

DataExploration

WingYan Lee

Code last run 2021-04-06.

Daily: Data as of February 10, 2021.

Neighbourhood: Data as of February 9, 2021.

Task 1: Daily cases

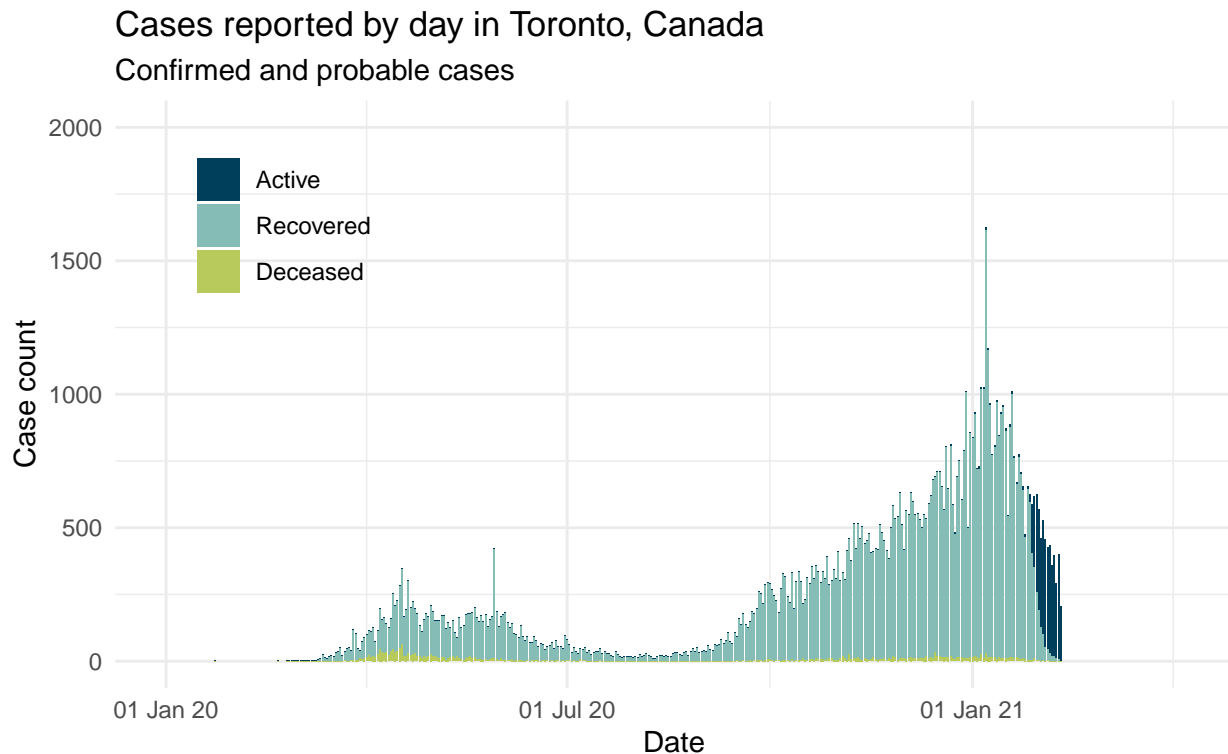
Data wrangling

```
reported <- reported_raw %>%
  mutate_if(is.numeric, replace_na, replace = 0) %>%
  mutate(reported_date = date(reported_date)) %>%
  rename(Active = active, Recovered = recovered, Deceased = deceased)

reported <- reported %>%
  pivot_longer(-reported_date, names_to = "Type", values_to = "Cases") %>%
  mutate(Type = fct_relevel(Type, "Active", "Recovered", "Deceased"))
```

Data visualization

```
reported %>%
  ggplot(aes(x = reported_date, y = Cases, fill = Type)) +
  geom_bar(stat = "identity") +
  scale_x_date(labels = scales::date_format("%d %b %y"),
               limits=c(date("2020-01-01"), Sys.Date())) +
  scale_y_continuous(limits = c(0, 2000), breaks = seq(0, 2000, by = 500)) +
  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: WingYan Lee 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.title = element_blank(), legend.position = c(0.15, 0.8)) +
  scale_fill_manual(values = c("#003F5C", "#86BCB6", "#B9CA5D"))
```



Created by: WingYan Lee for STA303/1002, U of T
Source: Ontario Ministry of Health, Integrated Public Health Information System and CORES
Data as of February 10, 2021

