

Data Visualization

Kashika Sharma

The report includes the comparisons of Galway west votes per party in the year 2016 and 2020 using the first preference votes of each candidate. These numbers are then compared with the national vote numbers for each party for 2016 and 2020. The data has already been provided for Galway west votes (2016 and 2020) and overall national votes data is taken from TCD political scientist Michael Gallagher's web site (2016 and 2020).

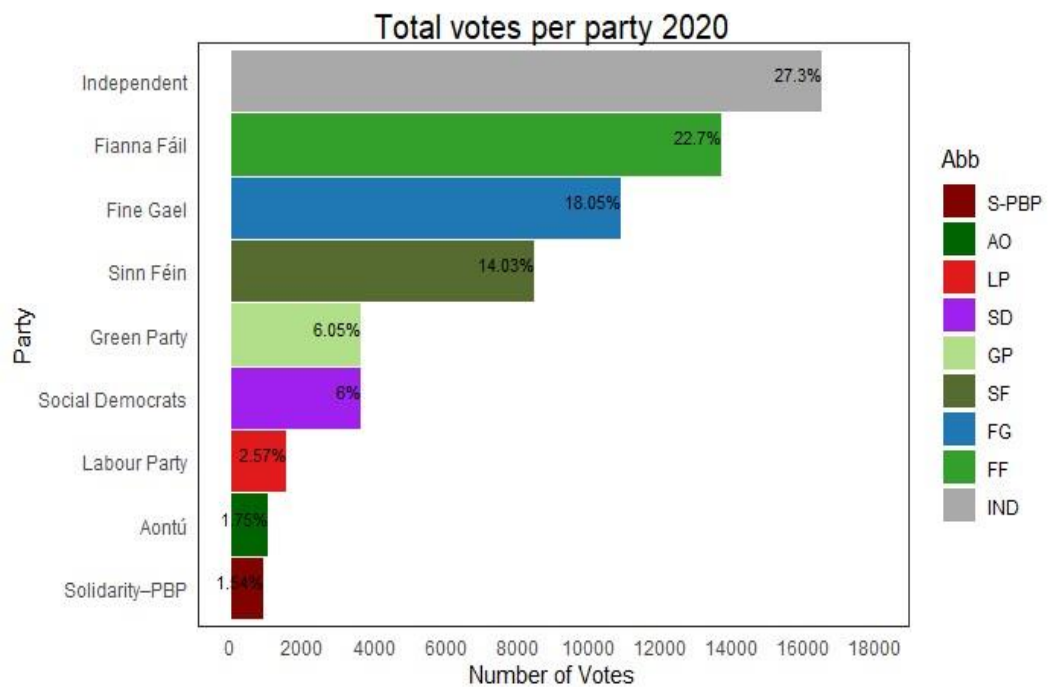
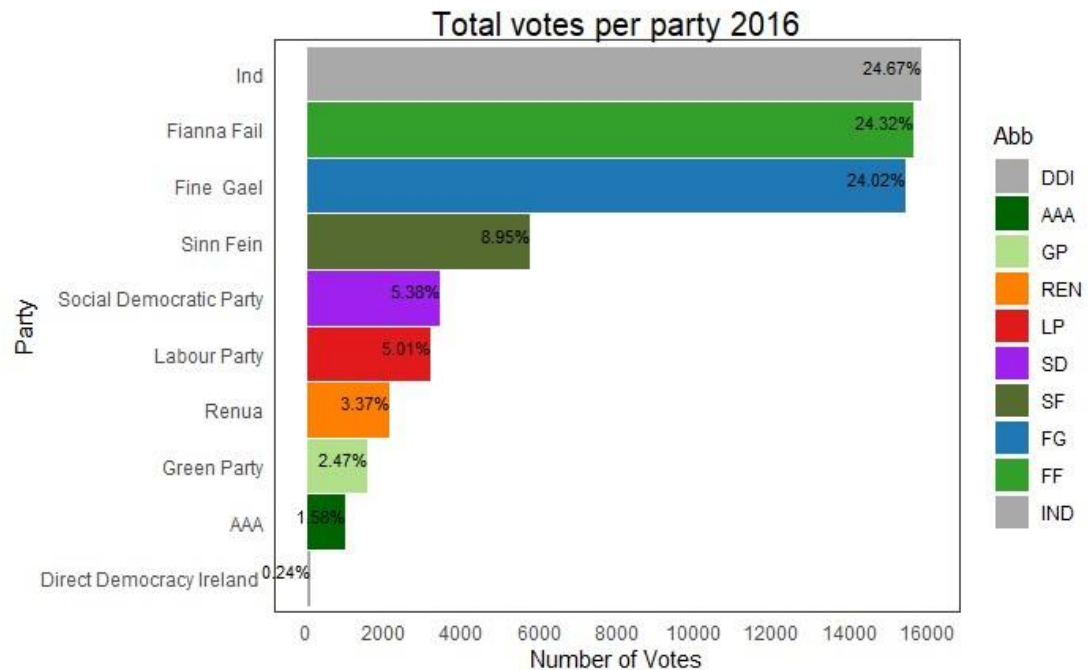
Data Cleaning:

