

Code last run 2021-02-16.

Daily: Data as of January 29, 2021.

Neighbourhood: Data as of January 28, 2021.

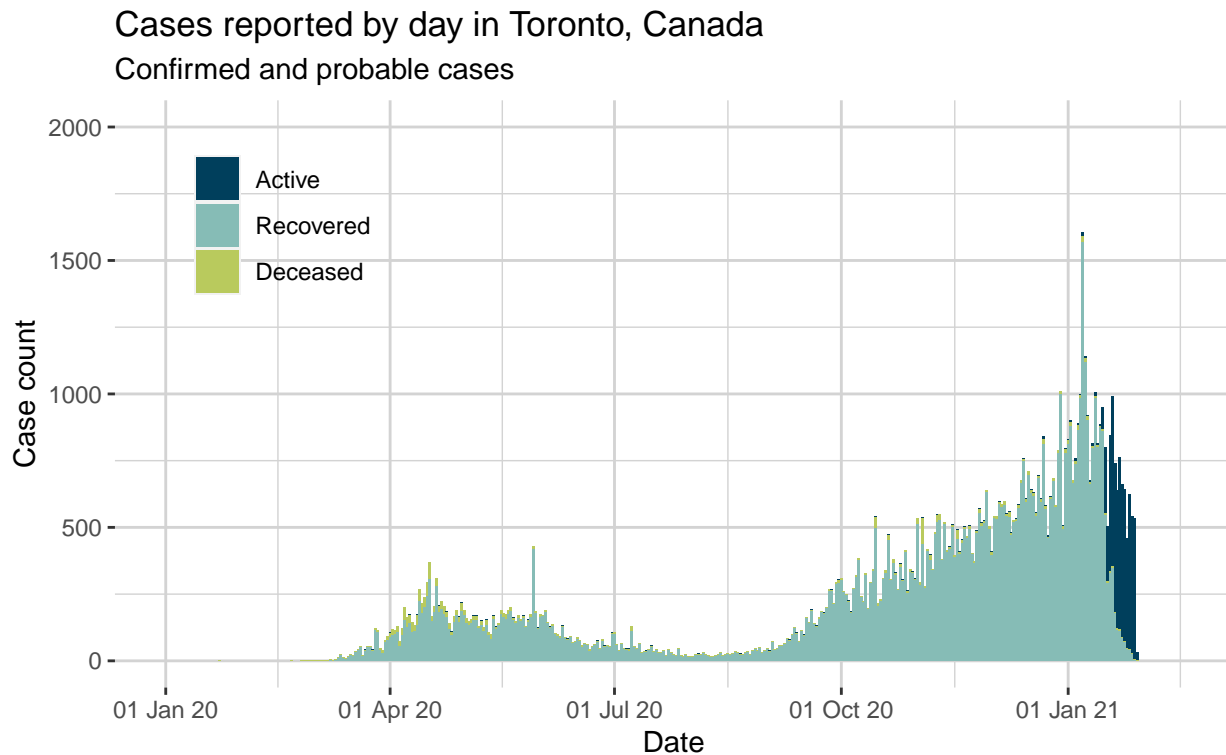
## Task 1: Daily cases

### Data wrangling

```
reported <- reported_raw %>%  
  mutate_if(is.numeric, replace_na, replace = 0) %>%  
  arrange(reported_date, active, recovered, deceased)  
reported$reported_date <- date(reported$reported_date)  
reported <- janitor::clean_names(reported, "title") %>%  
  pivot_longer(~`Reported Date`, names_to = "Status", values_to = "Count")
```

## Data visualization

```
reported %>%
  ggplot(aes(x = `Reported Date`, y = Count, fill = Status))+
  geom_bar(position = "stack", stat = 'identity', width= 1) +
  labs(title = "Cases reported by day in Toronto, Canada", subtitle = "Confirmed and probable cases",
       x= "Date", y= "Case count",
       caption = str_c("Created by: Gongen Zhong for STA302/1002 U of T\nSource: Ontario Ministry of Health",
                        "Integrated Public Health Information System and CORES"))+
  theme(legend.position = c(.15, .8), legend.title = element_blank(),
        legend.background = element_blank())+
  scale_y_continuous(limits = c(NA, 2000)) +
  scale_x_date(labels= scales::date_format("%d %b %y"), limits = c(date("2020-01-01"), Sys.Date())) +
  scale_fill_manual(values = c("#003F5C", "#86BCB6", "#B9CA5D"),
                    breaks = c("Active", "Recovered", "Deceased"))+
  theme(panel.background = element_rect('white'), panel.grid.major = element_line('lightgrey'),
        panel.grid.minor = element_line('lightgrey'))
```



