Code last run 2021-02-16.
Daily: Data as of January 29, 2021.
Neighbourhood: Data as of February 9, 2021.

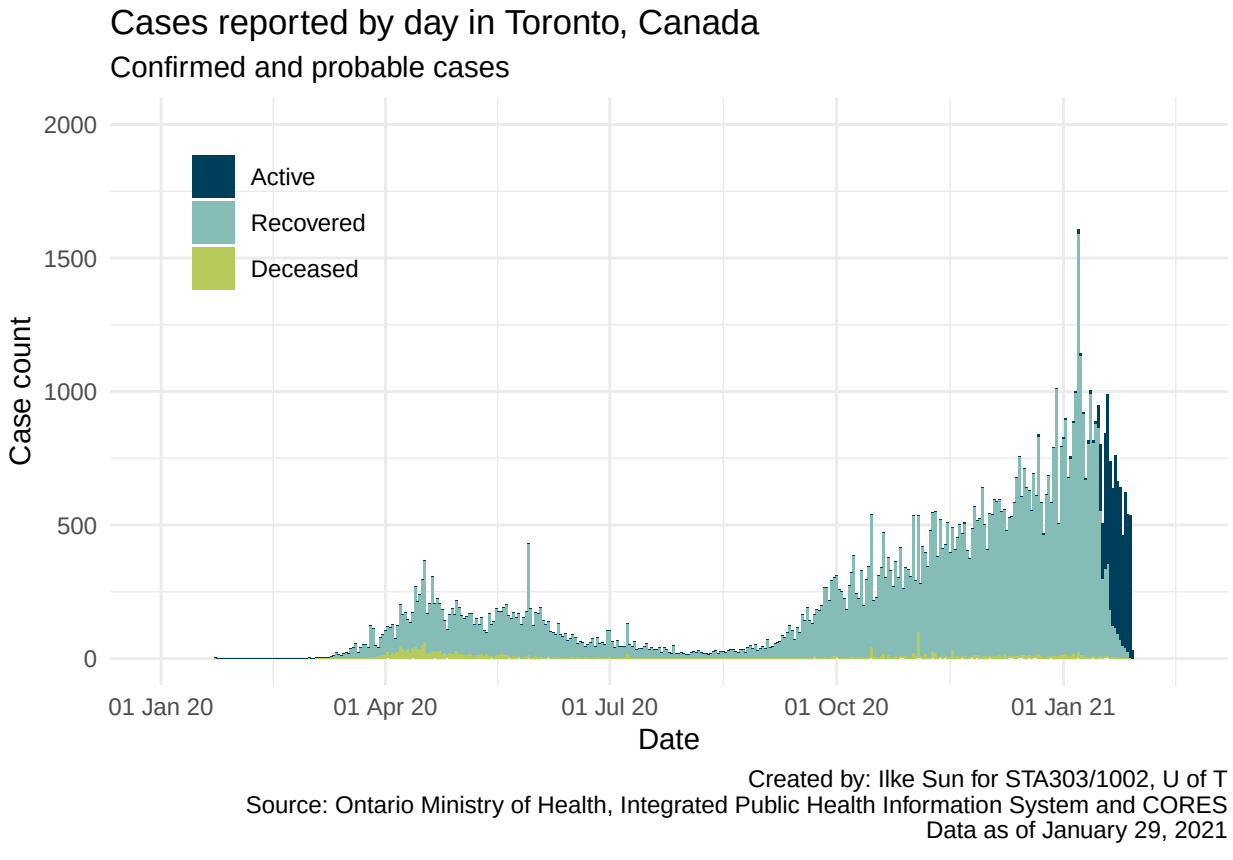
Task 1: Daily cases

Data wrangling

```
reported <- reported_raw %>%
  mutate(reported_date = date(reported_date)) %>%
  rename("Date" = "reported_date",
         "Recovered" = "recovered",
         "Active" = "active",
         "Deceased" = "deceased") %>%
  select(Date, Recovered, Deceased, Active) %>%
  complete(Date = seq.Date(min(Date), max(Date), by = "day")) %>%
  mutate_if(is.numeric, replace_na, replace = 0)
```

Data visualization

```
st <- date("2020-01-01")
nd <- Sys.Date()
reported %>%
  gather("Type", "Cases", -Date) %>%
  mutate(Type = factor(Type, levels = c("Active", "Recovered", "Deceased"))) %>%
  ggplot(aes(x = Date, y = Cases, fill = Type)) +
  geom_bar(position = "stack", 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: Ilke Sun for STA303/1002, U of T\n",
                       "Source: Ontario Ministry of Health, Integrated ",
                       "Public Health Information System and CORES\n",
                       date_daily)) +
  theme(legend.title = element_blank(), legend.position = c(.15, .8)) +
  scale_x_date(limits = c(st, nd), date_labels = "%d %b %y") +
  scale_y_continuous(limits = c(0, 2000)) +
  scale_fill_manual("legend", values = c("Active" = "#003F5C",
                                         "Recovered" = "#86BCB6",
                                         "Deceased" = "#B9CA5D"))
```




```
rm(reported_raw, st, nd)
```

