# R Notebook

This project is an exploratory data analysis that analyzes the influence and impact of the People's Party of Canada (PPC) on the 2021 Canadian parliamentary elections. They are a far right party started by Maxime Bernier that gained relative popularity in Canada as result of COVID-19 policies enacted by the presiding Liberal minority government lead by Prime Minister Justin Trudeau and general populist, reactionary sentiment. This EDA used tables, visualizations, and descriptive statistics to demonstrate how they shaped the election. In particular, the EDA, based on this given data, will discuss how the PPC may have acted as a spoiler for Canada's conservative party, the leading center-right party which Bernier broke off from in 2018 to form the PPC.

## Loading in Libraries

The mapcan package, and the geospatial data of canadian ridings that came with it by Andrew McCormack and Aaron Erlich was a useful tool I used to create some of the maps for this project.

```
library(tidyverse)
library(devtools)
#install_github("mccormackandrew/mapcan", build_vignettes = TRUE, force = TRUE)
library(ggthemes)
library(mapcan)
library(kableExtra)
```

## Data Wrangling

### Reading in data, removing duplicate variables or concantenating variables, and improving clarity of column names

The data comes from the Elections Canada, an official government website.

```
CA <- read_tsv("EventResults.txt", col_names = TRUE, skip = 1)

CA <- CA %>%
  rename_with(~ str_remove(.x, "-.*")) %>%
  mutate(Full_name = str_c(`Given name `, `Surname `, sep = " "), .keep = "unused") %>%
  select(-`Nom de la circonscription`, -`Type de résultats**`, -`Middle name(s) `, -`Appartenance politi
  filter(`Type of results*` == "validated") %>%
  select(-`Type of results*`) %>%
  mutate(Total_valid_votes = `Total number of ballots cast ` - `Rejected ballots `, .keep = "unused")

# Removing full name of political parties with acronyms

CA$`Political affiliation`[CA$`Political affiliation` == "NDP-New Democratic Party"] <- "NDP"
CA$`Political affiliation`[CA$`Political affiliation` == "People's Party - PPC"] <- "PPC"
```

```
CA$`Political affiliation`[CA$`Political affiliation` == "CFF - Canada's Fourth Front"] <- "CFF"

# Ranking candidates by the number of votes they recieved and assigning a variable to winners and loser


CA <- CA %>%
  group_by(`Electoral district name`) %>%
  mutate(Rank = rank(-`Votes obtained `)) %>%
  mutate(Winner = if_else(Rank == 1, "Winner", "Loser"))

### If I want to update % Votes obtained variable I'd put in %>% mutate(`% Votes obtained ` = (`Votes o

CA$Rank <- as_factor(CA$Rank)
CA$`Electoral district number ` <- as.double(CA$`Electoral district number `)

CA <- CA %>%
  rename(`Electoral district number` = `Electoral district number `,
         `Votes obtained` = `Votes obtained `,
         `Votes obtained %` = `% Votes obtained `)

# Minor adjustments to help with data manipulation in the future
```

## Updating election results as if the PPC never existed and their voters instead voted for the Conservatives

```
CA1 <- CA %>%
  group_by(`Electoral district name`) %>%
  mutate(New_PA = recode(`Political affiliation`, PPC = 'Conservative')) %>%
  group_by(`Electoral district name`, New_PA) %>%
  mutate(`New_PA` = str_c(`New_PA`),
            Votes_Obtained = sum(`Votes obtained`)) %>%
  filter(`Political affiliation` != "PPC") %>%
  select(- `Political affiliation`, - `Votes obtained`, - `Votes obtained %`, -`Total_valid_votes`) %>%
  # Once unneeded variables are removed and the PPC is filtered out, I added in the rank and Winner var
  group_by(`Electoral district name`) %>%
  mutate(Rank = rank(-Votes_Obtained)) %>%
  mutate(Winner = if_else(Rank == 1, "Winner", "Loser"))

CA1$Rank <- as_factor(CA1$Rank)

head(CA1)
```

```
## # A tibble: 6 x 7
## # Groups:   Electoral district name [2]
##   `Electoral dist~ `Electoral dist~ Full_name Rank  Winner New_PA Votes_Obtained
##              <dbl> <chr>            <chr>     <fct> <chr>  <chr>           <dbl>
## 1            10001 Avalon           Matthew ~ 2     Loser  Conse~          13385
## 2            10001 Avalon           Carolyn ~ 3     Loser  NDP              5151
## 3            10001 Avalon           Ken McDo~ 1     Winner Liber~          18608
## 4            10002 Bonavista--Buri~ Anne Mar~ 3     Loser  NDP              2484
## 5            10002 Bonavista--Buri~ Churence~ 1     Winner Liber~          13972
```

| Political affiliation | MP's |
|---|---|
| Bloc Québécois | 32 |
| Conservative | 141 |
| Green Party | 2 |
| Liberal | 144 |
| NDP | 19 |

| Political affiliation | MP's |
|---|---|
| Bloc Québécois | 33 |
| Conservative | 119 |
| Green Party | 2 |
| Liberal | 159 |
| NDP | 25 |

```
## 6           10002 Bonavista--Buri~ Sharon V~ 2     Loser  Conse~          13535
```

# Exploratory Data Analysis

## Comparing parliamentary results of CA (actual) and CA1 (result of all PPC voters going to the Conservatives)

```
CA1 %>%
  group_by(New_PA) %>%
  count(Winner) %>%
  filter(Winner == "Winner") %>%
  select(-Winner) %>%
  rename(c("MP's" = n, "Political affiliation" = New_PA)) %>%
  kable() %>%
  kable_styling(latex_options = "striped")
```

```
# Counting all instances of winner for a party and renaming that variable to MP's

CA %>%
  group_by(`Political affiliation`) %>%
  count(Winner) %>%
  filter(Winner == "Winner") %>%
  select(-Winner) %>%
  rename(c("MP's" = n))%>%
  kable() %>%
  kable_styling(latex_options = "striped")
```

```
# There is a discrepancy between this result and the current Canadian parliament due to two recounts. B
```

