Code last run 2021-02-16.
Daily: Data as of January 29, 2021.
Neighbourhood: Data as of January 31, 2021.

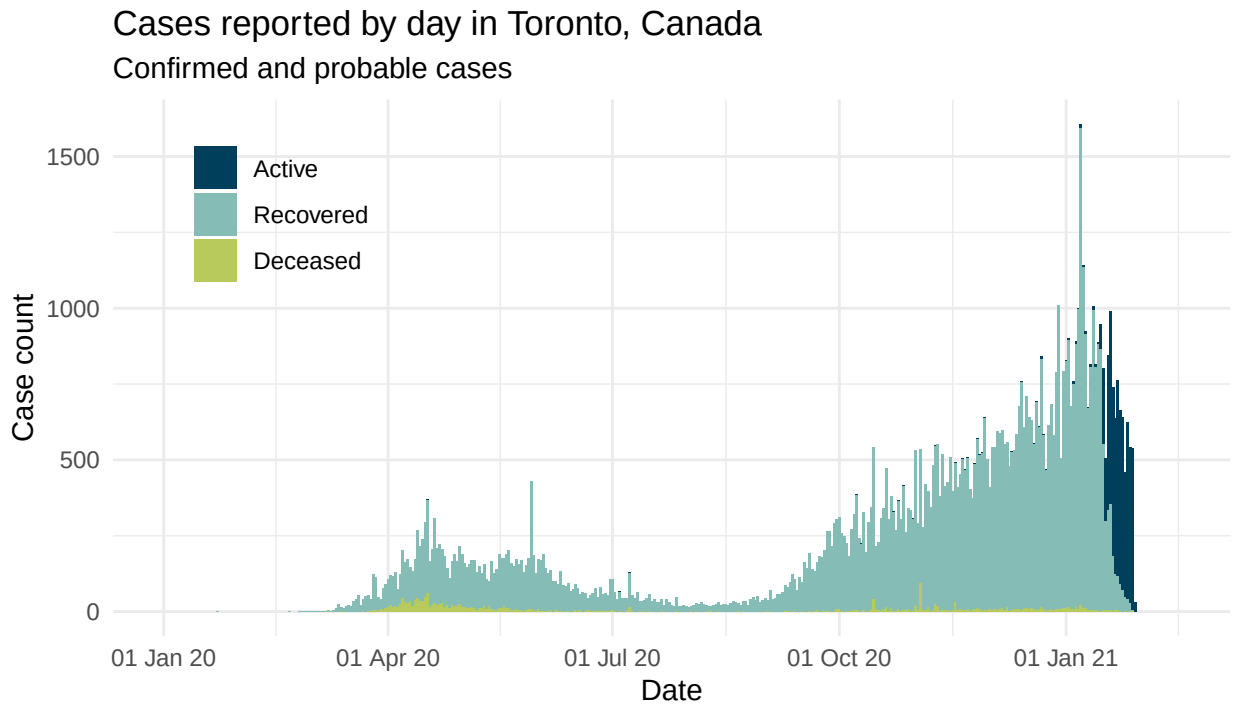
Task 1: Daily cases

Data wrangling

```
# converting NA values to 0
reported <- reported_raw %>% mutate_if(is.numeric, replace_na, replace=0) %>%
  mutate(reported_date = date(reported_date))
# tidying columns - every observation with its own row
reported <- reported %>% pivot_longer(c(recovered, active, deceased),
          names_to = "Case Outcomes") %>% uncount(value)
# capitalizing case outcomes to match original figure
reported$`Case Outcomes` <- str_to_title(reported$`Case Outcomes`)
# releveling categories
reported <- reported %>%
  count(reported_date,`Case Outcomes`) %>%
  mutate(`Case Outcomes` =
           fct_relevel(`Case Outcomes`, "Deceased", after = 2))
```

Data visualization

```
active <- "#003F5C"
recovered <- "#86BCB6"
deceased <- "#B9CA5D"
reported %>%
  ggplot(aes(x = reported_date, fill = `Case Outcomes`, y = n)) +
  geom_bar(stat = "identity", width = 1) +
  scale_x_date(limits = c(date("2020-01-01"), Sys.Date()),
  name = "Date", date_labels = "%d %b %y") + theme_minimal() +
  labs(title = "Cases reported by day in Toronto, Canada",
  subtitle = "Confirmed and probable cases", y = "Case count", caption
  = str_c("Created by: Jaffa Romain for STA303/1002, U of T \n Source:
Ontario Ministry of Health, Integrated Public Health Information System
and CORES \n Data as of ", format(Sys.Date(), "%B %d, %Y"))) +
  theme(legend.title = element_blank(), legend.position = c(.15, .8)) +
  scale_colour_manual(values = c(active, recovered, deceased),
                      aesthetics = "fill")
```



