Code last run 2021-02-16.
Daily: Data as of January 29, 2021.
Neighbourhood: Data as of January 28, 2021.

Task 1: Daily cases

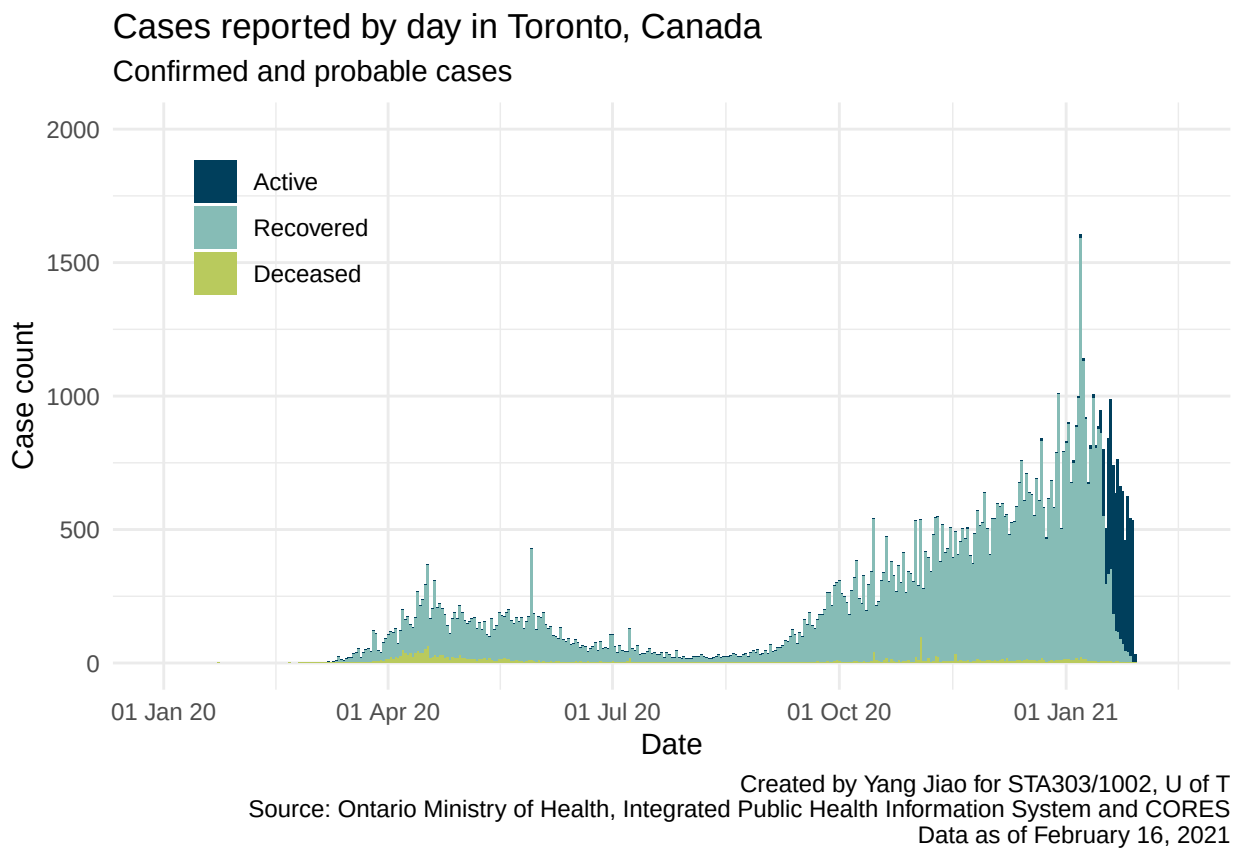
Data wrangling

```
reported <- reported_raw %>%
  mutate_if(is.numeric, replace_na,replace=0) %>%
  pivot_longer(!reported_date,
               names_to = "state", values_to = "count") %>%
  mutate_if(is_character, str_to_sentence) %>%
  mutate(reported_date=date(reported_date)) %>%
  mutate(state=as_factor(state)) %>%
  group_by(reported_date, state)%>%
  mutate(state = fct_relevel(state,"Active",after=0))

levels(reported$state)
```

```
## [1] "Active"    "Recovered" "Deceased"
```

