Data Visualisation: Assignment 3

Name: Marcel Aguilar Garcia Student Id: 20235620 Class: 2021-CT5136

1. Introduction

The aim of this report is to analyse the Irish election data for 2016 and 2020 for the Galway West constituency and to provide insights with the support of data visualisations.

2. The vote per party in each election for Galway West

In this first section, I analyse the votes of each party in Galway West for the election of 2016 and 2020. The votes are taken from *count 1* and have been grouped by adding up the votes of all candidates in each party.

As seen in Figure 1, in the election of 2016, the group of independent parties had the greater percentage of votes (26.2 %), followed by Fianna Fáil (25.8 %) and Fine Gael (25.5 %) respectively. These three parties together represented a total of 77.5 % of votes. The remaining votes were distributed between Sinn Féin (9.5 %), the Social Democratic Party (5.7 %), and the Labour Party (5.3 %). Finally, other small parties represented a 2 % of the total votes.

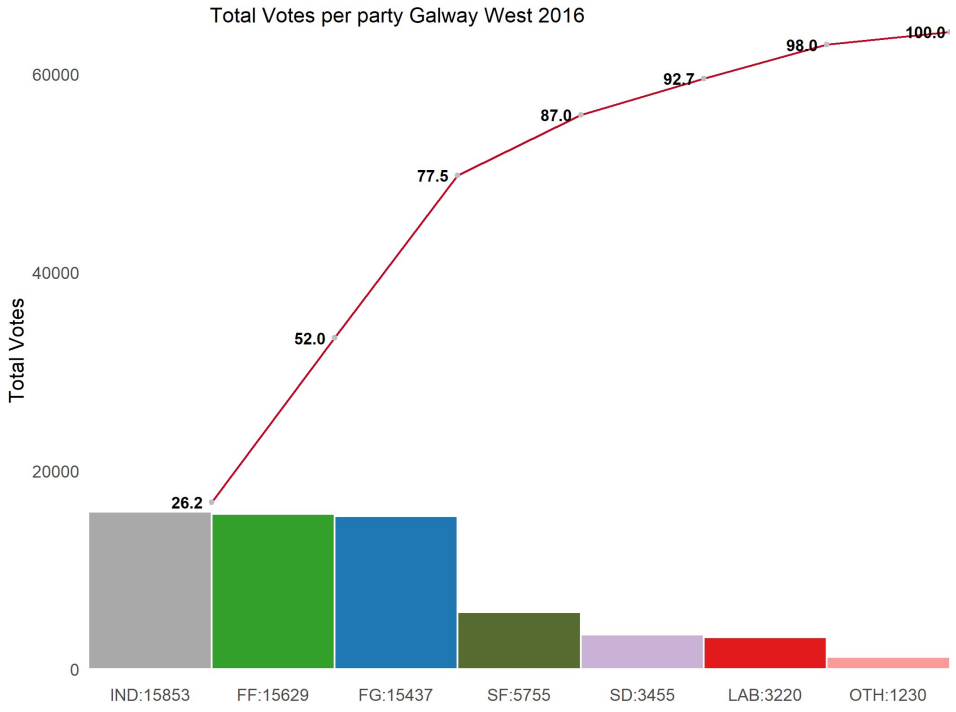

Figure 1

Figure 2 shows the election of 2020 in which the group of independent parties had the greater percentage of votes (28.4 %), followed by Fianna Fáil (23.6 %) and Fine Gael (18.8 %) respectively. These three parties

together represented a total of 70.8 % of votes. The remaining votes were distributed between Sinn Féin (14.6 %), the Green Party (6.3 %), and the Social Democratic Party (6.3 %). Finally, other small parties represented around 2 % of the total votes.

