

Unmasking the Grim Reality: Age Disparities in COVID-19 Fatality Rates in Toronto*

Kaavya Kalani

January 23, 2024

This paper presents a meticulous analysis of COVID-19 cases in Toronto, emphasizing the disproportionate impact on different age groups. Leveraging data from the City of Toronto’s Open Data Library, the study scrutinizes fatality rates, revealing a stark contrast between older and younger demographics. From the first reported case in January 2020 to January 2023, our analysis sheds light on the severity of outcomes, with a particular focus on the vulnerability of older individuals. The comprehensive exploration includes data preprocessing details, variable descriptions, and insightful visualizations. Findings underscore the urgency of targeted interventions for the elderly population in pandemic response strategies. The study contributes valuable insights for public health initiatives, emphasizing the need for age-specific considerations in managing and mitigating the impact of COVID-19.

Table of contents

1	Introduction	2
2	Data	2
2.1	About the dataset	3
2.2	Preprocessing and data cleaning	3
2.3	Variable of interest	3
2.4	Glimpse of the dataset	4
3	Results and Discussion	5
4	Conclusion	7

*Code and data are available at: https://github.com/kaavyakalani26/covid19_toronto

1 Introduction

The COVID-19 or the coronavirus pandemic is a global pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Wikipedia 2024). The novel virus was first identified in Wuhan, China in December 2019, and spread worldwide in early 2020. COVID-19 was declared a pandemic by the World Health Organization on March 11, 2020 (WHO 2020). This pandemic, to this date, has seen 700 million+ cases and around 7 million deaths (worldometers 2024).

We saw people from all over the world falling sick, people of all ages and health conditions. Some studies suggested that old people were more likely to be susceptible and in danger from this compared to younger. (Disease Control and Prevention 2023)

In this paper, we try to uncover these trends. We use the data of all the COVID-19 cases reported to Toronto Public Health since the first case which was reported in January 2020 until January 2023. My focus is on investigating age disparities to understand if older individuals exhibit a higher susceptibility to contracting the virus. Additionally, I analyze mortality rates across different age groups, with a specific emphasis on the elderly population to ascertain if age plays a significant role in COVID-19 severity. Further, I explore the correlation between age and the need for hospitalization or ICU admission, shedding light on healthcare utilization patterns. This study aims to contribute insights into the nuanced impact of COVID-19 on diverse age demographics, particularly emphasizing the vulnerability of older individuals. These findings intend to inform targeted public health interventions and strategies.

The paper is further organised into three sections: Data, Results, and Conclusion. In the Data section, I discuss how the dataset to be used for the analysis was obtained, pre-processed and cleaned. I will also explain the variables of interest for our analysis in the dataset. The Results section will then highlight and discuss the trends and associations found during the analysis. Lastly, the Conclusion section will summarize the major takeaways from our analysis.

2 Data

The data used for analysis in this paper was obtained from the City of Toronto's Open Data Library using the library `opendatatoronto` (Gelfand 2022). The dataset is called "COVID-19 Cases in Toronto" (Data 2024). Data was cleaned and analysed using the statistical programming software R (R Core Team 2023) along with the help of `tidyverse` (Wickham et al. 2019), `knitr` (Xie 2014), `ggplot2` (Wickham 2016), `here` (Müller 2020)

2.1 About the dataset

The “COVID-19 Cases in Toronto” dataset obtained for this analysis contains demographic, geographic, and severity information for all confirmed and probable cases reported to and managed by Toronto Public Health since the first case was reported in January 2020. The data are extracted from the provincial Case & Contact Management System (CCM).(Data 2024)

The dataset is completely refreshed and overwritten bi-weekly and the version I extracted was last refreshed on January 17, 2024

2.2 Preprocessing and data cleaning

The data was preprocessed and cleaned to contain only the entries that were essential for my analysis. This included me,

- Keeping only confirmed cases. The original dataset included probable cases also but for my analysis, we wanted to find trend among people who were confirmed to have COVID so for precision in trend, we omitted the probable cases.
- Keeping records of cases till Dec 31, 2022. This was because the dataset page documents how criteria and updating methods were changed multiple time throughout 2023. To ensure coherent and similarly-collected data for precision of result, this call was taken.
- Cases which are still active were removed as analysing the outcome of the case is one of my trends to observe which having active cases would not help.
- The last minor cleaning included removing empty age categories as that is a variable of interest and changing the datatype for the Reported date variable.