Task 2: Outbreak type

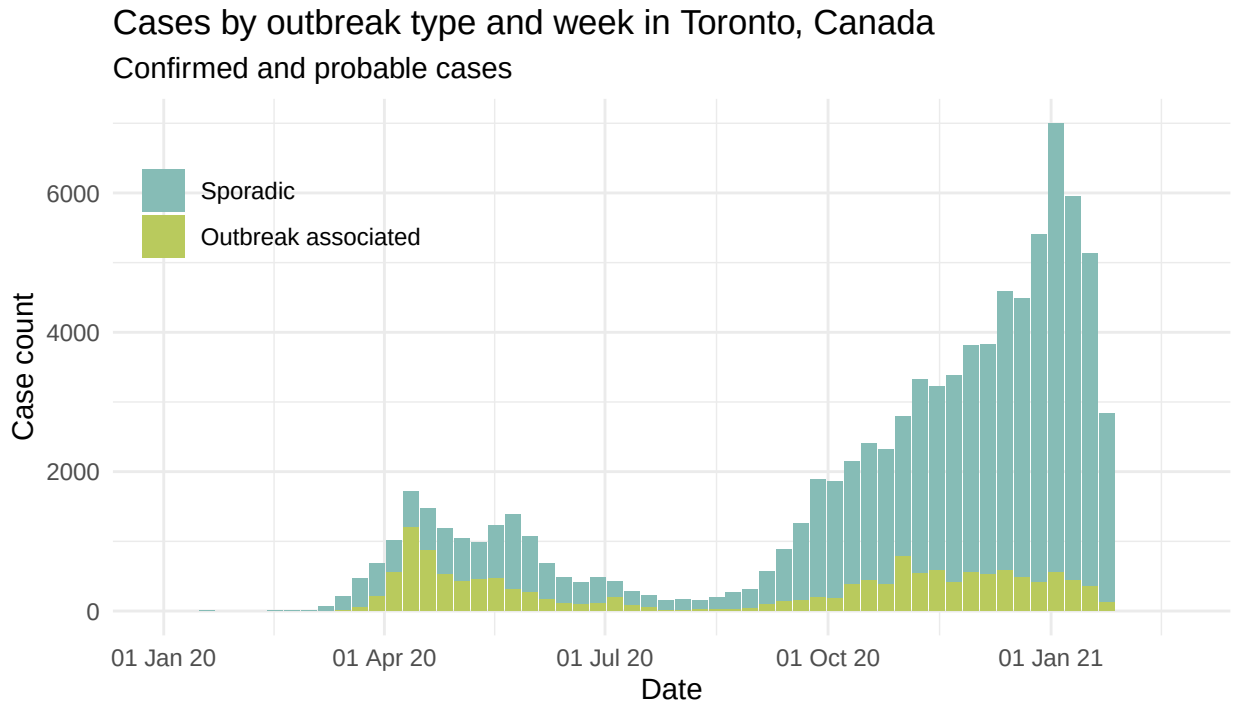
Data wrangling

```
outbreak <- outbreak_raw %>%
  mutate(reported_week = date(reported_week)) # changing date variable to have `date` type
outbreak <- outbreak %>%
  mutate(outbreak_or_sporadic =
           str_replace(outbreak_or_sporadic, "OB A","Outbreak a"))
outbreak$`Outbreak Type` <- outbreak$outbreak_or_sporadic
outbreak <- outbreak %>% select(!outbreak_or_sporadic) %>%
  mutate(`Outbreak Type` = fct_relevel(`Outbreak Type`,
          "Outbreak associated", after = 1))
# creating `total_cases` variable
outbreak <- outbreak %>%  group_by(reported_week) %>%
  summarize(total_cases = sum(cases), cases = cases,
            `Outbreak Type` = `Outbreak Type`)
```

```
## `summarise()` has grouped output by 'reported_week'. You can override using the `.groups` argument.
```

Data visualization

```
outbreak %>% ggplot(aes(x = reported_week,
        y = cases, fill = `Outbreak Type`)) +
  geom_bar(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: Jaffa Romain
for STA303/1002, U of T \n Source: Ontario Ministry of Health,
Integrated Public Health Information System and CORES \n Data as of ",
format(Sys.Date(), "%B %d, %Y"))) + scale_x_date(labels =
scales::date_format("%d %b %y"), limits = c(date("2020-01-01"),
(Sys.Date() + 7))) + lims(y = c(0, max(outbreak$total_cases))) +
theme_minimal() + scale_colour_manual(values = c("#86BCB6", "#B9CA5D"),
aesthetics = "fill") + theme(legend.title = element_blank(),
legend.position = c(.15, .8))
```