The data provided for Galway West and Nationals required some changes in the columns, some of them are:

- Added the Abbreviations for each party to compare the data from 2016-2020.
- Filtered only the votes with first preference.
- Added colors specific and appropriate for the parties (referred from live class notes).

1. Galway West votes 2016-2020:

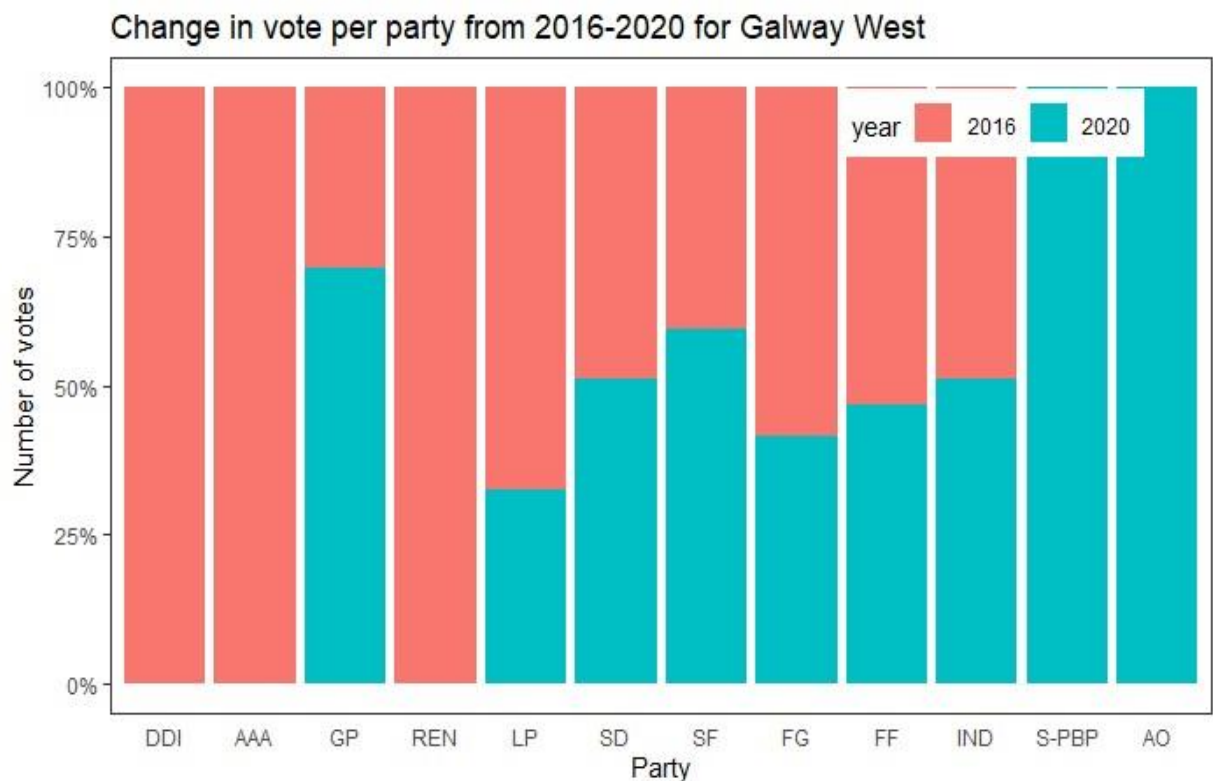
This section displays the total votes per party in 2016-2020. The plots for votes per party (only first preferences) for 2016 and 2020 are below. All the colors used to represent parties are appropriate as per their political domain. The horizontal bar chart is used to show the total votes per party as it is easy to compare the data this way. As clearly visible in the plots the Independents party has the highest votes in both 2016 and 2020 followed by FF, FG and SF after which the trend breaks in 2020. Along with the numbers on x axis, the percentage of votes are also shown as text on the bars for fast interpretation.