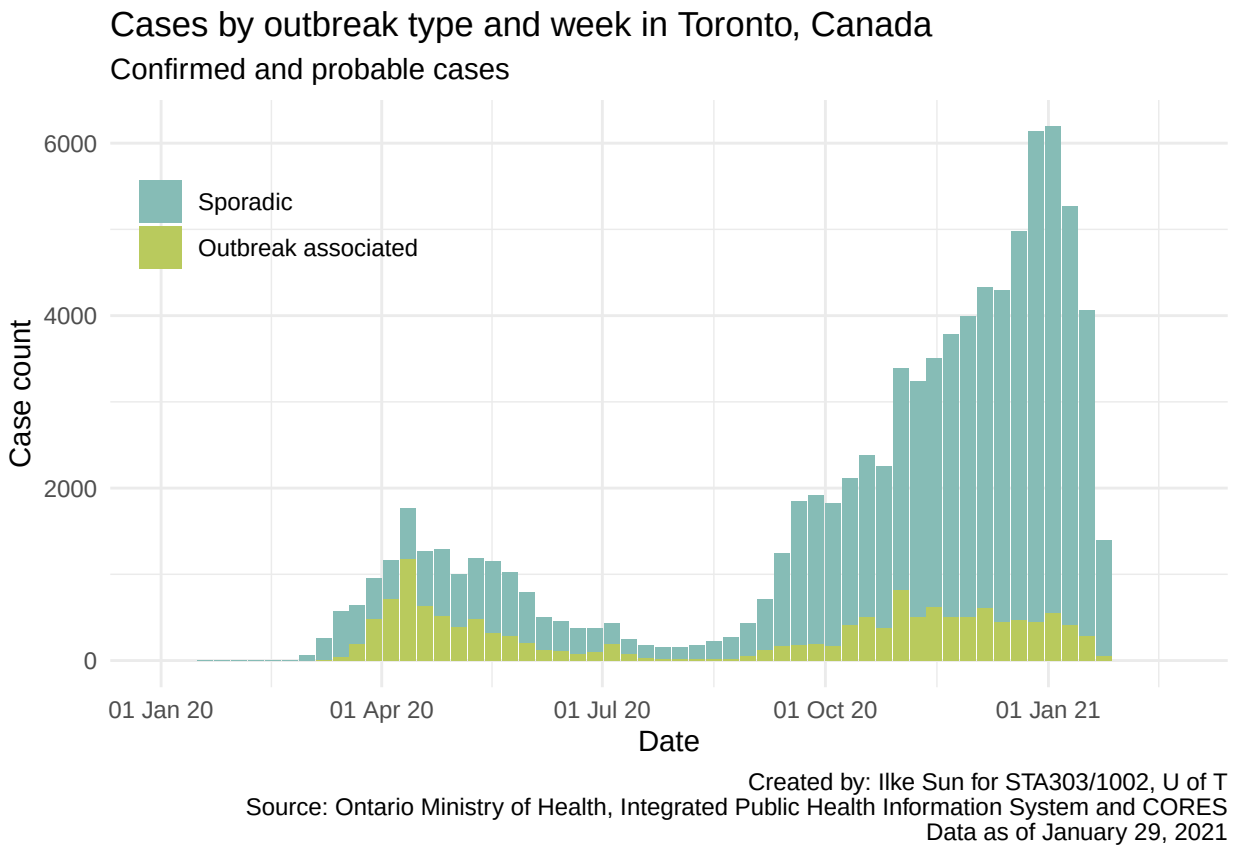
Task 2: Outbreak type

Data wrangling

```
st <- date(min(outbreak_raw$episode_week))
nd <- date(max(outbreak_raw$episode_week))

outbreak <- outbreak_raw %>%
  group_by(outbreak_or_sporadic) %>%
  mutate(episode_week = date(episode_week)) %>%
  complete(episode_week = seq.Date(st, nd, by = "week")) %>%
  rename("OS" = "outbreak_or_sporadic") %>%
  select(episode_week, OS, cases) %>%
  arrange(episode_week) %>%
  mutate(cases = replace_na(cases, replace = 0),
         OS = as.factor(OS))

levels(outbreak$OS) <- c("Outbreak associated", "Sporadic")
outbreak$OS <- factor(outbreak$OS,
                      levels = c("Sporadic", "Outbreak associated"))

outbreak_OB <- outbreak %>%
  filter(OS == "Outbreak associated") %>%
  mutate(total_cases = cases)

outbreak_OB <- outbreak_OB %>%
  rbind(outbreak_OB) %>%
  arrange(episode_week)

outbreak_SP <- outbreak %>%
  filter(OS == "Sporadic") %>%
  mutate(total_cases = cases)

outbreak_SP <- outbreak_SP %>%
  rbind(outbreak_SP) %>%
  arrange(episode_week)

outbreak$total_cases <- outbreak_OB$total_cases + outbreak_SP$total_cases

rm(outbreak_OB, outbreak_SP, st, nd)
```

Data visualization

```
st <- date("2020-01-01")
nd <- Sys.Date() + 7

outbreak %>%
  ggplot(aes(x = episode_week, y = cases, fill = OS)) +
  geom_bar(position = "stack", 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: Ilke Sun for STA303/1002, U of T\n",
