

Data Visualisation: Visualising the 2020 election results

By: Prakhar Gurawa

Introduction:

We will analyze and visualize the different aspects of the Irish voting system which uses a form of proportional representation (PR) called the single transferable votes. The voters indicate their preferences for the candidates in ascending order of preference where 1 is treated as the most preferable candidate than 2. So for our analysis, we will be considering only the votes with count 1 unless required for any other kind of analysis.

About the dataset:

The dataset is primarily divided into two parts, one solely for Galway west election and the others for overall national elections. Both regions have data for the years 2016 and 2020. The 2016 Galway West election data has been found from the Oireachtas website [1] which exists in CSV format. On the other hand, 2020 Galway West election data has been found from Galway West Fine Gael Part officials which exists in XLSX format. The overall national vote numbers for each party for 2016 and 2020 have been found on the website of TCD political scientist Micheal Gallagher [2], which has been copied and used as a CSV file for analysis of data.

List of packages used:

There are a number of excellent libraries like Matplotlib, Seaborn (Python visualization libraries) but I have used GGPlot (R's visualization library) which is based on the grammar of graphics which helps us create elegant and visually appealing outputs. Few other R packages used are dplyr, ggrepel, scales, readxl, ggforce, tidyverse, gridExtra, scales, ggthemes, knitr, colorblindr etc.

Visualizing the 2016/2020 election results:

Part 1: The vote per party for Galway west

The data used for this section is for Galway west elections for the years 2016 and 2020. The data for 2016 is in CSV format and required much less preprocessing and data cleaning as compared to 2020 data which exists in XLSX format.

- The 2016 data has columns as follows Constituency name, candidate surname, candidate first name, results, count number, transfers, votes, total votes, candidate ID, and party.
- Filtered the rows with count numbers equal to one as they are the votes that matter.
- Only total votes and party have been used for this task which is later aggregated on the basis of party name to calculate the total votes for a party.
- The plot used for this task is a bar graph as they are very efficient for quantitative data.
- Only a single color for all the parties, as different coloring schemes could have indicated a preference amongst the party.
- Figure 1 clearly shows the independent candidates have been very popular among the locals as they have been hugely supported. Also, the parties such as Fianna Fail and Fine Gael have performed well with over more than 15000 total votes for both.
- The data for 2020 Galway has columns as follows party name, candidate name, fpv, count 1 votes to count13 votes. But this dataset needed some data cleaning as was not in a direct format to process for visualizations. So dropped a few irrelevant columns and rows and overall performed a similar operation as that for 2016's data.

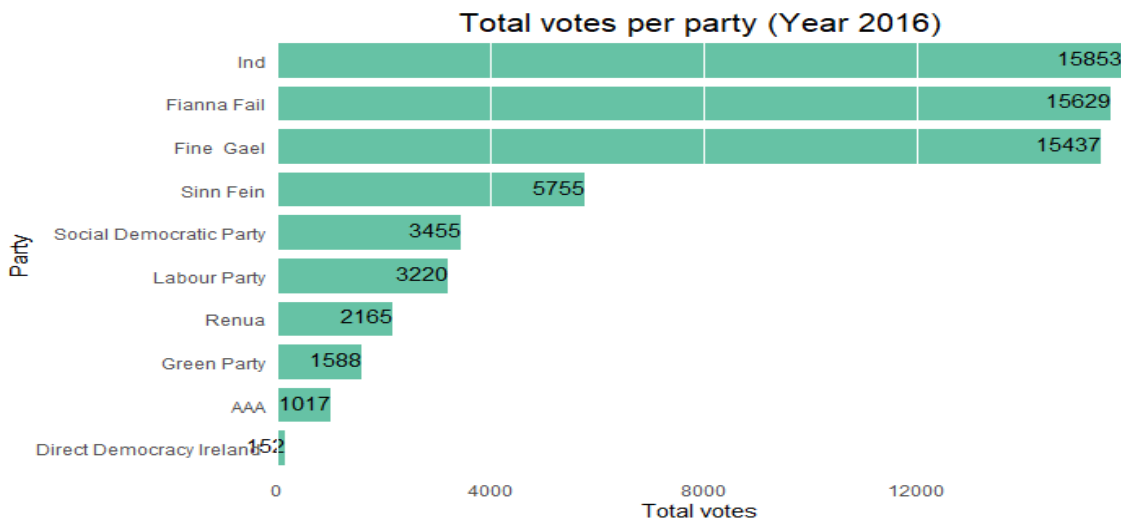


Fig 1. The vote distribution for Galway West elections of 2016.

- The color used for 2020's graph is comparatively darker as compared to 2016 as 2020's election is more significant and appealing for the viewers.

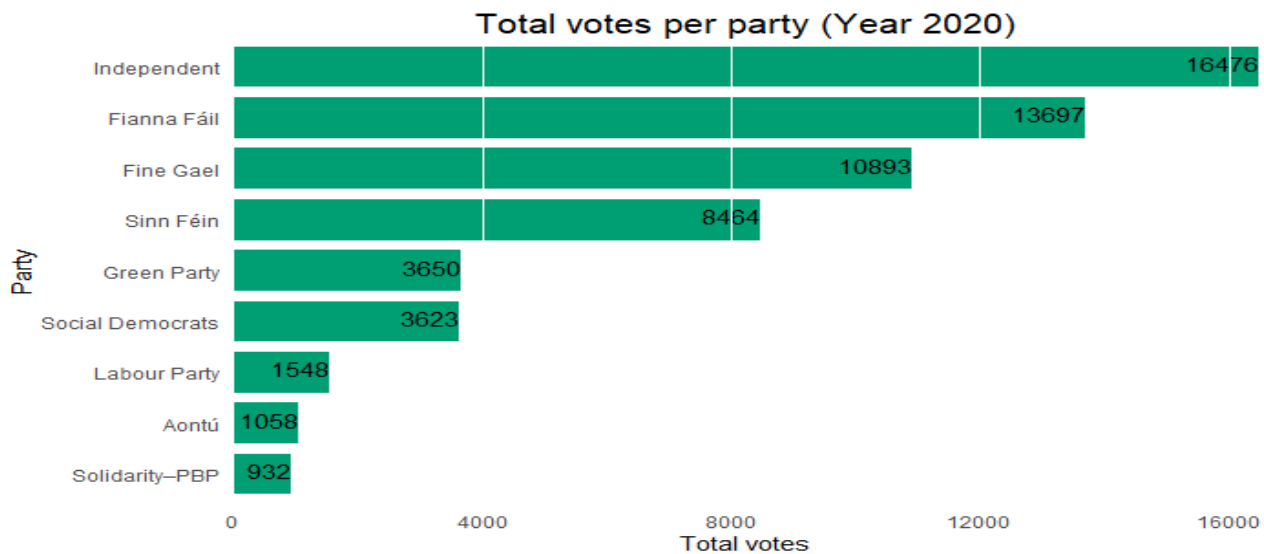


Fig 2. The vote distribution for Galway West elections of 2020.

- 2020's data showed similar patterns as that of 2016's results where independent candidates, Finna Fail and Fine Gael have topped the results.

Part 2: The change of vote per party from 2016-2020 for Galway West

- Side-by-side comparison for both the figures gave a number of interesting insights such as the Labour party has been more secure in the current elections as compared to the last ones where their total votes got doubled. With the reduction of votes for two prominent parties in the election of 2020s, it feels like voters preferring redistributions of power and votes.