2. Comparison of 2016-2020 votes Galway West:

This section shows the change in vote per party from 2016-2020. The stacked bar chart is used to represent the comparison between 2016 and 2020 for each party with y-axis scale in percentage as it is fast and easier to interpret the results. As we can see in the plot there is a vast change in the percentage of votes for parties GP, LP, SD, SF and FG.

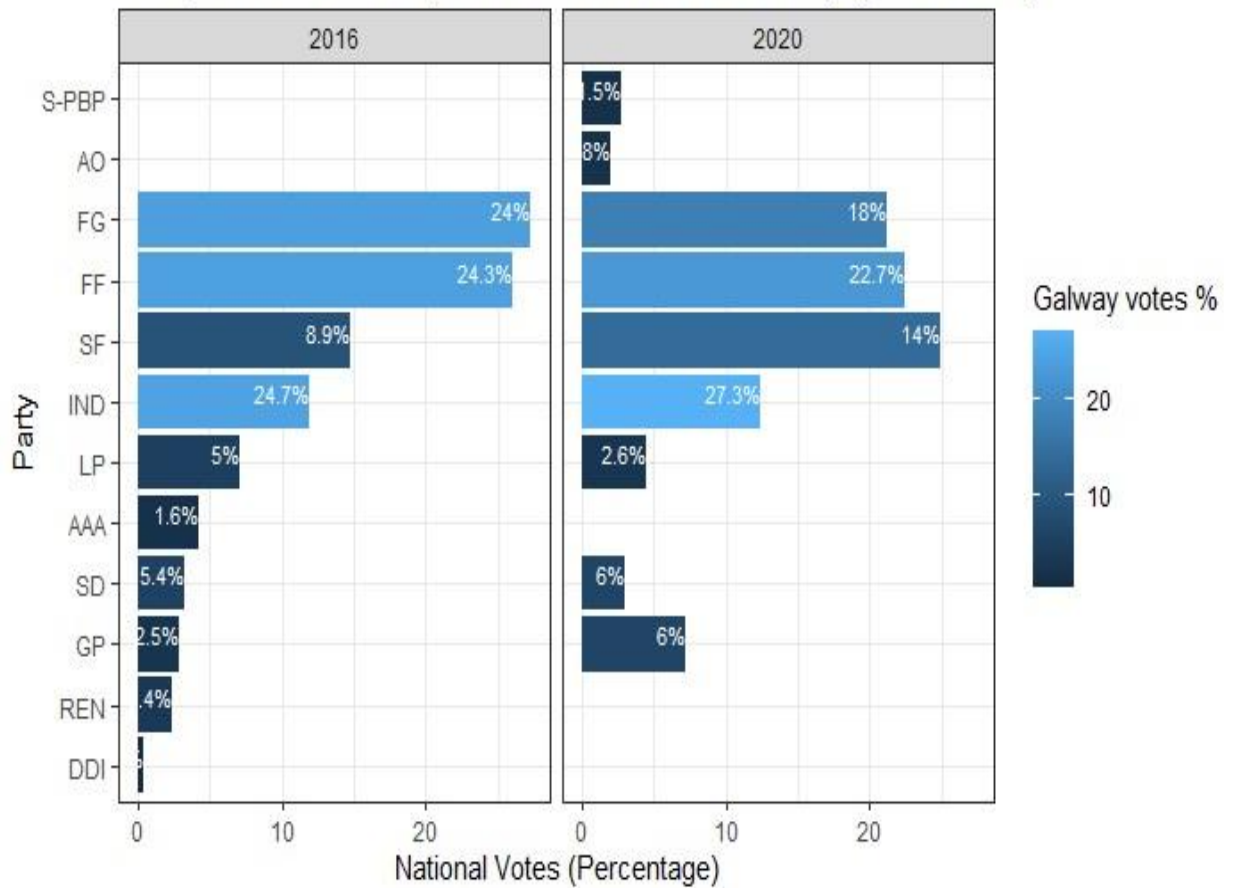
There are some parties which are not present in 2016(A-PBP and AO) and some in 2020(DDI, AAA). Therefore, the votes for A-PBP, AO, DDI and AAA parties cannot be compared.



