

STA304 2022 winter Paper1

Abstract

The data is about Polling data that is downloaded from opendatatoronto and the website is : <http://opentoronto.ca>. The data is imported from the Open Data Toronto Portal using the opendatatoronto package.

##Import data

```
#install the needed packages
#install.packages(tidyverse)
#install.packages(opendatatoronto)
#install.packages(dplyr)
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.6      v dplyr  1.0.7
## v tidyr   1.1.4      v stringr 1.4.0
## v readr   2.1.1      v forcats 0.5.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(opendatatoronto)
library(dplyr)
```

```
#import data
```

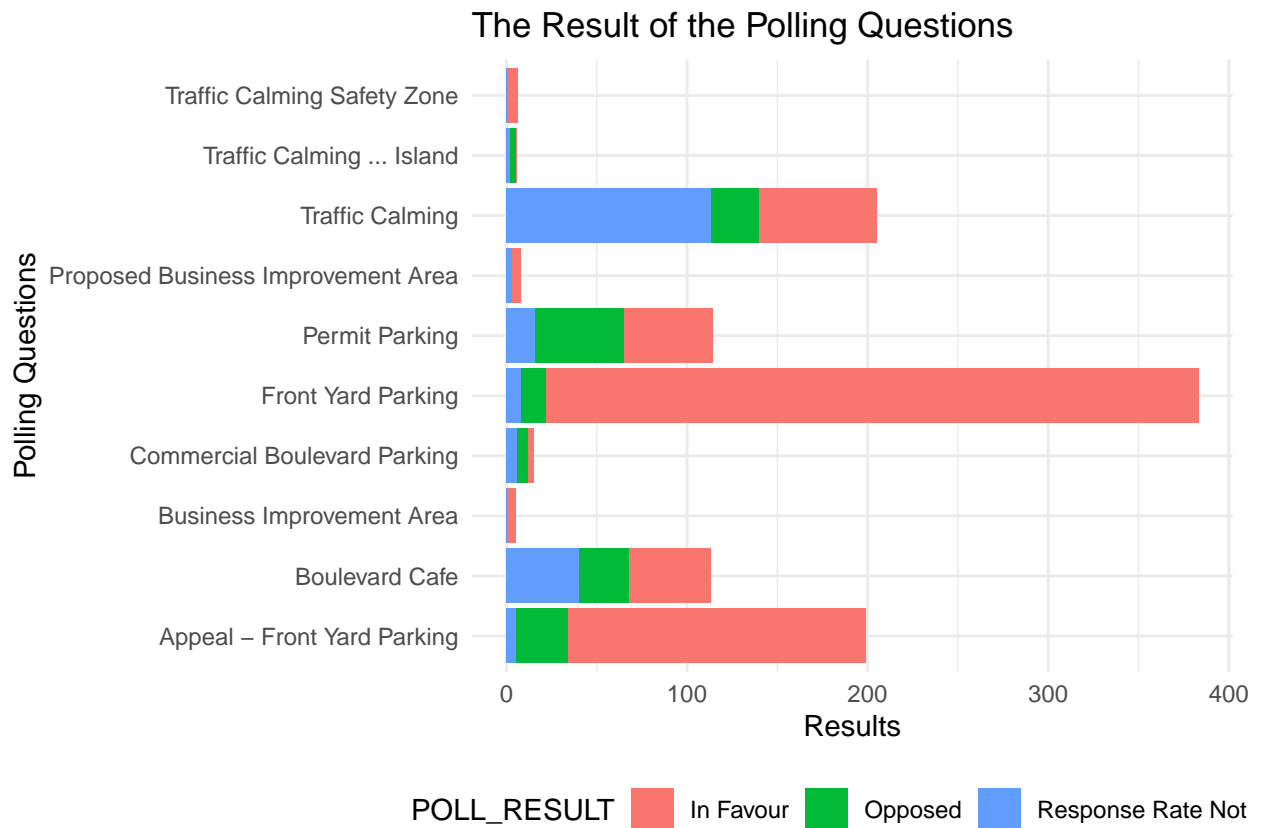
```
package <- show_package("7bce9bf4-be5c-4261-af01-abfbc3510309")
resources <- list_package_resources("7bce9bf4-be5c-4261-af01-abfbc3510309")
datastore_resources <- filter(resources, tolower(format) %in% c('csv', 'geojson'))
data <- filter(datastore_resources, row_number()==1) %>% get_resource()
```

```
#choose the columns and rows
```

```
data1 <- data %>% filter(!is.na(BALLOTS_IN_FAVOUR)) %>% filter(!is.na(BALLOTS_DISTRIBUTED))%>%
  select(APPLICATION_FOR, BALLOTS_CAST, BALLOTS_DISTRIBUTED, BALLOTS_IN_FAVOUR,POLL_RESULT)
```