Data visualization

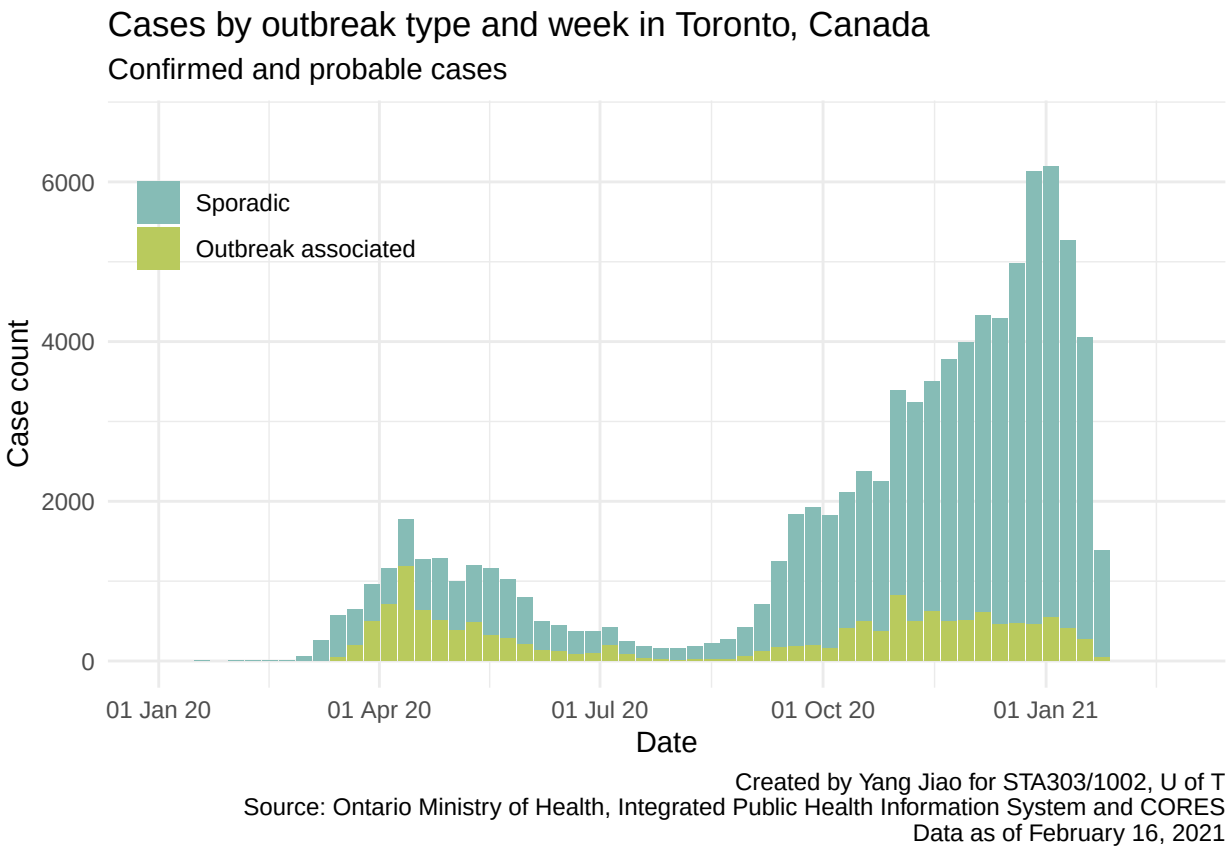


Task 2: Outbreak type

Data wrangling

```
outbreak <- outbreak_raw %>%
  mutate(
    episode_week=date(episode_week),
    outbreak_or_sporadic=str_replace(outbreak_or_sporadic,
                                     "OB Associated",
                                     "Outbreak associated"),
    outbreak_or_sporadic = as_factor(outbreak_or_sporadic)
           ) %>%
  group_by(episode_week, outbreak_or_sporadic)%>%
  mutate(
    outbreak_or_sporadic = fct_relevel(outbreak_or_sporadic,"Sporadic",after=0),
    total_cases=sum(cases))
```