Task 2: Outbreak type

Data wrangling

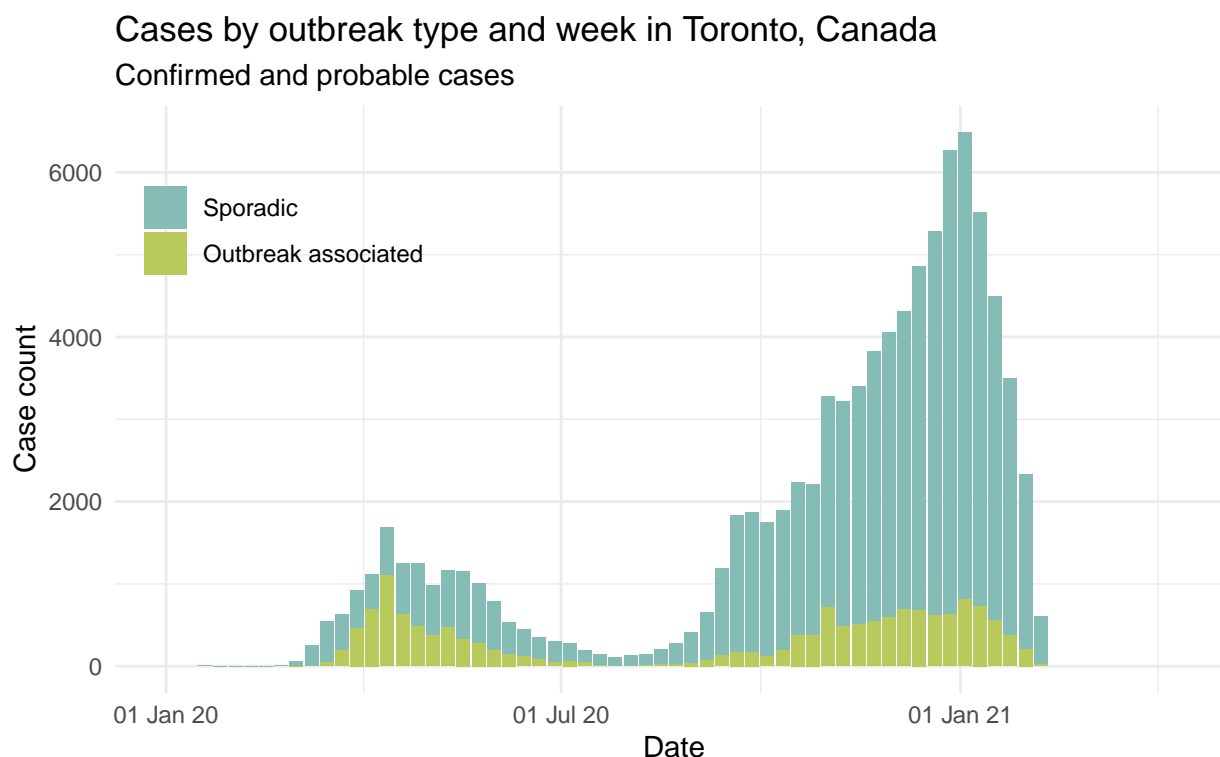
```
outbreak <- outbreak_raw %>%
  mutate(episode_week = date(episode_week),
         outbreak_or_sporadic = str_replace_all(outbreak_or_sporadic,
                                                  "OB Associated", "Outbreak associated"),
         outbreak_or_sporadic = fct_relevel(outbreak_or_sporadic, "Sporadic", after = 0))

outbreak_cases <- outbreak %>%
  group_by(episode_week) %>%
  summarise(total_cases = sum(cases))

outbreak <- outbreak %>%
  left_join(outbreak_cases, by = "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: WingYan Lee for STA303/1002, U of T\n",
                       "Source: Ontario Ministry of Health, Integrated Public Health Information System and CORES\n",
                       "Health Information System and CORES\n",
                       date_daily[1,1])) +
  scale_x_date(labels = scales::date_format("%d %b %y"),
              limits = c(date("2020-01-01"), Sys.Date() + 7)) +
  scale_y_continuous(limits = c(0, max(outbreak$total_cases))) +
  theme(legend.title = element_blank(), legend.position = c(0.15, 0.8)) +
  scale_fill_manual(values = c("#86BCB6", "#B9CA5D"))
```



Created by: WingYan Lee for STA303/1002, U of T
Source: Ontario Ministry of Health, Integrated Public Health Information System and CORES
Data as of February 10, 2021

Task 3: Neighbourhoods

Data wrangling: part 1

```
income <- nbhood_profile %>%
  filter(`_id` == 1143) %>%
  pivot_longer(-c(`_id`, Category, Topic, `Data Source`, Characteristic),
               names_to = "neighbourhood_name", values_to = "percentage") %>%
  filter(neighbourhood_name != "City of Toronto") %>%
  select(-`_id`, -Category, -Topic, -`Data Source`, -Characteristic) %>% #Remove useless columns
  mutate(percentage = parse_number(percentage))
```

