# Tutorial2

## Set up

```
library("janitor")
```

```
Attaching package: 'janitor'

The following objects are masked from 'package:stats':

    chisq.test, fisher.test
```

```
library("knitr")
library("lubridate")
```

```
Attaching package: 'lubridate'

The following objects are masked from 'package:base':

    date, intersect, setdiff, union
```

```
library("opendatatoronto")
library("tidyverse")
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr   1.1.4     v readr   2.1.5
v forcats 1.0.0     v stringr 1.5.1
v ggplot2 3.4.4     v tibble  3.2.1
v purrr   1.0.2     v tidyr   1.3.0

-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to beco
```

```
library(readr)
Tickets_Issued <- read_csv("Tickets Issued.csv")
```

```
Rows: 23204 Columns: 9
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (6): DIVISION, TICKET_TYPE, OFFENCE_CATEGORY, AGE_GROUP, HOOD_158, NEIGH...
dbl (3): _id, OFFENCE_YEAR, TICKET_COUNT

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

## Basic Cleaning of the Data

```
cleaned_tickets_issued <-
  clean_names(Tickets_Issued)
head(cleaned_tickets_issued)
```

```
# A tibble: 6 x 9
     id offence_year division ticket_type    offence_category age_group hood_158
  <dbl>        <dbl> <chr>    <chr>          <chr>            <chr>     <chr>
1     1         2015 D41      Prov Offence ~ Other HTA        Youth     121
2     2         2020 D42      Prov Offence ~ Other HTA        Youth     148
3     3         2020 D13      Prov Offence ~ Aggressive Driv~ Adult     107
4     4         2021 D22      Prov Offence ~ Other HTA        Adult     160
5     5         2014 D42      Prov Offence ~ Aggressive Driv~ Adult     116
6     6         2021 D32      Prov Offence ~ Distracted Driv~ Adult     40
# i 2 more variables: neighbourhood_158 <chr>, ticket_count <dbl>
```

##Create graph for

```
##Want to look at the ticket increase of the year in Etobicoke West Mall
  cleaned_tickets_issued |>

  filter(neighbourhood_158 == "Etobicoke West Mall",
         offence_year %in% c(2016, 2017, 2018, 2019, 2020))|>
  arrange(offence_year) |>
  select(offence_year, ticket_count)|>

  summarize(sum_per_year = sum(ticket_count, na.rm = TRUE),.by = offence_year) |>
  ggplot(mapping = aes(x = offence_year, y = sum_per_year)) +
  geom_bar(stat= 'identity') +
  labs(title = "Ticket per year Etobicoke West Mall", x = "Year", y = "total ticket")
```