My Courses

(2021 Winter) STA303H1 S LEC…
(2020 Fall) STA437H1 F LEC510…
(2020 Fall) STA347H1 F 20209:…
(2020 Summer) STA302/STA100…
(2020 Summer) ECO101H1 F L…
(2020 Winter) STA238H1 S LEC…
(2020 Winter) STA304H1 S LEC…
(2019 Fall-Winter) MAT235 All S…
(2019 Fall) STA237H1 F LEC010…
(2019 Summer) MAT223H1 F LE…

My grades for **Data exploration**

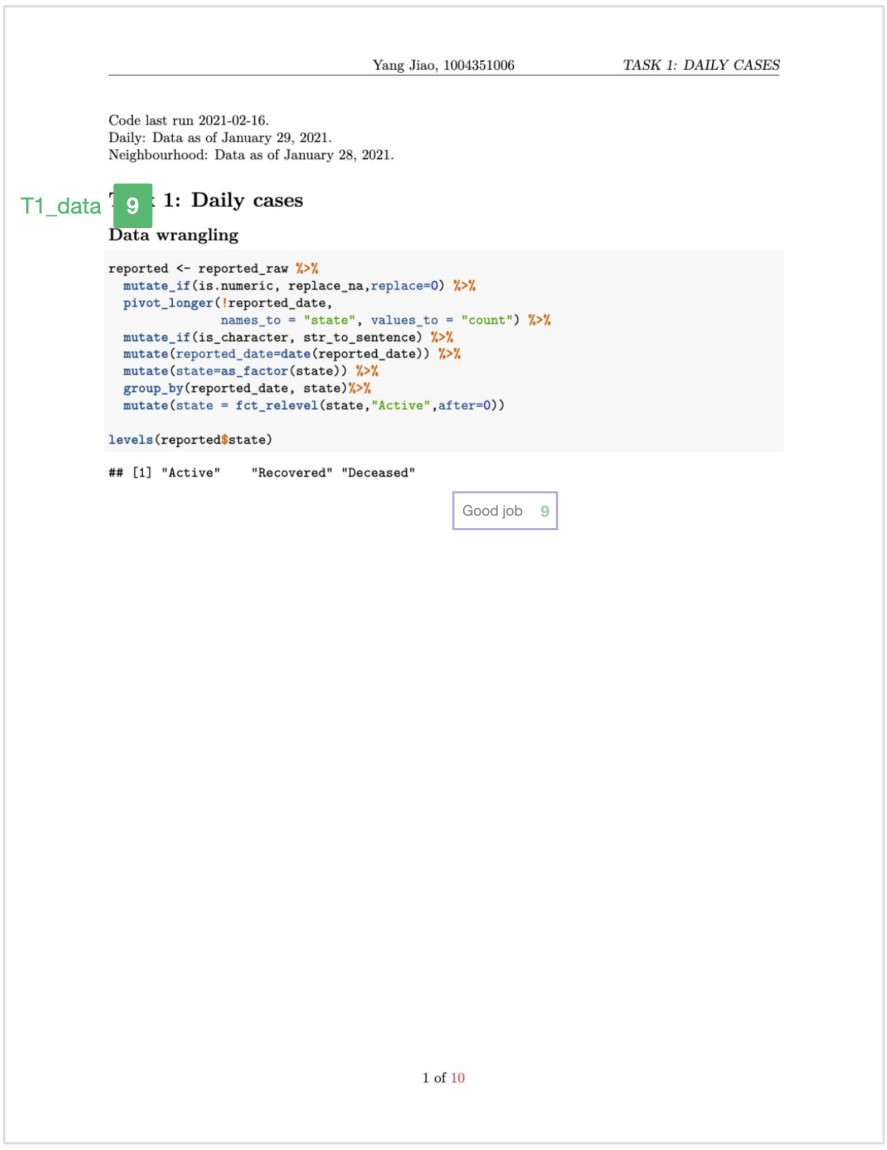
Print Get shareable link Hide feedback

Total score **94.6%** (70/74)

T1_data

/9 points

Upload the pages of your assessment that cover Task 1: Daily cases: Data wrangling.



