

# Exploring COVID-19 data for Toronto, Canada

STA303/1002 Data exploration assessment

Yifan Xu

Daily: Data as of January 29, 2021.

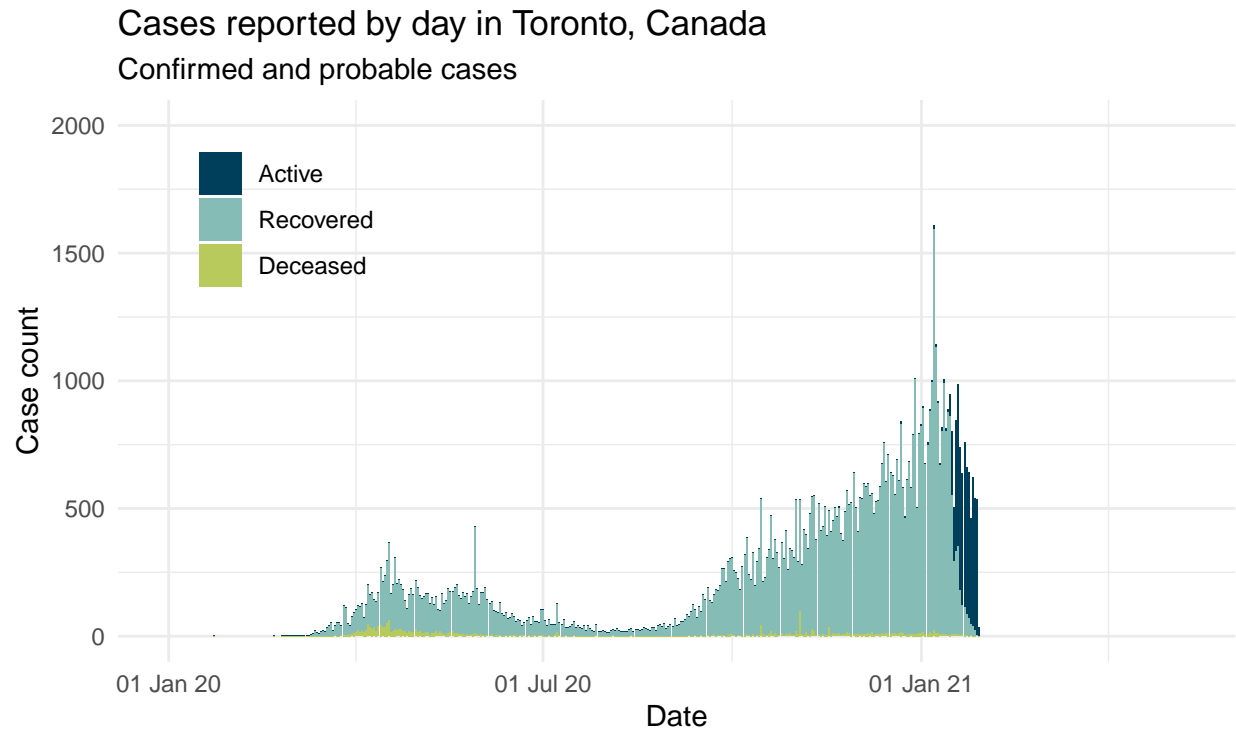
Neighbourhood: Data as of January 28, 2021.

## Task 1: Daily cases

```
reported <- reported_raw %>%
  mutate_if(is.numeric, replace_na, replace = 0) %>%
  mutate(reported_date = date(reported_date)) %>%
  pivot_longer(-c(reported_date),
               names_to = "type", values_to = "number") %>%
  mutate(type = str_to_sentence(type)) %>%
  mutate(type = fct_relevel(type, "Deceased", after = 2))
```

## Data visualization

```
reported %>%
  ggplot(aes(x = reported_date, y = number, fill = type)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(title = "Cases reported by day in Toronto, Canada",
       subtitle = "Confirmed and probable cases",
       x = "Date",
       y = "Case count",
       caption = str_c("Created by: <Yifan Xu> for STA303/1002, U of T\n",
                       "Source: Ontario Ministry of Health, Integrated\n",
                       "Public Health Information System and CORES\n",
                       date_daily[1,1])) +
  theme(legend.title = element_blank(), legend.position = c(0.15, 0.8)) +
  scale_y_continuous(limits = c(0, 2000)) +
  scale_fill_manual(values = c("#003F5C", "#86BCB6", "#B9CA5D")) +
  scale_x_date(labels = scales::date_format("%d %b %y"),
               limits = c(date("2020-01-01"), Sys.Date()))
```