The two tables (and maps included in the report) indicate that the PPC may have been somewhat of a spoiler party. If 100% of PPC voters had voted for the Conservatives instead of the PPC, then the Conservatives would have had 141 instead of 119 Members of Parliament (MP's), an increase of 22 MP's while the NDP would have lost 6 MP's and the Liberals 15 MP's. This result may have shaped future Canadian politics. While both simulated and actual outcomes put the Liberals outside of an outright majority, 170 MP's, the

Liberals would have had much more difficulty governing as a minority government if they had 144 opposed to 160 (it is 160 for reasons explained above) MP's, forcing the Liberals to coalesce with the NDP or BQ, meaning that members from those parties would likely need to be included in Trudeau's cabinet and making the Liberal's compromise on some of their platform policies. This may have have lead to a more fractious government as the government and opposition would have had a similar amount of MP's, forcing the Liberals to negotiate more often with all the other parties to pass legislation. However, it is unlikely that 100% of PPC members would have voted for the Conservatives if the PPC had not existed. The PPC gained steam by being an anti-establishment party. Many of its voters would just not have voted or voted for an insignificant fringe party if there was no alternative to the Conservatives. Many PPC voters were first time voters.

Thus, I created a function to simulate different possible outcomes for the distribution of votes by the PPC.

## Creating dataset to simulate PPC vote migrating to no one and/or different parties

```
CA2 <- CA %>%
  mutate(PPC_Vote = if_else(`Political affiliation` == "PPC", `Votes obtained`, 0))

# Duplicating PPC Vote into another column for further data manipulation

CA2$PPC_Vote[CA2$PPC_Vote == 0] <- NA

# Turned all 0's into NA so I can use the fill function

CA2 <- CA2 %>%
  group_by(`Electoral district name`) %>%
  fill(PPC_Vote, .direction = "updown") %>%
  mutate_all(~ replace(., is.na(.), 0))

# fill function adds the PPC Vote to all the other columns and then NA's are turned back to 0's
```

## Function allows users to put in the amount of votes Canadian political parties would have recieved if the PPC had not run

```
No_PPC <- function(Conservative, NDP, Liberal, Bloc_Quebecois){
  CA2 <- CA2 %>%
    mutate(`Votes obtained` = if_else(`Political affiliation` == "Conservative", `Votes obtained` + Cons
           `Votes obtained` = if_else(`Political affiliation` == "NDP", `Votes obtained` + NDP * PPC_Vo
           `Votes obtained` = if_else(`Political affiliation` == "Liberal", `Votes obtained` + Liberal =
           `Votes obtained` = if_else(`Political affiliation` == "  Bloc Québécois", `Votes obtained` +
    # Only major parties have the mutate ability since lesser parties do not have th clout to influence
    filter(`Political affiliation` != "PPC") %>%
    select(-PPC_Vote) %>%
    group_by(`Electoral district name`) %>%
    mutate(Rank = rank(-`Votes obtained`),
           Winner = if_else(Rank == 1, "Winner", "Loser"),
           `Votes obtained %` = round(`Votes obtained`/ Total_valid_votes * 100, digits = 2))
  CA2$`Votes obtained` <- round(CA2$`Votes obtained`)
  CA2
}
```

| Political affiliation | MP's |
|---|---|
| Bloc Québécois | 32 |
| Conservative | 141 |
| Green Party | 2 |
| Liberal | 144 |
| NDP | 19 |

**Example of Function usage**

```
j <- No_PPC(1, 0, 0, 0)
head(j)
```

```
## # A tibble: 6 x 9
## # Groups:   Electoral district name [2]
##   `Electoral district number` `Electoral dist~ `Political affi~ `Votes obtained`
##                         <dbl> <chr>            <chr>                        <dbl>
## 1                       10001 Avalon           Conservative                 13385
## 2                       10001 Avalon           NDP                           5151
## 3                       10001 Avalon           Liberal                      18608
## 4                       10002 Bonavista--Buri~ NDP                           2484
## 5                       10002 Bonavista--Buri~ Liberal                      13972
## 6                       10002 Bonavista--Buri~ Conservative                 13535
## # ... with 5 more variables: Votes obtained % <dbl>, Full_name <chr>,
## #   Total_valid_votes <dbl>, Rank <dbl>, Winner <chr>
```

## Creating a function to calculate number of MP's based on simulated election results

```
calc_mp <- function(dataset) {
  df <- dataset %>%
  group_by(`Political affiliation`) %>%
  count(Winner) %>%
  filter(Winner == "Winner") %>%
  select(-Winner) %>%
  rename("MP's" = n)
  df
}
```

**Example of Function usage**

```
calc_mp(No_PPC(1, 0, 0, 0))   %>%
  kable() %>%
  kable_styling(latex_options = "striped")
```

| Political affiliation | MP's |
|---|---|
| Bloc Québécois | 32 |
| Conservative | 127 |
| Green Party | 2 |
| Liberal | 153 |
| NDP | 24 |

### Comparing parliament results of CA (actual) and CA2 (Simulated results of election had PPC not been formed) using the No_PPC and calc_mp functions

According to an analysis of voter data from the 2019 Canadian parliamentary election and polling on the 2021 election, the Canadian Broadcasting Corporation (CBC) estimated that 59% of the PPC are former conservative voters, 24% voted for the PPC, 6% voted for the Liberals, 5% for the NDP, 4% for the Bloc, and 2% for the Greens. Let's use these numbers to simulate what the Canadian election may have looked like if the PPC did not exist. While this is not a perfect heuristic, people often change their politics, this data is supported by anecdotal and crosstab data from other pundits, analysts, and pollsters.

```
calc_mp(No_PPC(0.59, 0.05, 0.06, 0.04)) %>%
  kable() %>%
  kable_styling(latex_options = "striped")
```

This result indicates that the PPC did spoil some elections for the Conservatives. Compared to the actual results, the Conservatives gained +8 seats, the NDP lost -1 seats, and the Liberals lost -6 seats. This may have resulted in some media coverage indicating that the Conservatives were small victors of the election, but it would have likely had a small impact.