Data visualization



Task 3: Neighbourhoods

Data wrangling: part 1

```
income_temp <- nbhood_profile %>%
  filter(grepl("Income",Category))
#ncol(income_temp)

income <- nbhood_profile %>%
  janitor::clean_names() %>%
  filter(grepl(1143, id)) %>%
  mutate_at(6:146,parse_number) %>%
  pivot_longer(-c(id,category,topic,data_source,characteristic),
               names_to = "neighbourhood_name", values_to = "percentage") %>%
  rows_delete(tibble(neighbourhood_name = "city_of_toronto")) %>%
  mutate(
    neighbourhood_name=str_to_lower(neighbourhood_name),
    neighbourhood_name=str_replace_all(neighbourhood_name, "_", " "),
    neighbourhood_name=str_replace_all(neighbourhood_name, "-", " ")
    ) %>%
  select(neighbourhood_name,id, percentage)
```

Data wrangling: part 2

```
nbhood_raw_temp <- nbhood_raw %>%
  mutate(neighbourhood_name=str_to_lower(neighbourhood_name)) %>%
  mutate(neighbourhood_name=str_replace_all(neighbourhood_name, "_", " ")) %>%
  mutate(neighbourhood_name=str_replace_all(neighbourhood_name, "-", " ")) %>%
  mutate(rate_per_100000=rate_per_100_000_people) %>%
  select(neighbourhood_name,neighbourhood_id,rate_per_100000) %>%
  mutate(neighbourhood_name= str_replace_all(neighbourhood_name, "\\.","")) %>%
  mutate(neighbourhood_name=str_replace_all(neighbourhood_name, "'", " ")) %>%
  rows_delete(tibble(neighbourhood_name = "missing address/postal code")) %>%
  mutate(neighbourhood_name=str_replace_all(neighbourhood_name, "/", " ")) %>%
  mutate(neighbourhood_name=str_replace_all(neighbourhood_name, "\\(|\\)", "")) %>%
  mutate(neighbourhood_name=str_replace(neighbourhood_name,
                                        "yonge stclair",
                                        "yonge st clair")) %>%
  mutate(neighbourhood_name=str_replace(neighbourhood_name,
                                        "standrew windfields",
                                        "st andrew windfields"))
# back and make data corrections

nbhoods_temp <- nbhoods_shape_raw %>%
  mutate(neighbourhood_name= str_remove(AREA_NAME, "\\s\\(\\d+\\)$")) %>%
  select(neighbourhood_name,AREA_ID) %>%
  mutate(neighbourhood_name=str_to_lower(neighbourhood_name)) %>%
  mutate(neighbourhood_name=str_replace_all(neighbourhood_name, "_", " ")) %>%
  mutate(neighbourhood_name=str_replace_all(neighbourhood_name, "-", " ")) %>%
  mutate(neighbourhood_name=str_replace_all(neighbourhood_name, "\\.", " ")) %>%
  mutate(neighbourhood_name=str_replace_all(neighbourhood_name, "'", " ")) %>%
  mutate(neighbourhood_name=str_replace_all(neighbourhood_name, "/", " ")) %>%
  mutate(neighbourhood_name=str_replace_all(neighbourhood_name, "\\(|\\)", "")) %>%
```

```
  mutate(neighbourhood_name=str_replace(neighbourhood_name,
                                        "weston pellam park",
                                        "weston pelham park"))
# back and make data corrections


nbhoods_temp2 <- nbhoods_temp %>%
  full_join(nbhood_raw_temp, by="neighbourhood_name")
nbhoods_all <- nbhoods_temp2 %>%
  full_join(income, by="neighbourhood_name")

# checking duplicated rows
nrow(nbhoods_all)
```

```
## [1] 140
```

```
nrow(nbhood_raw_temp)
```

```
## [1] 140
```

```
nrow(income)
```

```
## [1] 140
```

```
nrow(nbhoods_temp)
```

```
## [1] 140
```

```
nrow(nbhoods_temp2)
```

```
## [1] 140
```

```
problems <- nbhoods_all %>%
  filter(is.na(neighbourhood_id) | is.na(AREA_ID) | is.na(id))
na <- nbhoods_all %>%
  filter_all(any_vars(is.na(.)))
```