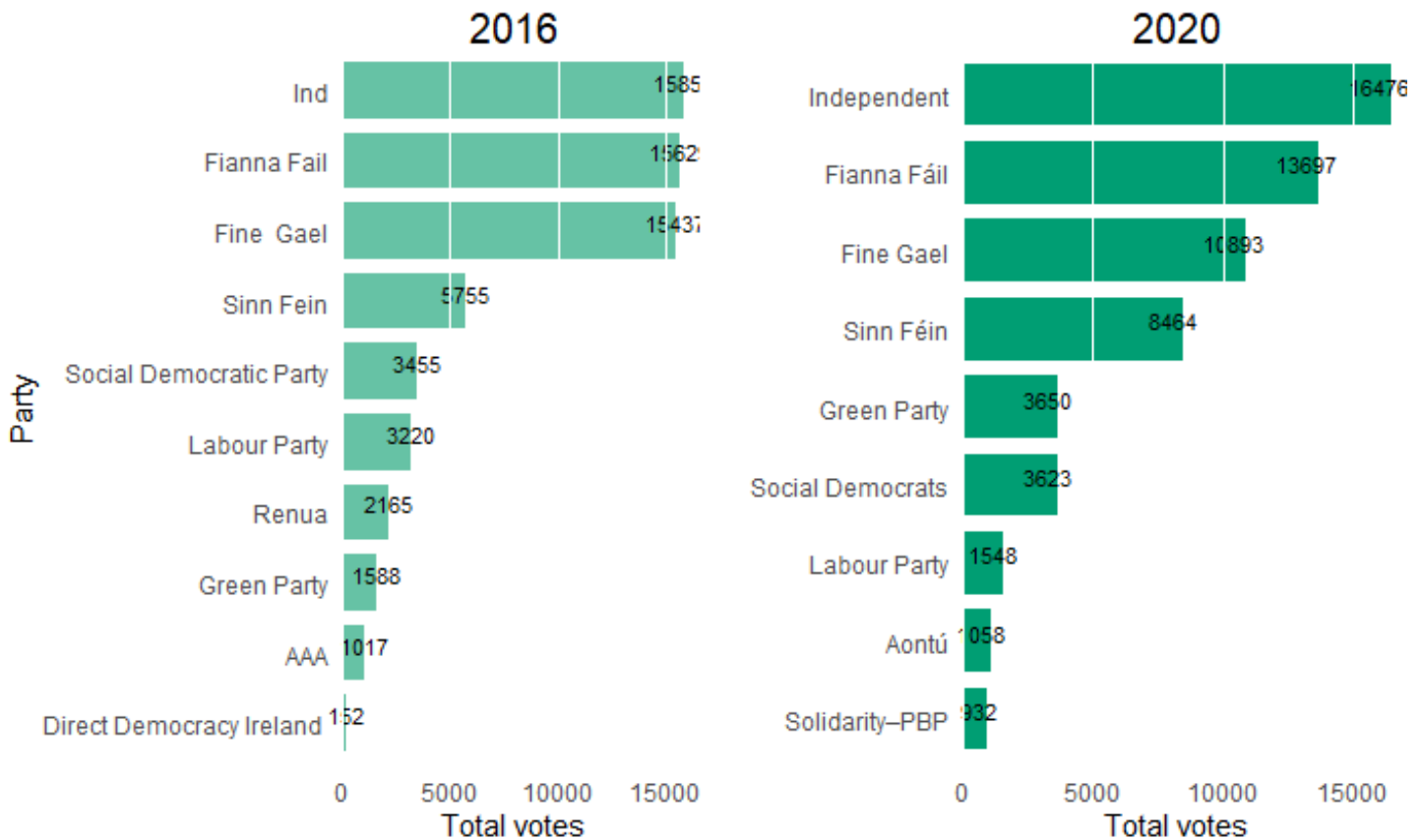


Fig 3. Side by Side comparison of parties Galway West elections of 2020/2016.

- An elegant way to visualize this problem is by using diverging dot plots, which provide much clear and understandable result. To create diverging chart first bound both data frames of the 2016 and 2020 elections to create a single data frame and spread the frame after which calculated the difference in vote for the political parties participating in the 2016 and 2020 elections
- Below figure 4 shows an increment in votes from the 2016s election using green color and decrement in votes using orange color. Also here I have not used color-blind-friendly palettes as color in this and all the above visuals are secondary aesthetics. Even a viewer with color blindness can understand these figures.
- From figure 4 we can point that the Fine Gael party has lost most votes between the two elections whereas Sinn Fein has shown the most improvement with an increment of 2709 votes. Other parties with improvements are Solidarity-PBP, Green Party, Aountu which shows that their PR teams have done a great job to improve their public image. The y axis represents the different parties and the x-axis as the difference in the votes.

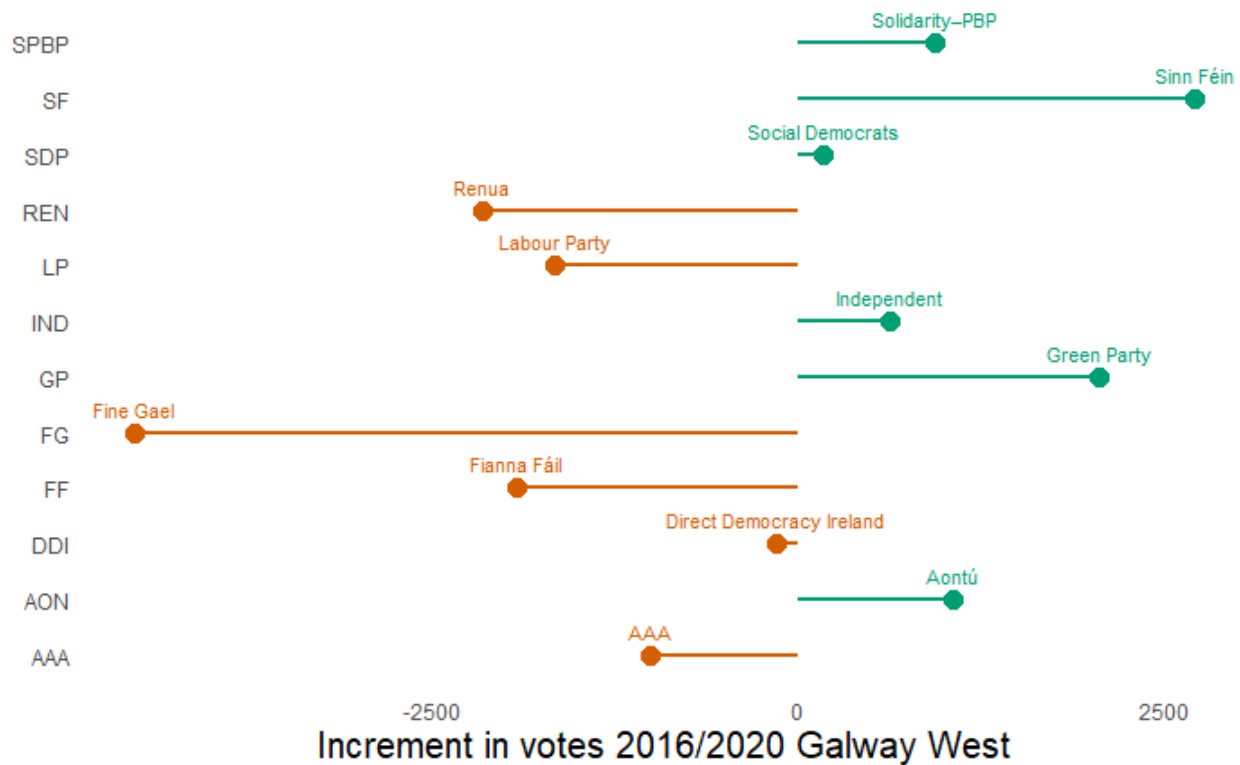


Fig 4. The increment in votes for political parties between the 2016 and 2020 elections

- Another way to visualize the performance of different parties during the two elections is depicted using bar charts. I have used a colorblind-friendly palette, as here color plays a crucial role to create a distinction between the year and convey a huge misconception among the color blind audience if not properly depicted.