## Visualizing the actual and simulated results of the Canadian Election

This section, especially the maps exists to visually reinforce what was being demonstrated earlier in the report. The bar plots indicate how votes were distributed proportionately and how they translated into electoral gain.

#### Manipulating voter data to merge with geospatial data

```
CA <- read_tsv("EventResults.txt", col_names = TRUE, skip = 1)

CA <- CA %>%
  rename_with(~ str_remove(.x, "-.*")) %>%
  mutate(Full_name = str_c(`Given name `, `Surname `, sep = " "), .keep = "unused") %>%
  select(-`Nom de la circonscription`, -`Type de résultats**`, -`Middle name(s) `, -`Appartenance politi
  filter(`Type of results*` == "validated") %>%
  select(-`Type of results*`) %>%
  mutate(Total_valid_votes = `Total number of ballots cast ` - `Rejected ballots `, .keep = "unused")

# Replacing full name of political parties with acronyms

CA$`Political affiliation`[CA$`Political affiliation` == "NDP-New Democratic Party"] <- "NDP"
CA$`Political affiliation`[CA$`Political affiliation` == "People's Party - PPC"] <- "PPC"
```

```
CA$`Political affiliation`[CA$`Political affiliation` == "CFF - Canada's Fourth Front"] <- "CFF"

# Ranking candidates by the number of votes they recieved and assigning a variable to winners and loser


CA <- CA %>%
  group_by(`Electoral district name`) %>%
  mutate(Rank = rank(-`Votes obtained `)) %>%
  mutate(Winner = if_else(Rank == 1, "Winner", "Loser"))

### If I want to update % Votes obtained variable I'd put in %>% mutate(`% Votes obtained ` = (`Votes o

CA$Rank <- as_factor(CA$Rank)
CA$`Electoral district number ` <- as.double(CA$`Electoral district number `)

CA <- CA %>%
  rename(`Electoral district number` = `Electoral district number `,
         `Votes obtained` = `Votes obtained `,
         `Votes obtained %` = `% Votes obtained `)

# Minor adjustments to help with data manipulation in the future
```

### Merging geospatial and voter data

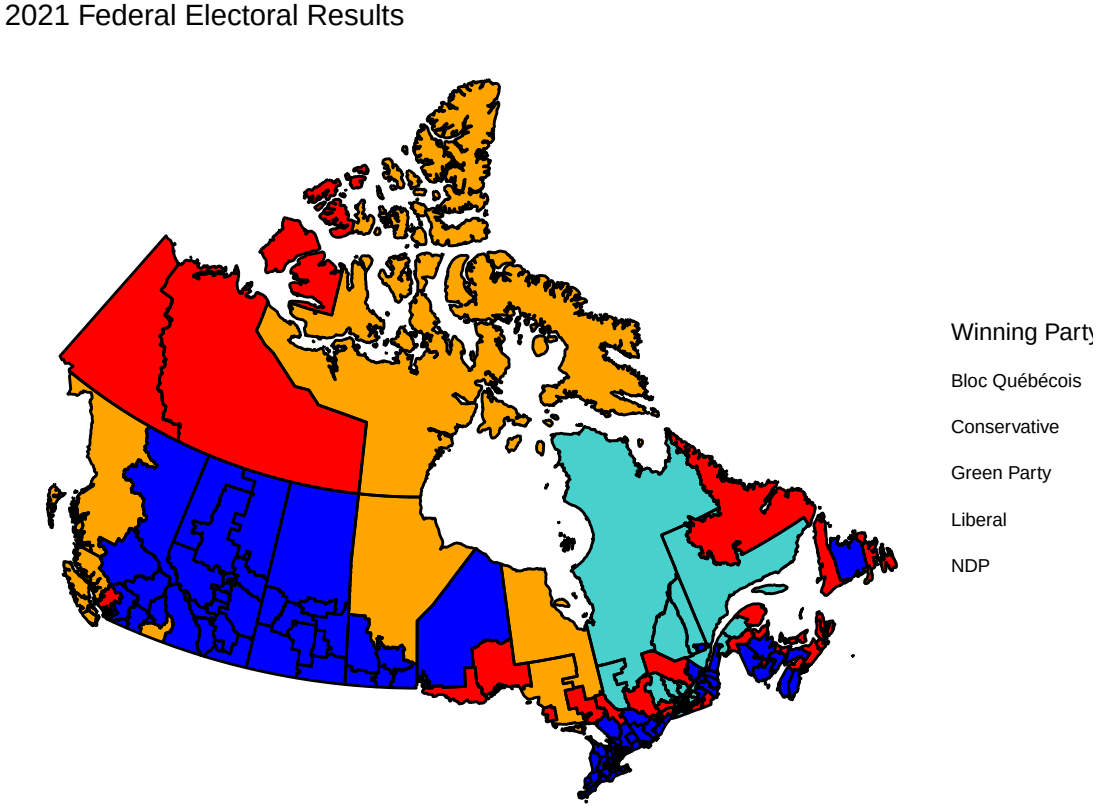
```
#read in geospatial data
pr_geographic <- mapcan(boundaries = ridings, type = standard)


#Fortify based on riding name
riding.points <- fortify(pr_geographic, region = "riding_code")
riding.points$riding_code <- toupper(riding.points$riding_code)

#Join geospatial and voter data
merged <- merge(riding.points, CA, by.x="riding_code", by.y="Electoral district number", all.x=TRUE) %>%
  select(-`riding_name_french`, -`pr_french`) %>%
  group_by(`riding_name_english`) %>%
  mutate(Winning_Party = case_when(`Winner` == "Winner" ~ `Political affiliation`)) %>%
  fill(Winning_Party, .direction = "updown")
```