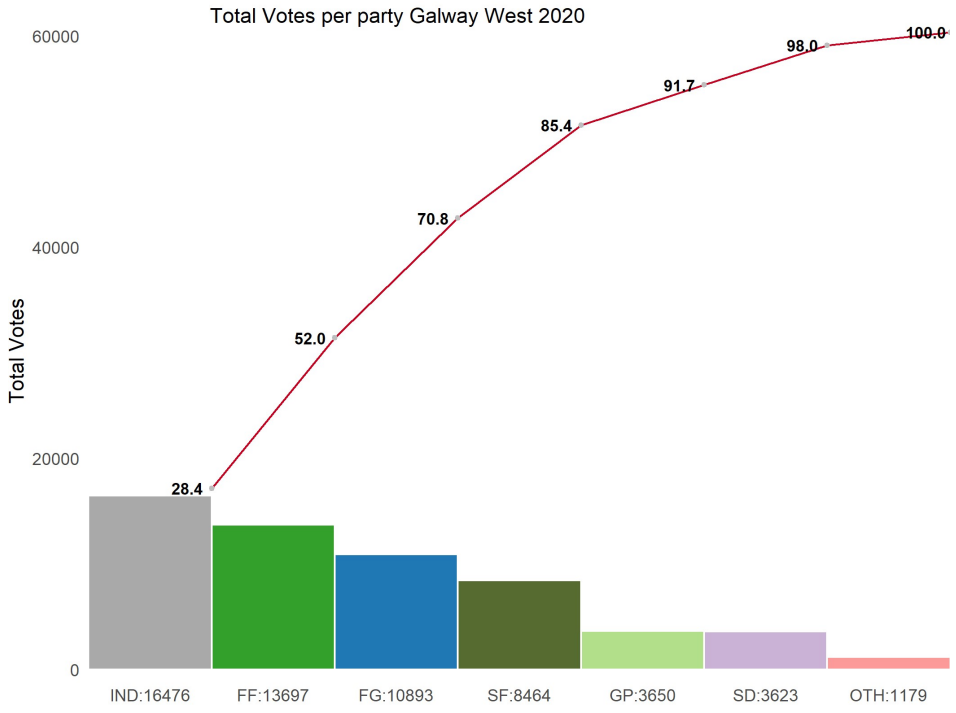

Figure 2

Design choices: I have decided to use a Pareto Chart to show the total number of votes and, at the same time, the percentage of votes in each election. By using the accumulative percentage, it is easier to see that, e.g., the three main parties had over 70 % of the total votes in each election. Axis x can be hidden as the labeling in the columns provides enough information. Axis y helps to visualize the total votes related to the accumulative percentage. I have used the party colours defined in the lectures. These colours are aligned with the parties and have been used for most of the visualisations in this assignment.

3. The change in vote per party from 2016-2020 for Galway West

Now that we have analysed the results in Galway West for 2016 and 2020, we can compare the results of both elections.

As seen in Figure 3, the party that has seen the greatest decrease of votes was Fine Gael, losing a total of 6.69 % of its representation in 2020, followed by the Labour Party that lost most of its votes in the last election. On the other hand, the Green Party went from being a minority party in Galway West in 2016 to have a total representation of 6.3 % of votes in 2020. In a similar way, Sinn Féin has seen a substantial increase of 5.1 % in comparison with 2016. Smaller changes can be seen to Fianna Fáil (-2.18 %), the Social Democratic Party (+0.55 %), and the group of independent parties (+2.3 %).