                       "Source: Ontario Ministry of Health, Integrated Public ",
                       "Health Information System and CORES\n",
                       date_daily)) +
  theme(legend.title = element_blank(), legend.position = c(.15, .8)) +
  scale_x_date(limits = date(c(st, nd)),
               date_labels = "%d %b %y") +
  scale_y_continuous(limits = c(0, max(outbreak$total_cases))) +
  scale_fill_manual("legend", values = c("Sporadic" = "#86BCB6",
                                         "Outbreak associated" = "#B9CA5D"))
```




```
rm(outbreak_raw, st, nd)
```

Task 3: Neighbourhoods

Data wrangling: part 1

```
income <- nbhood_profile %>%
  rename("id" = "_id") %>%
  filter(Topic == "Low income in 2015") %>%
  filter(id == 1075) %>%
  select(-id, -Category, -Topic, -`Data Source`,
         -Characteristic, -`City of Toronto`) %>%
  gather("neighbourhood_name", "percentage") %>%
  mutate(percentage = as.numeric(percentage)) %>%
  arrange(neighbourhood_name)

rm(nbhood_profile)
```

Data wrangling: part 2

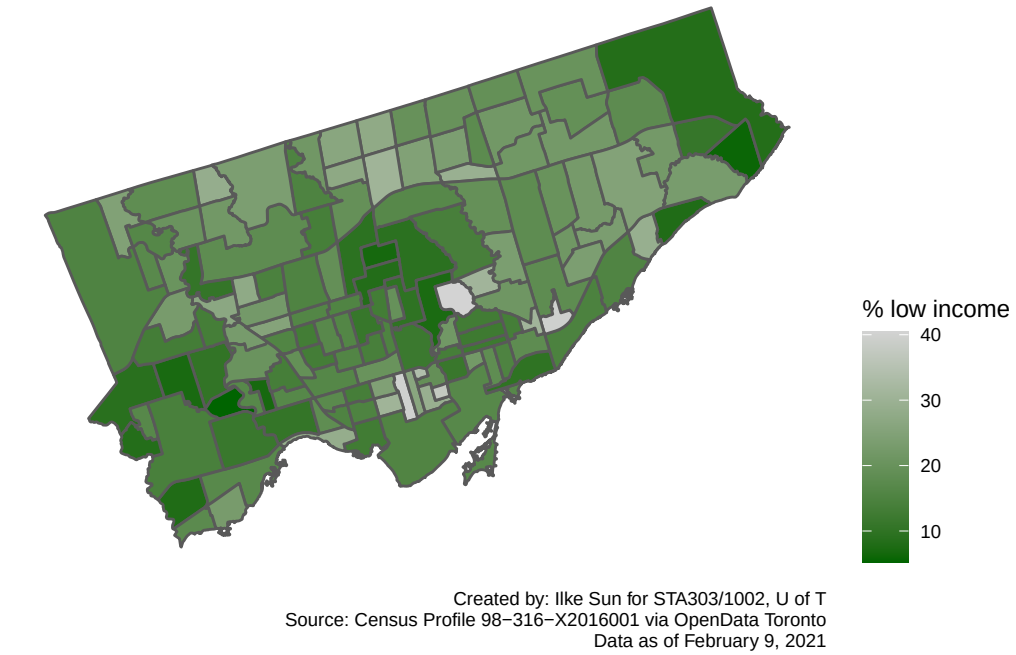
```
nbhood_raw <- nbhood_raw %>%
  filter(neighbourhood_name != "Missing Address/Postal Code") %>%
  arrange(neighbourhood_name)

nbhoods_all <- nbhoods_shape_raw %>%
  mutate(neighbourhood_name = str_remove(AREA_NAME, pattern = "\\(\\d+\\)")) %>%
  arrange(neighbourhood_name) %>%
  mutate(percentage = income$percentage,
         rate_per_100000 = nbhood_raw$rate_per_100_000_people)
```