## Plotting actual and simulated election results

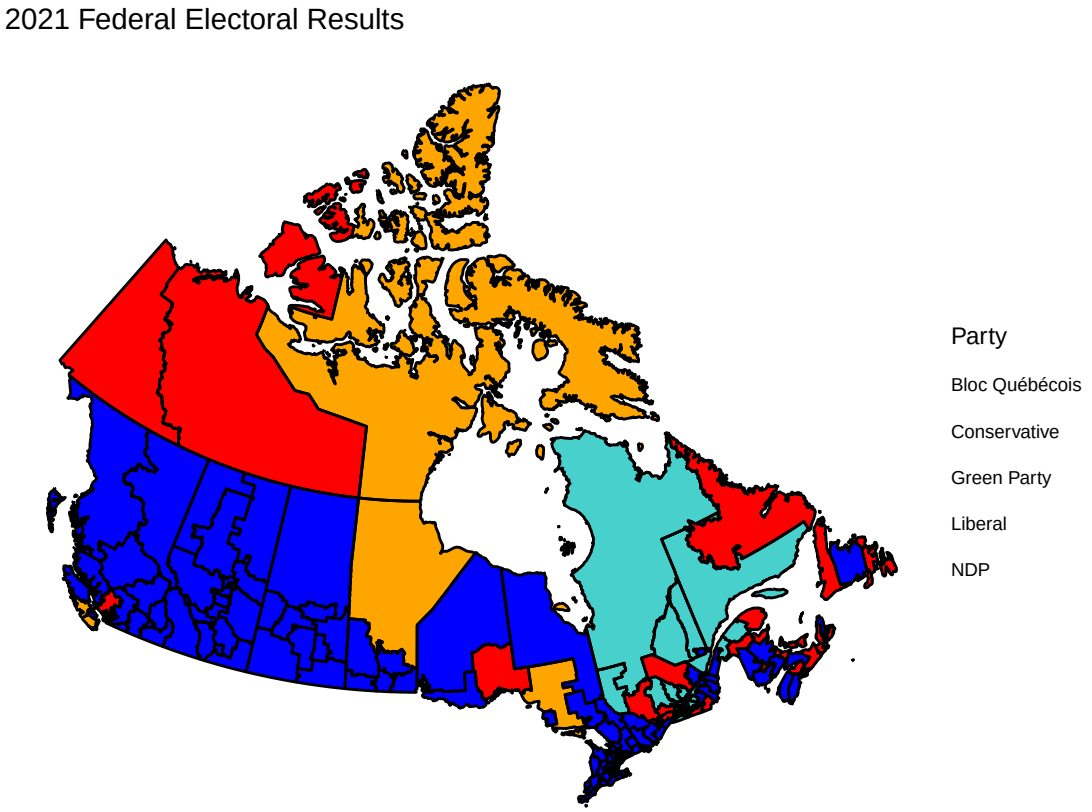
### Plotting actual results

```
ggplot() +
  # riding polygons
  geom_polygon(data = merged, aes(fill = Winning_Party,
                                  x = long,
                                  y = lat,
                                  group = group)) +
```

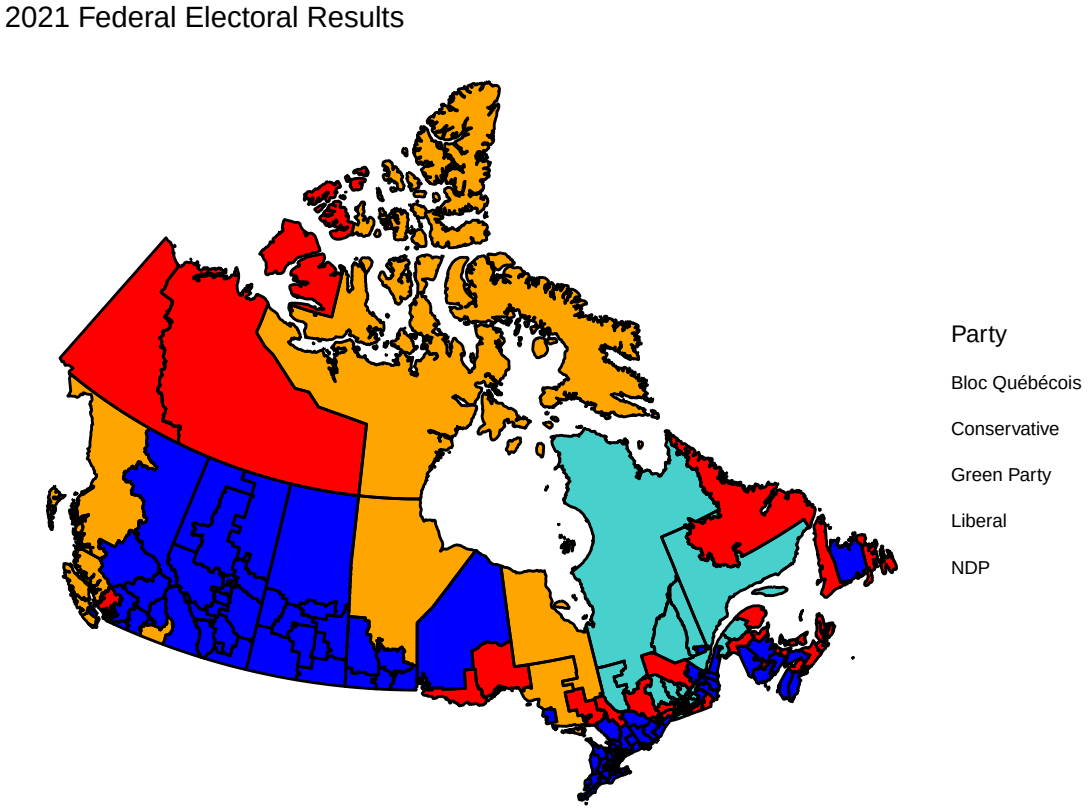
```r
# riding outline
geom_path(data = merged, aes(x = long, y = lat, group = group),
          color = "black", size = 0.5) +
coord_equal() +
# add the previously defined basic theme
theme_map() +
labs(x = NULL, y = NULL, fill = "Party") +
scale_fill_manual(name = "Winning Party",
                  values = c("mediumturquoise", "blue", "springgreen3", "red", "orange"),
                  guide = guide_legend(direction = "vertical",
                                       title.hjust = 0,
                                       title.vjust = 1,
                                       barheight = 30,
                                       label.position = "left",
                                       reverse = F,
                                       label.hjust = 0)) +
theme(legend.position = c(1, 0.3)) +
ggtitle("2021 Federal Electoral Results")
```



