Code last run 2021-02-16.
Daily: Data as of February 10, 2021.
Neighbourhood: Data as of February 9, 2021.

Task 1: Daily cases

Data wrangling

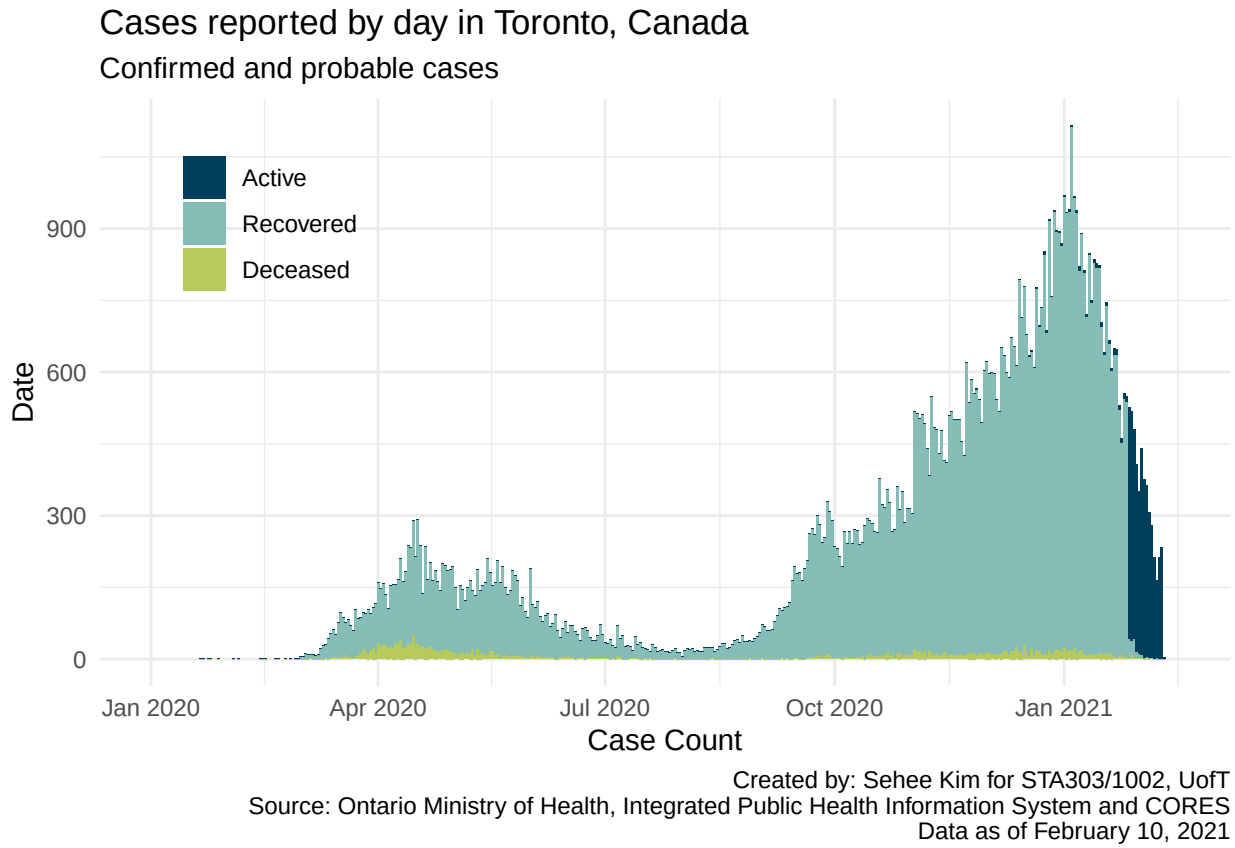
```
reported_raw %>% filter(is.na(episode_date))
```

```
## # A tibble: 0 x 4
## # ... with 4 variables: episode_date <dttm>, recovered <dbl>, active <dbl>,
## #   deceased <dbl>
```

```
# clean_reported <- reported_raw
# colnames(clean_reported) <- str_to_title(colnames(reported_raw))
# clean_reported$Episode_date <- date(clean_reported$Episode_date)

reported <- reported_raw %>%
  mutate(episode_date = date(reported_raw$episode_date)) %>%
  mutate_if(is.numeric,replace_na, replace=0) %>%
  gather("Type", "Value", -episode_date) %>%
  mutate(Type = case_when(
    Type== "active"~ "Active",
    Type== "recovered" ~"Recovered",
    Type == "deceased"~"Deceased"
  ) ) %>%
  mutate(Type = fct_relevel(Type, "Deceased", after = 2))
```