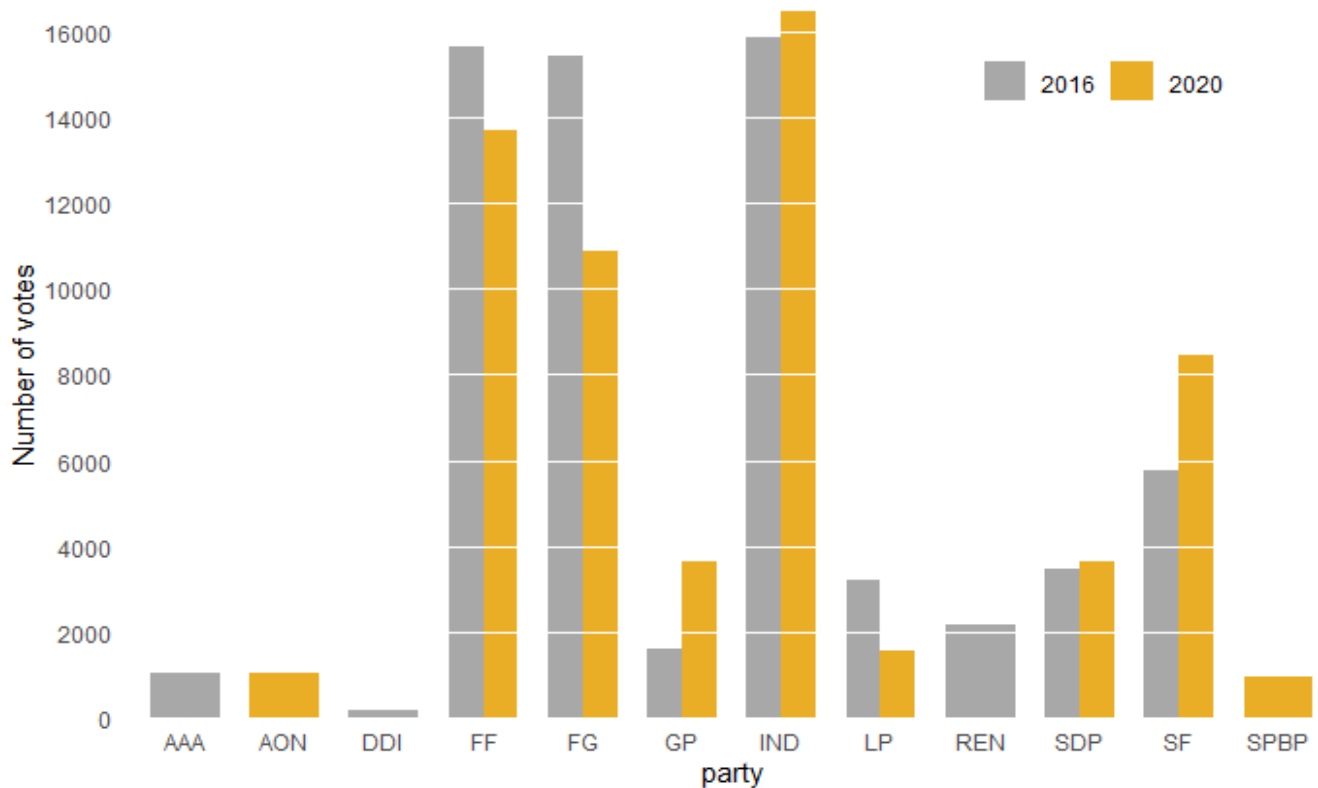
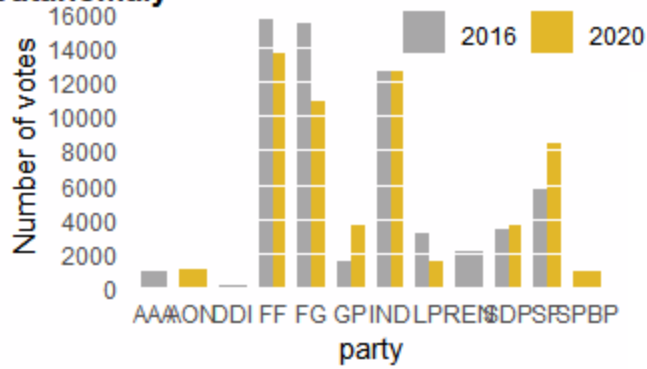
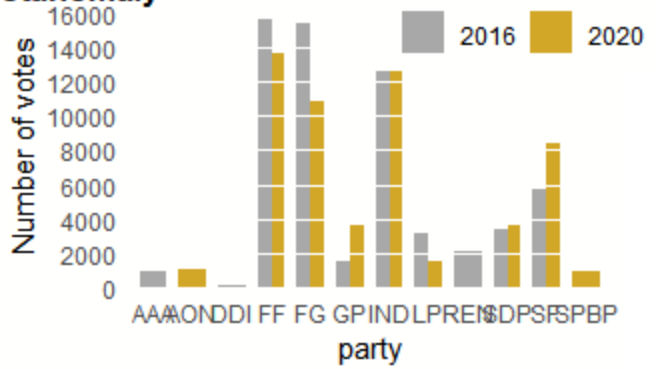


Fig 5. The increment in votes between the 2016 and 2020 elections depicted using a bar chart

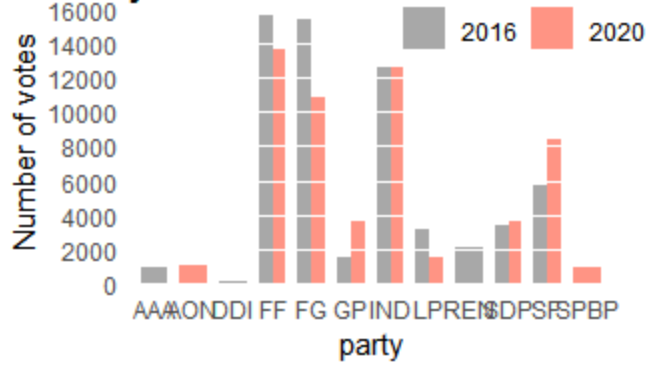
Deutanomaly



Protanomaly



Tritanomaly



Desaturated

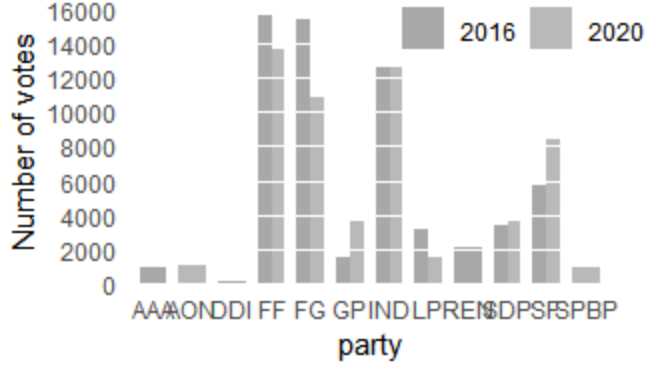


Fig 6. The alternate coloration for readers with CVD for the bar chart of figure 4

- The visuals for figure 6 have been created using the colorblindr library which shows that the coloring scheme is even suitable for a color-blind audience.

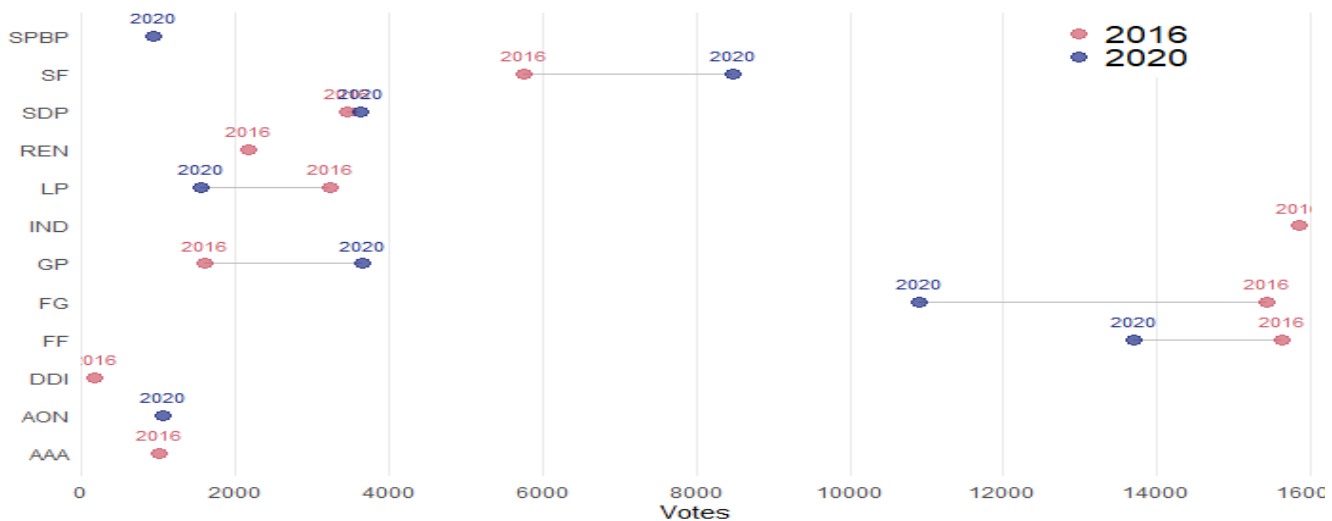


Fig 7. The difference in votes visualized using a point chart for a more lucid understanding

