Code last run 2021-02-12.
Daily:

numbers_are_accurate_as_of_download_date_and_are_subject_to_change
<chr>
Data as of January 29, 2021

1 row

.
Neighbourhood:

numbers_are_accurate_as_of_download_date_and_are_subject_to_change
<chr>
Data as of January 31, 2021

1 row

.

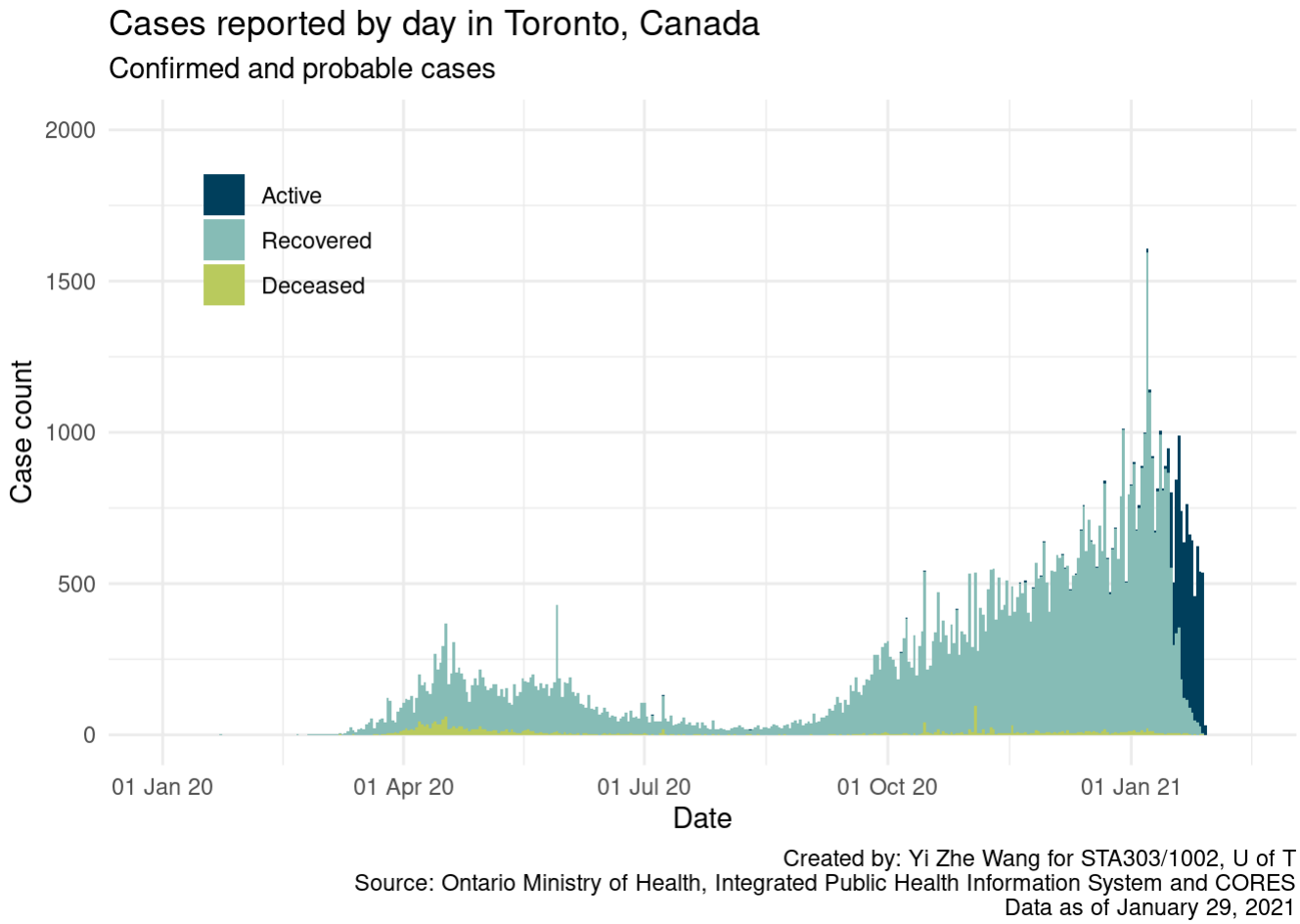
Task 1: Daily cases

Data wrangling

```
reported <- reported_raw %>%
  mutate_if(is.numeric, replace_na, replace=0) %>%
  mutate(reported_date = date(reported_date)) %>%
  janitor::clean_names() %>%
  pivot_longer(-reported_date, names_to = "status", values_to = "count") %>%
  mutate(status = case_when(
  status == "active" ~ "Active",
  status == "deceased" ~ "Deceased",
  status == "recovered" ~ "Recovered")) %>%
  mutate(status = fct_relevel(status, "Deceased", after = 2))
# glimpse(reported)
```