**Plotting map of Canadian election with simulated results created by the No_PPC function. This example assumes 100% of the PPC vote went to the Conservatives**



**Plotting map of election results according to CBC model of PPC voter distribution**



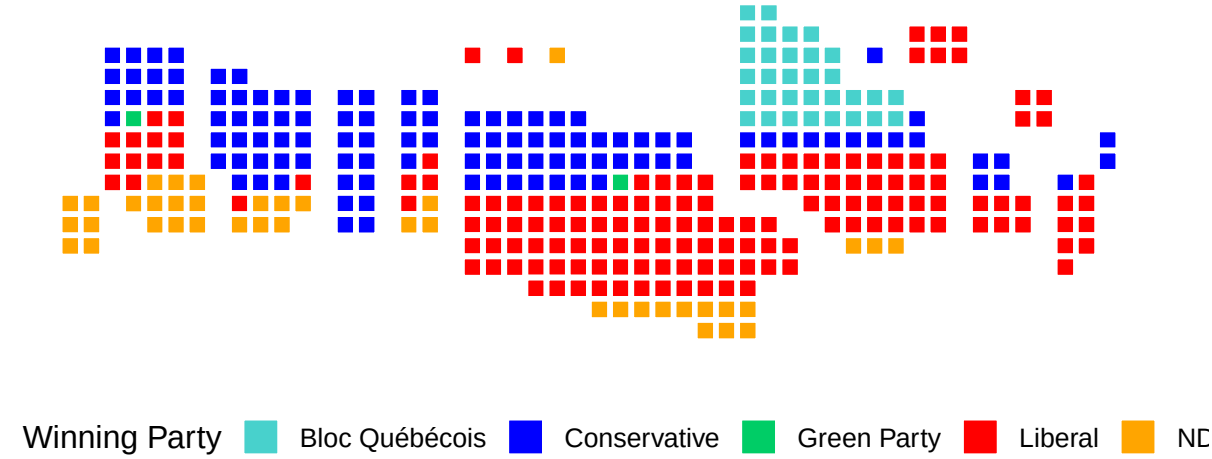

## Plotting tile maps of the Canadian election

```
pr_tile <- inner_join(federal_riding_bins, CA2, by = c("riding_code" = "Electoral district number"), ke
  select(-`riding_code`, -`pr_french`, -`pr_sgc_code`, -`representation_order`)
```

**Actual Result**

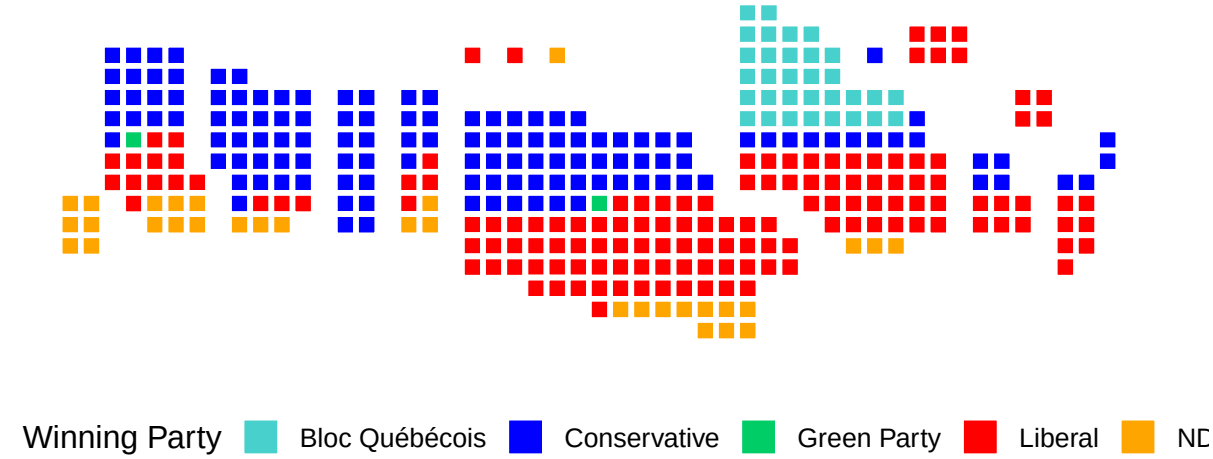
```
pr_tile %>%
  filter(Winner == "Winner") %>%
  riding_binplot(riding_col = `Electoral district number`,
                 value_col = `Political affiliation`,
                 arrange = TRUE,
                 continuous = FALSE) +
  scale_fill_manual(name = "Winning Party",
                    values = c("mediumturquoise", "blue", "springgreen3", "red", "orange")) +
  theme_mapcan() +
  ggtitle("Tile grid map of 2021 federal election results")
```

```
## Scale for 'fill' is already present. Adding another scale for 'fill', which
## will replace the existing scale.
```