Created by: Gongen Zhong for STA302/1002 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

### Data wrangling

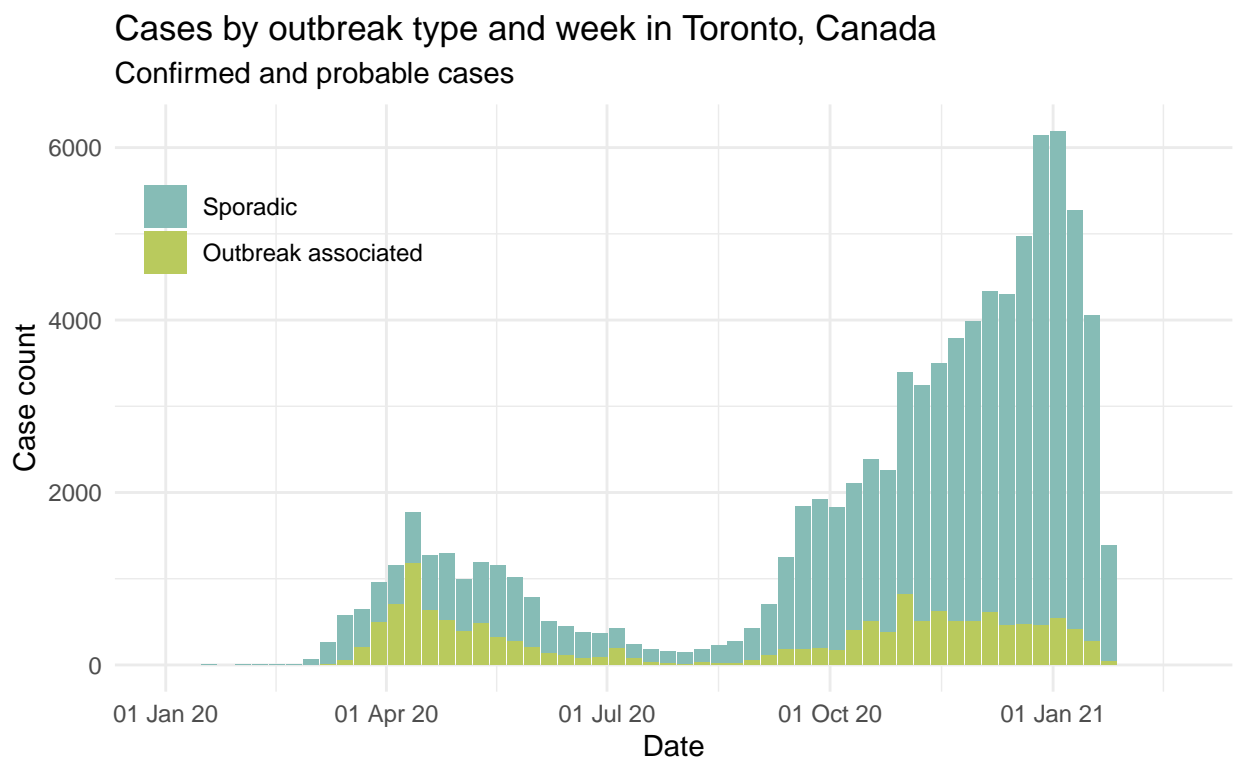
```
outbreak <- outbreak_raw %>%  
  mutate(outbreak_or_sporadic = str_replace(outbreak_or_sporadic, "OB A", "Outbreak a"))  
  
outbreak$episode_week <- as.Date(outbreak$episode_week)  
  
outbreak <- outbreak%>%  
  group_by(episode_week) %>%  
  mutate(total_cases= sum(cases))
```

## Data visualization

```

outbreak %>%
  ggplot(aes(x = episode_week, y = cases, fill = factor(outbreak_or_sporadic,
    levels = c("Sporadic", "Outbreak associated")))) +
  geom_bar(position = "stack", stat = 'identity') +
  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: Gongen Zhong for STA302/1002 U of T\n Source: Ontario Ministry of H
  theme_minimal() +
  theme(legend.title = element_blank(), legend.position = c(.15, .8)) +
  scale_x_date(labels = scales::date_format("%d %b %y"),
    limits = c(date("2020-01-01"), Sys.Date()+7)) +
  scale_fill_manual(values = c("#86BCB6", "#B9CA5D"), breaks = c("Sporadic", "Outbreak associated"))