2.3 Variable of interest

I have 6 variables I will be using in my analysis from the original dataset. My variables of interest are :

1. ID

Numerical variable which is a unique row identified for the database

2. Age Group

Age at time of illness (in years). This is a categorical variable which is split into 9 groups : 19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, 90+

3. Reported Date

The date on which the case was reported to Toronto Public Health.

4. Outcome

Categorical variable for the outcome of the case. This can either of two values :

- Fatal: Any case that has died and has been marked as Outcome equals ‘Fatal’ and Type of Death does not equal ‘Disease of Public Health Significance was unrelated to cause of death’ in the provincial reporting system (CCM). (Data 2024)
- Resolved: A case outcome description in CCM of ‘Recovered’ OR Case outcome description is equal to ‘Fatal’ AND Type of Death is equal to ‘Disease of Public Health Significance was unrelated to cause of death’ OR Today’s date is more than 14 days from episode date AND the case is not currently hospitalized/intubated/in ICU AND Case outcome description is not equal to ‘Fatal’ where Type of Death is not equal to ‘Disease of Public Health Significance was unrelated to cause of death’. (Data 2024)

5. Ever Hospitalized

Cases that were hospitalized related to their COVID-19 infection (includes cases that are currently hospitalized and those that have been discharged or are deceased). This is a categorical variable taking the value of either “yes” or “no”.

6. Ever in ICU

Cases that were admitted to the intensive care unit (ICU) related to their COVID-19 infection (includes cases that are currently in ICU and those that have been discharged or are deceased). This is a categorical variable taking the value of either “yes” or “no”.