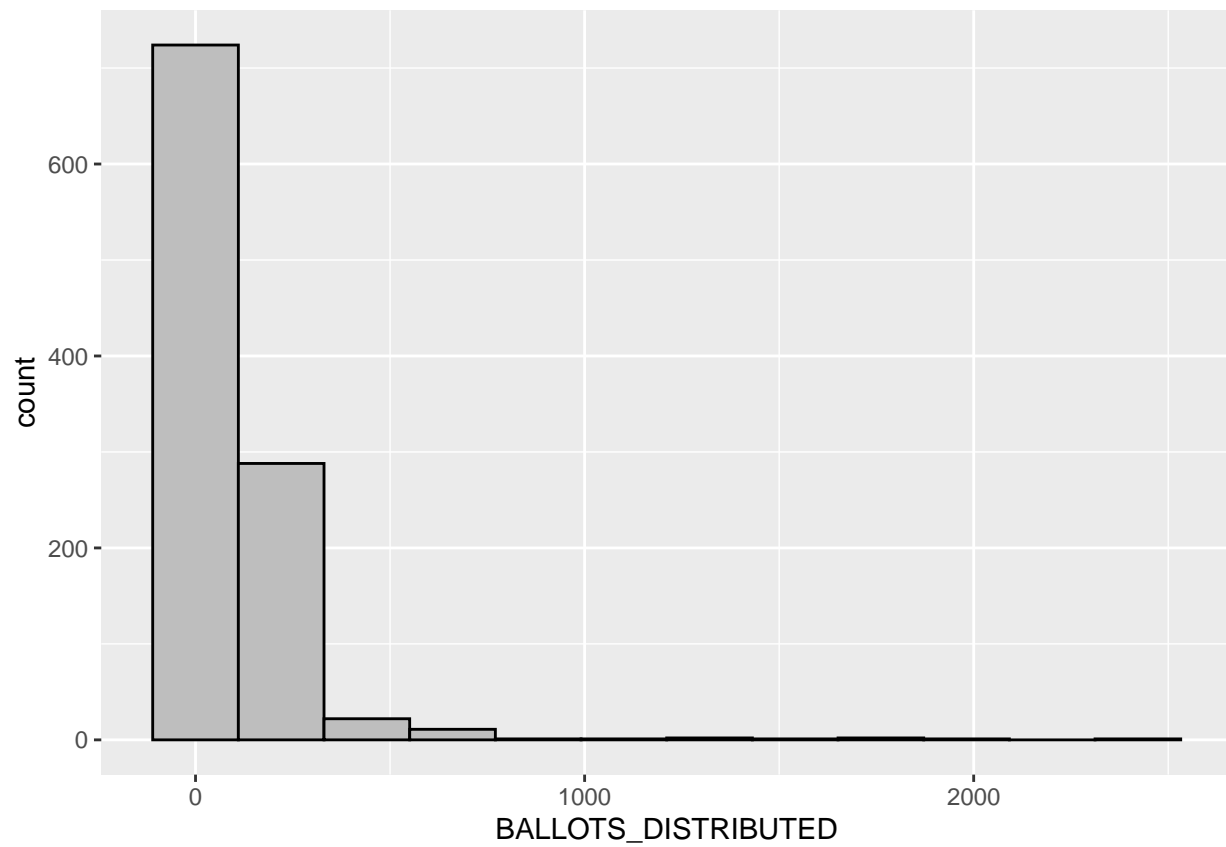
```
#plot
```

```
data1 %>%ggplot(aes(x= APPLICATION_FOR, fill = POLL_RESULT)) + geom_bar() +
labs(x = "Polling Questions", y = "Results", title = "The Result of the Polling Questions")+theme_minimal()
```