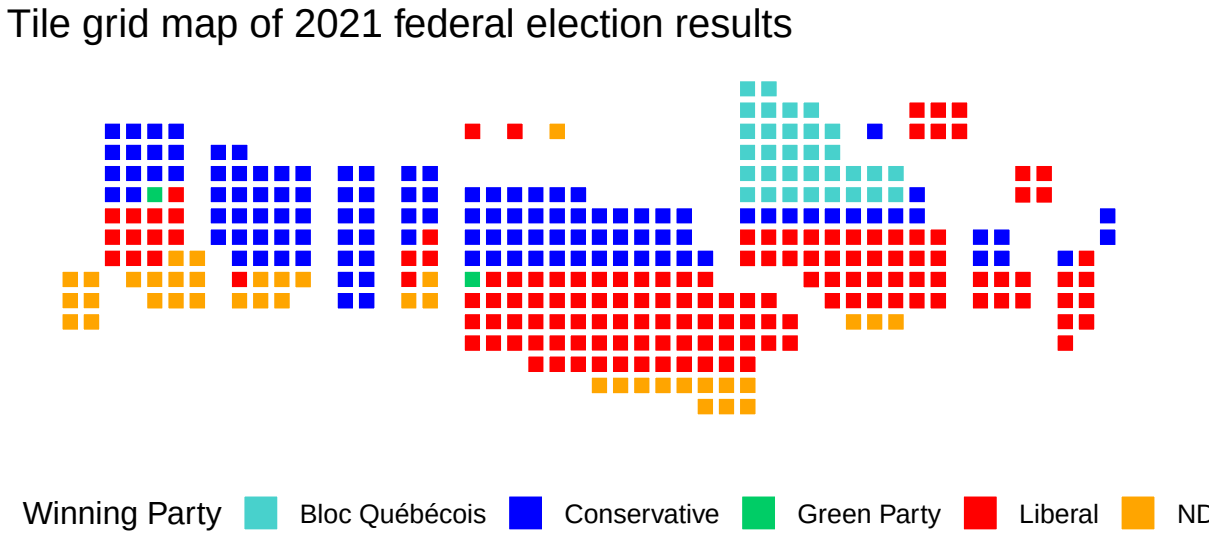

**Simulated Result: Conservatives gain entire PPC vote**

```
## Scale for 'fill' is already present. Adding another scale for 'fill', which
## will replace the existing scale.
```




**Simulated Result: Canadian parties gain PPC vote according to CBC/2019 voter data**

```
## Scale for 'fill' is already present. Adding another scale for 'fill', which
## will replace the existing scale.
```