Data wrangling: part 3

```
med_inc <- median(nbhoods_all$percentage);med_inc
```

```
## [1] 16.55
```

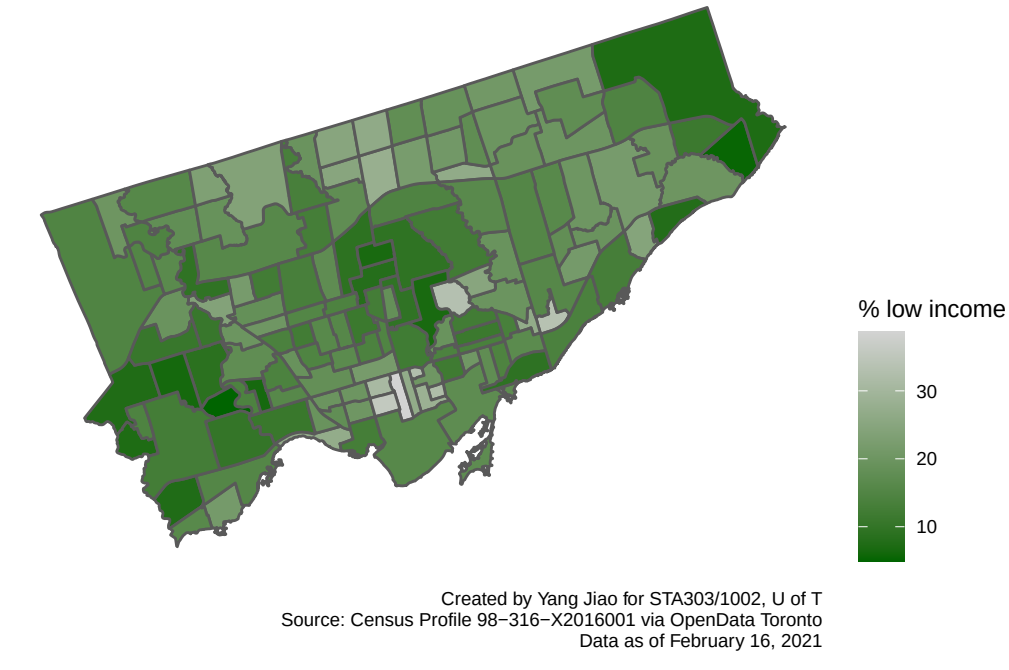
```
med_rate <- median(nbhoods_all$rate_per_100000);med_rate
```