3. Comparison of Galway west to National votes (2016-2020):

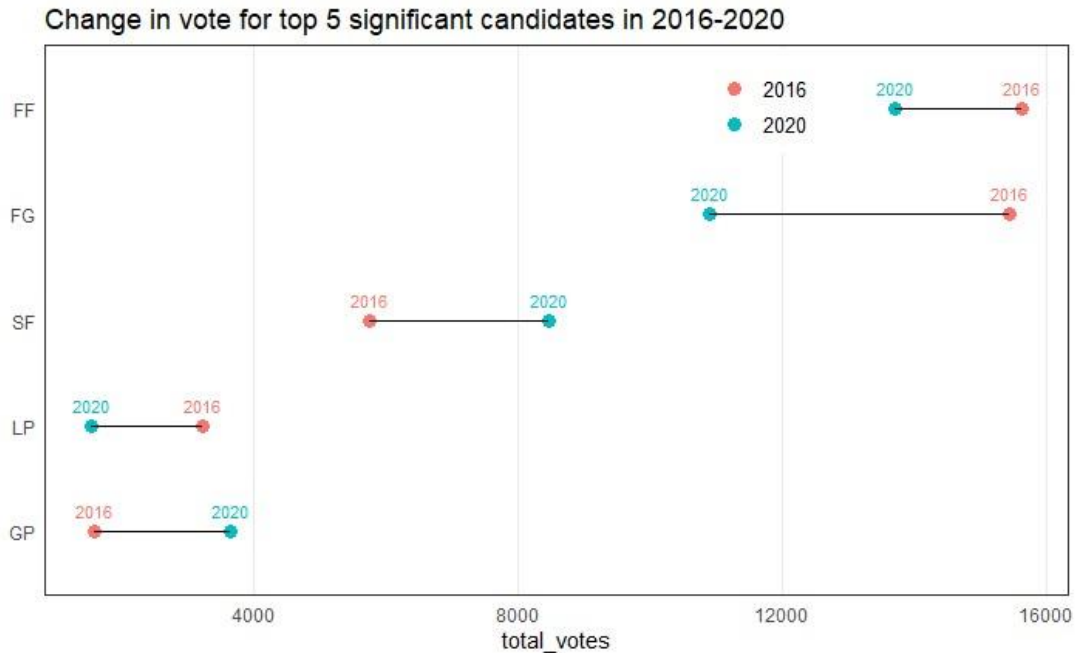
This section shows the comparison of votes between Galway and National. The overall national vote numbers (in terms of percentage) are on the x-axis and the color scale represents the Galway west votes percentage, the lighter the color more the votes, the percentage of Galway votes is also present in the form of text on the bars to make it easier to compare. From the plot below we can see that FG is leading in 2016 in both National votes with IND leading in Galway, whereas in 2020 SF is leading in National and IND in Galway.