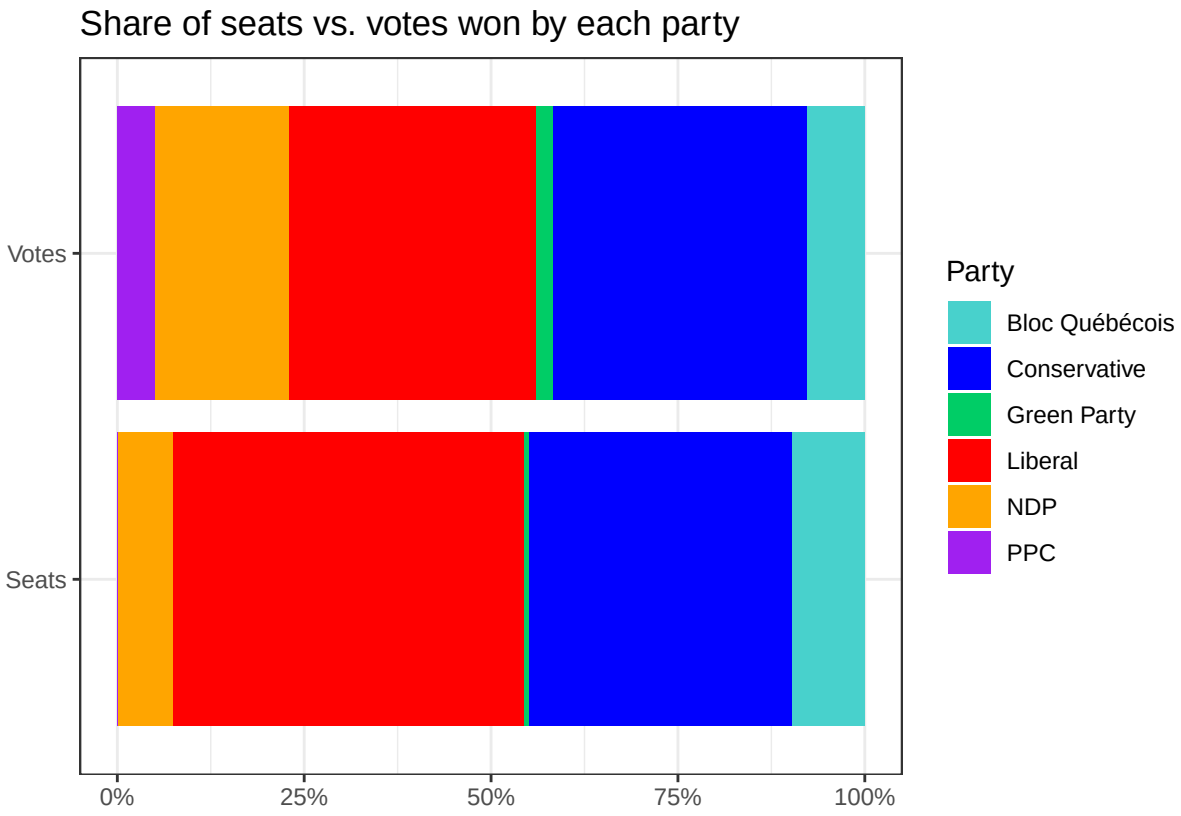



## Raw vote total vs MPs elected: Actual vs Simulated

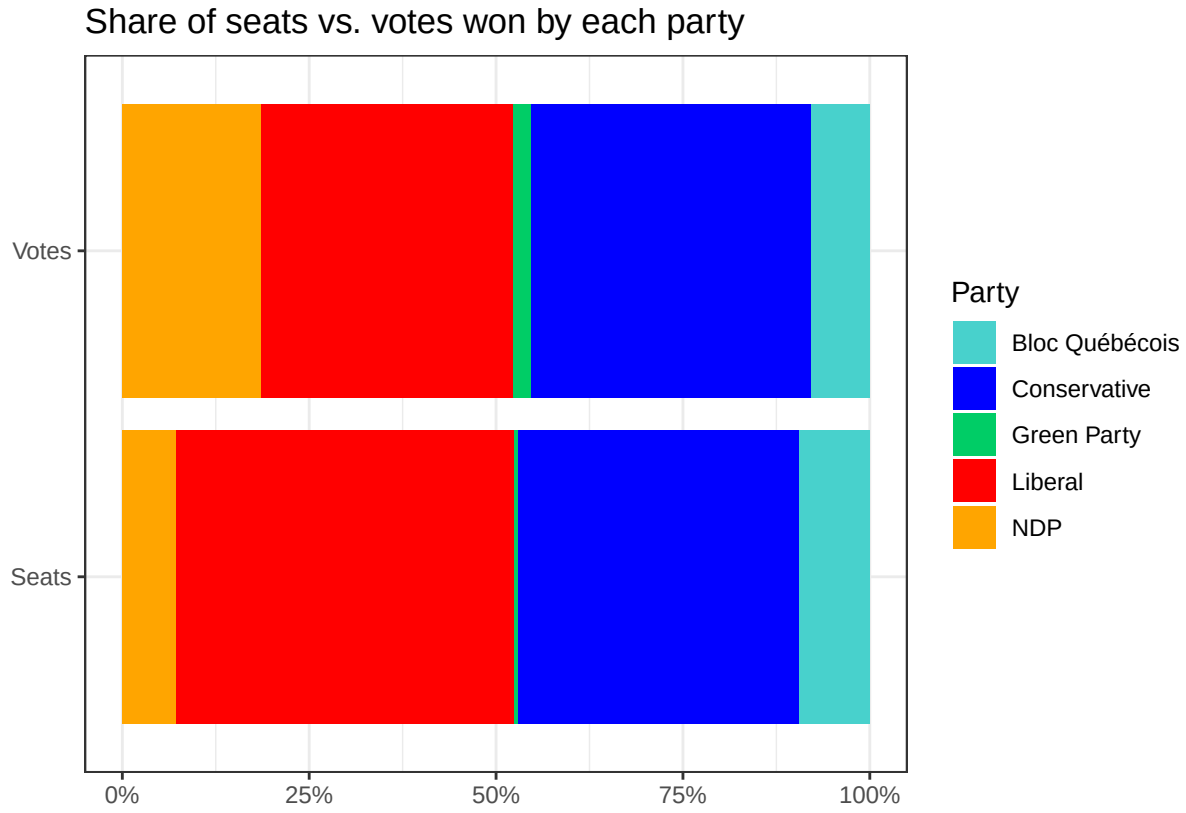
### Actual results

```
totals <- CA %>%
  group_by(`Political affiliation`) %>%
  summarise(Votes1 = sum(`Votes obtained`), `Seats` = sum(str_count(Winner, "Winner"))) %>%
  mutate(`Votes` = round(Votes1/sum(Votes1)*100, digits = 2)) %>%
  filter(`Seats` > 0 | Votes1 > 50000) %>%
  select(-Votes1) %>%
  pivot_longer(!`Political affiliation`, names_to = "x", values_to = "values")
```

```
totals %>%
  ggplot(aes(x, values, fill = `Political affiliation`)) +
  geom_col(position = "fill") +
  coord_flip() +
  scale_fill_manual(name = "Party",
                    values = c("mediumturquoise", "blue", "springgreen3", "red", "orange", "purple")) +
  scale_y_continuous(labels = scales::percent) +
  labs(x = "", y = "", title = "Share of seats vs. votes won by each party") +
  theme_bw()
```


Share of seats vs. votes won by each party
Votes
Seats
0%
25%
50%
75%
100%
Party
Bloc Québécois
Conservative
Green Party
Liberal
NDP
PPC

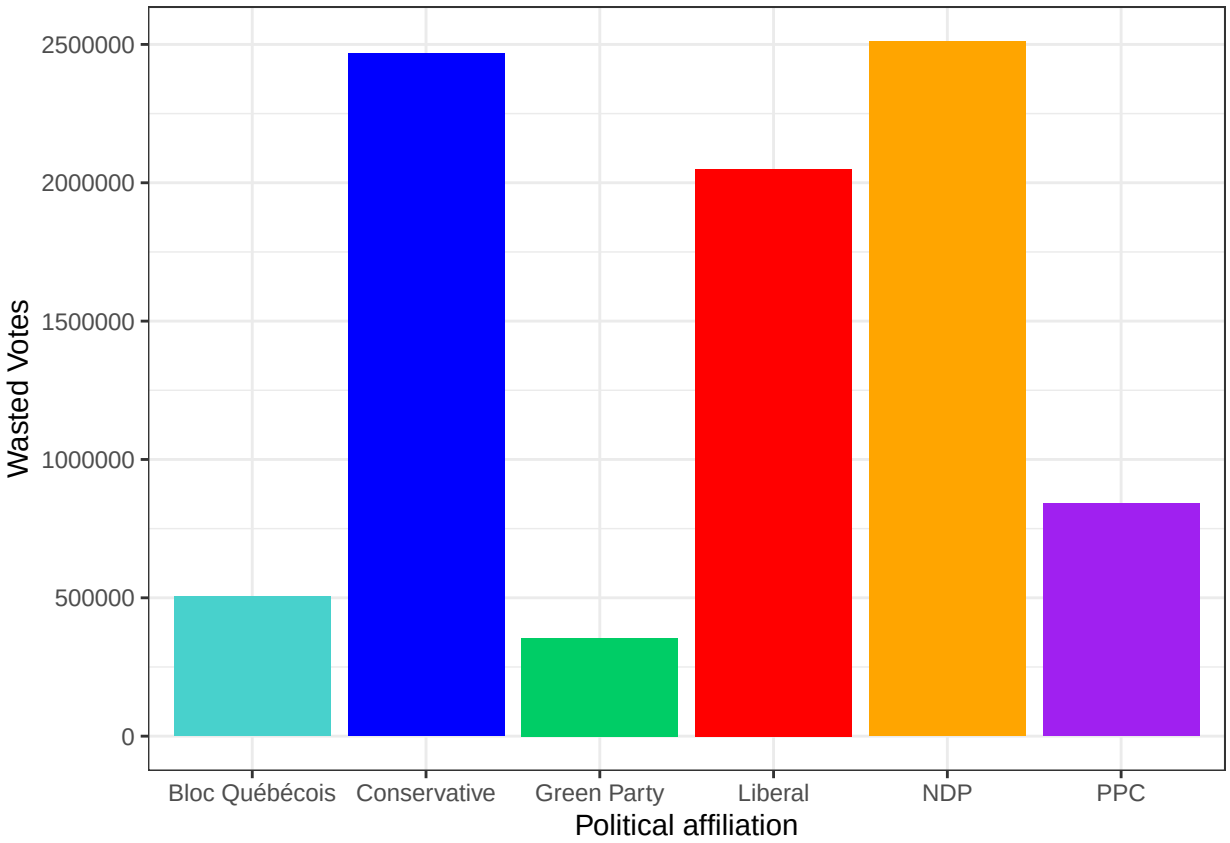
**Simulated according to CBC estimates**