Data visualization

```
reported %>%
  ggplot(aes(x=episode_date, y =Value, fill= Type)) +
  geom_bar(stat = "identity", width = 1)+
  theme_minimal()+
  theme(legend.position = c(0.15,0.8),
        legend.title=element_blank())+
  scale_x_date(limits = c(date("2020-01-01"),Sys.Date()))+
  scale_fill_manual(values = c("#003f5c","#86bcB6","#b9ca5d"))+
  labs(title="Cases reported by day in Toronto, Canada",
  subtitle="Confirmed and probable cases",
       x= "Case Count", y="Date",
       caption = str_c("Created by: Sehee Kim for STA303/1002, UofT\n",
"Source: Ontario Ministry of Health, Integrated Public Health Information System and CORES\n"
,date_daily[1,1]))
```



Task 2: Outbreak type

Data wrangling

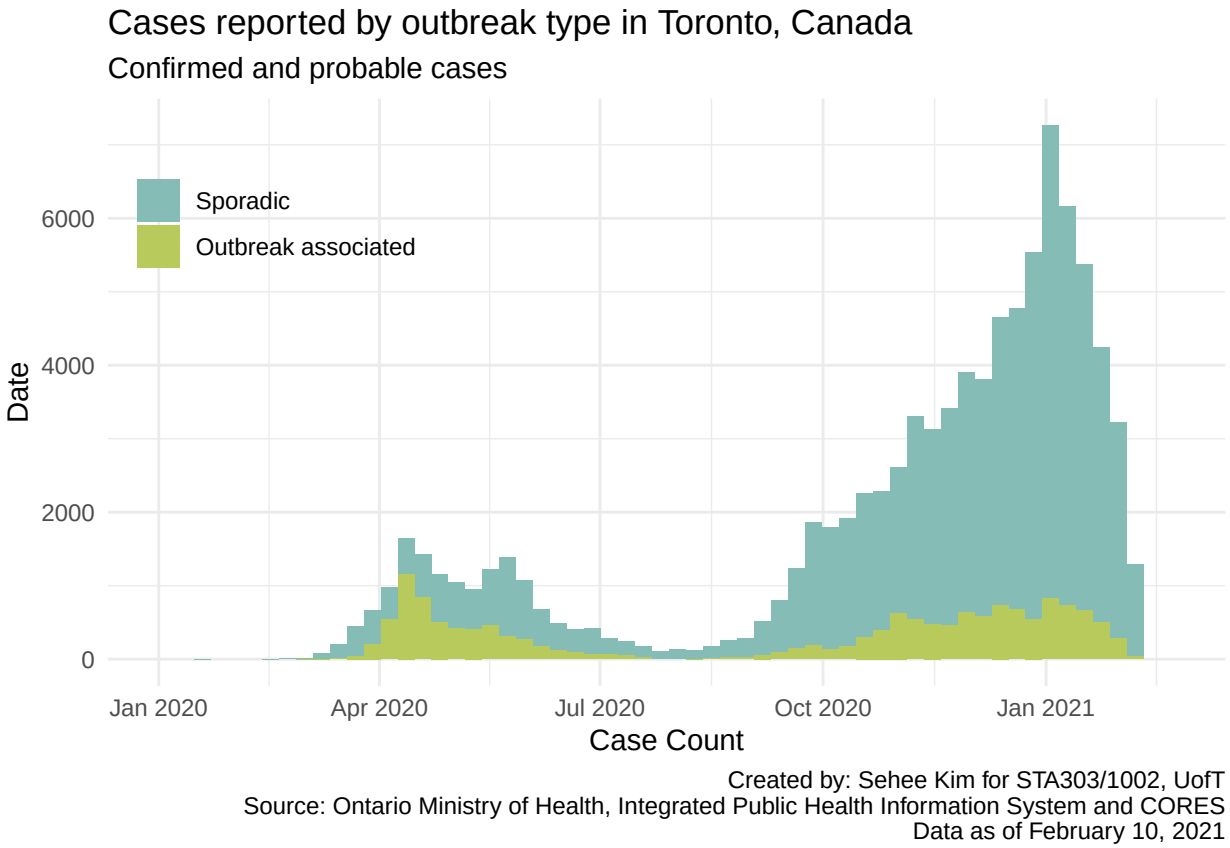
```
outbreak_raw %>%
  filter(is.na(outbreak_or_sporadic)) %>%
  filter(is.na(reported_week)) %>%
  filter(is.na(cases))
```

```
## # A tibble: 0 x 3
## # ... with 3 variables: outbreak_or_sporadic <chr>, reported_week <dttm>,
## #   cases <dbl>
```

```
outbreak <- outbreak_raw %>%
  mutate(reported_week = date(outbreak_raw$reported_week)) %>%
  mutate(outbreak_or_sporadic = str_replace_all(outbreak_or_sporadic, "OB Associated",
                                                "Outbreak associated")) %>%
  rename(Outbreak_type = outbreak_or_sporadic) %>%
  group_by(reported_week) %>%
  mutate(total_cases = sum(cases)) %>%
  mutate(Outbreak_type = fct_rev(Outbreak_type))

colnames(outbreak) <- str_to_title(colnames(outbreak))
```