Comparison of Galway West to the national average(2016-2020)

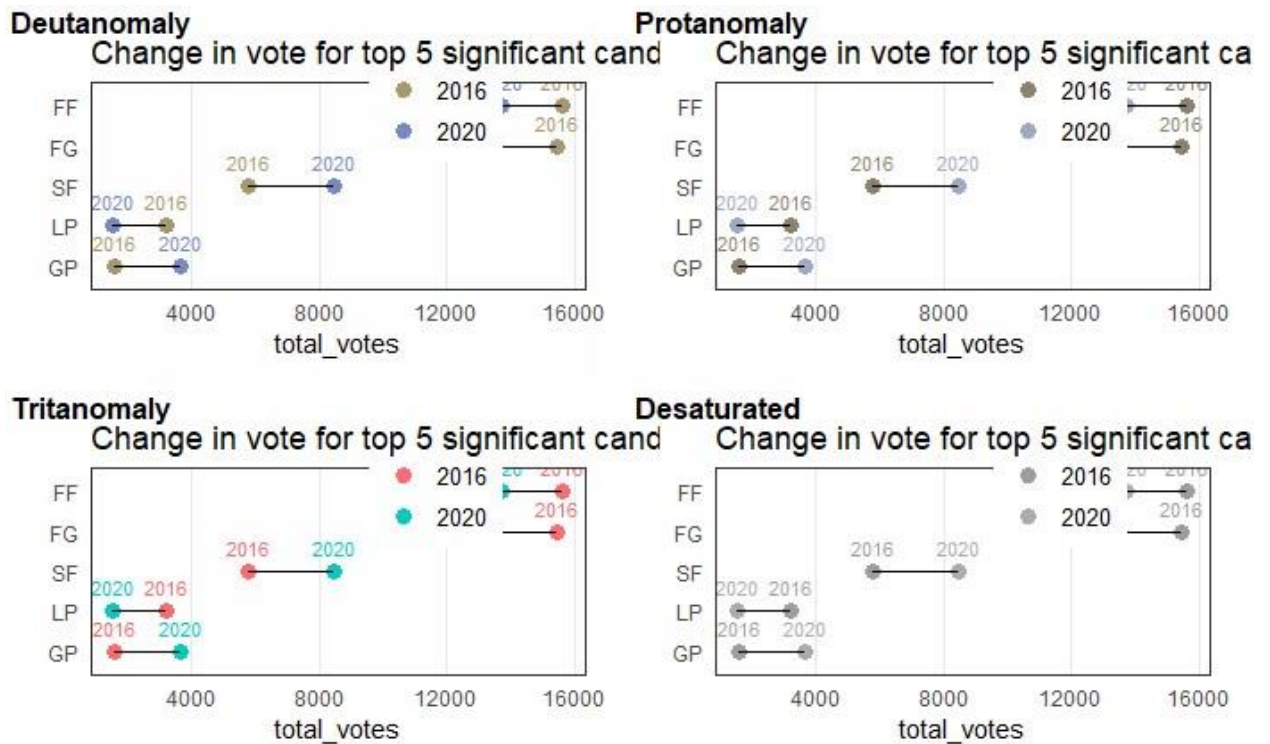


4. The change in vote for the most significant candidates in both elections in Galway West:

The dot plot below shows the change in votes from 2016-2020 for the top 5 most significant candidates. These candidates are selected based on the highest difference in the votes from 2016-2020. FG got 24% votes in 2016 which declined to 18% in 2020 whereas for party SF, the votes increased by 5% in 2020. The colors used in the plot are color blind friendly which can be seen in the CVD simulation below.



CVD Simulation of the above plot:



Appendix (Code):

```

Import      Libraries
library(ggplot2)
library      (dplyr)
  
