significant change in death licenses during the pandemic years (2019-2022), with the graph revealing extreme values in only three months. Notably, one of these peaks occurred in August 2020, but this spike did not indicate a consistent trend and subsequently fell back, suggesting it should be treated as an outlier and not considered statistically significant. Overall, the distribution of death licenses appears to align more closely with a uniform distribution.

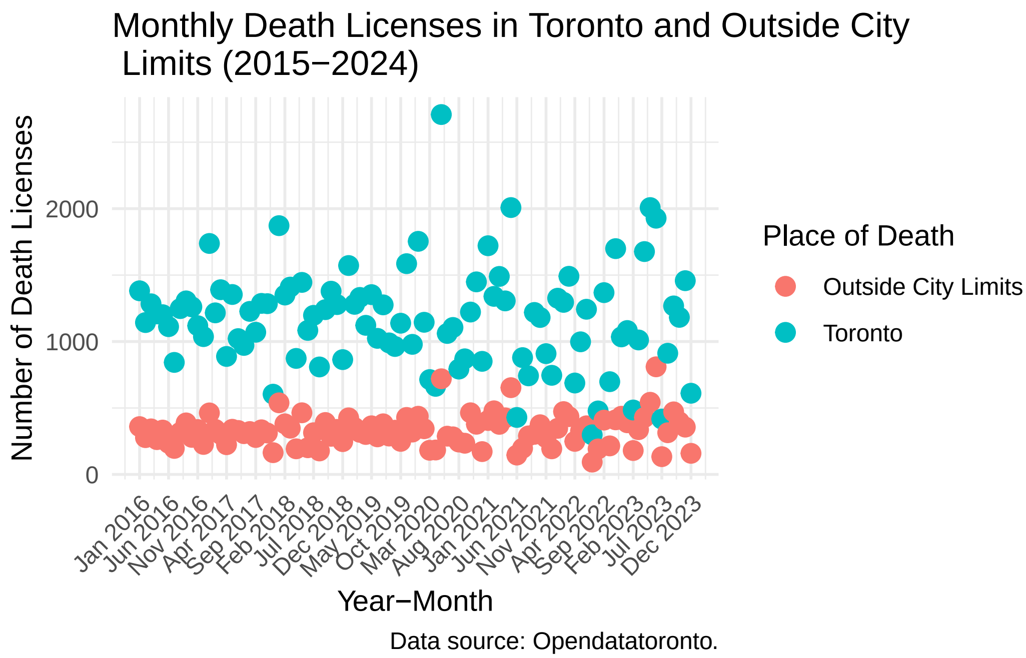

Figure 6: Monthly Death License Dot Plot

4 Discussion

This paper analyzes the number of death licenses registered by civic centers from 2011 to September 2024. By visualizing death licenses by year and month and employing time series forecasting, the study concludes that the peak in death licenses occurs in January, a month often associated with severe weather conditions, which may contribute to an increase in the number of death licenses issued.Despite the negative impacts of COVID-19 on Canada's economy, society, and healthcare, the distribution of death licenses each year does not reveal a noticeable trend, suggesting that it likely follows a uniform distribution. Consequently, the dataset does not support a conclusive positive or negative correlation with the pandemic.

The dataset consists only of numerical counts, civic centers, and localities, making it impossible to determine the causes of death. Moreover, since death licenses are not mandatory for all deaths and do not equate to the total number of deaths in the Toronto area, this data cannot provide a comprehensive analysis of total deaths or their causes. However, it is useful for government officials to understand how many deaths have occurred in the Toronto area and to facilitate the application for death licenses for administrative purposes.For government departments, death license statistics are valuable for registration and the improvement of health policies. As of the completion of this article, data for the period from September to December 2024 is not yet available. However, the main data analyzed avoids this specific time frame, and the projections made in the article are sufficiently described, minimizing potential biases in the analysis.

5 LLMs

Statement on LLM usage: LLMs were used in the making of this paper.

References

Chen, Hong, Jun Wang, Qiongsi Li, Abderrahmane Yagouti, Eric Lavigne, Richard Foty, Richard T Burnett, Paul J Villeneuve, Sabit Cakmak, and Ray Copes. 2016. "Assessment of the Effect of Cold and Hot Temperatures on Mortality in Ontario, Canada: A Population-Based Study." *CMAJ Open* 4 (1): E48–58. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4866918/.

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/.

Hyndman, Rob J, and Yeasmin Khandakar. 2008. "Automatic Time Series Forecasting: The Forecast Package for R." *Journal of Statistical Software* 27 (3): 1–22. https://doi.org/10.18637/jss.v027.i03.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

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

Office, City Clerk's. n.d. "Open Data Dataset." City of Toronto Open Data Portal. https://open.toronto.ca/dataset/death-registry-statistics/.

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, City of. 2022. "Statistics on Extreme Cold Weather Alerts in Toronto." City of Toronto. https://www.toronto.ca/city-government/data-research-maps/research-reports/housing-and-homelessness-research-and-reports/statistics-on-extreme-cold-weather-alerts-in-toronto/.

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

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.