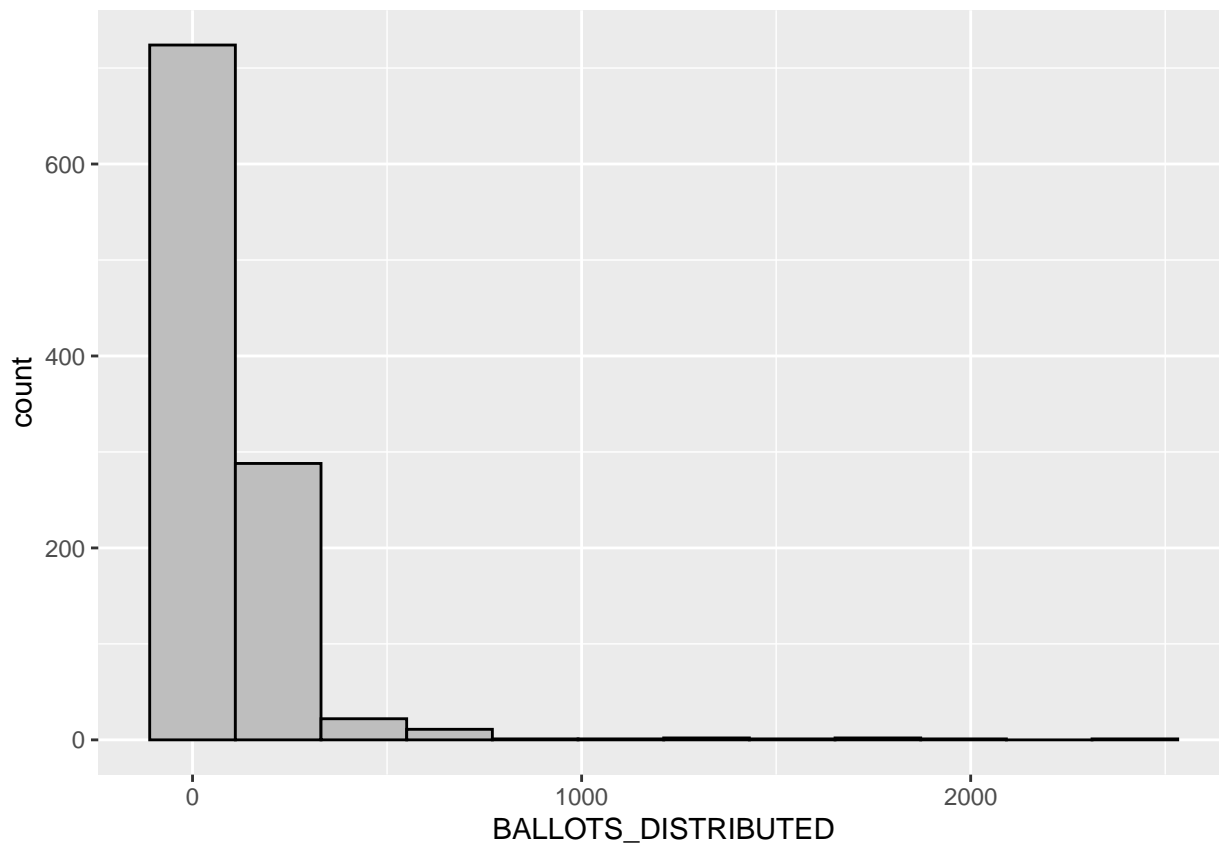


```
data1<- data1 %>%mutate(percent =BALLOTS_CAST/BALLOTS_DISTRIBUTED)
```

```
#plot
data1 %>%ggplot(aes(x=BALLOTS_DISTRIBUTED)) +
geom_histogram(bins=12, color="black", fill="gray")
```



```
#plot  
data1 %>%ggplot(aes(x=BALLOTS_DISTRIBUTED)) +  
geom_histogram(bins=12, color="black", fill="gray")
```



##the distribution of ballots

#summary table 1

```
table1 <- data1 %>% summarise(
  min = min(BALLOTS_DISTRIBUTED),
  Q1 = quantile(BALLOTS_DISTRIBUTED,0.25),
  median = median(BALLOTS_DISTRIBUTED),
  Q3 = quantile(BALLOTS_DISTRIBUTED,0.75),
  max = max(BALLOTS_DISTRIBUTED),
  mean = mean(BALLOTS_DISTRIBUTED),
  sd = sd(BALLOTS_DISTRIBUTED),
)
```

```
knitr::kable(table1, caption = "Ballots Number Summary Table")
```

Table 1: Ballots Number Summary Table

min	Q1	median	Q3	max	mean	sd
2	56	85	125	2424	118.4734	166.7614

#summary table 2

```
table2 <- data1 %>% summarise(
  min = min(BALLOTS_CAST),
  Q1 = quantile(BALLOTS_CAST,0.25),
  median = median(BALLOTS_CAST),
  Q3 = quantile(BALLOTS_CAST,0.75),
```

```

        max = max(BALLOTS_CAST),
        mean = mean(BALLOTS_CAST),
        sd = sd(BALLOTS_CAST),
    )

knitr::kable(table2, caption = "Ballots Cast Summary Table")

```

Table 2: Ballots Cast Summary Table

min	Q1	median	Q3	max	mean	sd
0	25	37	54	971	46.62334	50.14024

##Citation for R package