Part 3: Comparison of Galway West to the national average for party share of the vote for 2016/2020

- The dataset for national votes has been collected from TCD political scientist Michael Gallagher's website where we will analyze how a party performs nationally and on the local level (Galway).
- Both the national data of 2020 and 2016 had columns as a party name, the number of candidates, votes, votes percentage, change since last election, seats, change in seats, change in seats since last elections.
- The read data needed some data cleaning such as removing a few last rows and change in names that occurred due to file encoding differences.
- The 2020s Galway data is bonded to 2020s national data to create a combined visual and analysis where both the data are distinguished using an extra column added region which is either Galway or Ireland.
- Finally, the combined data is used to plot a point chart which consists of the party at the y axis and the percentage of the vote in that particular election at the x-axis (or we can say the dominance of a party in an election at the x-axis). Similar operations are done for both years 2016 and 2020.

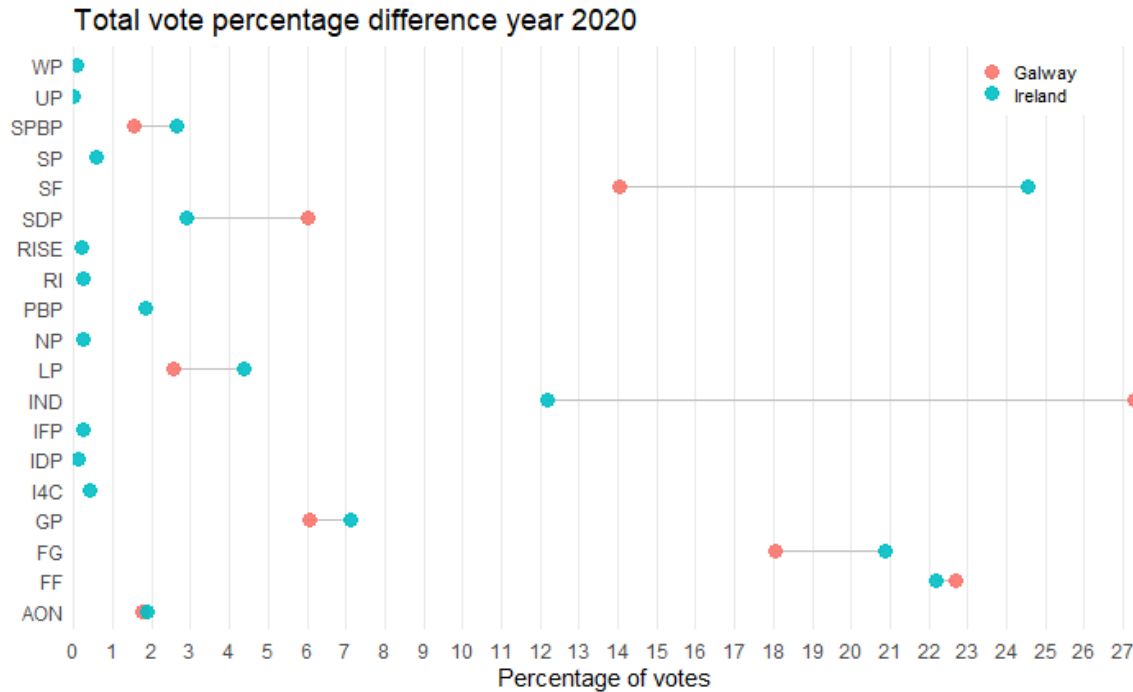


Fig 8. Comparison of performance of political parties for the year 2020 at national and Galway level

- The above figure shows parties like Sinn-Fein, Solidarity–PBP, Fine-Gael, Labour-Party are more powerful at the national labor than at the local level. Also, Galway promotes independent candidates more as compared to that at the national level which is a good sign of neutrality and political awareness among the voters of Galway. The other party that performs well at the Galway level in 2020 is Social-Democrats.
- Similar operations of reading and data cleaning are done to create a comparison of 2016 elections at the national and Galway level. The parties are provided with manual labels to not create confusion and errors will merging as there is a name difference even for the same party in the national and Galway data such as “Fianna Fáil” in Galway 2020 dataset and “Fianna-Fail” in the national 2020 dataset.

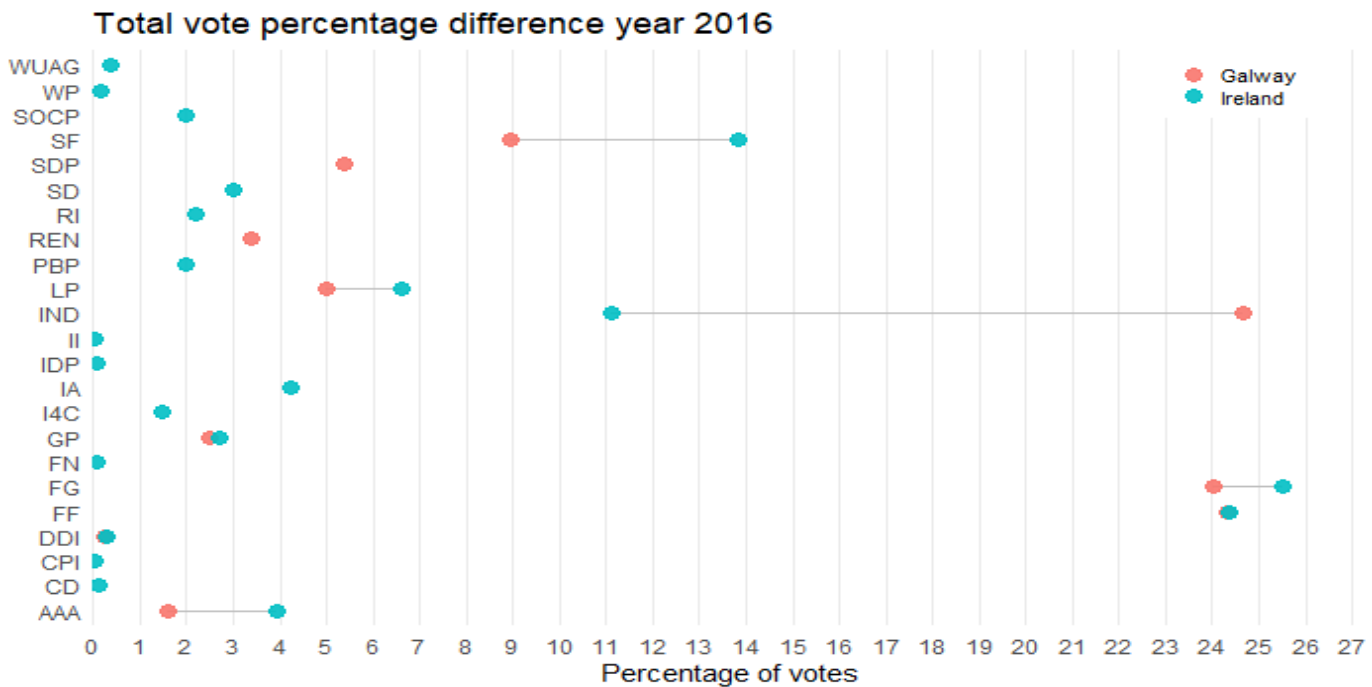


Fig 9. Comparison of performance of political parties for the year 2016 at national and Galway level

- The above figure indicated that parties such as Sinn-Fein, Labour-Party, AAA, and Fine-Gael performed well at the national level as compared. Similar to 2020 the Galway have preferred Independent candidates more as compared to the national level.