Data visualization

```
reported %>%
  ggplot(aes(x = reported_date, y = count, fill = status)) +
  geom_bar(stat = "identity", width = 1) +
  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)) +
  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: Yi Zhe Wang for STA303/1002, U of T
  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"))
```



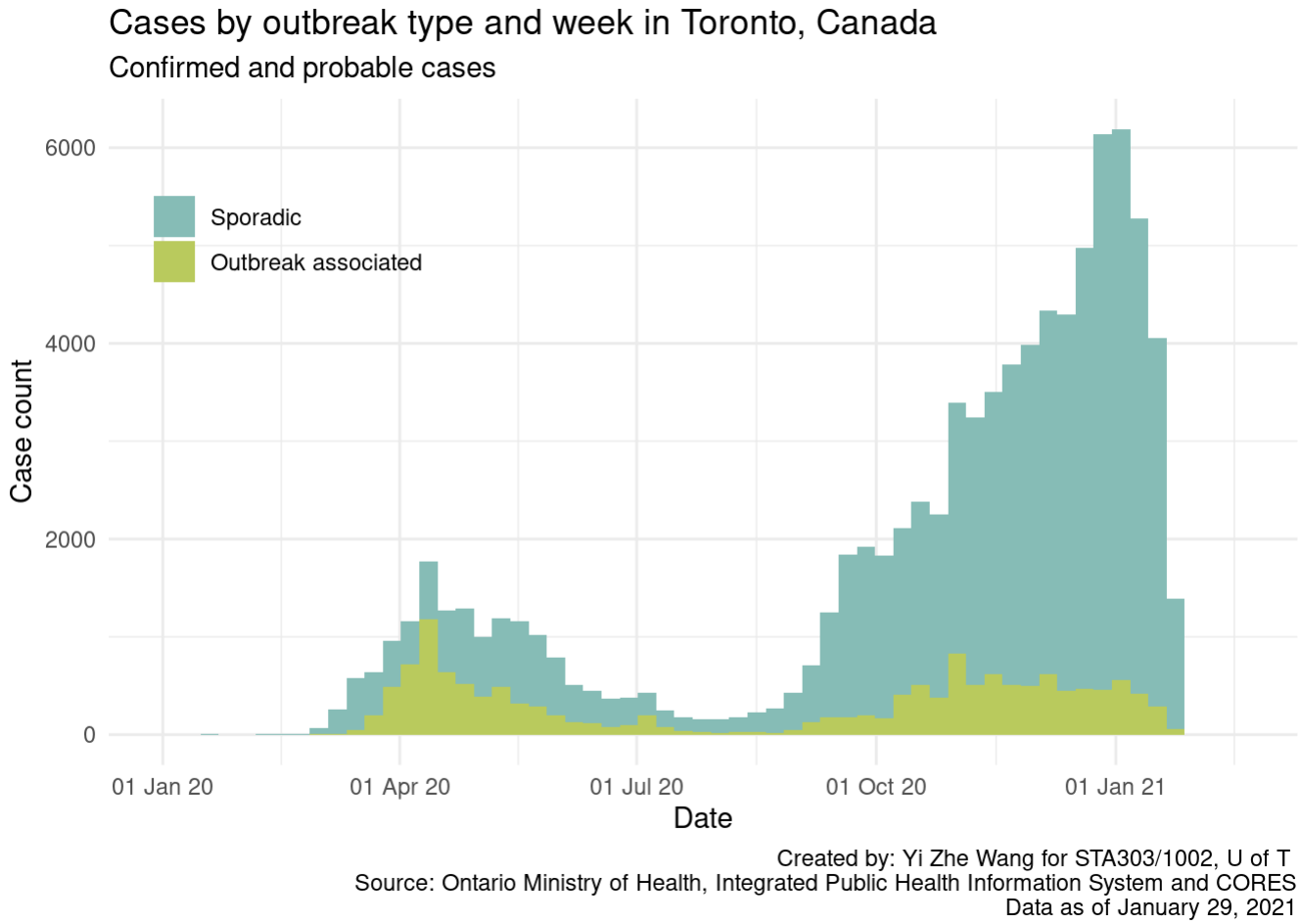
Task 2: Outbreak type

Data wrangling

```
outbreak <- outbreak_raw %>%
  mutate_if(is.numeric, replace_na, replace = 0) %>%
  mutate(episode_week = date(episode_week)) %>%
  mutate(outbreak_or_sporadic = case_when(
  outbreak_or_sporadic == "OB Associated" ~ "Outbreak associated",
  outbreak_or_sporadic == "Sporadic" ~ "Sporadic")) %>%
  janitor::clean_names() %>%
  group_by(episode_week) %>%
  mutate(total_cases = sum(cases)) %>%
  mutate(outbreak_or_sporadic = fct_rev(outbreak_or_sporadic))
# glimpse(outbreak)
```