Task 3: Neighbourhoods

Data wrangling: part 1

```
income <- nbhood_profile %>% filter(Category == "Income",
           Topic == "Low income in 2015", Characteristic ==
             "  18 to 64 years (%)", `_id` == 1143 )
# tidy data
income <- income %>%
  pivot_longer(-c(Category, `_id`, Topic, Characteristic, `Data Source`),
               names_to = "neighbourhood_name", values_to = "low_income")

income$low_income = as.numeric(as.character(income$low_income))
```

Data wrangling: part 2

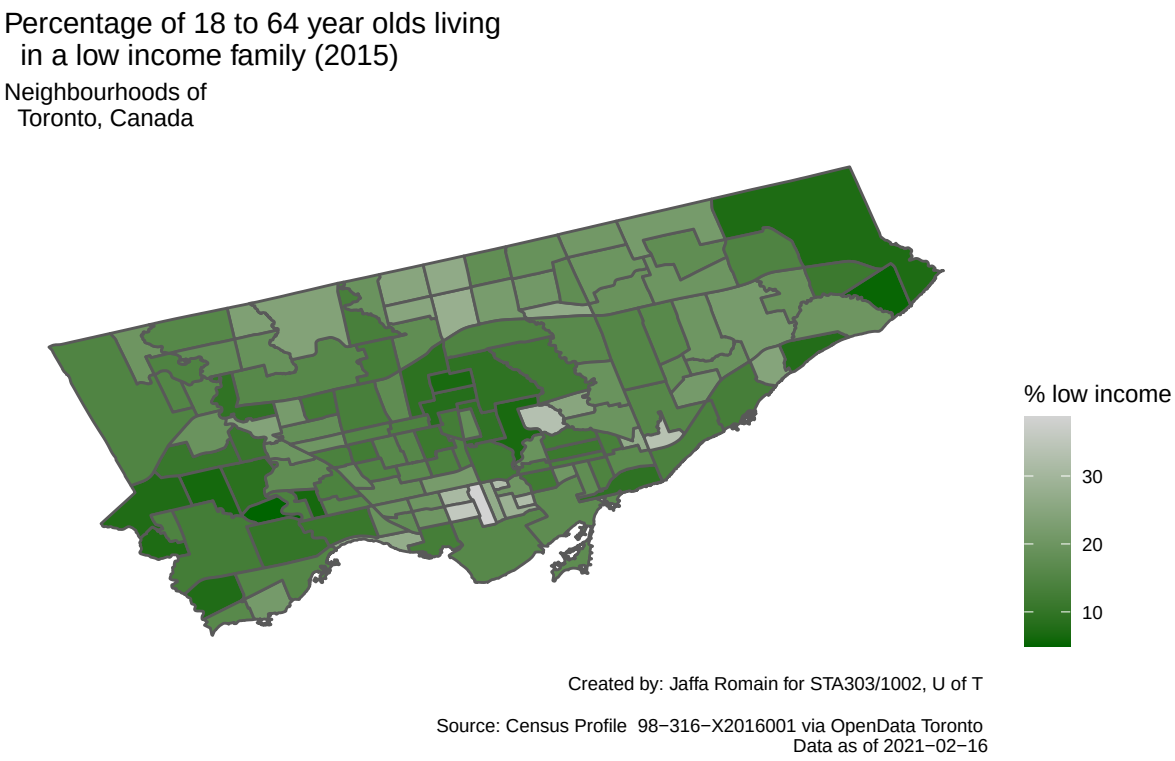
```
nbhoods_all <- nbhoods_shape_raw %>%
  mutate(neighbourhood_name = str_remove(AREA_NAME, "\\s\\(\\d+\\)$"))
# tidy to match neighbourhood names in other data set
nbhoods_all <- nbhoods_all %>%
  mutate(neighbourhood_name =
           str_replace(neighbourhood_name, "St.Ja", "St. Ja")) %>%
  mutate(neighbourhood_name = str_replace(neighbourhood_name, "Pellam", "Pelham"))


nbhoods_all <- full_join(nbhood_raw, nbhoods_all, by="neighbourhood_name")
nbhoods_all <- full_join(nbhoods_all, income, by="neighbourhood_name")
nbhoods_all <- nbhoods_all %>% filter(neighbourhood_name != "City of Toronto")

# remove NA values to get 140 distinct neighbourhoods
nbhoods_all <- nbhoods_all %>% drop_na(low_income)
```