Part 4: The change in votes for candidates in both election of Galway West 2016/2020

In this part, we will find which candidates were impacted most in the Galway West region during the two elections of 2016 and 2020. For this part, we have used the same dataset as used in part 1 and considered only the count 1 votes. As in the 2016 dataset, the candidate's first name and last name were a different column, I combined both to create a new column as full name as this will be used as the foreign key between the two datasets of 2016 and 2020.

- Similar data preparation and cleaning operations are performed as that of task 1. Also, a new column is added to both the tables named candidate_short which represents a short notation for every candidate. This column helps us merge the two data frames as the names are inconsistent between the two data frames
- A new column pos is added to the merged data frame which is either true if the vote difference for a candidate is positive or else false if negative (candidate have performed worse in 2020 elections as compared to 2016 elections).
- As color is a secondary aesthetic here I have used a combination of green and red to create a sense of improvement or deterioration for each candidate. Although this combination is highly discouraged as it creates issues for the color blind audience.
- The figure 10 indicates Hildegarde Naughton has seen the best improvement in the number of votes from 2016 to 2020. Whereas Sean Kyne's performance was unsatisfactory and downgraded in 2020 as compared to that of 2016.

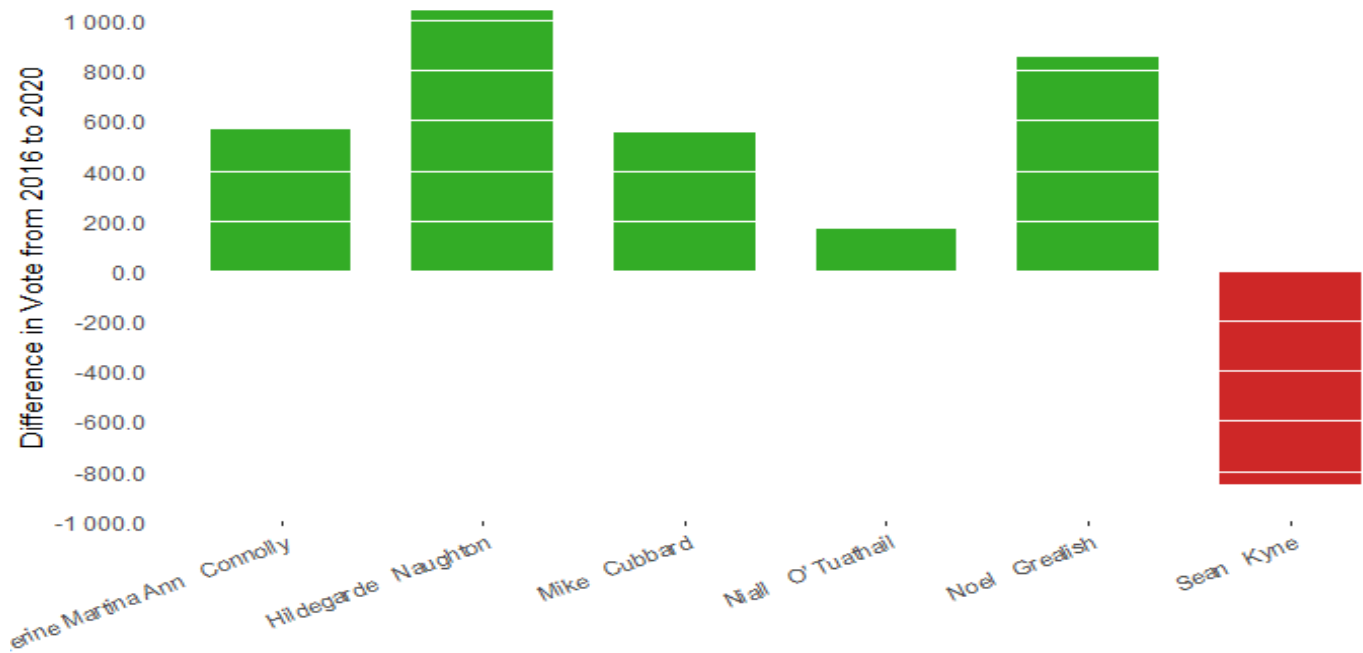


Fig 10. Comparison of performance of political parties candidates for the year 2016/2020

References:

1. Oireachtas website: <https://www.oireachtas.ie/en/elections/>
2. Micheal Gallagher website for 2016 and 2020:
https://www.tcd.ie/Political_Science/people/michael_gallagher/Election2016.php
https://www.tcd.ie/Political_Science/people/michael_gallagher/Election2020.php
3. https://clayford.github.io/dwir/dwr_05_combine_merge_rehsape_data.html
4. <https://uc-r.github.io/tidyr#gather>
5. GGPlot Notes from Dr. Conor Hayes
6. <https://color.adobe.com/create/color-wheel>
7. <https://github.com/claustwilke/colorblindr>
8. <https://blog.datawrapper.de/partycolors/>

Code Appendix:

```
title: "Assignment3"
author: "Prakhar"
date: "27/03/2021"
output: html_document
```

```
```${r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```
```

Installing relevant libraries

```
```${r libraries, message=FALSE, warning=FALSE}
library(ggplot2)
library(dplyr)
library(ggrepel)
library(scales)
library(readxl)
library(ggforce)
library(tidyverse)
library(gridExtra)
library(scales)
library(ggthemes)
library(stringr)
library(knitr)
library(kableExtra)
library(ggthemes)
#library(colorblindr)
```
```

Part 1 : Vote per party in each election of Galway West

Using the data for party election for the region of Galway West constituency for the year of 2016 and 2020 which have been scraped from <https://www.oireachtas.ie/en/elections/> (2016 Galway west election data) and Galway West Fine Gael Party officials (2020 Galway west election data) we will find vote per part for both the year.

Read and plot 2016 Galway West consituency dataset

```
```${r read 2016 galway}
The data for 2016 Galway west consituency is in csv format
data_2016 <- read.csv("2016-04-28_general-election-count-details-galway-west-csv_en.csv",col.names
=c("consituency_name","candidate_surname","candidate_first_name","result","count_number","transfers","votes","total_votes","candidate_id","party"))

head shows the first few rows of the data
head(data_2016)
```

```${r process 2016 galway}
```

```
read rows with count number as 1 as they are the only votes that matter/considered for voting
data_2016 <- data_2016 %>% filter(count_number==1)
ignore irrelevant columns
data_2016 <- data_2016 %>% select(total_votes,party,)

sum rows based on same political parties to calculate cumulative sum of votes reference
:https://stackoverflow.com/questions/1660124/how-to-sum-a-variable-by-group
data_2016 <- aggregate(data_2016$total_votes, by=list(Category=data_2016$party), FUN=sum)
data_2016 <- data_2016 %>% rename(party=Category,votes=x)

head shows the first few rows of the data
head(data_2016)
...

```

```
```{r plot 2016 galway}
# PLOT 1a : vote for year 2016 (Bar charts)
plot_2016<-ggplot(data_2016, (aes(x= reorder(party,votes), y=votes))) +
  geom_col(fill="#66c2a5", width =0.8) +
  scale_y_continuous(
    breaks = c(0,4000,8000,12000,16000),
    expand = c(0, 0),
    name = "Total votes") +
  geom_text(aes(label=votes), size = 4, vjust=0.25,hjust=1) +
  ggtitle("Total votes per party (Year 2016)") +
  xlab(label = "Party") +
  coord_flip(clip = "off") +

  theme(
    axis.line.y = element_blank(),
    axis.ticks.y = element_blank(),
    axis.line.x = element_blank(),
    axis.ticks.x = element_blank(),
    #axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1),
    plot.title = element_text(vjust = -1, hjust = 0.5, size = 15),
    panel.background = element_blank(),
    panel.grid.major.x = element_line(size = 0.4, linetype = 'solid', colour = "white"),
    panel.ontop = TRUE)

```

```
plot(plot_2016)
...

```

Read and plot 2020 Galway West constituency dataset

```
```{r read 2020 galway}
The data for 2020 Galway west constituency is in xlsx format
data_2020 <-
read_excel("GalwayWest-2020.xlsx",col_names=c("dummy","party","candidate_name","fpv","c1","c2","c3","c4","c5","
c6","c7","c8","c9","c10","c11","c12","c13"))

Data cleaning : the first column and first two rows are of no use so we will drop it
data_2020$dummy <-NULL
data_2020 = data_2020[-1,]
data_2020 = data_2020[-1,]

```

```
head shows the first few rows of the data
head(data_2020)
...