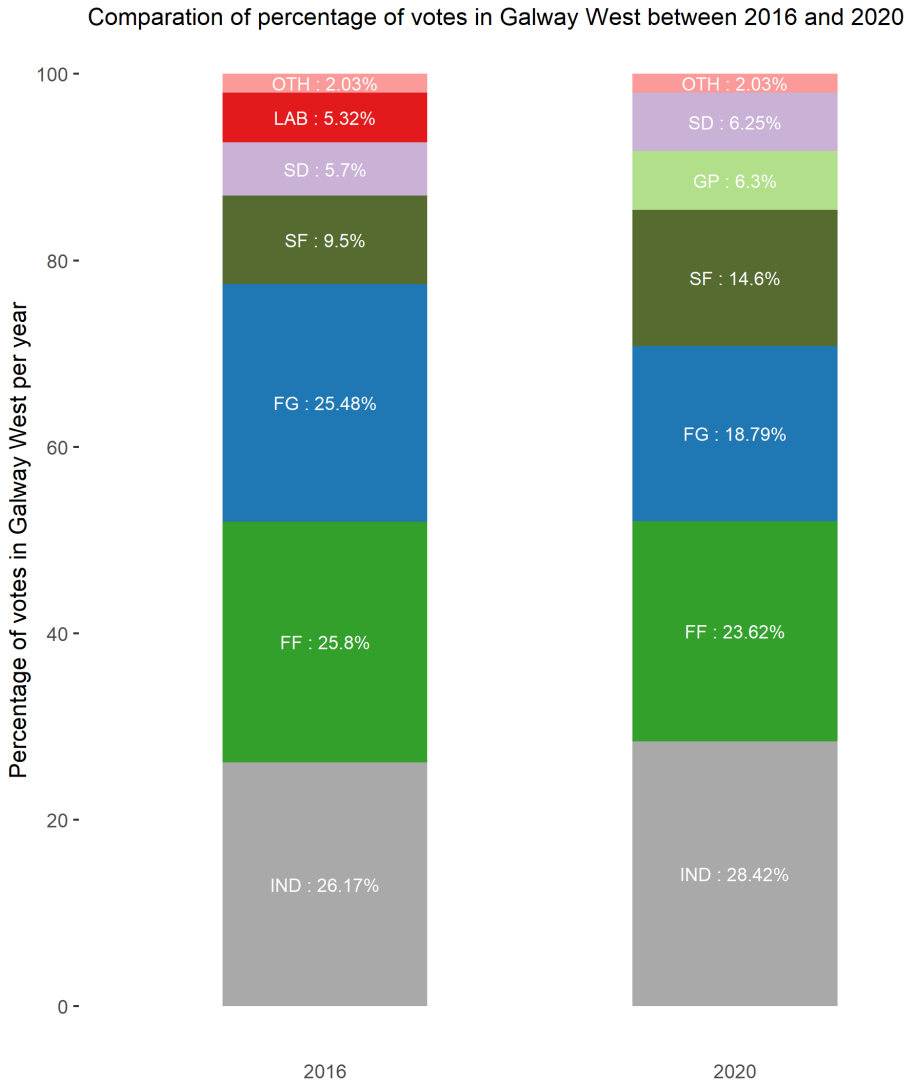

Figure 3

Design choices: As this is a comparison of both elections, I could have based the data visualisation on the difference of votes or percentages. However, by doing so, the visualisation would just show this difference. I have decided to use a Grouped Vertical Bar per year to help the reader visualise the number of votes of the party and not just the difference. E.g, while the Green Party had a substantial increase, I want to show that the representation in 2020 is relatively low compared to the other parties. I tried to have a high data-ink ratio by hiding axis x and having a very simple axis y. The only reason why I kept axis y was to show the accumualtive percentage of the parties.

4. A comparison of Galway West to the national average for party share of the vote for 2016 and 2020

In this section, I aim to analyse the difference of votes between the national average and the Galway average for 2016 and 2020.

Figure 4 and Figure 5 shows that there are many similarities when comparing Galway West to national

averages in 2016 and 2020. In both years, Galway West had a much larger representation of independent parties than the national averages, as much as 15.04 % in 2016, and 16.22 % in 2020. On the other hand, Sinn Féin had a much lower percentage of votes in Galway West compared to the national average. This difference has increased from 4.35 % up to 9.93 %. Other important parties such as Fianna Fáil and Fine Gael do not present a substantial difference compared to the national averages. In both cases, the difference is lower than 3 %.