Yang Jiao, 1004351006 *TASK 1: DAILY CASES*

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

T1_data 9

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"
```

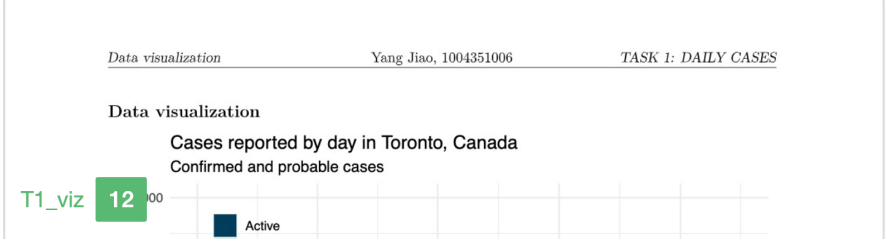
Good job 9

1 of 10

T1_viz

/12 points

Upload the pages of your assessment that cover Task 1: Daily cases: Data visualisation.



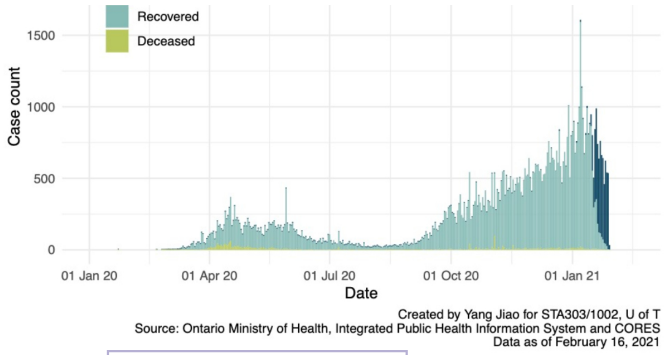
Data visualization Yang Jiao, 1004351006 *TASK 1: DAILY CASES*

Data visualization

Cases reported by day in Toronto, Canada
Confirmed and probable cases

T1_viz 12

Active



Please add on the code for data visualization as well 12

T2_data

/7 points

Upload the pages of your assessment that cover Task 2: Outbreak type: Data wrangling.

Task 2: Outbreak type

Data wrangling

T2_data 7

```
ak <- outbreak_raw %>%
te(
isode_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))
```

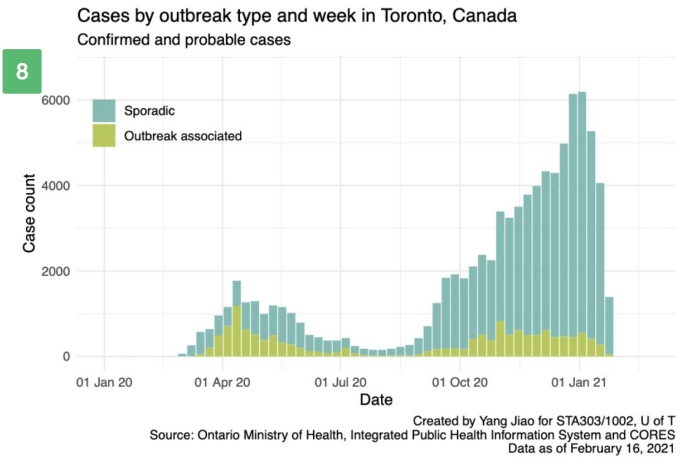
Nice work 7

T2_viz

/12 points

Upload the pages of your assessment that cover Task 2: Outbreak type: Data visualisation.

Data visualization



T2_viz 8

codes? -4

T3_data

/16 points

Upload the pages of your assessment that cover Task 3: Neighbourhoods: Data wrangling. Parts 1, 2 and 3.

Task 3: Neighbourhoods

Data wrangling: part 1