Data wrangling: part 2

```
income <- income %>%
  mutate(neighbourhood_name = str_replace(neighbourhood_name, "St. James", "St.James"),
         neighbourhood_name = str_replace(neighbourhood_name, "Pelham", "Pellam"))

nbhood_raw <- nbhood_raw %>%
  mutate(neighbourhood_name = str_replace(neighbourhood_name, "St. James", "St.James"),
         neighbourhood_name = str_replace(neighbourhood_name, "Pelham", "Pellam"))

nbhoods_all <- nbhoods_shape_raw %>%
  mutate(neighbourhood_name = str_remove(AREA_NAME, "\\s\\(\\d+\\)$")) %>%
  full_join(income, by = "neighbourhood_name") %>%
  full_join(nbhood_raw, by = "neighbourhood_name") %>%
  filter(neighbourhood_name != "Missing Address/Postal Code" &
         neighbourhood_name != "City of Toronto") %>%
  rename(rate_per_100000 = rate_per_100_000_people)
```

Data wrangling: part 3

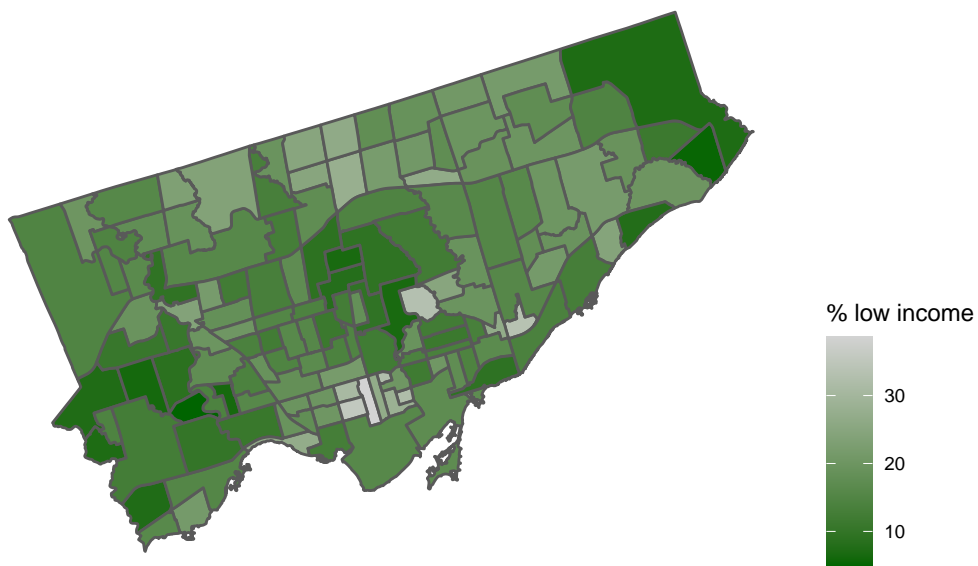
```
nbhoods_final <- nbhoods_all %>%
  mutate(med_inc = median(percentage),
         med_rate = median(rate_per_100000),
         nbhood_type = ifelse(percentage >= med_inc,
                               ifelse(rate_per_100000 >= med_rate,
                                     "Higher low income rate,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

Data visualization