Created by: <Yifan Xu> for STA303/1002, U of T  
Source: Ontario Ministry of Health, Integrated  
Public Health Information System and CORES  
Data as of January 29, 2021

## Task 2: Outbreak type

```

outbreak <- outbreak_raw %>%
  mutate(episode_week = date(episode_week)) %>%
  mutate(outbreak_or_sporadic = str_replace(outbreak_or_sporadic,
                                             "OB Associated", "Outbreak associated")) %>%
  mutate(outbreak_or_sporadic = fct_relevel(outbreak_or_sporadic,
                                             "Outbreak associated", after = 1))

outbreak2 <- outbreak %>%
  group_by(episode_week) %>%
  summarise(total_cases = sum(cases), .groups = "drop")

# add the variable "total_cases" to the outbreak data set
outbreak <- outbreak %>%
  left_join(outbreak2, by = c("episode_week" = "episode_week"))

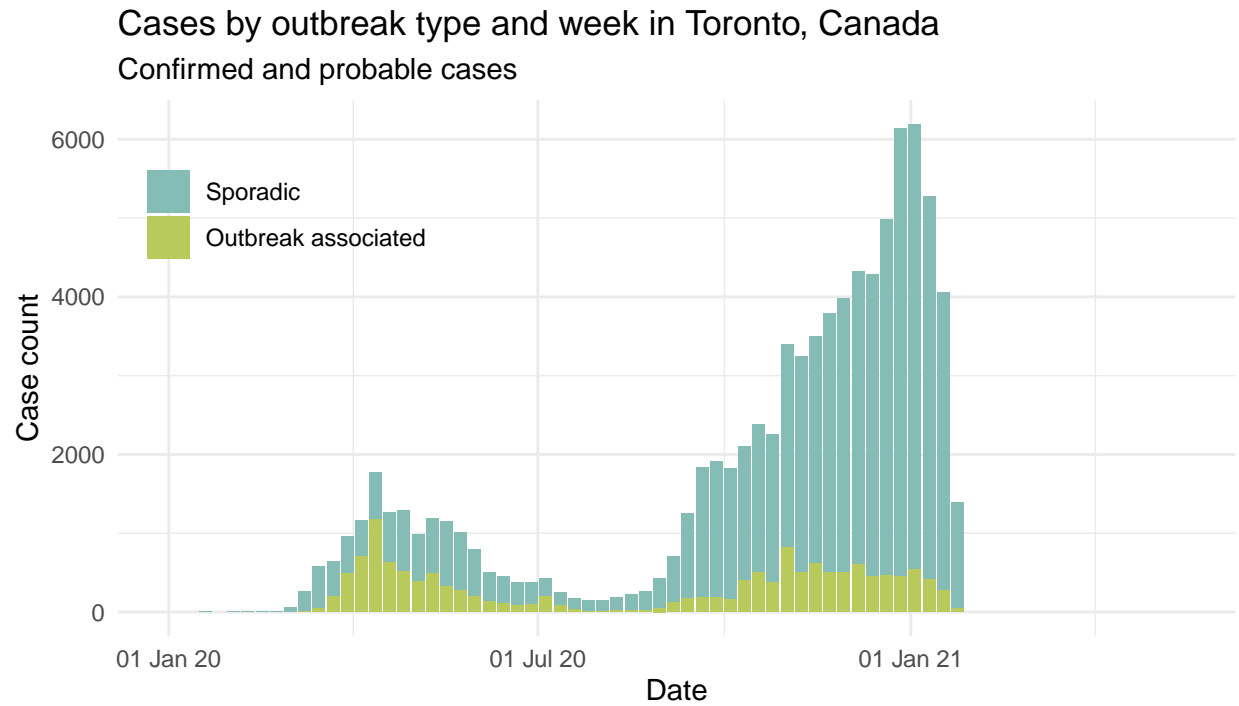
```