```
temp <- nbhood_profile %>%
  (grepl("Income",Category))
ncome_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),
```

T3_data 16

```
                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"
))
```

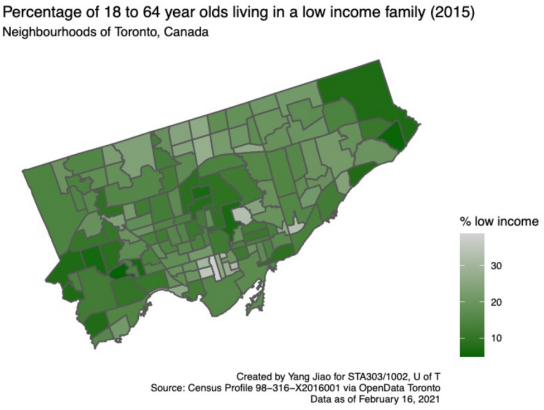

T3_viz1

/6 points

Upload the pages of your assessment that cover Task 3: Neighbourhoods: Data visualisation. Graph 1 only!

Data visualization

T3_viz1 6

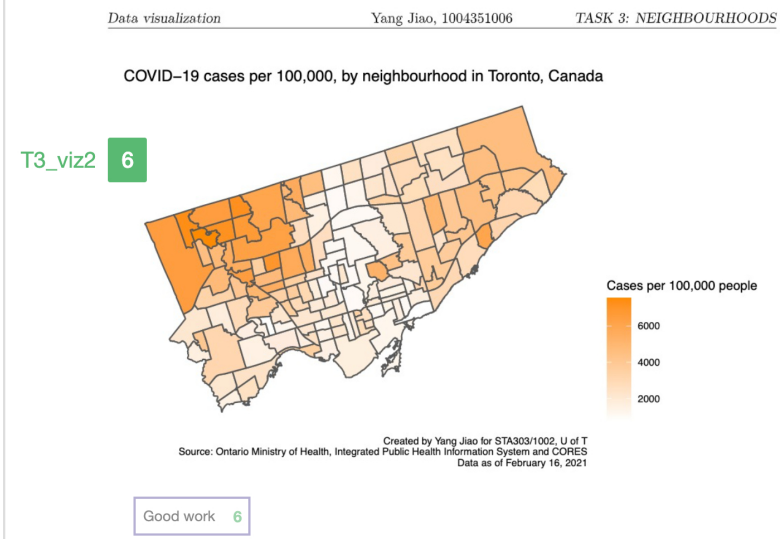


Good work 6

T3_viz2

/6 points

Upload the pages of your assessment that cover Task 3: Neighbourhoods: Data visualisation. Graph 2 only!



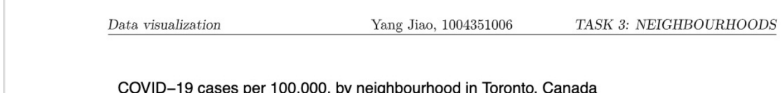
T3_viz2 6

Good work 6

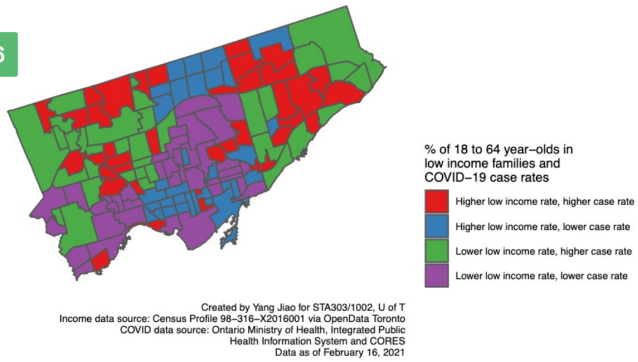
T3_viz3

/6 points

Upload the pages of your assessment that cover Task 3: Neighbourhoods: Data visualisation. Graph 3 only!



T3_viz3 6



Good work 6

Back to top

Help