```
ggplot(data = nbhoods_final) +  
  geom_sf(aes(fill = percentage)) +  
  theme_map() +  
  theme(legend.position = "right") +  
  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: WingYan Lee for STA303/1002, U of T\n",  
                          "Source: Census Profile 98-316-X2016001 via OpenData Toronto\n",  
                          date_nbhood[1,1]))
```

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

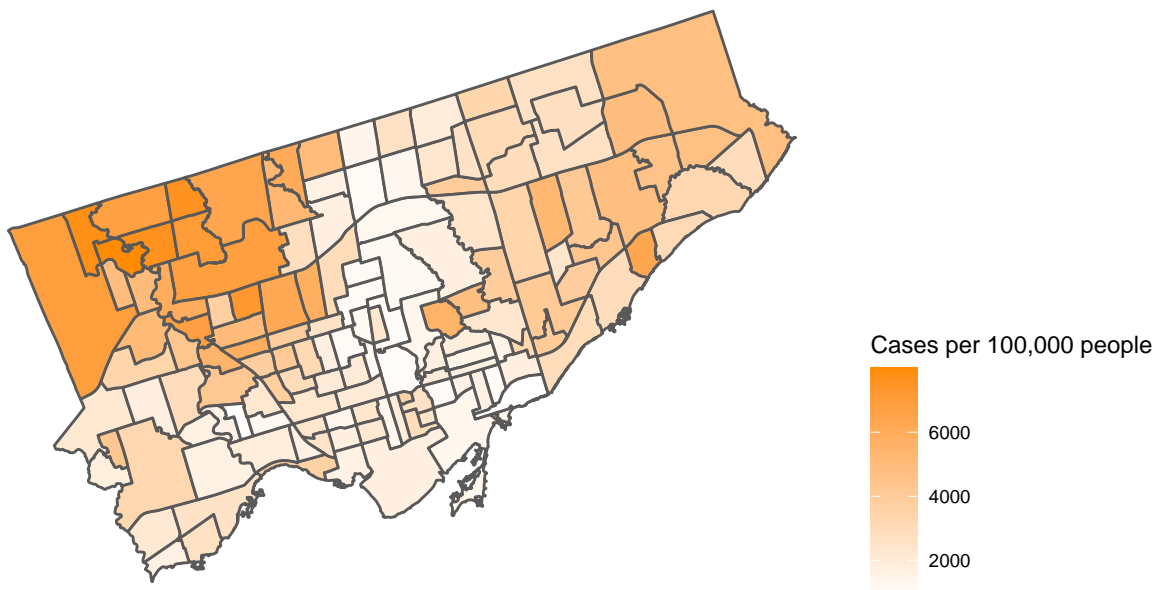
Neighbourhoods of Toronto, Canada



Created by: WingYan Lee for STA303/1002, U of T
Source: Census Profile 98-316-X2016001 via OpenData Toronto
Data as of February 9, 2021

```
ggplot(data = nbhoods_final) +
  geom_sf(aes(fill = rate_per_100000)) +
  theme_map() +
  theme(legend.position = "right") +
  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: WingYan Lee for STA303/1002, U of T\n",
                       "Source: Ontario Ministry of Health, Integrated Public Health Information System and CORES\n",
                       "date_nbhood[1,1]"))
```

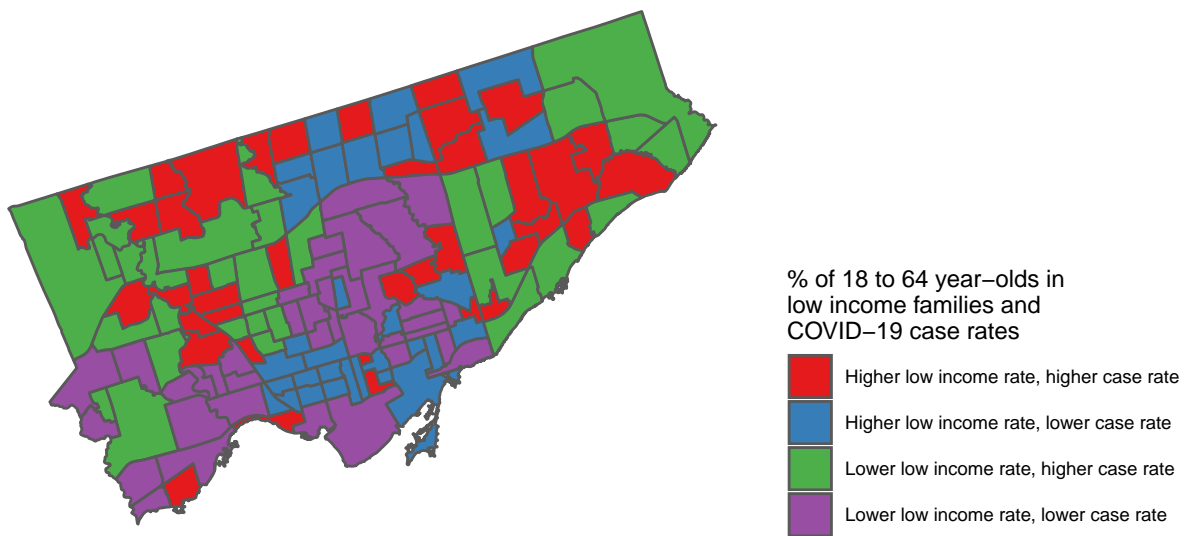
COVID-19 cases per 100,000, by neighbourhood in Toronto, Canada



Created by: WingYan Lee for STA303/1002, U of T
 Source: Ontario Ministry of Health, Integrated Public Health Information System and CORES
 Data as of February 9, 2021

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

COVID-19 cases and low income status by neighbourhood in Toronto, Canada



Created by: WingYan Lee 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 February 9, 2021