## Data visualization

```

outbreak %>%
  ggplot(aes(x = episode_week, y = cases, fill = outbreak_or_sporadic)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(title = "Cases by outbreak type and week in Toronto, Canada",
       subtitle = "Confirmed and probable cases",
       x = "Date",
       y = "Case count",
       caption = str_c("Created by: <Yifan Xu> for STA303/1002, U of T\n",
                       "Source: Ontario Ministry of Health,\n",
                       "Integrated Public Health Information System\n",
                       "and CORES\n",
                       date_daily[1,1])) +
  theme(legend.title = element_blank(), legend.position = c(0.15, 0.8)) +
  scale_y_continuous(limits = c(0, max(outbreak$total_cases))) +
  scale_fill_manual(values = c("#86BCB6", "#B9CA5D")) +
  scale_x_date(labels = scales::date_format("%d %b %y"),
              limits = c(date("2020-01-01"), Sys.Date()+7))

```



Created by: <Yifan Xu> for STA303/1002, U of T  
Source: Ontario Ministry of Health,  
Integrated Public Health Information System  
and CORES  
Data as of January 29, 2021

## Task 3: Neighbourhoods

```
income <- nbhood_profile %>%
  filter(`_id` == 1143) %>%
  pivot_longer(2:146, names_to = "neighbourhood_name", values_to = "percentage")
```

```
# choose "neighbourhood_name" and "percentage" columns
income <- income[, c(2, 3)]
income <- income[5:145, ]
```

```
income <- income %>%
  mutate(percentage = as.numeric(percentage))
```

```
# mutate the name
nbhoods_all <- nbhoods_shape_raw %>%
  mutate(neighbourhood_name = str_remove(nbhoods_shape_raw$AREA_NAME, "\\s\\((\\d+\\))$"))
```

```
# change into the same neighbourhood names for 3 neighbourhoods
income[91, 1] <- "North St.James Town"
income[21, 1] <- "Cabbagetown-South St.James Town"
nbhoods_all[46, 19] <- "Weston-Pelham Park"
nbhood_raw[32, 2] <- "North St.James Town"
nbhood_raw[59, 2] <- "Cabbagetown-South St.James Town"
```

```
# merge 3 data sets
nbhoods_all <- nbhoods_all %>%
  left_join(income, by = c("neighbourhood_name" = "neighbourhood_name" )) %>%
  left_join(nbhood_raw, by = c("neighbourhood_name" = "neighbourhood_name" ))
```

```
# delete NA columns to make the data set tidy
nbhoods_all <- nbhoods_all[, -c(4, 9, 10, 11, 12)]
nbhoods_all <- nbhoods_all[, -c(12, 13)]
```

```
# rename
names(nbhoods_all)[names(nbhoods_all) == "rate_per_100_000_people"] <- "rate_per_100000"
```

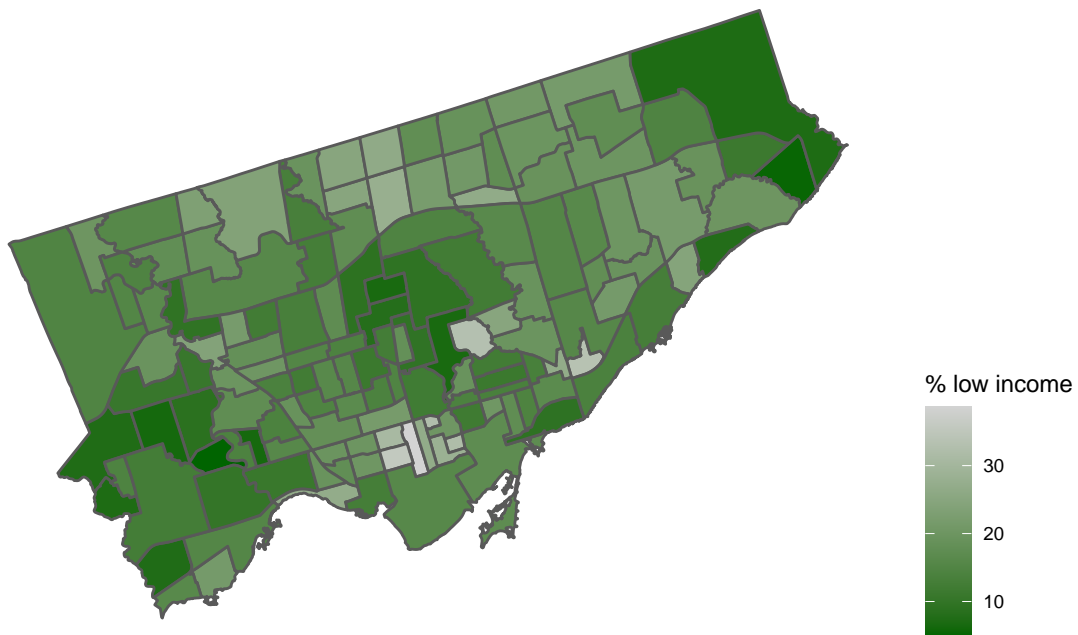
```
nbhoods_final <- nbhoods_all %>%
  mutate(med_inc = median(percentage)) %>%
  mutate(med_rate = median(rate_per_100000)) %>%
  mutate(nbhood_type = case_when(
    percentage >= med_inc & rate_per_100000 >= med_rate ~ "Higher low income rate, higher case rate",
    percentage >= med_inc & rate_per_100000 < med_rate ~ "Higher low income rate, lower case rate",
    percentage < med_inc & rate_per_100000 >= med_rate ~ "Lower low income rate, higher case rate",
    percentage < med_inc & rate_per_100000 < med_rate ~ "Lower low income rate, lower case rate"
  ))
```