```{r process 2020 galway}
# selecting only relevant columns part and count 1 votes
data_2020 <- data_2020 %>% select(party,c1,)
# converting c1 datatype to numeric for calculation
data_2020$c1 <- as.numeric(as.character(data_2020$c1))
# Remove last row with NA Reference: https://stat.ethz.ch/pipermail/r-help/2012-December/343413.html
data_2020<- data_2020[-nrow(data_2020),]
# aggregative rows on basis on basis of party
data_2020 <- aggregate(data_2020$c1, by=list(Category=data_2020$party), FUN=sum)
data_2020 <- data_2020 %>% rename(party=Category,votes=x)
# head shows the first few rows of the data
head(data_2020)
...

```

```
```{r plot 2020 galway}
plot_2020<-ggplot(data_2020, (aes(x= reorder(party,votes), y=votes))) +
 geom_col(fill="#009E73", width =0.8) +
 scale_y_continuous(
 breaks = c(0,4000,8000,12000,16000),
 expand = c(0, 0),
 name = "Total votes") +
 geom_text(aes(label=votes), size = 4, vjust=0.25,hjust=1) +
 ggtitle("Total votes per party (Year 2020)") +
 xlab(label = "Party") +
 coord_flip(clip = "off") +

 theme(
 axis.line.y = element_blank(),
 axis.ticks.y = element_blank(),
 axis.line.x = element_blank(),
 axis.ticks.x = element_blank(),
 #axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1),
 plot.title = element_text(vjust = -1, hjust = 0.5, size = 15),
 panel.background = element_blank(),
 panel.grid.major.x = element_line(size = 0.4, linetype = 'solid', colour = "white"),
 panel.ontop = TRUE)

```

```
plot(plot_2020)
...

```

## Analysis of change of votes from 2016-2020 for Galway West

```
```{r compare 2016 2020 Galway}

plot_galway_2016<-ggplot(data_2016, (aes(x= reorder(party,votes), y=votes))) +
  geom_col(fill="#66c2a5", width =0.8) +
  scale_y_continuous(
    breaks = c(0,5000,10000,15000),
    expand = c(0, 0),
    name = "Total votes") +

```

```
geom_text(aes(label=votes), size = 3, vjust=0) +
ggtitle("2016") +
xlab(label = "Party") +
coord_flip(clip = "off") +

theme(
  axis.line.y = element_blank(),
  axis.ticks.y = element_blank(),
  axis.line.x = element_blank(),
  axis.ticks.x = element_blank(),
  #axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1),
  plot.title = element_text(vjust = -1, hjust = 0.5, size = 15),
  panel.background = element_blank(),
  panel.grid.major.x = element_line(size = 0.4, linetype = 'solid', colour = "white"),
  panel.ontop = TRUE)
```

```
plot_galway_2020<-ggplot(data_2020, (aes(x= reorder(party,votes), y=votes))) +
geom_col(fill="#009E73", width =0.8) +
scale_y_continuous(
  breaks = c(0,5000,10000,15000),
  expand = c(0, 0),
  name = "Total votes") +
geom_text(aes(label=votes), size = 3, vjust=0,hjust=0.7) +
ggtitle("2020") +
xlab(label = "") +
coord_flip(clip = "off") +
```

```
theme(
  axis.line.y = element_blank(),
  axis.ticks.y = element_blank(),
  axis.line.x = element_blank(),
  axis.ticks.x = element_blank(),
  #axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1),
  plot.title = element_text(vjust = -1, hjust = 0.5, size = 15),
  panel.background = element_blank(),
  panel.grid.major.x = element_line(size = 0.4, linetype = 'solid', colour = "white"),
  panel.ontop = TRUE)
```

```
grid.arrange(plot_galway_2016, plot_galway_2020, nrow = 1)
```

```
```
```

```
A more elegant way for comparision of votes for 2016-2020 Galway west
```

```
```{r data 2016 2020 Galway}
```

```
data_2016_vote <- data_2016
```

```
# Providing short notations to part for Galway West 2016
```

```
data_2016_vote$short <- c("AAA","DDI","FF","FG","GP","IND","LP","REN","SF","SDP")
```

```
data_2016_vote$year <- 2016
```

```
data_2020_vote <- data_2020
```

```
# Providing short notations to part for Galway West 2020
```

```
data_2020_vote$short <- c("AON","FF","FG","GP","IND","LP","SF","SDP","SPBP")
```

```
data_2020_vote$year <-2020
```

```
colnames(data_2020_vote)<- c("party","votes","short","year")
```

```

# head shows the first few rows of the data
head(data_2016_vote)
# head shows the first few rows of the data
head(data_2020_vote)
...

```{r compare}
Bind both data for 2016 and 2020 in a single dataframe Reference :
https://stackoverflow.com/questions/8169323/r-concatenate-two-dataframes
galway_data_2016_2020 <- rbind(data_2016_vote, data_2020_vote)
cast year to numeric
galway_data_2016_2020$year <- as.character(as.numeric(galway_data_2016_2020$year))
print type for galway combined data
sapply(galway_data_2016_2020, typeof)
drop the party name
galway_data_2016_2020 <- galway_data_2016_2020[c(2,3,4)]

#To find the difference, make the data untidy
combine_data_16_20 <- galway_data_2016_2020 %>% spread(key = year,value = votes,convert = TRUE)
Change column names
colnames(combine_data_16_20) <- c("Short", "Year2016", "Year2020")
#Replace NA to 0
combine_data_16_20[is.na(combine_data_16_20)] <- 0
#take the difference of the votes in new data frame
combine_diff_16_20 <- combine_data_16_20 %>% mutate(vote_diff = Year2020 - Year2016) %>% select("Short",
"vote_diff")

head(combine_diff_16_20)
...

Plot the vote difference for 2016-2020 Galway West
```{r}

p<-rev(c("Solidarity-PBP","Sinn Féin","Social Democrats","Renua","Labour Party","Independent","Green
Party","Fine Gael","Fianna Fáil","Direct Democracy Ireland","Aontú","AAA"))

# separate colors for increase/decrease in cotes
combine_diff_16_20$color <- ifelse(combine_diff_16_20$vote_diff >=0, "#009E73", "#D55E00")

galway_2016_2020_plot<-ggplot(combine_diff_16_20, aes(x=vote_diff, y=Short, colour = Short)) +
  geom_segment(aes(x = 0, y = Short, xend = vote_diff, yend = Short), size = 1, color =
combine_diff_16_20$color) +
  geom_point(size =3.5) +

  geom_text(aes(label=p), vjust=-1.0, size=3, na.rm = TRUE, show.legend = FALSE)+
  scale_colour_manual(values= combine_diff_16_20$color) +
  scale_x_continuous( name = "Increment in votes 2016/2020 Galway West" ) +
  theme(axis.line.y = element_blank(),
        panel.grid.major.x = element_line(size=0, colour = "white"),
        axis.line.x = element_blank(),
        axis.ticks.y = element_blank(),
        axis.ticks.x = element_blank(),
        axis.title.y = element_blank(),

```

```
axis.title.y.left = element_blank(),
legend.position = "None",
axis.title.x = element_text(size = 15))
```

```
plot(galway_2016_2020_plot)
...

```

Visualisation using Bar charts

Also I have used the below visualisation to explain how to create visuals for color blind audience too. As for the next visual the color plays a crucial role as compared to all the visual before where we were using a single ink color or the visuals were even understandable even if the viewer not correctly guess the color.