Data visualization

```
#present_day_seven = parse_date_time(date_daily[1,1], orders= c("ymd", "dmy", "mdy"))

outbreak %>%
  ggplot(aes(x=Reported_week, y= Cases, fill = Outbreak_type))+
  geom_bar(stat="identity", width=7)+
  theme_minimal()+
  theme(legend.position = c(0.15,0.8),
        legend.title=element_blank())+
  scale_x_date(limits = c(date("2020-01-01"),Sys.Date()+7))+
  scale_fill_manual(values = c("#86bcB6","#b9ca5d"))+
  labs(title="Cases reported by outbreak type in Toronto, Canada",
       subtitle="Confirmed and probable cases",
       x= "Case Count", y="Date",
       caption = str_c("Created by: Sehee Kim for STA303/1002, UofT\n",
"Source: Ontario Ministry of Health, Integrated Public Health Information System and CORES\n"
,date_daily[1,1]))
```



Task 3: Neighbourhoods

Data wrangling: part 1

```
income <- filter(nbhood_profile, nbhood_profile$Topic == "Low income in 2015", nbhood_profile$`_id` ==1
  pivot_longer(-c(Characteristic,`_id`,Category,Topic,`Data Source`),
               names_to = "neighbourhood_name", values_to="Percentage") %>%
  select(-`_id`, -Category, -Topic, -`Data Source`, -Characteristic) %>%
  mutate(Percentage=parse_number(Percentage, locale=locale(decimal_mark = ".", grouping_mark = " ")))
```

Data wrangling: part 2

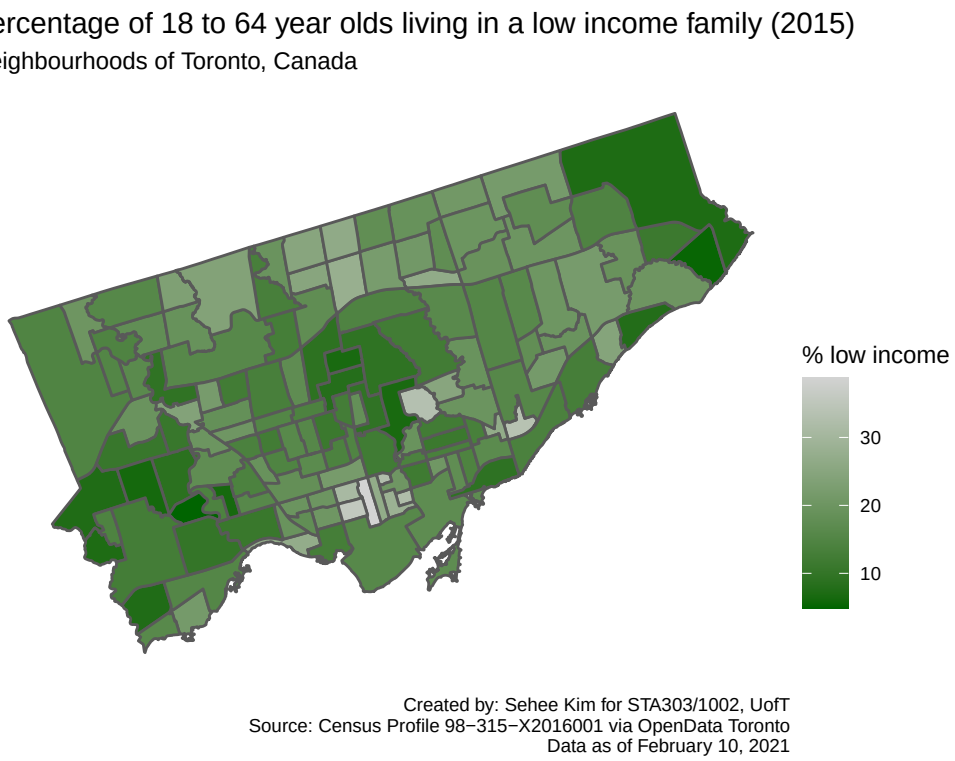
```
#glimpse(nbhoods_shape_raw)

#first see what happens if we join the data right away
nbhoods_all_try <- nbhoods_shape_raw %>%
  mutate(neighbourhood_name = str_remove(AREA_NAME, "\\s\\(\\d+\\)*$")) %>%
  left_join(income, by = "neighbourhood_name")

problems <- nbhoods_all_try %>% #typos
  filter(is.na(Percentage))

Rate_per_cases <- nbhood_raw %>%
  select(-neighbourhood_id,-case_count) %>%
  rename(rate_per_100000 = rate_per_100_000_people)

nbhoods_all <-nbhoods_shape_raw %>%
  mutate(neighbourhood_name = str_remove(AREA_NAME, "\\s\\(\\d+\\)*$")) %>%
  mutate(neighbourhood_name = case_when(
    str_detect(neighbourhood_name, "St.James")
    ~ str_replace_all(neighbourhood_name, "St.James", "St. James"),
    neighbourhood_name == "Weston-Pellam Park"~ "Weston-Pelham Park",
    TRUE ~ neighbourhood_name
  )) %>%
  left_join(income, by = "neighbourhood_name") %>%
  left_join(Rate_per_cases, by = "neighbourhood_name") %>%
  select(-X, -Y, -LONGITUDE,-LATITUDE,-PARENT_AREA_ID)

#check <- nbhoods_all %>%
  #filter(is.na(Percentage))
#colnames(nbhoods_all) <- str_to_title(colnames(nbhoods_all))
```

