STA304 Tutorial Week 3

Tianyi Zhang 1005156607

23/01/2022

Setup

(Gelfand 2020)

In this tutorial we use R Core Team (2020)

Data Cleaning

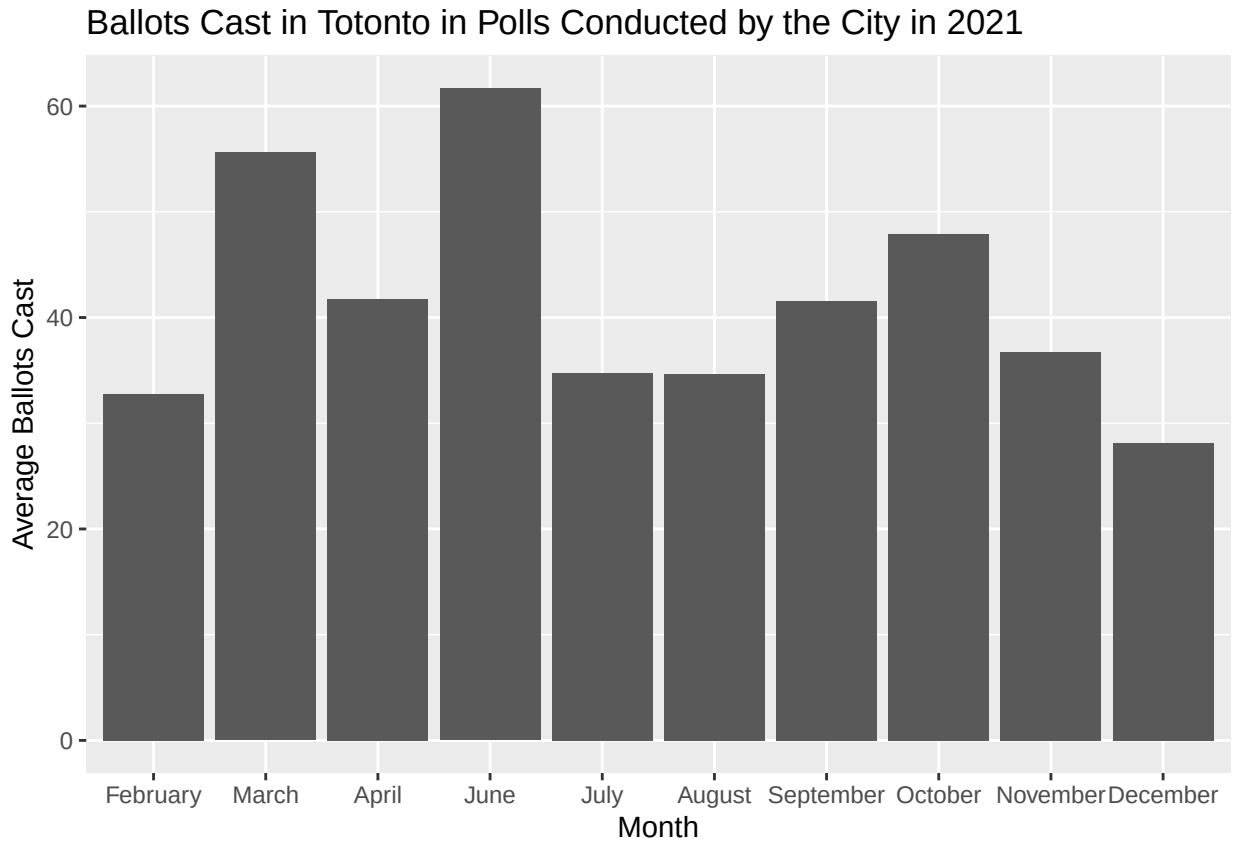
```
#Polls conducted by the city
polls <-
  list_package_resources("7bce9bf4-be5c-4261-af01-abfbc3510309") %>%
  filter(tolower(format) %in% c('csv', 'geojson')) %>%
  # filter(row_number()==1) %>%
  get_resource()

#Clean dataset
polls_clean <-
  polls %>%
  clean_names() %>%
  select(id, ballots_cast, open_date) %>% # Select only the columns needed
  mutate(open_date = as_date(open_date)) %>% # change all open_date values to date object format
  filter(open_date >= as_date("2021-01-01")) # filter the dataset so that only the
```

Graph

Below is a histogram that showcases the average ballots cast in city-conducted polls in Toronto, 2021.

```
# Make a graph of the dataset
polls_clean %>%
  mutate(open_date = month(open_date,
                           label = TRUE,
                           abbr = FALSE)) %>%
 drop_na(open_date) %>% # Drop the NA values
  group_by(open_date) %>%
  summarise(number_cast = mean(ballots_cast)) %>%
  ggplot(aes(x=open_date,y=number_cast)) + # use ggplot to graph
  geom_col() +
  labs(x="Month",
```

```
      y="Average Ballots Cast",
      title="Ballots Cast in Totonto in Polls Conducted by the City in 2021")
```



GitHub Repo

https://github.com/zhan7818/STA304-Winter-2022.git

Citations

Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal.* https://CRAN.R-project.org/package=opendatatoronto.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.