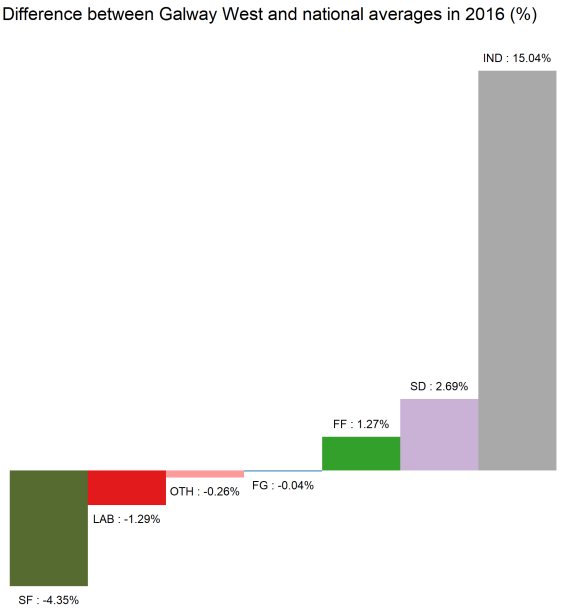

Figure 4

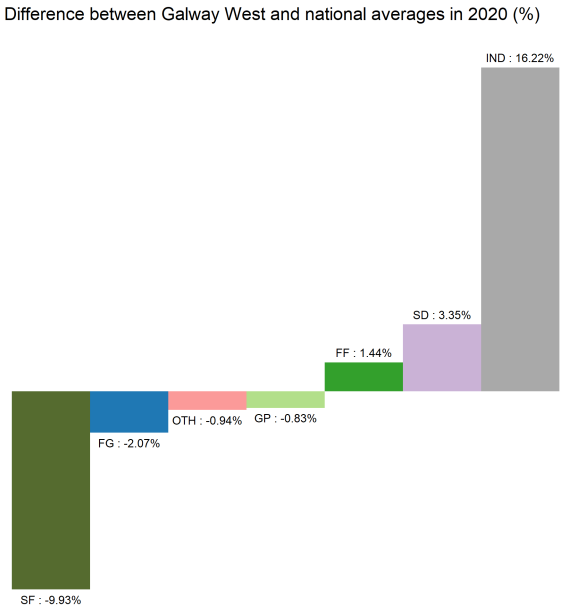

Figure 5

Design choices: I first thought of using a Group Bar Chart with two categories, national average and Galway West average. However, I did not like the idea of using the same party colour for both categories or not using the party colour at all. I finally used a bar chart representing just the variable for the difference between the national average and Galway West average. By placing both plots next to each other, it is easy to compare the proportions of each party between 2016 and 2020. I did not find a need of using axis as all information can be summarised in the label. As some bars were too short, the label was unreadable so I decided to place it outside.

5. The change in vote for the most significant candidates in both elections in Galway West Group Bar Chart

In this final section, I would like to show the difference of votes with the most important candidates in Galway West between 2016 and 2020. I am just considering candidates that have at least 4000 votes in each election. These candidates are Eamon O'Cuiv (Fianna Fáil), Noel Grealish (Independent), Sean Kyne (Fine Gael), Catherine Connolly (Independent), and Hildergarde Naugthon (Fine Gael).

As seen in Figure 6, Catherine Connolly and Noel Grealish, both independent candidates have seen an increase of votes in 2020. Hildergarde Naugton (Fine Gael) has raised two positions compared to the previous year and it is the only non independent candidate that has seen an increase of votes. The two candidates with most votes from Fianna Fáil and Fine Gael, Eamon O'Cuiv (Fianna Fáil) and Sean Kyne (Fine Gael), had a considerable decrease of votes in 2020.

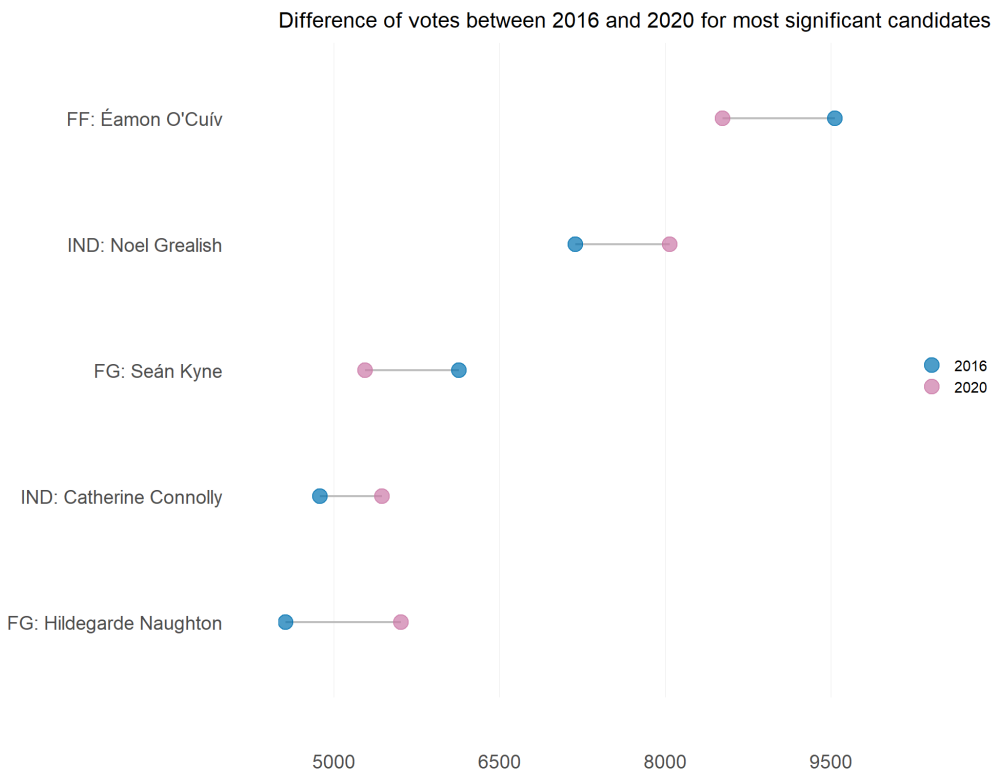

Figure 6: Colours selected from Kabe and Ito Palette

Design choices: In this section, the main focus is the difference of votes for each candidate. This dot plot with two categories is able to summarise the number of votes in each year and the difference between them. The line connecting both dots helps to compare this difference between all candidates. This plot has an alternative colouration for readers with CVD. **The colour selection is from Kabe and Ito Palette.**. This plot would be completely useless if someone is not able to distinct the colour of the years as would not recognise if a candidate had an increase or decrease of votes.