```

```
library(forcats)
library(readxl)
library(colorblindr)
```

PART 1

```
#Read and do the necessary data cleaning of 2016 Galway data data_2016
<- read.csv("2016general.csv")
data_2016 <- data_2016[which(data_2016$Count.Number == 1),] #only consider first preference

data_2016_v1 <- data_2016 %>% #group by Party and count total votes
select(Party,Votes) %>% group_by(Party) %>%
  summarise(total_votes = sum(Votes))

data_2016_v1 <- data_2016_v1 %>%
  mutate(prop = round(total_votes/sum(total_votes),4)*100) #to calculate the percentage

#Set the abbreviations
data_2016_v1$Abb <- c("AAA","DDI","FF","FG","GP","IND","LP","REN","SF",
,"SD")
data_2016_v1<-data_2016_v1%>%mutate(Abb = fct_reorder(Abb, total_votes
))%>%arrange(desc(total_votes))

#Assign party colors party.colours <- c('FG' = '#1f78b4', 'FF' =
'#33a02c', 'SF' = 'darkoli vegreen', 'GP' = '#b2df8a', 'SD'
='purple', 'IND' = 'darkgrey', 'REN' = '#ff7f00', 'DDI'
='darkgrey', 'AAA' = 'darkgreen', "LP" = "#e31a1c", 'A0 ' = 'darkgreen', 'S-
PBP' = '#820200')
theme_set(theme_bw())

#Plot1 : Vote per party in 2016 Galway West
ggplot(data_2016_v1, (aes(x= reorder(Party,total_votes), y=total_votes
, fill = Abb))) +
geom_col(colour="white", width =1) +
scale_y_continuous(limits = c(0, 16000),
                    breaks = seq(0,16000, by = 2000),
name = "Number of Votes") + scale_fill_manual(values =
party.colours) +
```

```

    geom_text(aes(label = paste0(prop, "%")), size = 3, vjust=0,hjust =
1) +
  ggtitle("Total votes per party 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(),
plot.title = element_text(vjust = -1, hjust = 0.5, size = 15),
panel.background = element_blank(),
panel.grid.major =
element_blank(),
panel.grid.minor = element_blank(),
panel.ontop = TRUE)

```

Read 2020 data

```

#Read and do the data cleaning of 2020 Galway west data data_2020
<- read_excel("GalwayWest-2020.xlsx")

```

```
## New names:
```

```
## * `` -> ...2
```

```
## * `` -> ...3
```

```
## * `` -> ...4
```

```
## * `` -> ...5
```

```
## * `` -> ...6 ##
```

```
* ...
```

```
data_2020[1,2] <- 'Party'
```

```
colnames(data_2020) <- data_2020[1,] data_2020
```

```
<- data_2020[-1, -1]
```

Data cleaning for 2020 political data

```
#select count 1 column for first preference votes
```

```
data_2020$`Count 1` <- as.numeric(as.character(data_2020$`Count 1`))
```

```
data_2020<- data_2020[-16,] data_2020_v1 <- data_2020 %>%
```

```
select(Party,`Count 1`) %>% group_by(Party) %>%
```

```
summarise(total_votes = sum(`Count 1`))
```

```
data_2020_v1 <- data_2020_v1 %>%
```

```
mutate(prop = round(total_votes/sum(total_votes),4)*100) #for percen
tage
```

```
data_2020_v1$Abb <- c("AO", "FF", "FG", "GP", "IND", "LP", "SF", "SD", "S-PBP")
```

```
)
data_2020_v1<-data_2020_v1%>%mutate(Abb = fct_reorder(Abb, total_votes
))%>%arrange(desc(total_votes))
```

Visualisation of 2020 data

```
#Plot2 : Total votes per party for 2020 galway west
ggplot(data_2020_v1, (aes(x= reorder(Party,total_votes), y=total_votes
,fill = Abb))) +
  geom_col(colour="white", width =1) +
  scale_fill_manual(values = party.colours) +
  scale_y_continuous(limits = c(0, 18000),
                     breaks = seq(0,18000, by = 2000),
name = "Number of Votes") +
  geom_text(aes(label = paste0(prop, "%")), size = 3, vjust=0,hjust =
1) +
  ggtitle("Total votes per party 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(),
plot.title = element_text(vjust = -1, hjust = 0.5, size = 15),
panel.background = element_blank(), panel.grid.major =
element_blank(), panel.grid.minor = element_blank(),
panel.ontop = TRUE)
```

PART 2 : Comparision between 2016 and 2020 data

```
data_2016_v1$year <- 2016 data_2020_v1$year
<-2020
```

#Combine the 2016 and 2020 data

```
combined <- rbind(data_2016_v1, data_2020_v1)
combined$year <- as.character(as.numeric(combined$year))
```