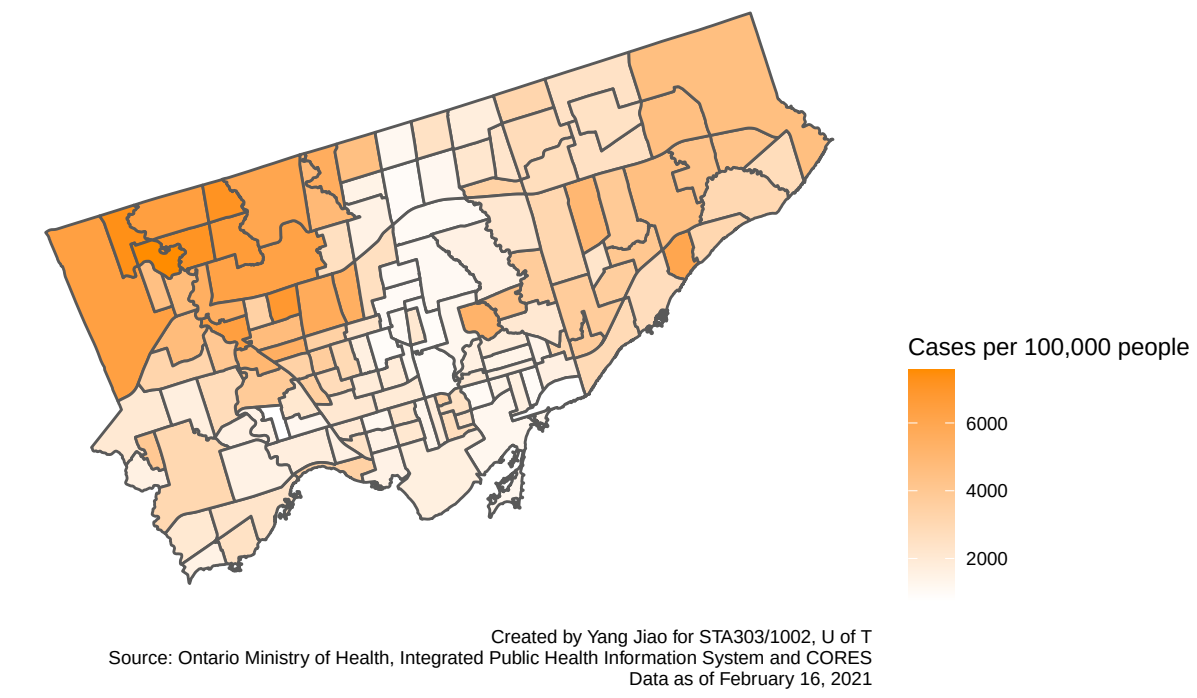
```
## [1] 2486.039
```

```
nbhoods_final <- nbhoods_all %>%
  select(neighbourhood_name,percentage,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))
```

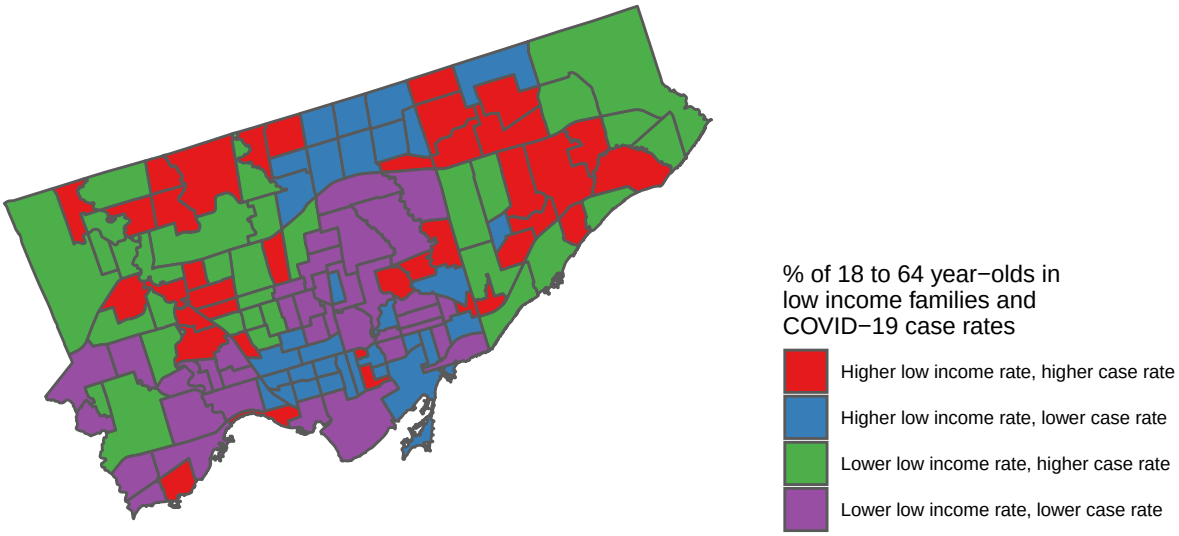
```
    ~ "Lower low income rate, lower case rate"
))
```

Data visualization




COVID−19 cases per 100,000, by neighbourhood in Toronto, Canada
Cases per 100,000 people
6000
4000
2000
Created by Yang Jiao for STA303/1002, U of T
Source: Ontario Ministry of Health, Integrated Public Health Information System and CORES
Data as of February 16, 2021

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



Created by Yang Jiao 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
Data as of February 16, 2021