## Data visualization

```
ggplot(data = nbhoods_final) +
  geom_sf(aes(fill = percentage)) +
```

```
theme_map() +  
scale_fill_gradient(name= "% low income", low = "darkgreen", high = "lightgrey") +  
labs(title = "Percentage of 18 to 64 year olds living in a low income family (2015)",  
      subtitle = "Neighbourhoods of Toronto, Canada",  
      caption = str_c("Created by: <Yifan Xu> for STA303/1002, U of T\n",  
                      "Source: Census Profile 98-316-X2016001\n",  
                      "via OpenData Toronto\n",  
                      date_daily[1,1])) + theme(legend.position="right")
```

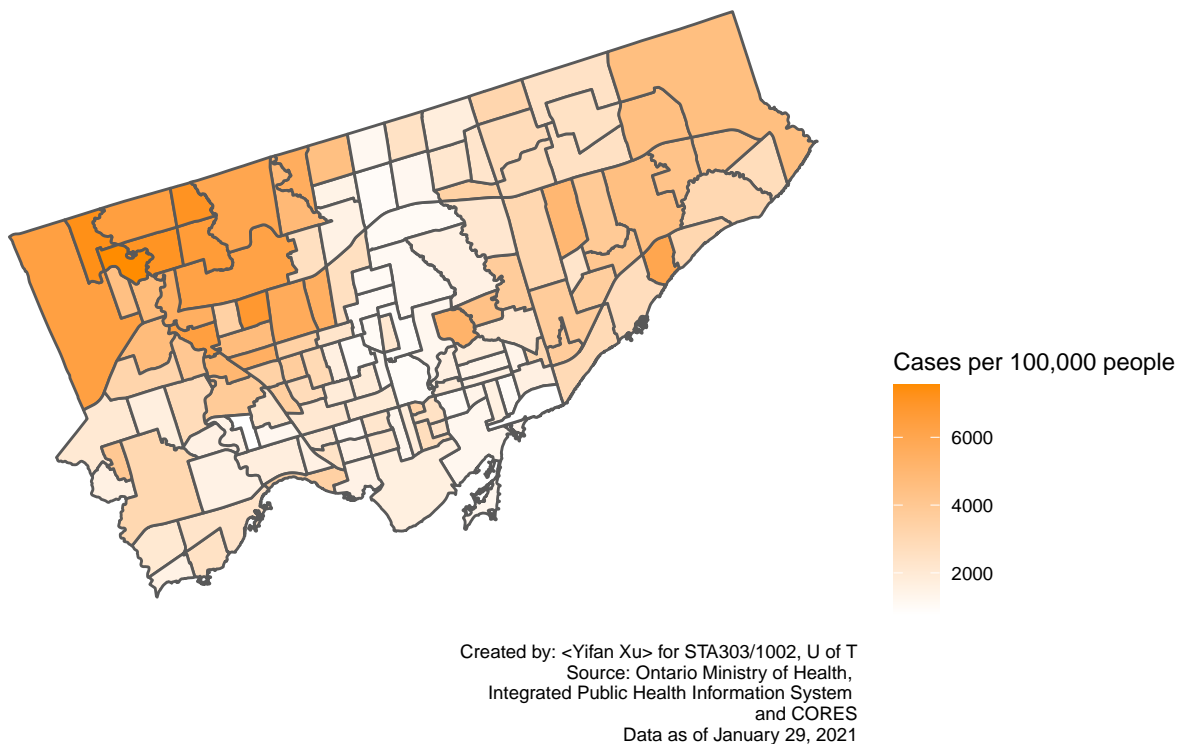
Percentage of 18 to 64 year olds living in a low income family (2015)  
Neighbourhoods of Toronto, Canada



