POL40950: Introduction to Statistics

Homework 1

Nilesh Nayak

Deadline: 7 October 2021

Instructions

- Rename this file to your LASTNAME_FIRSTNAME_homework01.Rmd and insert your name at the header of this script (see `INSERT YOUR NAME`).
- If an answer to a question requires code, add the code in the code block below. For questions that need interpretations or explanations, write your answer in *italics* (using _ and _ at the beginning and end of your answer) below the question.
- **This script will only knit if `data_which_candidate.rds` and `dat_elections.csv` are stored in the same folder as this RMarkdown file! Both datasets are in the "Homeworks" folder on Brightspace.**
- Please knit this file as an .html document and upload the .html document to the assignment folder on Brightspace.
- If the code for one of the questions is not working, leave the code in the chunk, but change the beginning of the chunk to `{r,eval=FALSE}`. Only do this as a last resort, though.

install.packages("dplyr")

```
# load required package
library(tidyverse)
library(dplyr)
library(ggplot2)
```

Working with and Wrangling Survey Data (100 points + 5 bonus points)

In this homework, we work with responses to the Voting Advice Application "Which Candidate". You can see the raw data at: https://doi.org/10.7910/DVN/OEYRN7.

We use a cleaned version of the original file. Let's load the file:

```
# make sure to store the datsets in the same folder as this .Rmd file!
dat_vaa <- readRDS("data_which_candidate.rds")
```

1. How many observations does the dataset consist of? What is the unit of analysis? And which variables are included in this dataset (check out `names()`)? [5 points]

```
dim(dat_vaa)
```

```
## [1] 141936      4
```

There are 141936 observations and 4 variables before removing missing values

```
head(dat_vaa,10)
```

```
## # A tibble: 10 x 4
## # Groups:   partyvote [5]
##    immig partyvote        lr_selfplacement age
##    <dbl> <chr>                       <dbl> <fct>
##  1    NA <NA>                           NA <NA>
##  2    NA <NA>                           NA <NA>
##  3    NA <NA>                           NA <NA>
##  4     0 Social Democrats                2 18-24
##  5     0 <NA>                           NA <NA>
##  6     0 <NA>                           NA <NA>
##  7     1 <NA>                            3 25-34
##  8     1 Fianna Fáil                     2 25-34
##  9     0 Fine Gael                       4 35-44
## 10    NA Green Party                     5 35-44
```

```
summary(dat_vaa)
```

```
##      immig        partyvote         lr_selfplacement      age
##  Min.   :0.000   Length:141936      Min.   : 0.00    25-34  :30080
##  1st Qu.:0.000   Class :character   1st Qu.: 3.00    35-44  :22187
##  Median :0.000   Mode  :character   Median : 4.00    18-24  :19227
##  Mean   :0.272                      Mean   : 3.92    45-54  : 9573
##  3rd Qu.:1.000                      3rd Qu.: 5.00    55-65  : 4975
##  Max.   :1.000                      Max.   :10.00    (Other): 1967
##  NA's   :28892                      NA's   :61591    NA's   :53927
```

```
names(dat_vaa)
```

```
## [1] "immig"            "partyvote"        "lr_selfplacement" "age"
```

There are 4 variables in the dataset : immig, partyvote, lr_selfplacement, age

```
dat_vaa_filtered = dat_vaa %>% drop_na()
dim(dat_vaa_filtered)
```

```
## [1] 45880     4
```

There are 45880 observations and 4 variables after removing missing values

```
summary(dat_vaa_filtered)
```

```
##      immig         partyvote         lr_selfplacement       age
##  Min.   :0.0000   Length:45880       Min.   : 0.000   under 18:    0
##  1st Qu.:0.0000   Class :character   1st Qu.: 2.000   18-24   : 9547
##  Median :0.0000   Mode  :character   Median : 4.000   25-34   :15214
##  Mean   :0.2407                      Mean   : 3.753   35-44   :11696
##  3rd Qu.:0.0000                      3rd Qu.: 5.000   45-54   : 5319
##  Max.   :1.0000                      Max.   :10.000   55-65   : 2899
##                                                       over 65 : 1205
```

```
head(dat_vaa_filtered,10)
```

```
## # A tibble: 10 x 4
## # Groups:   partyvote [5]
##    immig partyvote        lr_selfplacement age
##    <dbl> <chr>                       <dbl> <fct>
##  1     0 Social Democrats                2 18-24
##  2     1 Fianna Fáil                     2 25-34
##  3     0 Fine Gael                       4 35-44
##  4     0 Green Party                     3 18-24
##  5     0 S-PBP                           0 18-24
##  6     0 Social Democrats                3 55-65
##  7     0 S-PBP                           1 18-24
##  8     0 Green Party                     2 18-24
##  9     0 Green Party                     2 35-44
## 10     0 Green Party                     3 25-34
```

Unit of analysis is -> Party vote share per age group at each country level which support more/less restrictive immigration policies

