Toronto Covid Cases Statistics and Analysis*

A Case Studie with Graghs and Tables

Xiaoyu Miao

January 24, 2024

This dataset is about covid-19 cases in Toronto published by Toronto Public Health, and it's classified by age, gender, neighborhood, covid status, etc. In this data analysis, we find out that people aged 20-29 in Toronto are the group with the highest reported number of Covid-19 cases, whereas people aged 90 and above are the group with the lowest reported number of covid-19 cases. We also find out that covid cases are almost equally distributed in neibourhoods around the city.

1 Introduction

In responding to the COVID-19 pandemic, Toronto Public Health (TPH) has compiled a comprehensive dataset, capturing the spread and impact of the virus in Toronto since January 2020. This dataset is rich with details, including demographic, geographic, and the severity of all confirmed and probable COVID-19 cases. Such detailed information is crucial for understanding how the pandemic unfolds in the community, covering both individual cases and larger outbreaks. Yet, as the pandemic has evolved, so too has this dataset, reflecting the dynamic nature of the situation and the shifting strategies in public health response.

Significant changes in the dataset, particularly from February 2023 onwards, illustrate the challenges in maintaining consistent and reliable data during a global health crisis. Key data fields like "currently hospitalized" and "currently in ICU" were removed due to difficulties in obtaining accurate discharge information. These changes, coupled with alterations in data entry practices, have created gaps and inconsistencies in the data, posing challenges for a clear and continuous understanding of the pandemic's trajectory.

This paper presents a focused analysis of the TPH dataset, navigating these changes to extract meaningful insights about COVID-19's impact in Toronto. Despite the hurdles in data reporting, the study identifies evolving patterns in the virus's spread and its effects on different

*Code and data are available at: LINK.

demographics. The importance of this analysis lies in its contribution to a more comprehensive understanding of the pandemic, offering valuable insights for public health officials and policymakers. By analyzing the dataset, discussing its implications, and suggesting future strategies, this research provides a crucial lens through which to view the management of health data during an ongoing global health crisis.

```
# A tibble: 10 x 1
   Age.Group
   <chr>
 1 50 to 59 Years
 2 20 to 29 Years
 3 60 to 69 Years
 4 80 to 89 Years
 5 70 to 79 Years
 6 30 to 39 Years
 7 40 to 49 Years
 8 19 and younger
 9 <NA>
10 90 and older
```

```
# A tibble: 141 x 1
   Neighbourhood.Name
   <chr>
 1 Willowdale East
 2 Parkwoods-Donalda
 3 Church-Yonge Corridor
 4 Newtonbrook West
 5 Milliken
 6 Willowdale West
 7 Henry Farm
 8 Don Valley Village
 9 Lawrence Park South
10 Bridle Path-Sunnybrook-York Mills
# i 131 more rows
```

```
# A tibble: 10 x 2
   Age.Group          n
   <chr>          <int>
 1 20 to 29 Years 77574
 2 30 to 39 Years 73479
 3 40 to 49 Years 57156
 4 19 and younger 55205
```

```
 5 50 to 59 Years 52834
 6 60 to 69 Years 35647
 7 70 to 79 Years 23839
 8 80 to 89 Years 23046
 9 90 and older   14272
10 <NA>             422
```

```
# A tibble: 10 x 1
   Age.Group
   <chr>
 1 50 to 59 Years
 2 20 to 29 Years
 3 60 to 69 Years
 4 80 to 89 Years
 5 70 to 79 Years
 6 30 to 39 Years
 7 40 to 49 Years
 8 19 and younger
 9 <NA>
10 90 and older
```

```
# A tibble: 141 x 1
   Neighbourhood.Name
   <chr>
 1 Willowdale East
 2 Parkwoods-Donalda
 3 Church-Yonge Corridor
 4 Newtonbrook West
 5 Milliken
 6 Willowdale West
 7 Henry Farm
 8 Don Valley Village
 9 Lawrence Park South
10 Bridle Path-Sunnybrook-York Mills
# i 131 more rows
```

```
# A tibble: 141 x 2
   Neighbourhood.Name                   n
   <chr>                            <int>
 1 <NA>                             13608
 2 Waterfront Communities-The Island  9101
 3 Woburn                             8787
```

```
 4 Downsview-Roding-CFB                7657
 5 Rouge                               7361
 6 West Humber-Clairville              7335
 7 Malvern                             7311
 8 Mount Olive-Silverstone-Jamestown   6641
 9 Islington-City Centre West          6270
10 Glenfield-Jane Heights              5882
# i 131 more rows
```