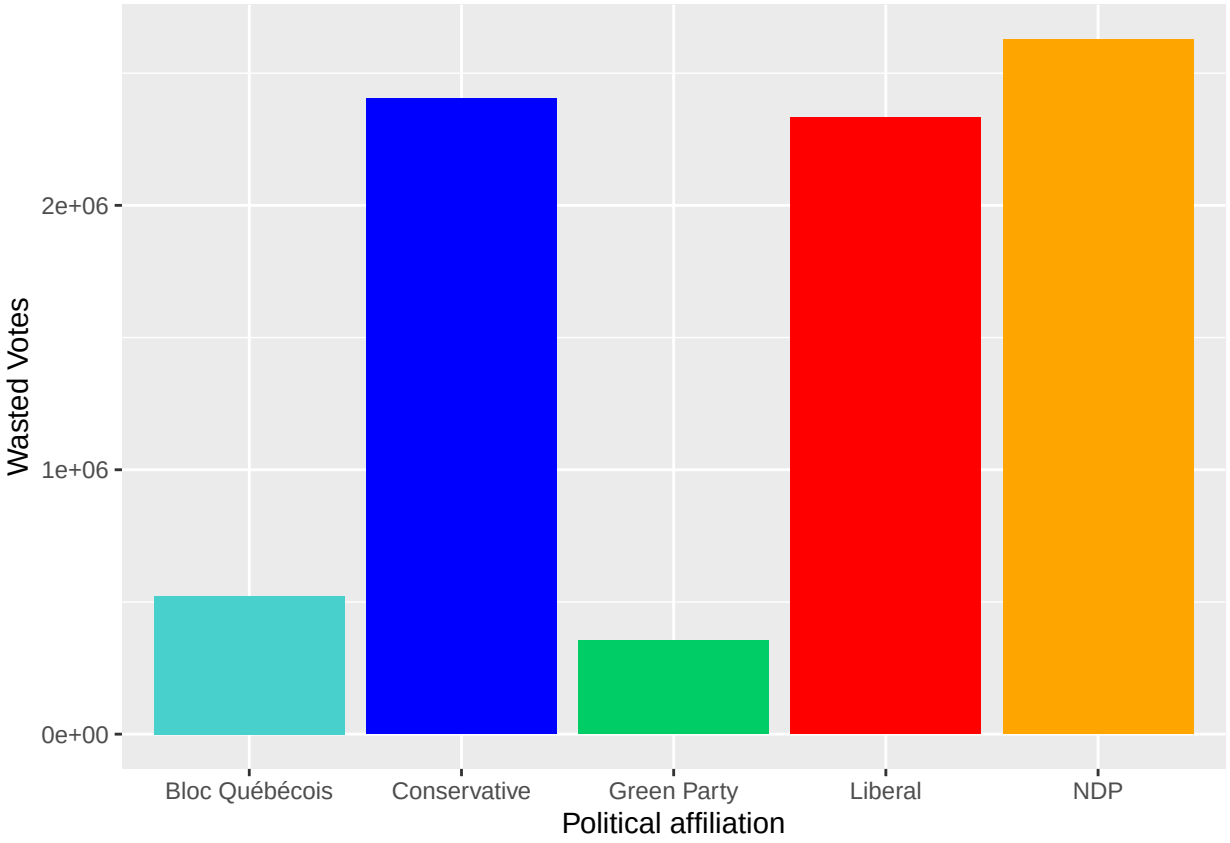

These graphs indicate that Canadian politics are not as representative as they can be. The Liberal and Bloc parties are over represented while the PPC, NDP, and Conservative parties are under represented in the parliament. Without the PPC vote, not much changes. This disparity between votes and seats is likely a product of Canada's first past the post system. It is also likely a byproduct of inefficient votes. This graph raises interesting questions about the Canadian electoral system overall. If Canada adopts a proportional representation system, it would likely be more ideologically diverse, with both the far right PPC and progressive NDP and Green parties having more representation. It also raises the question about voter efficiency. The Labor and BQ parties had relatively more efficient votes because they were able to get more members into parliament than in a proportional system. The following section examines voter inefficiency and the PPC's impact on this aspect of Canada's electoral system.

## Finding wasted votes

```
CA %>%
  mutate(`Wasted Votes` = ifelse(Winner == "Winner", 0, `Votes obtained`)) %>%
  group_by(`Political affiliation`) %>%
  summarise(`Wasted Votes` = sum(`Wasted Votes`)) %>%
  filter(`Wasted Votes` > 50000) %>%
  ggplot(aes(`Political affiliation`, `Wasted Votes`)) +
  geom_bar(stat = "identity", fill = c("mediumturquoise", "blue", "springgreen3", "red", "orange", "pur
  theme_bw()
```




```
No_PPC(1, 0, 0, 0) %>%
  mutate(`Wasted Votes` = ifelse(Winner == "Winner", 0, `Votes obtained`)) %>%
  group_by(`Political affiliation`) %>%
  summarise(`Wasted Votes` = sum(`Wasted Votes`)) %>%
  filter(`Wasted Votes` > 50000) %>%
  ggplot(aes(`Political affiliation`, `Wasted Votes`)) +
  geom_bar(stat = "identity", fill = c("mediumturquoise", "blue", "springgreen3", "red", "orange"))
```


Wasted Votes
2e+06
1e+06
0e+00
Bloc Québécois
Conservative
Green Party
Liberal
NDP
Political affiliation