References

[1] Galway West data 2016 https://www.oireachtas.ie/en/elections/

[2] Galway West data 2020 *Galway West Fine Gael Party officials*

[3] National Average data 2016 https://www.tcd.ie/Political_Science/people/michael_gallagher/Election2016.php

[4] National Average data 2020 https://www.tcd.ie/Political_Science/people/michael_gallagher/Election2020.php

Appendix: R Code

```
#Loading required libraries

library(readxl)
library(tidyverse)
library(ggplot2)
library(viridis)
library(scales)
library(forcats)
library(dplyr)
library(knitr)


#Defining column format for 2020 dataset
column_data_types <- c("text","text","text","numeric","numeric","numeric","numeric",
                       "numeric","numeric","numeric","numeric","numeric","numeric",
                       "numeric","numeric","numeric","numeric")

#Loading 2020 dataset from excel file - ignoring header (there is an extra line)
data2020 <- read_excel('GalwayWest-2020.xlsx',skip=1,col_types = column_data_types)
data2020$Party <- NULL
data2020 <- head(data2020,-1)

#Renaming columns to more understandable/intuitive name
#I just need Party and Votes for each year
data2020_by_party <- data2020 %>%
  dplyr::rename(
    Party = '...2',
    FPvGalway = `FPv%`,
    Candidate = 'Candidate',
    Votes = 'Count 1',
    Count2 = 'Count 2',
    Count3 = 'Count 3',
    Count4 = 'Count 4',
    Count5 = 'Count 5',
    Count6 = 'Count 6',
    Count7 = 'Count 7',
    Count8 = 'Count 8',
    Count9 = 'Count 9',
    Count10 = 'Count 10',
    Count11 = 'Count 11',
    Count12 = 'Count 12',
    Count13 = 'Count 13'
  ) %>%
  select('Party','Votes') %>%
  group_by(Party)

#Adding all the votes from each party together - independent parties
#will be treated as one party
data2020 <- data2020_by_party %>%
  dplyr::summarise(totalVotes= sum(Votes))


#Renaming parties for consistency with 2016 dataset
data2020[5,'Party'] <- 'Independents'

#Loading National Averages from 2020 and joining with dataset for
```

```
#Future comparaition
nationalaverage2020 <- read_csv('NationalAverage2020.csv')
data2020 <- data2020 %>% inner_join(nationalaverage2020)


#Loading 2016 dataset from csv
data2016 <- read_csv('2016-04-28_general-election-count-details-galway-west-csv_en.csv')

#Renaming columns to more understandable/intuitive name
#I just need Party and Votes for each year
data2016_by_party <- data2016 %>%
  dplyr::rename(
    Constituency = 'Constituency Name',
    Surname = 'Candidate surname',
    FirstName = 'Candidate First Name',
    Result = 'Result',
    CountNumber = 'Count Number',
    Transfers = 'Transfers',
    Votes = 'Votes',
    totalVotes = 'Total Votes',
    CandidateId = 'Candidate Id',
    Party = 'Party',
  ) %>%
  filter(CountNumber == 1) %>%
  select('Party','CountNumber','totalVotes')%>%
  group_by(Party)

#Adding all the votes from each party together - independent parties
#will be treated as one party
data2016 <- data2016_by_party %>%
  dplyr::summarise(totalVotes = sum(totalVotes))

#Renaming parties for consistency with 2020 dataset
data2016[3,'Party'] <- 'Fianna Fáil'
data2016[4,'Party'] <- 'Fine Gael'
data2016[6,'Party'] <- 'Independents'
data2016[9,'Party'] <- 'Sinn Féin'
data2016[10,'Party'] <- 'Social Democrats'
nationalaverage2016 <- read_csv('NationalAverage2016.csv')

#Loading National Averages from 2016 and joining with dataset for
#Future comparaition
data2016 <- data2016 %>% inner_join(nationalaverage2016)

#Joining rows from 2016 and 2020 datasets in one tibble with
#an extra column with their year 2016/2020
data2016['year'] = "2016"
data2020['year'] = "2020"
data <- bind_rows(data2016,data2020)

#Grouping all parties that have less than 2500 votes into a
#category for "Others"
data <-data %>% mutate(Party = ifelse(totalVotes < 2500, 'Others',Party))
data_by_party_year<-data %>% group_by(Party,year)
data <- data_by_party_year%>%
  dplyr::summarise(
    totalVotes = mean(totalVotes),
    nationalAverage = mean(NationalAverage)
  )
```

```
data$totalVotes <-as.integer(data$totalVotes)


#Adding a short name for all parties - original names are too long to add in Plot