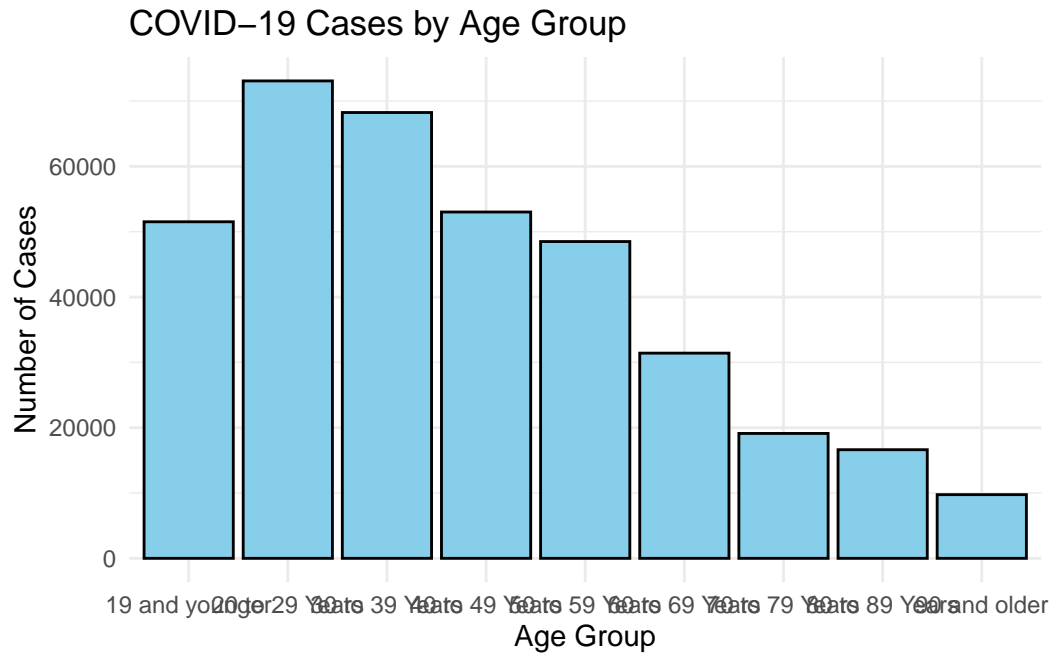
2.4 Glimpse of the dataset

Table 1: Sample of cleaned Covid-19 data

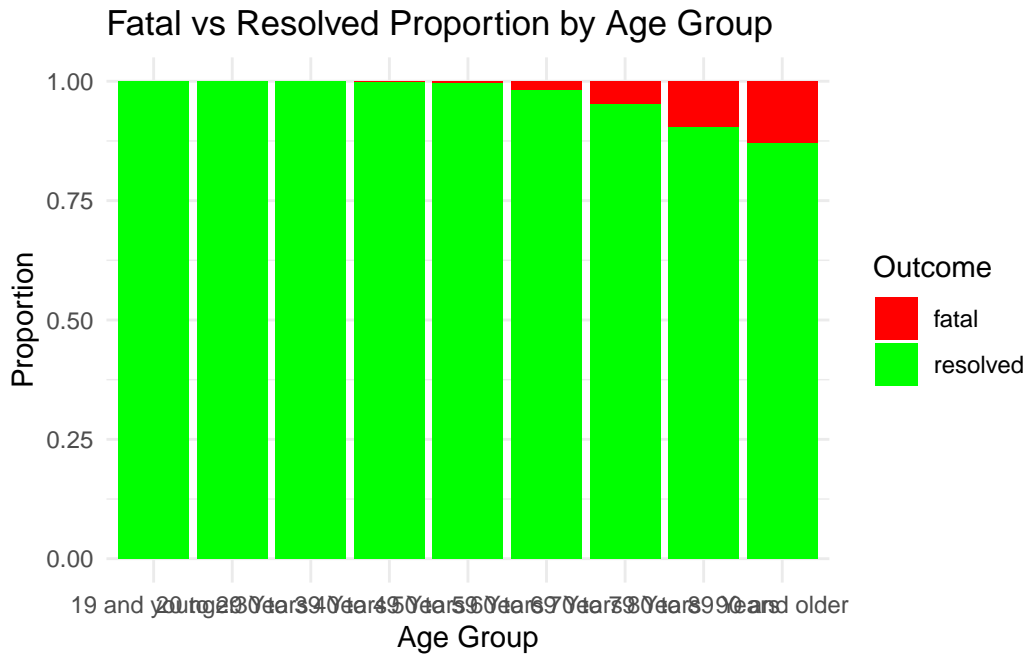
ID	Age Group	Reported Date	Outcome	Ever Hospitalized	Ever in ICU
1	50 to 59 Years	2020-01-23	resolved	no	no
2	50 to 59 Years	2020-01-23	resolved	yes	no
3	20 to 29 Years	2020-02-21	resolved	no	no
4	60 to 69 Years	2020-02-25	resolved	no	no
5	60 to 69 Years	2020-02-26	resolved	no	no
6	50 to 59 Years	2020-02-27	resolved	no	no

Table 1 provides a glimpse of our our cleaned COVID-19 data. We will now use the data to uncover trends and make associations.