```
```{r galway 2016 2020 Bar}
A colorblind friendly palatte
cbPalette <- c("#999999", "#E69F00", "#56B4E9", "#009E73", "#F0E442")
#c("#fc8d62", "#66c2a5")
theme_set(theme_classic())

galway_data_2016_2020 <- rbind(data_2016_vote, data_2020_vote)
cast year to numeric
galway_data_2016_2020$year <- as.character(as.numeric(galway_data_2016_2020$year))

party_2016_2020_vote_diff<-ggplot(galway_data_2016_2020, (aes(x= short, y=votes, fill=year))) +
 geom_col(position="dodge", alpha=0.85,width = 0.7) +
 scale_y_continuous(
 breaks = c(0, 2000, 4000,6000,8000, 10000,12000,14000,16000),
 expand = c(0, 0),
 name = "Number of votes")+
 xlab(label = "party") +
 scale_fill_manual(values = cbPalette , name = NULL)+
 theme(
 axis.line.y = element_blank(),
 axis.ticks.y = element_blank(),
 axis.line.x = element_blank(),
 axis.ticks.x = element_blank(),
 panel.background = element_blank(),
 panel.grid.major.y = element_line(size = 0.4, linetype = 'solid', colour = "white"),
 panel.ontop = TRUE,
 legend.position= c(0.8, 0.9), legend.direction="horizontal")
```

```
plot(party_2016_2020_vote_diff)
...
```{r}
#cvd_grid(party_2016_2020_vote_diff)
...

```

```
```{r galway 2016 2020 lollypop , message=FALSE, warning=FALSE}
ggplot(galway_data_2016_2020, aes(x = votes, y= short)) +
 geom_line(aes(group = short), colour = "grey", size=0.5) +
```

```

geom_point(aes(colour = year), size = 3, alpha = 0.7) +
geom_text(aes(label=year, colour = year), vjust=-1.2, size=3, na.rm = TRUE, show.legend = FALSE)+
scale_colour_manual(values= c("#ce5a6c", "#212f85"), name = "Year") +

scale_x_continuous(limits = c(0, 16000),
 expand = c(0, 0),
 breaks = seq(0, 16000, by = 2000),
 name = "Votes")+
theme(axis.title.y = element_blank(),
 panel.grid.major.x = element_line(size=0.03),
 panel.grid.minor.x = element_blank(),
 panel.grid.major.y = element_blank(),
 axis.line.y = element_blank(),
 axis.line.x = element_blank(),
 axis.ticks.y = element_blank(),
 axis.ticks.x = element_blank(),
 legend.position= c(0.85, 0.94),
 legend.text = element_text(size = 14),
 legend.key.size = unit(0.7, "lines"),
 legend.title = element_blank())
...

```

## Comparision of Galway West to the national average for party share of votes for year 2016/2020

The data for national average for 2016 has been collected from TCD political scientist Michael Gallagher's web site:  
[https://www.tcd.ie/Political\\_Science/people/michael\\_gallagher/Election2016.php](https://www.tcd.ie/Political_Science/people/michael_gallagher/Election2016.php) (2016 data) and  
[https://www.tcd.ie/Political\\_Science/people/michael\\_gallagher/Election2020.php](https://www.tcd.ie/Political_Science/people/michael_gallagher/Election2020.php) (2020 data)

### Analysis of vote difference between Galway and National year 2020

```

```{r national 2020}
#Read the CSV of national averages for 2020
national_2020 <- read.csv("national-2020-results.csv") %>% select(i..2020.election.result,Votes,X..vote)
colnames(national_2020)<-c("party","votes","votes_percent")
national_2020$votes <- as.numeric(gsub(",","", national_2020$votes))
#set region as it will be again used in plot and calculation
national_2020$region <- "Ireland"
#remove last few row
national_2020<-national_2020[c(-20,-21,-22),]
head(national_2020)
...

```

The data scrapped has a number of issues in part names due to encoding difference and needs to be cleaned for better and accurate processing and visualisation

```

```{r cleaning national 2020}
national_2020$party[national_2020$party %in% "People Before Profit"] <- "People-Before-Profit"
national_2020$party[national_2020$party %in% "Green Party"] <- "Green-Party"
national_2020$party[national_2020$party %in% "Fianna Fáil"] <- "Fianna-Fail"
national_2020$party[national_2020$party %in% "Labour"] <- "Labour-Party"
national_2020$party[national_2020$party %in% "Fine Gael"] <- "Fine-Gael"
national_2020$party[national_2020$party %in% "Social Democrats"] <- "Social-Democrats"
national_2020$party[national_2020$party %in% "Sinn Féin"] <- "Sinn-Fein"

```

```

national_2020$party[national_2020$party %in% "Solidarityâ\200"PBP (Solidarityâ\200" People Before Profit
Alliance)"]<-"Solidarity-PBP"
national_2020$party[national_2020$party %in% "AontÃ"] <- "Aontu"
head(national_2020)
...

```{r galway 2020}
# modifying Galway 2020 data for next task
galway_2020_data<-data_2020_vote %>% select(party,votes,short)
galway_2020_data$region <- "Galway"
galway_2020_data$votes_percent <- round(galway_2020_data$votes * 100 / sum(galway_2020_data$votes), 2)
head(galway_2020_data)
...

```{r national 2020}
modifying country 2020 data for next task
ireland_2020_data <- national_2020 %>% select(party,votes,region,votes_percent)
ireland_2020_data$short <-
c("SF","FF","FG","GP","LP","SDP","SPBP","PBP","SP","RISE","AON","I4C","IFP","RI","NP","IDP","WP","UP","IND")
head(ireland_2020_data)
...

```{r plot national galway 2020}
combined_2020_ireland_galway_data <- rbind(galway_2020_data,ireland_2020_data)

theme_set(theme_classic())

#plot to show the values of all the data form 2016.2020, national
compare_national_local_plot<-ggplot(combined_2020_ireland_galway_data , mapping = aes(x = votes_percent, y =
short)) +
  geom_line(aes(group = short), colour = "grey", size = 0.5) +
  geom_point(aes(colour = region), size = 3, alpha = 0.9) +
  #Set the x-axis scale
  scale_x_continuous(
    limits = c(0, max(combined_2020_ireland_galway_data$votes_percent)),
    expand = c(0, 0),
    breaks = seq(0, max(combined_2020_ireland_galway_data$votes_percent)+3, by = 1),
    name = "Percentage of votes"
  ) + ggtitle("Total vote percentage difference year 2020") +
  #Set the theme
  theme(
    axis.title.y = element_blank(),
    panel.grid.major.x = element_line(size = 0.03),
    panel.grid.minor.x = element_blank(),
    panel.grid.major.y = element_blank(),
    axis.line.y = element_blank(),
    axis.line.x = element_blank(),
    axis.ticks.y = element_blank(),
    axis.ticks.x = element_blank(),
    legend.position = c(0.90, 0.95),
    legend.text = element_text(size = 8),
    legend.key.size = unit(0.7, "lines"),

```



```

    legend.title = element_blank()
  )

plot(compare_national_local_plot)
...

#### Analysis of vote difference between Galway and National year 2016

```{r national 2016}
#Read the CSV of national averages for 2016
national_result_2016 <- read.csv("national-2016-results.csv") %>% select(i..2016.election.result,Votes, X..vote)
colnames(national_result_2016) <- c("party","votes","votes_percent")
national_result_2016$votes <- as.numeric(gsub(",","", national_result_2016$votes))
#set region as it will be again used in plot and calculation
national_result_2016$region <- "Ireland"

#remove last few row
national_result_2016 <- national_result_2016[c(-22),]
Data cleaning
national_result_2016$party[national_result_2016$party %in% "Solidarity\200"PBP (Solidarity\200" People Before Profit Alliance)"] <- "Solidarity-PBP"
national_result_2016$party[national_result_2016$party %in% "Fianna Fáil"] <- "Fianna-Fail"
national_result_2016$party[national_result_2016$party %in% "Sinn Féin"] <- "Sinn-Fein"
national_result_2016$party[national_result_2016$party %in% "People Before Profit"] <- "People-Before-Profit"
national_result_2016$party[national_result_2016$party %in% "Green Party"] <- "Green-Party"
national_result_2016$party[national_result_2016$party %in% "Fine Gael"] <- "Fine-Gael"
national_result_2016$party[national_result_2016$party %in% "Social Democrats"] <- "Social-Democrats"
national_result_2016$party[national_result_2016$party %in% "Aontú"] <- "Aontu"
national_result_2016$party[national_result_2016$party %in% "Labour"] <- "Labour-Party"

head(national_result_2016)
...