data$short <- plyr::mapvalues(data$Party,from=c("Fianna Fáil", "Fine Gael", "Green Party",
                                    "Independents","Labour Party","Others","Sinn Féin","Social
                                        Democrats"),
                      to=c("FF", "FG", "GP","IND","LAB","OTH","SF","SD"))


#Defining the different colours for each party - I used same colours as in Week 6 exercise
#which seem consistent with the media colours
party.colours <- c('FG' = '#1f78b4', 'FF' = '#33a02c', 'SF' = 'darkolivegreen',
                   'LAB' = '#e31a1c', 'GP' = '#b2df8a', 'SD' ='#cab2d6','SAD'='#cab2d6', 'IND' =
                       'darkgrey',
                   'REN' = '#ff7f00', 'DDI' ='darkgrey','AAA' ='darkgrey', "OTH"='#fb9a99',"S/PBP"=
                       '#fdbf6f')



#Calculating percentage of votes in GalwayWest per party and assigning to new column
data<-data %>% group_by(year) %>% dplyr::mutate(galwayAverage = 100*totalVotes/sum(totalVotes))

#Setting an order based on the Galway average
data$short <- fct_reorder(data$short,data$galwayAverage)
data$short2 <- data$short

#Difference of percentage between national average and Galway average
data$difference = data$galwayAverage-data$nationalAverage

#Assigning label names for the different plots
data$label_percentages <- paste0(data$short2, " : " ,round(data$difference,2), "%")
data$label_galwayAverage <- paste0(data$short2, " : " ,round(data$galwayAverage,2), "%")
data$label_votes <- paste0(data$short2,":",data$totalVotes)

#Creating plot of comparation of percentages between 2016 and 2020 in Galway West
g_galway_2016_2020 <-
  ggplot(data, aes(x=year,y=galwayAverage, fill=short)) +
  geom_col(width = 0.5) +
  geom_text(aes(label=label_galwayAverage, group = short), colour="white", size =3, position =
      position_stack(vjust = .5))+
  scale_y_continuous(limits = c(0, 100),
                     breaks = seq(0,100, by = 20),
                     name = "Percentage of votes in Galway West per year") +
  scale_fill_manual(values = party.colours)+
  ggtitle("Comparation of percentage of votes in Galway West between 2016 and 2020")+
  theme(
    legend.position = "none",
    axis.line.y = element_blank(),
    axis.line.x = element_blank(),
    axis.ticks.x = element_blank(),
    axis.title.x = element_blank(),
    plot.title = element_text(vjust = -2, hjust = 0.1, size = 11),
    plot.margin = margin(6, 6, 3, 3)
  )

#Creating plot - difference between Galway West and national averages in 2016
g_difference_2016 <- ggplot(data = subset(data,data$year==2016), aes(y = difference, x =
    reorder(short,difference), fill = short)) +
```

```
  geom_col(width=1,position = "dodge")+ scale_fill_manual(values = party.colours)+
  geom_text(aes(y = difference + 0.5 * sign(difference),label=label_percentages, group = short),
      colour="black", size =3,position = position_dodge(width = 0.9))+
  ggtitle("Difference between Galway West and national averages in 2016 (%)")+
  scale_y_continuous(name = "Difference of votes (%)")+
  theme(
    legend.position = "none",
    axis.line.y = element_blank(),
    axis.title.y = element_blank(),
    axis.title.x.bottom = element_blank(),
    axis.text.x.bottom = element_blank(),
    axis.text.y = element_blank(),
    axis.ticks.y = element_blank(),
    axis.line.x = element_blank(),
    axis.ticks.x = element_blank(),
    axis.text.x = element_text(vjust = 5),
    plot.margin = margin(20, 20, 20, 20),
    panel.background = element_blank() )

#Creating plot - difference between national averages and galway west in 2020
g_difference_2020 <- ggplot(data = subset(data,data$year==2020), aes(y = difference, x =
    reorder(short,difference), fill = short)) +
  geom_col(width=1,position = "dodge")+ scale_fill_manual(values = party.colours)+
  geom_text(aes(y = difference + 0.5 * sign(difference),label=label_percentages, group = short),
      colour="black", size =3,position = position_dodge(width = 0.9))+
  ggtitle("Difference between Galway West and national averages in 2020 (%)")+
  scale_y_continuous(name = "Difference of votes (%)")+
  theme(
    legend.position = "none",
    axis.line.y = element_blank(),
    axis.title.y = element_blank(),