Data visualization

```
outbreak %>%
  ggplot(aes(x = episode_week, y = cases, fill = outbreak_or_sporadic)) +
  geom_bar(stat = "identity", width = 7) +
  scale_x_date(labels = scales::date_format("%d %b %y"),
               limits = c(date("2020-01-01"), Sys.Date()+7)) +
  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: Yi Zhe Wang for STA303/1002, U of T
  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("#86BCB6", "#B9CA5D"))
```



Task 3: Neighbourhoods

Data wrangling: part 1

```
income <- nbhood_profile %>%
  slice(1143) %>%
  pivot_longer(-c(`_id`, Category, Topic, `Data Source`, Characteristic, `City of Toronto`),
               names_to = "neighbourhood_name", values_to = "percentage") %>%
  mutate(percentage = parse_number(percentage)) %>%
  group_by(neighbourhood_name)

income <- income[-c(1:6)]
# glimpse(income)
```

Data wrangling: part 2

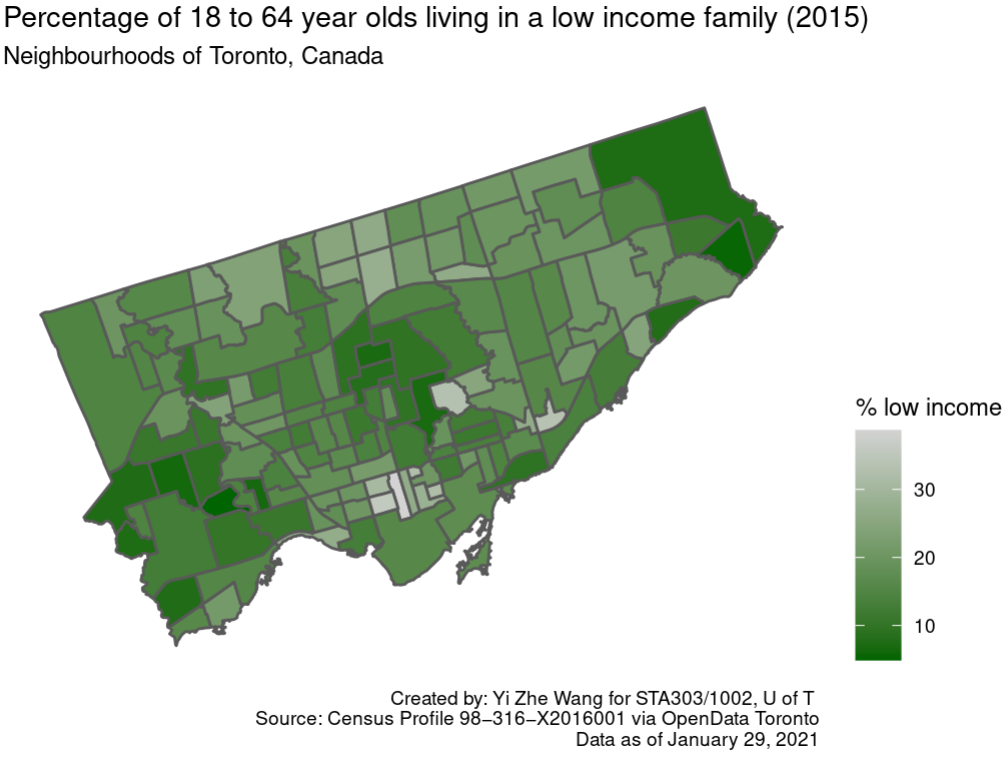
```
nbhoods_all <- nbhoods_shape_raw %>%
  mutate(neighbourhood_name = str_remove(AREA_NAME, "\\s\\(\\d+\\)$")) %>%
  mutate(neighbourhood_name = str_replace(
  neighbourhood_name, "Weston-Pellam Park", "Weston-Pelham Park")) %>%
  mutate(neighbourhood_name = str_replace(
  neighbourhood_name, "North St.James Town", "North St. James Town")) %>%
  mutate(neighbourhood_name = str_replace(
  neighbourhood_name, "Cabbagetown-South St.James Town", "Cabbagetown-South St. James Town")) %
>%
  left_join(income, by = "neighbourhood_name") %>%
  left_join(nbhood_raw, by = "neighbourhood_name") %>%
  rename(rate_per_100000 = rate_per_100_000_people)
# glimpse(nbhoods_all)
```