```{r galway 2016}
# modifying Galway 2016 data for next task
galway_2016_data <- data_2016_vote %>% select(party,votes,short)
galway_2016_data$region <- "Galway"
galway_2016_data$votes_percent <- round(galway_2016_data$votes * 100 / sum(galway_2016_data$votes), 2)
head(galway_2016_data)
...

```{r}
modifying country 2016 data for next task
ireland_2016_data <- national_result_2016 %>% select(party,votes,region,votes_percent)
ireland_2016_data$short <-
c("FG","FF","SF","LP","AAA","PBP","SOCP","SD","GP","RI","WUAG","DDI","WP","CD","FN","IDP","CPI","II","IA","I4C","IND")
ireland_2016_data
...

```{r plot galway national 2016}
combined_2016_ireland_galway_data <- rbind(galway_2016_data,ireland_2016_data)
theme_set(theme_classic())

```

```

#plot to show the values of all the data form 2016.2020, national
compare_national_local_plot_2016<-ggplot(combined_2016_ireland_galway_data , mapping = aes(x =
votes_percent, y = short)) +
  geom_line(aes(group = short), colour = "grey", size = 0.5) +
  geom_point(aes(colour = region), size = 3, alpha = 0.9) +
  #Set the x-axis scale
  scale_x_continuous(
    limits = c(0, max(combined_2020_ireland_galway_data$votes_percent)),
    expand = c(0, 0),
    breaks = seq(0, max(combined_2020_ireland_galway_data$votes_percent)+3, by = 1),
    name = "Percentage of votes"
  ) + ggtitle("Total vote percentage difference year 2016") +
  #Set the theme
  theme(
    axis.title.y = element_blank(),
    panel.grid.major.x = element_line(size = 0.03),
    panel.grid.minor.x = element_blank(),
    panel.grid.major.y = element_blank(),
    axis.line.y = element_blank(),
    axis.line.x = element_blank(),
    axis.ticks.y = element_blank(),
    axis.ticks.x = element_blank(),
    legend.position = c(0.90, 0.95),
    legend.text = element_text(size = 8),
    legend.key.size = unit(0.7, "lines"),
    legend.title = element_blank()
  )

plot(compare_national_local_plot_2016)

```

Change in vote for candidate in both election of Galway West

We will try to find which candidates were impacted most during both the elections of Galway West 2016 and 2020

```

```{r change candidate 2016 2020}

```

# We will read 2016 and 2020 data again for Galway

# 2016 galway data

```

data_cand_2016 <- read.csv("2016-04-28_general-election-count-details-galway-west-csv_en.csv",col.names
=c("consituency_name","candidate_surname","candidate_first_name","result","count_number","transfers","votes","t
otal_votes","candidate_id","party"))

```

# read rows with count number as 1 as they are the only votes that matter

```

data_cand_2016 <- data_cand_2016 %>% filter(count_number==1)

```

# Candidate name = first name + last name

```

data_cand_2016$candidate_name <- paste(data_cand_2016$candidate_first_name,"
",data_cand_2016$candidate_surname)

```

# ignore irrelevant columns

```

data_cand_2016 <- data_cand_2016 %>% select(total_votes,party,candidate_name)

reference : https://stackoverflow.com/questions/1660124/how-to-sum-a-variable-by-group
data_cand_2016 <- aggregate(data_cand_2016$total_votes,
by=list(Category=data_cand_2016$candidate_name,data_cand_2016$party), FUN=sum)

data_cand_2016 <- data_cand_2016 %>% rename(Candidate=Category,party=Group.2,votes=x)
data_cand_2016$year <-2016
head shows the first few rows of the data
head(data_cand_2016)

data_cand_2020 <-
read_excel("GalwayWest-2020.xlsx",col_names=c("dummy","party","candidate_name","fpv","c1","c2","c3","c4","c5","c6","c7","c8","c9","c10","c11","c12","c13"))

the first column and first two rows are of no use so we will drop it
data_cand_2020$dummy <-NULL
data_cand_2020 = data_cand_2020[-1,]
data_cand_2020 = data_cand_2020[-1,]

data_cand_2020 <- data_cand_2020 %>% select(party,c1,candidate_name)
converting c1 datatype to numeric for calculation
data_cand_2020$c1 <- as.numeric(as.character(data_cand_2020$c1))

Remove last row with NA Reference: https://stat.ethz.ch/pipermail/r-help/2012-December/343413.html
data_cand_2020<- data_cand_2020[-nrow(data_cand_2020),]

data_cand_2020 <- aggregate(data_cand_2020$c1,
by=list(Category=data_cand_2020$candidate_name,data_cand_2020$party), FUN=sum)
data_cand_2020<- data_cand_2020 %>% rename(Candidate=Category,party=Group.2,votes=x)
data_cand_2020$year <-2020
head shows the first few rows of the data
head(data_cand_2020)

...

```{r combine 2016 2020 galway candidate data}

#data_cand_2016_2020 <- rbind(data_cand_2016, data_cand_2020)
data_cand_2016$candidate_short <-
c("TH","RON","AOC","JC","MH","HN","JOM","SK","SS","CMAC","FHE","MC","NG","PF","TR","DN","JC","ND","TC","N
OT")
data_cand_2020$candidate_short <-
c("COC","EOC","OC","HN","SK","POR","CMAC","DOF","MC","NG","NMN","MF","NOT","CB","JL")

data_cand_2016_2020 <- merge(data_cand_2016,data_cand_2020,by="candidate_short")
data_cand_2016_2020 <- data_cand_2016_2020 %>% select(candidate_short,Candidate.x,votes.x,party.y,votes.y)
colnames(data_cand_2016_2020)<- c("candidate_short","candidate_name","vote_2016","party","vote_2020")
# reference :
https://stackoverflow.com/questions/19838860/to-find-the-difference-between-two-column-elements-in-a-data-frame/19838952

```

```
data_cand_2016_2020$diff_vote <- data_cand_2016_2020$vote_2020 - data_cand_2016_2020$vote_2016
data_cand_2016_2020$pos <- factor(ifelse(data_cand_2016_2020$diff_vote >= 0, "TRUE", "FALSE"), levels =
c("TRUE", "FALSE"))
data_cand_2016_2020
...
```

```
```{r plot candidate 2016 2020 galway}
```

# As color is a secondary aesthetic in this case i have used a combination of green and red which is in general highly discouraged as it creates problem for color blind audience

```
cb2 <- c("#0F9D00", "#C50101") # created using https://color.adobe.com/create/color-wheel
```

```
ggplot(data_cand_2016_2020, aes(x = candidate_name, y = diff_vote, fill=pos)) +
 geom_col(position = "identity", alpha = 0.85, width=0.7) +
```

```
 scale_y_continuous(limits = c(-1000, 1100),
 breaks = seq(-1000, 1500, by= 200) ,
 name = "Difference in Vote from 2016 to 2020",
 expand=c(0,0),
 labels = scales::number_format(accuracy = 0.1)) + scale_fill_manual(values=cb2) +
```

```
 theme(
 axis.line.y = element_blank(),
 axis.ticks.y = element_blank(),
 axis.line.x = element_blank(),
 #axis.ticks.x = element_blank(),
 axis.text.x = element_text(angle = 25, vjust = 1, hjust = 1),
 axis.title.x = element_blank(),
 plot.margin = margin(6, 6, 3, 3),
 panel.background=element_blank(),
 panel.grid.major.y = element_line(size = 0.2, linetype = 'solid', colour = "white"),
 panel.ontop = TRUE,
 legend.position = "none"
)
)
```

```
...
```

#### Few References:

1. [https://clayford.github.io/dwir/dwr\\_05\\_combine\\_merge\\_rehsape\\_data.html](https://clayford.github.io/dwir/dwr_05_combine_merge_rehsape_data.html)
2. [https://www.tcd.ie/Political\\_Science/people/michael\\_gallagher/Election2016.php](https://www.tcd.ie/Political_Science/people/michael_gallagher/Election2016.php)
3. [https://www.tcd.ie/Political\\_Science/people/michael\\_gallagher/Election2020.php](https://www.tcd.ie/Political_Science/people/michael_gallagher/Election2020.php)
4. <https://uc-r.github.io/tidyr#gather>
5. GGPlot Notes
6. <https://color.adobe.com/create/color-wheel>
7. <https://github.com/clauswilke/colorblindr>
8. <https://blog.datawrapperr.de/partycolors/>