Visualisation for comparision

```
#Plot3 : Change in vote per party from 2016-2020 for Galway West
theme_set(theme_bw())
```

```
ggplot(combined, (aes(x= Abb, y=total_votes, fill=year))) +
  geom_col(position="fill") +
  scale_y_continuous(labels = scales::percent,
name = "Number of votes")+ xlab(label =
"Party") +
```



```

theme(  axis.line.y =
element_blank(),  axis.line.x =
element_blank(),  axis.ticks.x =
element_blank(),  panel.background =
element_blank(),  panel.ontop =
TRUE,
        panel.grid.major = element_blank(),
panel.grid.minor = element_blank(),

        legend.position= c(0.8, 0.9), legend.direction="horizontal")+
ggtitle("Change in vote per party from 2016-2020 for Galway West")
## PART 3 : A comparison of Galway West to the national average for party share of the vote for
2016 and 2020

```

#2016 National data

```

Ovr_2016 <- read.csv("Ovr_Nat2016.csv")
#Used NA for parties which are not necessary
Ovr_2016$Abb <- c("FG","FF","SF","LP","AAA","NA","NA","SD","GP","REN",
"NA", "DDI", "NA", "NA","NA","NA","NA","NA", "NA", "NA","IND","Total")

#Reference :http://www.endmemo.com/r/gsub.php
Ovr_2016$Votes <- as.numeric(gsub(",", "", Ovr_2016$Votes)) merged_2016 <-
left_join(data_2016_v1,Ovr_2016[,c("X2016.election.resu
lt","Votes","Abb")],by = 'Abb') merged_2016 <- merged_2016 %>%
mutate(prop_ovr = round(Votes/sum(Votes),4)*100)

```

#2020 National data

```

Ovr_2020 <- read.csv("Ovr_Nat2020.csv")
Ovr_2020$Abb <- c("SF","FF","FG","GP","LP","SD","S-PBP","NA","NA","NA"
, "AO", "NA", "NA", "NA","NA","NA","NA","NA", "IND","Total")
Ovr_2020$Votes <- as.numeric(gsub(",", "", Ovr_2020$Votes))

```

```

merged_2020 <- left_join(data_2020_v1,Ovr_2020[,c("X2020.election.resu
lt","Votes","Abb")],by = 'Abb') merged_2020 <- merged_2020 %>%
mutate(prop_ovr = round(Votes/sum(Votes),4)*100)

```

```

merged_2020$X2020.election.result <- NULL
merged_2020<-merged_2020%>%mutate(Abb = fct_reorder(Abb, Votes))%>%arr
ange(desc(Votes))

```

```

merged_2016$X2016.election.result <- NULL
merged_2016<-merged_2016%>%mutate(Abb = fct_reorder(Abb, Votes))%>%arr
ange(desc(Votes))

```

```

#Combine the data for 2016 and 2020 of both National and Galway West
merged <- rbind(merged_2016,merged_2020)

#Plot 4 : Comparison of Galway West to the national average(2016-2020)
ggplot(merged,aes(Abb,prop_ovr,fill=prop))+
  geom_bar(stat="identity",position="dodge")+ facet_wrap(~year)+
  ggtitle("Comparison of Galway West to the national average(2016-2020
)")+
  ylab(label = "National Votes (Percentage)")+
  xlab(label = "Party")+ coord_flip(clip =
"off")+ labs(fill = "Galway votes %")+
  geom_text(aes(label = paste0(round(prop,1), "%")), size = 3, vjust=0
,hjust = 1,color = "white")

```

PART 4 : The change in vote for the most significant candidates in both elections in Galway West

```

#Select Top 5 most significant candidates in 2016 and 2020 com
<- full_join(data_2016_v1,data_2020_v1,by = 'Abb') com$diff <-
com$total_votes.y - com$total_votes.x
com_1 <- combined %>% filter (Abb %in% c('FG', 'SF','GP','FF','LP'))

#Plot5 : Change in vote for top 5 significant candidates in 2016-2020
plot5 <- ggplot(com_1, aes(x = total_votes, y= Abb)) +
  geom_point(aes(colour = year),size=3) + geom_line(aes(group = Abb))
+
  geom_text(aes(label=year, colour = year),size =3,vjust =-1)+
  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(0.70,
0.90), legend.text = element_text(size = 10)
, legend.title = element_blank())+
  ggtitle("Change in vote for top 5 significant candidates in 2016-2020"
) plot5

```

```
#plot 6 : CVD simulation plot  
plot_cvd <- cvd_grid(plot5) plot_cvd
```