Tutorial1

Ximing Shen

```
#### Preamble ####
# Purpose: Read in data from the 2021 Canada Federal Election and make
# a graph of the number of ridings each party won.
# Author: Ximing Shen
# Email: ximing.shen@mail.utoronto.ca
# Date: 9 January 2024
# Prerequisites: ---
```

```
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 beco
```

```
library(janitor)
```

```
Attaching package: 'janitor'

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

    chisq.test, fisher.test
```

```r
#### Simulation ####
simulated_data <-
  tibble(
    "Riding" = 1:338,
    "Party" = sample(
      x = c("Liberal", "Conservative", "Bloc Québécois", "New Democratic",
            "Green", "Other"),
      size = 338,
      replace = TRUE
    )
  )
simulated_data
```

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

```r
#### Read in the data ####
raw_elections_data <-
  read_csv(
    file = "table_tableau11.csv",
    show_col_types = FALSE,
  )
write_csv(
  x = raw_elections_data,
  file = "Canada_Federal_Election.csv"
)

head(raw_elections_data)
```

```
# A tibble: 6 x 13
  Province              Electoral District N~1 Electoral District N~2 Population
  <chr>                 <chr>                                   <dbl>      <dbl>
1 Newfoundland and Lab~ Avalon                                  10001      86494
2 Newfoundland and Lab~ Bonavista--Burin--Tri~                  10002      74116
3 Newfoundland and Lab~ Coast of Bays--Centra~                  10003      77680
4 Newfoundland and Lab~ Labrador                                10004      27197
5 Newfoundland and Lab~ Long Range Mountains                    10005      86553
6 Newfoundland and Lab~ St. John's East/St. J~                  10006      85697
# i abbreviated names: 1: `Electoral District Name/Nom de circonscription`,
#   2: `Electoral District Number/Numéro de circonscription`
# i 9 more variables: `Electors/Électeurs` <dbl>,
#   `Polling Stations/Bureaux de scrutin` <dbl>,
#   `Valid Ballots/Bulletins valides` <dbl>,
#   `Percentage of Valid Ballots /Pourcentage des bulletins valides` <dbl>,
#   `Rejected Ballots/Bulletins rejetés` <dbl>, ...
```

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

```
cleaned_elections_data <-
  clean_names(raw_elections_data)
```

```
cleaned_elections_data <-
  cleaned_elections_data |>
  select(electoral_district_name_nom_de_circonscription,
         elected_candidate_candidat_elu
  )
head(cleaned_elections_data)
```

```
# A tibble: 6 x 2
  electoral_district_name_nom_de_circonscription elected_candidate_candidat_elu
  <chr>                                          <chr>
1 Avalon                                         McDonald, Ken Liberal/Libéral
2 Bonavista--Burin--Trinity                      Rogers, Churence Liberal/Libér~
3 Coast of Bays--Central--Notre Dame             Small, Clifford Conservative/C~
```

```
4 Labrador                                    Jones, Yvonne Liberal/Libéral
5 Long Range Mountains                        Hutchings, Gudie Liberal/Libér~
6 St. John's East/St. John's-Est              Thompson, Joanne Liberal/Libér~
```

```
cleaned_elections_data <-
  cleaned_elections_data |>
  rename(
    riding = electoral_district_name_nom_de_circonscription,
    elected_candidate = elected_candidate_candidat_elu
  )

head(cleaned_elections_data)
```

```
# A tibble: 6 x 2
  riding                           elected_candidate
  <chr>                            <chr>
1 Avalon                           McDonald, Ken Liberal/Libéral
2 Bonavista--Burin--Trinity        Rogers, Churence Liberal/Libéral
3 Coast of Bays--Central--Notre Dame Small, Clifford Conservative/Conservateur
4 Labrador                         Jones, Yvonne Liberal/Libéral
5 Long Range Mountains             Hutchings, Gudie Liberal/Libéral
6 St. John's East/St. John's-Est   Thompson, Joanne Liberal/Libéral
```

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

head(cleaned_elections_data)
```

```
# A tibble: 6 x 2
  riding                           party
  <chr>                            <chr>
1 Avalon                           Libéral
2 Bonavista--Burin--Trinity        Libéral
```