```



Created by: Gongen Zhong for STA302/1002 U of T  
Source: Ontario Ministry of Health, Integrated Public Health Information System and CORES  
Data as of January 29, 2021

## Task 3: Neighbourhoods

### Data wrangling: part 1

```
income <- nbhood_profile %>%
  filter(`_id` == 1143) %>%
  pivot_longer(cols = -c(`_id`:Characteristic), names_to = "neighbourhood_name",
               values_to = "percentage")%>%
  select(-(`_id`:Characteristic))%>%
  mutate(percentage = parse_number(percentage))
income <- income[-c(1),]
```

### Data wrangling: part 2

```
nbhoods_all <- nbhoods_shape_raw %>%
  mutate(neighbourhood_name = str_replace(AREA_NAME, "\\s\\s(\\d+\\s)", ""))%>%
  left_join(nbhood_raw)%>%
  left_join(income, by = "neighbourhood_name")%>%
  mutate_if(is.numeric, replace_na, replace = 0)%>%
  rename(rate_per_100000 = rate_per_100_000_people)
```

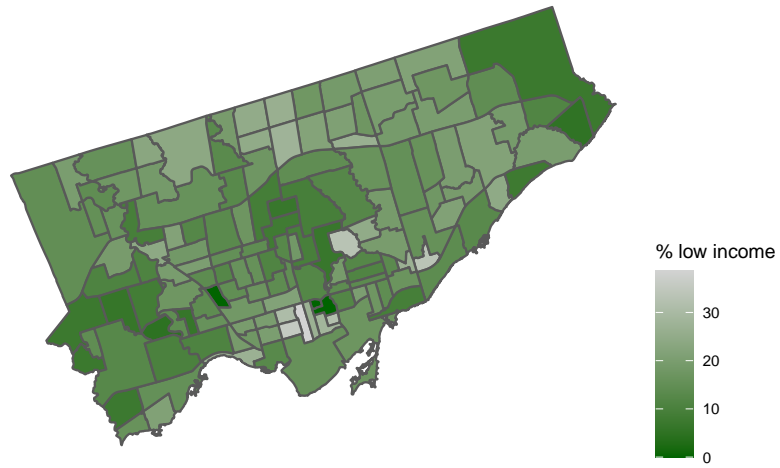
```
## Joining, by = "neighbourhood_name"
```

### Data wrangling: part 3

```
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, Higher case rate",
  ))
```

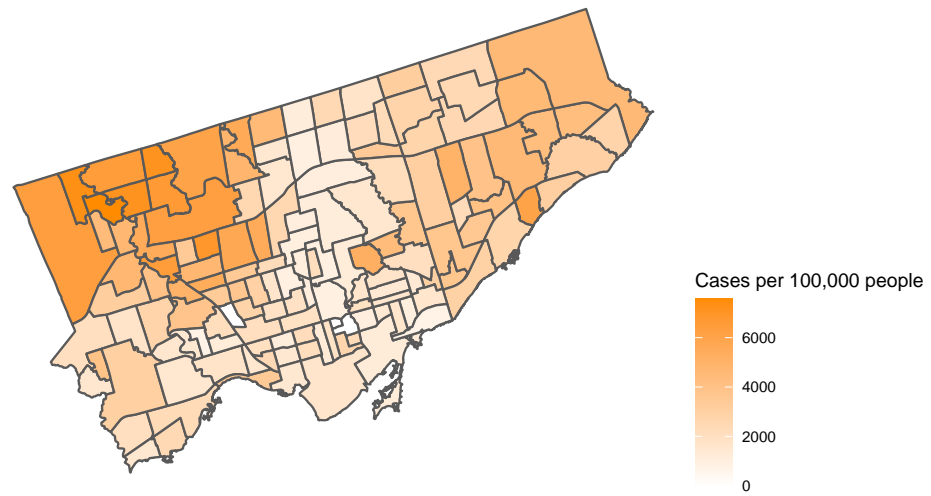
## Data visualization

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



Created by: Gongen Zhong for STA302/1002 U of T  
Source: Census Profile 98-316-X2016001 via OpenData Toronto  
Data as of January 29, 2021

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



Created by: Gongen Zhong for STA302/1002 U of T  
Source: Ontario Ministry of Health, Integrated Public Health Information System and CORES  
Data as of January 29, 2021

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