    axis.title.x.bottom = element_blank(),
    axis.text.x.bottom = element_blank(),
    axis.text.y = element_blank(),
    axis.ticks.y = element_blank(),
    axis.line.x = element_blank(),
    axis.ticks.x = element_blank(),
    axis.text.x = element_text(vjust = 5),
    plot.margin = margin(20, 20, 20, 20),
    panel.background = element_blank() )



data$short <- as.factor(data$short)
data$short = fct_reorder(data$short, -data$totalVotes)
data <- data %>% arrange(-galwayAverage, short)
data$cumulative <- ave(data$galwayAverage, data$short, FUN=cumsum)
data <- transform(data, cumulative=ave(galwayAverage,year,FUN=cumsum))
total_votes_2016 <- sum(data2016$totalVotes)
total_votes_2020 <- sum(data2020$totalVotes)

d16 <- subset(data,data$year==2016)
theme_set(theme_classic())

#Creating plot - final result, votes per party in Galway West 2016
g_galwaywest_2016 <- ggplot(d16, (aes(x= short, y=totalVotes, fill=short))) +
  geom_col(width=1, colour="white") +
```

```
  geom_line(aes(x = short, y =(cumulative*total_votes_2016)/100), position=position_nudge(x = 0.5,
      y = 0), group = 1, col="#ca0020") +
  geom_point(aes(x = short, y = (cumulative*total_votes_2016)/100), position=position_nudge(x =
      0.5, y = 0), size =1, col="grey") +
  geom_text(aes(fontface = 2, x = short, y = (cumulative*total_votes_2016)/100, label =
      sprintf("%.1f", cumulative)), size = 3, nudge_x = 0.3,nudge_y = 6) +
  scale_x_discrete(labels = d16$label_votes, expand = c(0,0.1))+
  scale_y_continuous(name = "Total Votes")+
  scale_fill_manual(values = party.colours)+
  ggtitle("Total Votes per party Galway West 2016") +
  theme(
    legend.position = "none",
    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(vjust = 8),
    axis.title.x = element_blank(),
    plot.title = element_text(vjust = -8, hjust = 0.25, size = 11),
    panel.background = element_blank(),
    panel.grid.major.y= element_blank(),
    panel.ontop = TRUE,
    plot.margin = margin(100, 0, 0, 0))

d20<-subset(data,data$year==2020)
#Creating plot - final result, votes per party in Galway West 2020
g_galwaywest_2020 <- ggplot(d20, (aes(x= short, y=totalVotes, fill=short))) +
  geom_col(width=1, colour="white") +
  geom_line(aes(x = short, y =(cumulative*total_votes_2020)/100), position=position_nudge(x = 0.5,
      y = 0), group = 1, col="#ca0020") +
  geom_point(aes(x = short, y = (cumulative*total_votes_2020)/100), position=position_nudge(x =
      0.5, y = 0), size =1, col="grey") +
  geom_text(aes(fontface = 2, x = short, y = (cumulative*total_votes_2020)/100, label =
      sprintf("%.1f", cumulative)), size = 3, nudge_x = 0.3,nudge_y = 6) +
  scale_x_discrete(labels = d20$label_votes, expand = c(0,0.1))+
  scale_y_continuous(name = "Total Votes")+
  scale_fill_manual(values = party.colours)+
  ggtitle("Total Votes per party Galway West 2020") +
  theme(
    legend.position = "none",
    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(vjust = 8),
    axis.title.x = element_blank(),
    plot.title = element_text(vjust = -8, hjust = 0.25, size = 11),
    panel.background = element_blank(),
    panel.grid.major.y= element_blank(),
    plot.margin = margin(100, 0, 0, 0),
    panel.ontop = TRUE
  )


#Defining column format for 2020 dataset
column_data_types <- c("text","text","text","numeric","numeric","numeric","numeric",
                       "numeric","numeric","numeric","numeric","numeric","numeric",
                       "numeric","numeric","numeric","numeric")
```

```r
#Loading 2020 dataset from excel file - ignoring header (there is an extra line)
data2020 <- read_excel('GalwayWest-2020.xlsx',skip=1,col_types = column_data_types)
data2020$Party <- NULL
data2020 <- head(data2020,-1)

#Renaming columns to more understandable/intuitive name
#I just need Party, Candidate, and Votes for each year
data2020_by_party <- data2020 %>%
  dplyr::rename(
    Party = '...2',
    FPvGalway = `FPv%`,
    candidate = 'Candidate',
    '2020' = 'Count 1',
    Count2 = 'Count 2',
    Count3 = 'Count 3',
    Count4 = 'Count 4',
    Count5 = 'Count 5',
    Count6 = 'Count 6',
    Count7 = 'Count 7',
    Count8 = 'Count 8',