2. Group the data frame by `partyvote` (check `group_by()`) and get the absolute frequency (`count()`) of respondents per party. [5 points]

```
dat_vaa_group_partyvote_count <- dat_vaa_filtered %>% group_by(partyvote) %>% count()
names(dat_vaa_group_partyvote_count)[2] <- "count"
dat_vaa_group_partyvote_count
```

```
## # A tibble: 9 x 2
## # Groups:   partyvote [9]
##   partyvote        count
##   <chr>            <int>
## 1 Aontú              424
## 2 Fianna Fáil       5169
## 3 Fine Gael         6647
## 4 Green Party      10514
## 5 Labour            2994
## 6 Other/Ind         4175
## 7 S-PBP             2489
## 8 Sinn Féin         8577
## 9 Social Democrats  4891
```

3. Calculate the relative frequencies of respondents (=proportions) in each `partyvote` group. Note: you will find suggestions on how to calculate proportions per group online at StackOverFlow. [10 points]

```
dat_vaa_group_partyvote_freq = dat_vaa_filtered %>%  group_by(partyvote) %>% summarise(count = n()) %>%
  arrange(desc(relative_frequency))

dat_vaa_group_partyvote_freq
```

```
## # A tibble: 9 x 3
##   partyvote        count relative_frequency
```

```
##   <chr>            <int>             <dbl>
## 1 Green Party      10514              22.9
## 2 Sinn Féin         8577              18.7
## 3 Fine Gael         6647              14.5
## 4 Fianna Fáil       5169              11.3
## 5 Social Democrats  4891              10.7
## 6 Other/Ind         4175               9.1
## 7 Labour            2994               6.5
## 8 S-PBP             2489               5.4
## 9 Aontú              424               0.9
```

4. Get the first-preference vote shares for Irish parties in 2020 from the ParlGov dataset (`dat_elections`). Are voters from certain parties over- or under-represented in the Voting Advice Application data compared with the official election results? Is this dataset representative of Irish voters? [15 points]

```
# load ParlGov election data
# removed fileEncoding = "UTF-8"
#dat_elections <- read.csv("dat_elections.csv",fileEncoding = "UTF-8")

dat_elections <- read.csv("https://parlgov.org/data/parlgov-development_csv-utf-8/view_election.csv")
dim(dat_elections)
```

```
## [1] 8673   16
```

```
dat_elections = dat_elections %>% drop_na()
dim(dat_elections)
```

```
## [1] 7159   16
```

```
head(dat_elections)
```

```
##   country_name_short country_name election_type election_date vote_share seats
## 1                AUS    Australia    parliament    1903-12-16       29.7    26
## 2                AUS    Australia    parliament    1903-12-16       34.4    25
## 3                AUS    Australia    parliament    1903-12-16       31.0    23
## 4                AUS    Australia    parliament    1906-12-02       38.2    27
## 5                AUS    Australia    parliament    1906-12-02       36.6    26
## 6                AUS    Australia    parliament    1906-12-02       16.4    16
##   seats_total party_name_short             party_name     party_name_english
## 1          75               PP    Protectionist Party    Protectionist Party
## 2          75              FTP       Free Trade Party       Free Trade Party
## 3          75              ALP Australian Labor Party Australian Labor Party
## 4          75              FTP       Free Trade Party       Free Trade Party
## 5          75              ALP Australian Labor Party Australian Labor Party
## 6          75               PP    Protectionist Party    Protectionist Party
##   left_right country_id election_id previous_parliament_election_id
## 1     7.4000         33         730                             731
## 2     6.0000         33         730                             731
## 3     3.8833         33         730                             731
## 4     6.0000         33         725                             730
## 5     3.8833         33         725                             730
## 6     7.4000         33         725                             730
```

```
##   previous_cabinet_id party_id
## 1                 997     1898
## 2                 997     1938
## 3                 997     1253
## 4                1000     1938
## 5                1000     1253
## 6                1000     1898
```

```
newdata <- subset(dat_elections, country_name=="Ireland" & substr(election_date,1,4)=="2020")
dim(newdata)
```

```
## [1]  9 16
```

```
newdata
```

```
##      country_name_short country_name election_type election_date vote_share
## 4105                IRL      Ireland    parliament    2020-02-08      22.18
## 4106                IRL      Ireland    parliament    2020-02-08      24.53
## 4107                IRL      Ireland    parliament    2020-02-08      20.86
## 4108                IRL      Ireland    parliament    2020-02-08       7.13
## 4109                IRL      Ireland    parliament    2020-02-08       4.38
## 4110                IRL      Ireland    parliament    2020-02-08       2.90
## 4111                IRL      Ireland    parliament    2020-02-08       2.63
## 4112                IRL      Ireland    parliament    2020-02-08       1.90
## 4113                IRL      Ireland    parliament    2020-02-08       0.39
##      seats seats_total party_name_short
## 4105    38         160               FF
## 4106    37         160               SF
## 4107    35         160               FG
## 4108    12         160            Green
## 4109     6         160              Lab
## 4110     6         160               DS
## 4111     5         160            D-PRB
## 4112     1         160                A
## 4113     1         160               IC
##                                    party_name
## 4105                             Fianna FÃ¡il
## 4106                               Sinn FÃ©in
## 4107                                Fine Gael
## 4108           Green Party  â\200" Comhaontas Glas
## 4109                               Labour Party
## 4110                  Daonlathaigh ShÃ³isialta
## 4111 DlÃºthphartÃocht â\200" Pobal Roimh BhrabÃºs
## 4112                                   AontÃº
## 4113                       Independents 4 Change
##                      party_name_english left_right country_id election_id
## 4105                         Fianna Fail     6.0713         37        1088
## 4106                           Sinn Fein     2.7935         37        1088
## 4107   Fine Gael (Familiy of the Irish)     6.4372         37        1088
## 4108                         Green Party     2.4350         37        1088
## 4109                        Labour Party     3.6252         37        1088
## 4110                   Social Democrats     3.3000         37        1088
## 4111 Solidarity -- People Before Profit     1.3000         37        1088
```

```
## 4112                              Aontu     7.4000         37        1088
## 4113              Independents 4 Change     1.3000         37        1088
##      previous_parliament_election_id previous_cabinet_id party_id
## 4105                            1002                1511      280
## 4106                            1002                1511     2217
## 4107                            1002                1511     1393
## 4108                            1002                1511     1573
## 4109                            1002                1511      318
## 4110                            1002                1511     2619
## 4111                            1002                1511     2776
## 4112                            1002                1511     2795
## 4113                            1002                1511     2621
```

```
newdata %>% select(party_name_english, vote_share)
```

```
##                      party_name_english vote_share
## 4105                         Fianna Fail      22.18
## 4106                           Sinn Fein      24.53
## 4107   Fine Gael (Familiy of the Irish)      20.86
## 4108                         Green Party       7.13
## 4109                        Labour Party       4.38
## 4110                   Social Democrats        2.90
## 4111 Solidarity -- People Before Profit       2.63
## 4112                              Aontu       1.90
## 4113              Independents 4 Change       0.39
```

```
cbind(newdata %>% select(party_name_english, vote_share), partyvote = dat_vaa_group_partyvote_freq$party
```

```
##                      party_name_english vote_share        partyvote
## 4105                         Fianna Fail      22.18      Green Party
## 4106                           Sinn Fein      24.53        Sinn Féin
## 4107   Fine Gael (Familiy of the Irish)      20.86        Fine Gael
## 4108                         Green Party       7.13      Fianna Fáil
## 4109                        Labour Party       4.38 Social Democrats
## 4110                   Social Democrats        2.90        Other/Ind
## 4111 Solidarity -- People Before Profit       2.63           Labour
## 4112                              Aontu       1.90            S-PBP
## 4113              Independents 4 Change       0.39            Aontú
##      relative_frequency
## 4105               22.9
## 4106               18.7
## 4107               14.5
## 4108               11.3
## 4109               10.7
## 4110                9.1
## 4111                6.5
## 4112                5.4
## 4113                0.9
```