Data wrangling: part 3

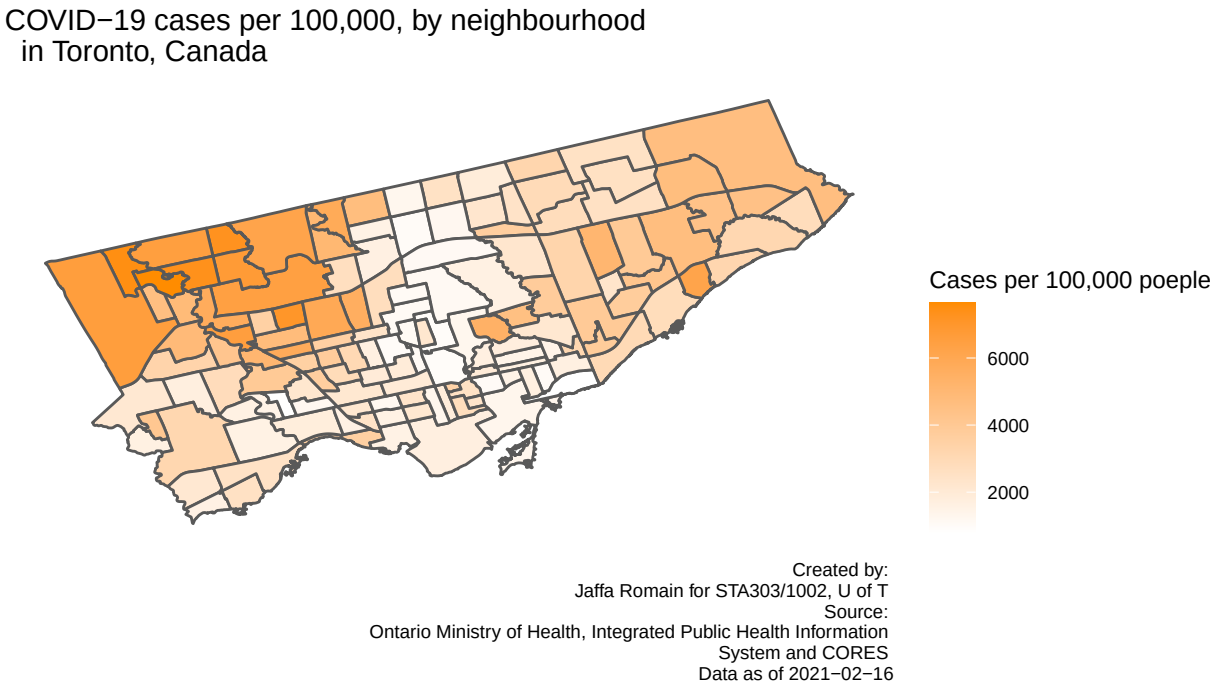
```
nbhoods_final <- nbhoods_all %>%
  mutate(med_inc = median(low_income),
         med_rate= median(rate_per_100_000_people))
nbhoods_final <- nbhoods_final %>% mutate(nbhood_type =
case_when(low_income >= med_inc & rate_per_100_000_people
          >= med_rate~"Higher low income rate, higher case rate",
          low_income >= med_inc & rate_per_100_000_people
          <= med_rate~"Higher low income rate, lower case rate",
          low_income <= med_inc & rate_per_100_000_people
          >= med_rate~"Lower low income rate, higher case rate",
        low_income <= med_inc & rate_per_100_000_people
        <= med_rate~"Lower low income rate, lower case rate" ))
```

Data visualization

```
ggplot(data = nbhoods_final) +
  geom_sf(aes(geometry = geometry, fill = low_income)) +
  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: Jaffa Romain for STA303/1002, U of T
        \n Source: Census Profile  98-316-X2016001 via OpenData Toronto \n Data as of ",
  Sys.Date())) + theme(legend.position = "right", legend.justification = c("right", "bottom"))
```



```
ggplot(data = nbhoods_final) +
  geom_sf(aes(geometry = geometry, fill =
  rate_per_100_000_people)) + theme_map() +
  scale_fill_gradient(name="Cases per 100,000 poeple",
  low = "white", high = "darkorange") +
  labs(title = "COVID-19 cases per 100,000, by neighbourhood
  in Toronto, Canada", caption = str_c( "Created by:
  Jaffa Romain for STA303/1002, U of T \n Source:
  Ontario Ministry of Health, Integrated Public Health Information
  System and CORES \n Data as of ", Sys.Date())) +
  theme(legend.position = "right", legend.justification = c("right", "bottom"))
```



```
ggplot(data = nbhoods_final) +
  geom_sf(aes(geometry = geometry, fill = nbhood_type)) +
  theme_map() + scale_fill_brewer(palette = "Set1",
  name="% of 18 to 64 year-olds in \n low income families and \n COVID-19 case rates", ) + labs(title =
caption = str_c("Created by: Jaffa Romain for STA303/1002, U of T \n Income data source: Census Profile
Sys.Date())) +
  theme(legend.position = "right",
        legend.justification = c("right", "bottom") )
```