Data wrangling: part 3

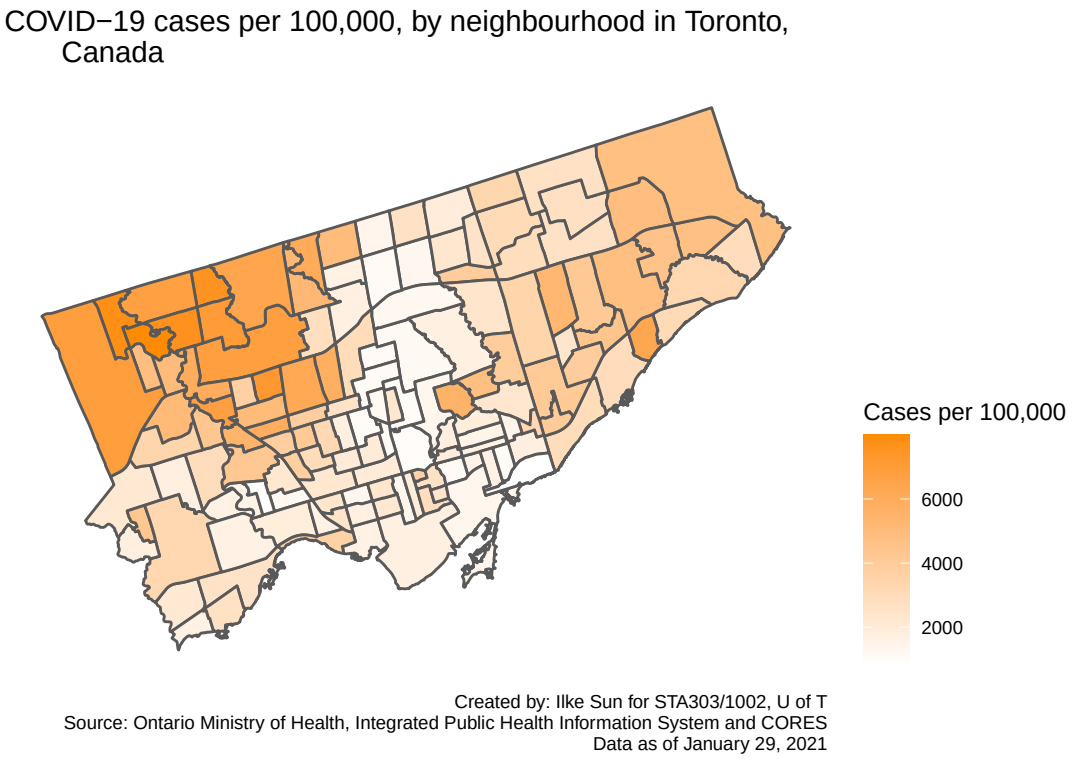
```
nbhoods_final <- nbhoods_all %>%
  mutate(med_inc = median(percentage),
         med_rate = median(rate_per_100000)) %>%
  mutate(nbhood_type = case_when(
    percentage >= med_inc & rate_per_100000 >= med_rate ~
      "Higher low income rate, higher case rate",
    percentage >= med_inc & rate_per_100000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

rm(nbhood_raw, nbhoods_all)
```

Data visualization

```
nbhoods_final %>%
  ggplot(aes(fill = percentage)) +
  geom_sf() +
  theme_map() +
  theme(legend.position = "right") +
  labs(title = str_c("Percentage of 18 to 64 year olds living in a low income ",
                     "family (2015)"),
       subtitle = "Neighbourhoods of Toronto, Canada",
       caption = str_c("Created by: Ilke Sun for STA303/1002, U of T\n",
                       "Source: Census Profile 98-316-X2016001 ",
                       "via OpenData Toronto\n",
                       date_nbhood)) +
  scale_fill_gradient(name = "% low income",
                      low = "darkgreen",
                      high = "lightgrey")
```



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



```
nbhoods_final %>%
  ggplot(aes(fill = nbhood_type)) +
  geom_sf() +
  theme_map() +
  theme(legend.position = "right") +
  labs(title ="COVID-19 cases per 100,000, by neighbourhood in Toronto, Canada",
       caption = str_c("Created by: Ilke Sun 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_daily)) +
  scale_fill_brewer(name = str_c("% of 18 to 64 year-olds in\n",
                      "low income families and\n",
                      "COVID-19 case rates"), palette = "Set1")
```