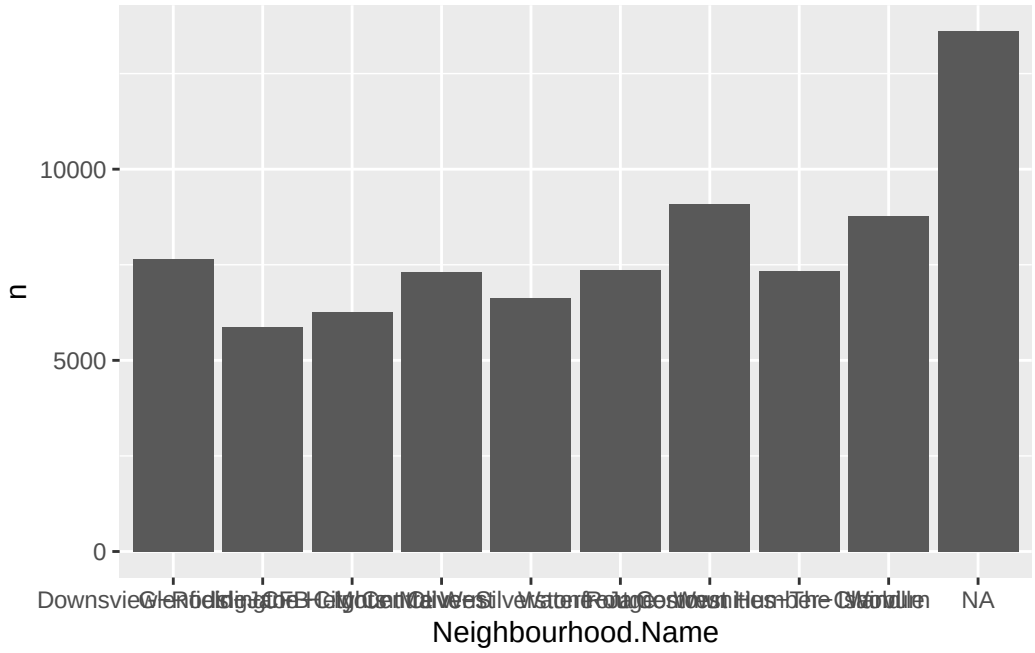

Figure 1: By Neibourhoods

```
# A tibble: 413,474 x 15
    X_id Assigned_ID Outbreak.Associated Age.Group      Neighbourhood.Name FSA
   <dbl>       <dbl> <chr>               <chr>          <chr>              <chr>
 1     1           1 NO                  50 to 59 Years Willowdale East    M2N
 2     2           2 NO                  50 to 59 Years Willowdale East    M2N
 3     3           3 NO                  20 to 29 Years Parkwoods-Donalda  M3A
 4     4           4 NO                  60 to 69 Years Church-Yonge Corr~ M4W
 5     5           5 NO                  60 to 69 Years Church-Yonge Corr~ M4W
 6     6           6 NO                  50 to 59 Years Newtonbrook West   M2R
 7     7           7 NO                  80 to 89 Years Milliken           M1V
 8     8           8 NO                  60 to 69 Years Willowdale West    M2N
 9     9           9 NO                  50 to 59 Years Willowdale East    M2N
```