    Count9 = 'Count 9',
    Count10 = 'Count 10',
    Count11 = 'Count 11',
    Count12 = 'Count 12',
    Count13 = 'Count 13'
  ) %>% select("Party","candidate","2020")


#Finally, I will reload the datasets for 2016 and 2020 to generate the plot
#for the candidates. I thought it is better to reload the whole dataset
#as the data has to be presented in a different way- not grouping by party.

#Reloading 2016 dataset from csv
data2016 <- read_csv('2016-04-28_general-election-count-details-galway-west-csv_en.csv')

#Renaming columns to more understandable/intuitive name
#I just need Party, Candidate, and Votes for each year
data2016_by_party <- data2016 %>%
  dplyr::rename(
    Constituency = 'Constituency Name',
    Surname = 'Candidate surname',
    FirstName = 'Candidate First Name',
    Result = 'Result',
    CountNumber = 'Count Number',
    Transfers = 'Transfers',
    '2016' = 'Votes',
    totalVotes = 'Total Votes',
    CandidateId = 'Candidate Id',
    Party = 'Party',
  ) %>%
  filter(CountNumber == 1)

#Similar data clean up than in FILE 1 but for candidates
data2016_by_party$candidate <- paste0(data2016_by_party$FirstName," ",data2016_by_party$Surname)

data2016_by_party <- data2016_by_party %>% select("Party","candidate","2016")

data2020_by_party[1,'candidate'] <- "Éamon O'Cuív"
data2016_by_party[8,'candidate'] <- "Seán Kyne"
```

```
data2016_by_party[15,'candidate'] <- "Catherine Connolly"

data_candidates <-data2016_by_party %>% inner_join(data2020_by_party, by = "candidate") %>%
    select ("Party.y","candidate","2016","2020")


data_candidates$short <- plyr::mapvalues(data_candidates$Party.y,from=c("Fianna Fáil", "Fine
    Gael", "Green Party",
                                                                    "Independent","Labour
                                                                        Party","Others","Sinn
                                                                        Féin","Social Democrats"),
                                         to=c("FF", "FG", "GP","IND","LAB","OTH","SF","SD"))


data_candidates$candidate <- paste0(data_candidates$short,": ",data_candidates$candidate)
data_candidates <- gather(data_candidates,key="year",value="votes",c("2016","2020"))

data_candidates <- data_candidates %>% filter(votes>4000)
data_candidates <-data_candidates %>% arrange(votes)

data_candidates$candidate <- as.factor(data_candidates$candidate)
data_candidates$candidate = fct_reorder(data_candidates$candidate, data_candidates$votes)

#Ploting candidates dot plot
g_candidates <- ggplot(data_candidates , aes(x = votes, y= candidate)) +
  geom_line(aes(group = candidate), colour = "grey", size=0.5) +
  geom_point(aes(colour = year), size = 3, alpha = 0.7) +
  scale_colour_manual(values= c("#0072B2","#CC79A7"), name = "") +
  scale_x_continuous(limits = c(4500, 10000),
                     expand = c(0, 0),
                     breaks = seq(5000, 10000, by = 1500),
                     name = "") +
  ggtitle("Difference of votes between 2016 and 2020 for most significant candidates")+
  theme(axis.title.y = element_blank(),
        panel.grid.major.x =element_line(size=0.03),
        panel.grid.major.y = element_blank(),
        panel.background = element_blank(),
        axis.line.y = element_blank(),
        axis.line.x = element_blank(),
        axis.ticks.y = element_blank(),
        axis.ticks.length = unit(.85,"cm"),
        axis.ticks.x = element_blank(),
        legend.text = element_text(size = 7), # legend text was a little large
        legend.key.size = unit(0.7, "lines"),
        legend.title = element_blank(),
        legend.background = element_blank(),
        legend.key = element_blank(),
        plot.title = element_text(size=10),
        plot.margin = margin(75, 0, 75, 0)
  )


#Saving all the visualisations
ggsave("g_galwaywest_2016.png",plot=g_galwaywest_2016)
ggsave("g_galwaywest_2020.png",plot=g_galwaywest_2020)
ggsave("percentage_per_year_galway.png",plot=g_galway_2016_2020)
ggsave("difference_2016.png",plot=g_difference_2016)
ggsave("difference_2020.png",plot=g_difference_2020)
ggsave('candidates.png',plot = g_candidates)
```