minipaper2-opendatatoronto

Yichen Ji (Selina)

Figure 1

```
library(dplyr)
library(knitr)
library(opendatatoronto)
library(tidyverse)
library(purrr)
library(ggplot2)
packages <- list_packages(limit = 10)

ttc_packages <- search_packages("ttc")

ttc_delay_packages <- search_packages("TTC Subway Delay")

ttc_delay_resources <- ttc_delay_packages %>%
  list_package_resources()

ttc_delay_statistics <- ttc_delay_resources %>%
```

Table 1: Gender Distribution by Year

year_of_death	Female	Male	Unknown	Sum
2017	25	75	0	100
2018	21	73	0	94
2019	34	92	1	127
2020	28	115	0	143
2021	50	161	6	217
2022	39	146	4	189
2023	9	69	1	79
Total	206	731	12	949

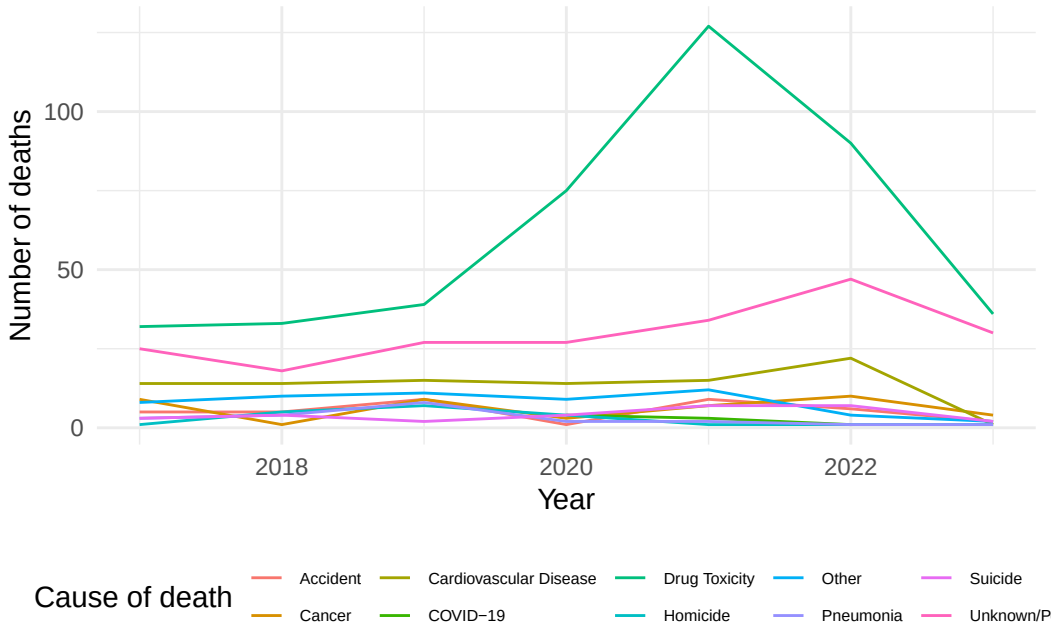


Figure 1: Trend of causes of death from 2017 to 2023

```
  slice(5) %>%
  get_resource()

ttc_tibble <- map_dfr(ttc_delay_statistics, ~tibble(
  station = .x$Station,
  code = .x$Code,
  min_delay = .x$`Min Delay`
))

ttc_tibble
```

```
# A tibble: 20,737 x 3
   station               code  min_delay
   <chr>                 <chr>     <dbl>
 1 SHEPPARD WEST STATION MUATC        10
 2 DUNDAS STATION        MUNCA         0
 3 MUSEUM STATION        MUSC          0
 4 BAY LOWER             EUOE          0
 5 MUSEUM STATION        MUO           6
 6 BLOOR DANFORTH SUBWAY MUGD          0
 7 KIPLING STATION       MUSAN         3
```

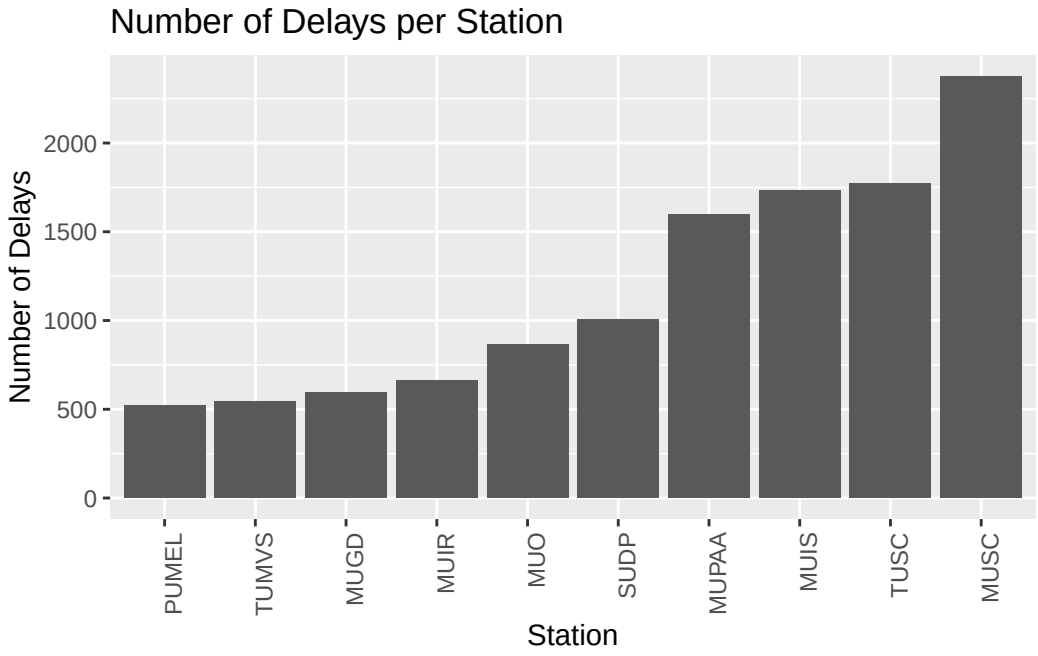
```
 8 UNION STATION         MUIS          0
 9 COLLEGE STATION       SUDP          7
10 KIPLING STATION       SUDP          0
# i 20,727 more rows
```

```
ttc_tibble%>%filter(min_delay > 0)
```

```
# A tibble: 7,168 x 3
   station               code  min_delay
   <chr>                 <chr>     <dbl>
 1 SHEPPARD WEST STATION MUATC        10
 2 MUSEUM STATION        MUO           6
 3 KIPLING STATION       MUSAN         3
 4 COLLEGE STATION       SUDP          7
 5 WARDEN STATION        MUI           3
 6 DONLANDS STATION      TUNOA         4
 7 KEELE STATION         TUNOA         4
 8 DONLANDS STATION      TUNOA         4
 9 KENNEDY SRT STATION   ERTC          7
10 VICTORIA PARK STATION EUDO          6
# i 7,158 more rows
```

```
station_delays <- ttc_tibble%>%
  group_by(code) %>%
  summarise(Delay_Count = n())%>%
  ungroup() %>%
  top_n(10, Delay_Count) %>%
  arrange(Delay_Count)

# Plot the histogram
ggplot(station_delays, aes(x = reorder(code, Delay_Count), y = Delay_Count)) +
  geom_col() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(title = "Number of Delays per Station", x = "Station", y = "Number of Delays")
```


Number of Delays per Station
Number of Delays
2000
1500
1000
500
0
PUMEL
TUMVS
MUGD
MUIR
MUO
SUDP
MUPAA
MUIS
TUSC
MUSC
Station