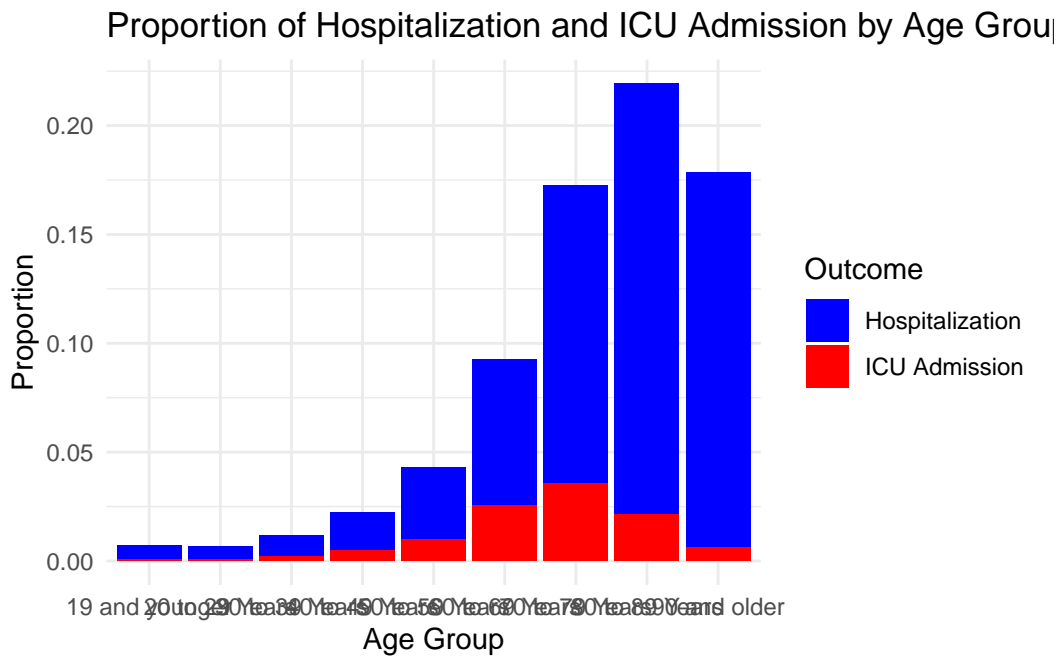
3 Results and Discussion

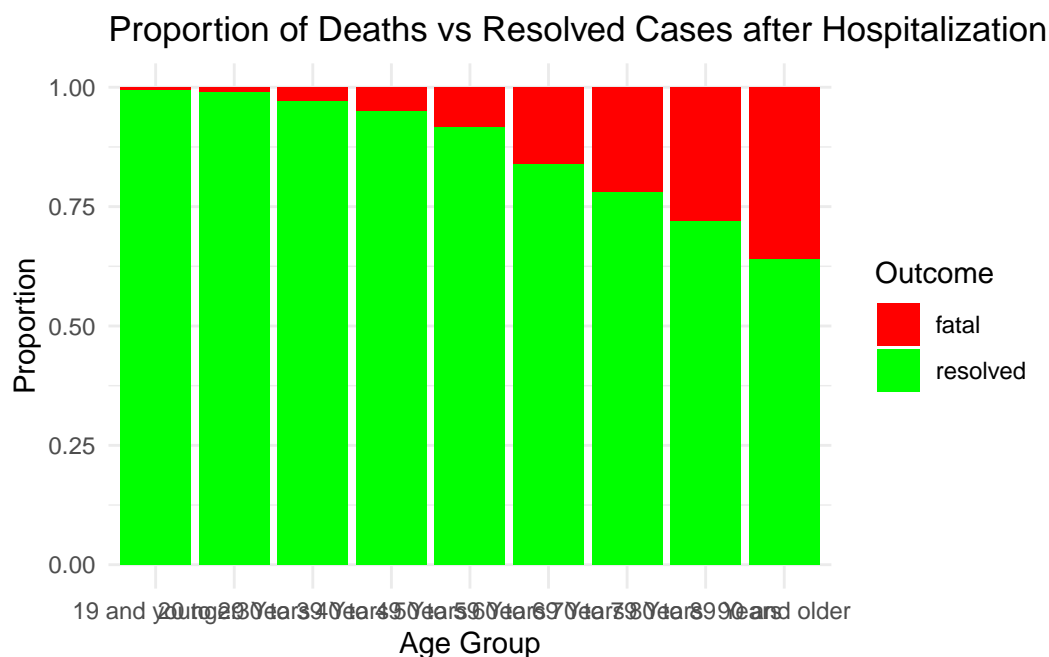


```
# A tibble: 2 x 3
  outcome Count Proportion
  <chr>   <int>     <dbl>
1 fatal    4721     0.0127
2 resolved 366564    0.987
```



	Outcome	Proportion
1	Hospitalization	0.044475268
2	ICU Admission	0.007737991





4 Conclusion

In conclusion, our comprehensive analysis underscores a discernible pattern revealing that older individuals exhibit a heightened vulnerability to a fatal outcome upon contracting COVID-19. This assertion is substantiated by the noteworthy increase in mortality rates observed in higher age categories. Moreover, our findings illuminate a parallel trend, indicating an elevated likelihood of ICU admission as age progresses. The cumulative evidence strongly suggests that advancing age is a critical determinant in the severity of COVID-19 outcomes, emphasizing the imperative for targeted public health measures and specialized care strategies for the elderly population.

References

- Data, Toronto Open. 2024. "COVID-19 Cases in Toronto." <https://open.toronto.ca/dataset/covid-19-cases-in-toronto/>.
- Disease Control, Centers for, and Prevention. 2023. "COVID-19 Risks and Information for Older Adults." <https://www.cdc.gov/aging/covid19/index.html>.
- Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. <https://sharlagelfand.github.io/opendatatoronto/>.
- Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. <https://here.r-lib.org/>.

- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- WHO. 2020. “WHO Director-General’s Opening Remarks at the Media Briefing on COVID-19 - 11 March 2020.” <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>.
- 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 Golemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wikipedia. 2024. “COVID-19 Pandemic.” https://en.wikipedia.org/wiki/COVID-19_pandemic.
- worldometers. 2024. “Coronavirus Cases.” <https://www.worldometers.info/coronavirus/>.
- Xie, Yihui. 2014. “Knitr: A Comprehensive Tool for Reproducible Research in R.” In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. <http://www.crcpress.com/product/isbn/9781466561595>.