Data wrangling: part 3

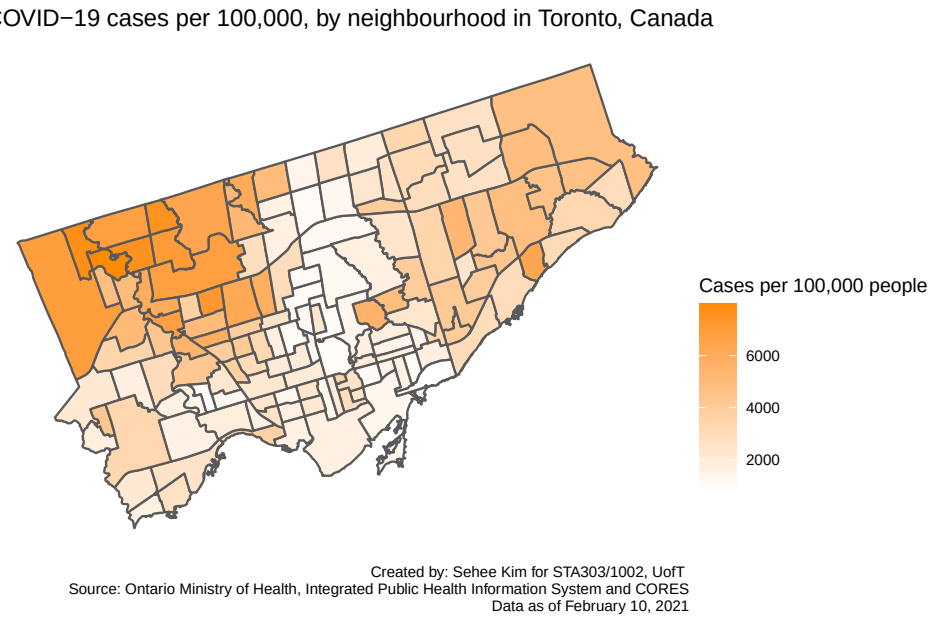
```
nbhoods_final <- nbhoods_all %>%
  mutate(med_inc = median(Percentage)) %>%
  mutate(med_rate = median(rate_per_100000)) %>%
  mutate(nbhood_type = case_when(
    Percentage <= med_inc & rate_per_100000 <= med_rate ~ "Higher low income rate, higher case rate",
    Percentage <= med_inc & rate_per_100000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

```
ggplot(data = nbhoods_final) +
  geom_sf(position = "identity", aes(fill= Percentage))+
  theme_map()+
  theme(legend.position = c(1.0,0.1))+
  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: Sehee Kim for STA303/1002, UofT\n",
       "Source: Census Profile 98-315-X2016001 via OpenData Toronto\n"
       ,date_daily[1,1]))
```



```
ggplot(data = nbhoods_final) +
  geom_sf(aes(fill = rate_per_100000)) +
  theme_map() +
  theme(legend.position = c(1.0,0.1))+
  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: Sehee Kim for STA303/1002, UofT \n",
 "Source: Ontario Ministry of Health, Integrated Public Health Information System and CORES\n",
 date_daily[1,1]))
```



```
ggplot(data = nbhoods_final) +
  geom_sf(aes(fill = nbhood_type)) +
  theme_map() +
  theme(legend.position = c(1.0,0.0))+
  scale_fill_brewer(palette = "Set1", name = "% of 18 to 64 year-olds in \n low income families and \n C
  labs(title ="COVID-19 cases per 100,000, by neighbourhood in Toronto, Canada",
       caption = str_c("Created by: Sehee Kim for STA303/1002, UofT \n",
      "Income data source: Census Profile 98-316-x2016001 via OpenData Toronto \n",
      "Source: Ontario Ministry of Health, Integrated Public Health Information System and CORES\n"
      ,date_daily[1,1]))
```