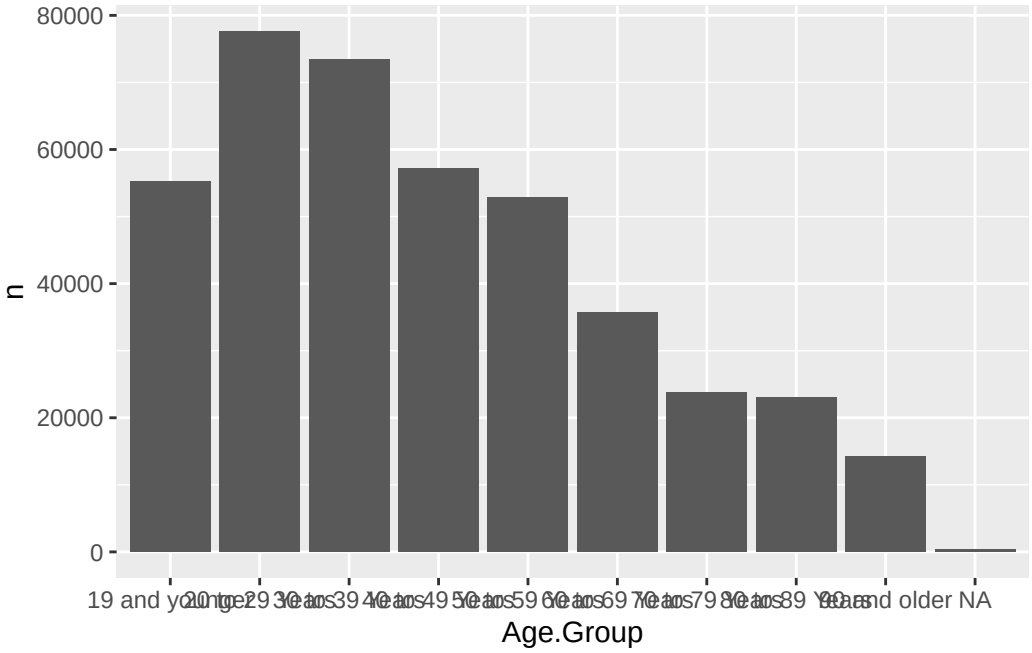

Figure 2: By Neibourhoods

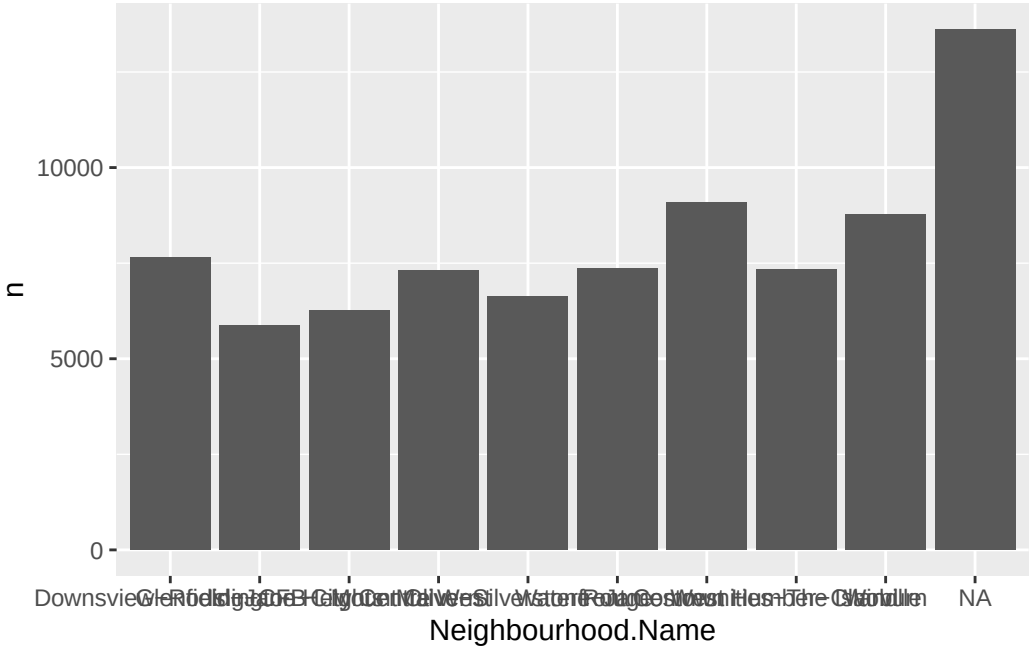

Figure 3: By Neibourhoods

```
10    10          10 NO                   60 to 69 Years Henry Farm         M2J
# i 413,464 more rows
# i 9 more variables: Source.of.Infection <chr>, Classification <chr>,
#   Episode.Date <date>, Reported.Date <date>, Client.Gender <chr>,
#   Outcome <chr>, Ever.Hospitalized <chr>, Ever.in.ICU <chr>,
#   Ever.Intubated <chr>
```

Age: In this graph, we sorted the dataset by age group. We can find that out of all people who caught Covid in Toronto, age ranged from 20-29 years old has the highest number of people, and 30-39 years old has the second highest number. 90 and older has the lowest number.

District: In this graph, we sorted the dataset by district in Toronto. Because there are way too many districts in the dataset, we only chose 10 of them to make the graph. We can find that Covid cases is almost equally distributed in every district.

Related Literature: According to Covid-19 Cases Age Distribution Canada on Statistica website, people aged 20-20 years old are indeed the largest group who caught Covid-19. This is also shown on record article by Public Health Ontario. The same thing about different communities/neighborhoods is also shown in the report. Reference is included.

2 References

#LLM Usage: README.md

No LLM is used in this paper. All the contents are based on learnings from the course materials and classmates discussions.

references.bib

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.