Created by: <Yifan Xu> for STA303/1002, U of T  
Source: Census Profile 98-316-X2016001  
via OpenData Toronto  
Data as of January 29, 2021

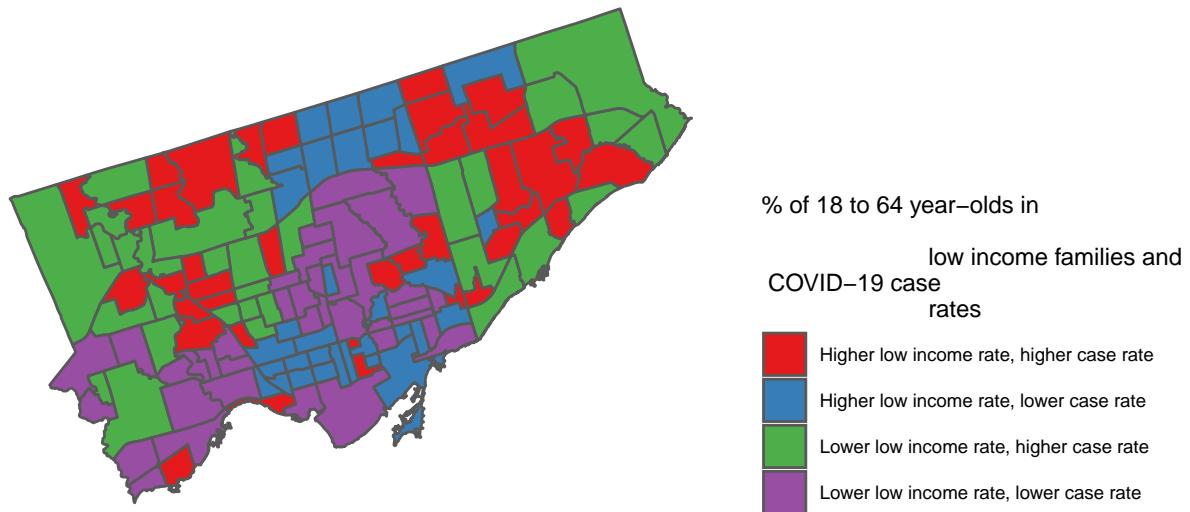
```
ggplot(data = nbhoods_final) +  
  geom_sf(aes(fill = rate_per_100000)) +  
  theme_map() +  
  scale_fill_gradient(name= "Cases per 100,000 people", low = "white", high = "darkorange") +  
  labs(title = "COVID-19 cases per 100,000, by neighbourhood in Toronto, Canada",  
        caption = str_c("Created by: <Yifan Xu> for STA303/1002, U of T\n",  
                          "Source: Ontario Ministry of Health,  
                          Integrated Public Health Information System  
                          and CORES\n",  
                          date_daily[1,1])) + theme(legend.position="right")
```

COVID-19 cases per 100,000, by neighbourhood in Toronto, Canada



```
ggplot(data = nbhoods_final) +
  geom_sf(aes(fill = nbhood_type)) +
  theme_map() +
  labs(title = "COVID-19 cases per 100,000, by neighbourhood in Toronto, Canada",
        caption = str_c("Created by: <Yifan Xu> for STA303/1002, U of T\n",
                          "Income data source: Census Profile 98-316-X2016001\n",
                          "via OpenData Toronto\n",
                          "COVID data source: Ontario Ministry of Health, Integrated Public\n",
                          "Health Information System and CORES\n",
                          date_daily[1,1])) +
  scale_fill_brewer(palette = "Set1") +
  theme(legend.position="right") +
  guides(fill=guide_legend(title="% of 18 to 64 year-olds in\n",
                           low income families and\n",
                           COVID-19 case\n",
                           rates"))
```

COVID-19 cases per 100,000, by neighbourhood in Toronto, Canada



Created by: <Yifan Xu> for STA303/1002, U of T  
Income data source: Census Profile 98-316-X2016001  
via OpenData Toronto  
COVID data source:  
Ontario Ministry of Health, Integrated Public  
Health Information System and CORES  
Data as of January 29, 2021