Exploring the 2019 Canadian Election

Shuyang Qiu

```
#### Preamble ####
# Purpose: Read in data from the 2019 Canadian Election and make
# a graph of the number of seats each party won.
# Author: Shuyang Qiu
# Email: shuyang.qiu@mail.utoronto.ca
# Date: 9 January 2023
# Prerequisites: Know where to get Canadian elections data.

#### Workspace setup ####
#install.packages("tidyverse")
#install.packages("janitor")

library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.4.4     v tibble    3.2.1
v lubridate 1.9.3     v tidyr     1.3.0
v purrr     1.0.2
-- 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 becor
```

```
library(janitor)
```

```
Attaching package: 'janitor'

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

```
    chisq.test, fisher.test
```

Plan

Simulate

```
simulated_data <-
  tibble(
    # Use 1 through to 338 to represent each division
    "District" = 1:338,
    # Randomly pick an option, with replacement, 151 times
    "Party" = sample(
      x = c("Liberal", "Conservative", "Bloc Québécois", "New Democratic", "Green", "Other"),
      size = 338,
      replace = TRUE
    )
  )

simulated_data
```

```
# A tibble: 338 x 2
   District Party
      <int> <chr>
 1        1 Green
 2        2 New Democratic
 3        3 Other
 4        4 Other
 5        5 Green
 6        6 Conservative
 7        7 Bloc Québécois
 8        8 Green
 9        9 Conservative
10       10 Conservative
# i 328 more rows
```

Acquire

```
#### Read in the data ####
raw_elections_data <-
  read_csv(
    file =
      "https://www.elections.ca/res/rep/off/ovr2021app/53/data_donnees/table_tableau11.csv",
    show_col_types = FALSE
  )

# We have read the data from the elections.ca website. We may like to save
# it in case something happens or they move it.
write_csv(
  x = raw_elections_data,
  file = "canadian_voting.csv"
)
```

```
#### Basic cleaning ####
raw_elections_data <-
  read_csv(
    file = "canadian_voting.csv",
    show_col_types = FALSE
  )

# Make the names easier to type
cleaned_elections_data <-
  clean_names(raw_elections_data)

# Select columns
cleaned_elections_data <-
  cleaned_elections_data |>
  select(
    electoral_district_name_nom_de_circonscription,
    elected_candidate_candidat_elu
  )

# Rename columns
cleaned_elections_data <-
  cleaned_elections_data |>
  rename(
    district = electoral_district_name_nom_de_circonscription,
    elected_candidate = elected_candidate_candidat_elu
  )
```

```
# Keep only party information
cleaned_elections_data <-
  cleaned_elections_data |>
  separate(
    col = elected_candidate,
    into = c("Other", "party"),
    sep = "/"
  ) |>
  select(-Other)

# Map French names back to English
cleaned_elections_data <-
  cleaned_elections_data |>
  mutate(
    party =
      case_match(
        party,
        "Libéral" ~ "Liberal",
        "Conservateur" ~ "Conservative",
        "NPD-Nouveau Parti démocratique" ~ "NDP-New Democratic Party",
        "Bloc Québécois" ~ "Bloc Québécois",
        "Parti Vert" ~ "Green Party"
      )
  )

# Write cleaned data
write_csv(
  x = cleaned_elections_data,
  file = "cleaned_elections_data.csv"
)
```

Explore

```
#### Read in the data ####
cleaned_elections_data <-
  read_csv(
    file = "cleaned_elections_data.csv",
    show_col_types = FALSE
  )
```

```
# Get number of seats each party won
cleaned_elections_data |>
  count(party)
```

```
# A tibble: 5 x 2
  party                        n
  <chr>                    <int>
1 Bloc Québécois              32
2 Conservative               119
3 Green Party                  2
4 Liberal                    160
5 NDP-New Democratic Party    25
```

```
# Graph data
cleaned_elections_data |>
  ggplot(aes(x = party)) +
  geom_bar() +
  theme_minimal() + # Make the theme neater
  labs(x = "Party", y = "Number of seats") # Make labels more meaningful
```