```
3 Coast of Bays--Central--Notre Dame Conservateur
4 Labrador                           Libéral
5 Long Range Mountains               Libéral
6 St. John's East/St. John's-Est     Libéral
```

```
cleaned_elections_data$party |>
  unique()
```

```
[1] "Libéral"                          "Conservateur"
[3] "Bloc Québécois"                   "NPD-Nouveau Parti démocratique"
[5] "Parti Vert"
```

```
#### Recode Party Name ####
cleaned_elections_data <-
  cleaned_elections_data |>
  mutate(
    party =
      case_match(
        party,
        "Libéral" ~ "Liberal",
        "Conservateur" ~ "Conservative",
        "Liberal" ~ "Liberal",
        "Bloc Québécois" ~ "Bloc Québécois",
        "NPD-Nouveau Parti démocratique" ~ "New Democratic",
        "Parti Vert"  ~ "Green",

      )
  )

head(cleaned_elections_data)
```

```
# A tibble: 6 x 2
  riding                             party
  <chr>                              <chr>
1 Avalon                             Liberal
2 Bonavista--Burin--Trinity          Liberal
3 Coast of Bays--Central--Notre Dame Conservative
4 Labrador                           Liberal
5 Long Range Mountains               Liberal
6 St. John's East/St. John's-Est     Liberal
```

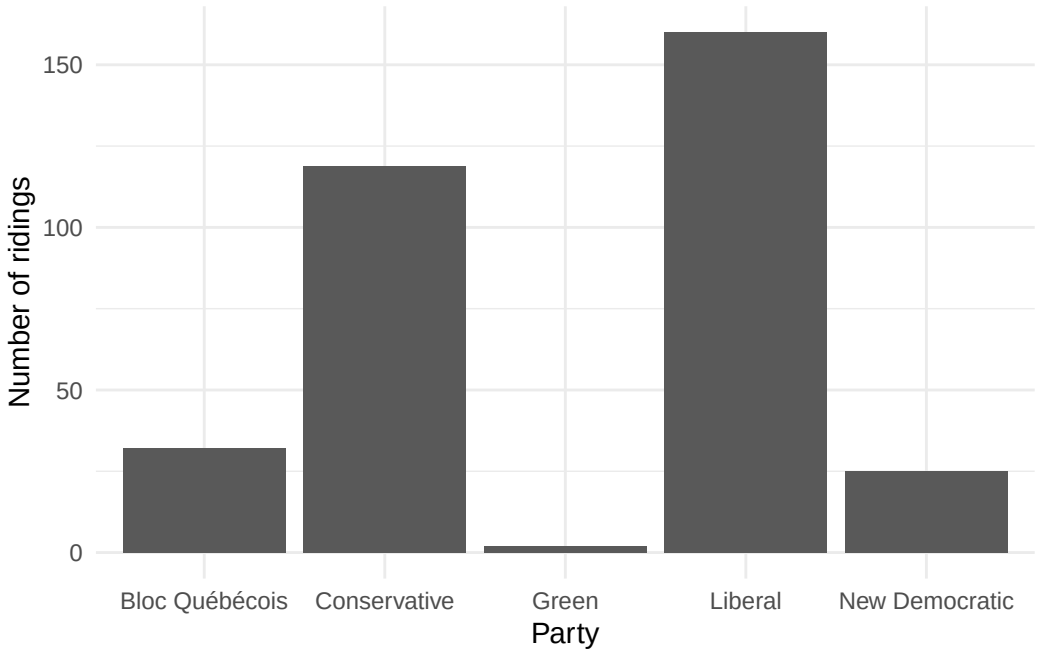
```
write_csv(
  x = cleaned_elections_data,
  file = "Cleaned_Canada_Federal_Election.csv"
)
```

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

cleaned_elections_data |>
  count(party)
```

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

```
cleaned_elections_data |>
  ggplot(aes(x = party)) +
  geom_bar() +
  theme_minimal() +
  labs(x = "Party", y = "Number of ridings")
```


150
100
50
0
Number of ridings
Bloc Québécois
Conservative
Green
Liberal
New Democratic
Party