Data wrangling: part 3

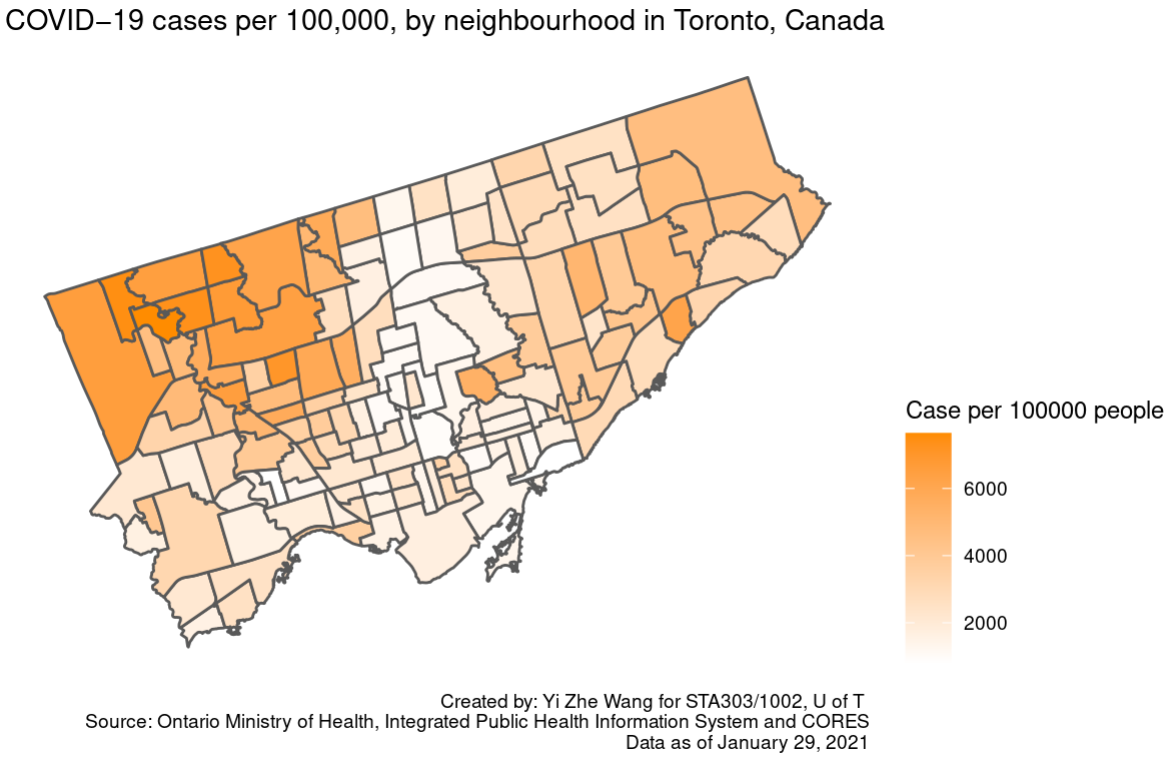
```
nbhoods_final <- nbhoods_all %>%
  mutate(med_inc = median(sort(percentage))) %>%
  mutate(med_rate = median(sort(rate_per_100000))) %>%
  mutate(nbhood_type =
           case_when(percentage >= med_inc & rate_per_100000 >= med_rate
                     ~ "Higher low income rate, higher case rate",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

Data visualization

```
ggplot(data = nbhoods_final) +
  geom_sf(aes(fill = percentage)) +
  theme_map() +
  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: Yi Zhe Wang for STA303/1002, U of T
  Source: Census Profile 98-316-X2016001 via OpenData Toronto\n",
  date_daily[1,1])) +
  scale_fill_gradient(name = "% low income", low = "darkgreen", high = "lightgrey") +
  theme(legend.position = "right")
```



```
ggplot(data = nbhoods_final) +
  geom_sf(aes(fill = rate_per_100000)) +
  theme_map() +
  labs(title = "COVID-19 cases per 100,000, by neighbourhood in Toronto, Canada",
  caption = str_c("Created by: Yi Zhe Wang for STA303/1002, U of T
  Source: Ontario Ministry of Health, Integrated Public Health Information System and CORES\n",
  date_daily[1,1])) +
  scale_fill_gradient(name = "Case per 100000 people", low = "white", high = "darkorange") +
  theme(legend.position = "right")
```



```
ggplot(data = nbhoods_final) +
  geom_sf(aes(fill = nbhood_type)) +
  theme_map() +
  labs(title = "COVID-19 cases and low-income status by neighbourhood in Toronto, Canada",
  caption = str_c("Created by: Yi Zhe Wang 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\n",
  date_daily[1,1])) +
  scale_fill_brewer(name = "% of 18 to 64 year-olds in\n low income families and\n COVID-19 case
rates", palette = "Set1") +
  theme(legend.position = "right")
```