Green Party and Social Democrats are over represented as survey respondents are very high, but vote share received is lesser.

Fianna Fail , Fine Gael, Sinn Fein are under represented as survey respondents are lower, but vote share received is higher.

5. The variable `immig` takes the value 1 if a respondent expressed that immigration "should be more restrictive". (The original question is: "Should immigration into Ireland be made more restrictive or less restrictive?"). What proportion of *all* participants agrees that immigration should be more restrictive? [5 points]

```
dat_vaa %>%  group_by(immig) %>% summarise(cnt = n()) %>% mutate(freq = round(cnt / sum(cnt)*100, 3)) %
```

```
## # A tibble: 3 x 3
##   immig   cnt  freq
##   <dbl> <int> <dbl>
## 1     0 82252  58.0
## 2     1 30792  21.7
## 3    NA 28892  20.4
```

21.69% agree that immigration should be more restrictive

I have considered the "dat_vaa" dataframe which has NA's. I think the NA's add the impact in the "all participants list" since these respondents could voluntarily chose to not provide an input in the survey

6. Calculate these proportions separately for supporters from each party (using `partyvote` as the grouping variable). [5 points]

```
dat_vaa_filtered %>%  group_by(partyvote, immig) %>%  summarise(cnt = n()) %>% mutate(freq = round(cnt /
```

```
## `summarise()` has grouped output by 'partyvote'. You can override using the `.groups` argument.
```

```
## # A tibble: 18 x 4
## # Groups:   partyvote [9]
##    partyvote        immig   cnt  freq
##    <chr>            <dbl> <int> <dbl>
##  1 Aontú                0   209 49.3
##  2 Aontú                1   215 50.7
##  3 Fianna Fáil          0  3162 61.2
##  4 Fianna Fáil          1  2007 38.8
##  5 Fine Gael            0  4888 73.5
##  6 Fine Gael            1  1759 26.5
##  7 Green Party          0  9739 92.6
##  8 Green Party          1   775  7.37
##  9 Labour               0  2469 82.5
## 10 Labour               1   525 17.5
## 11 Other/Ind            0  2532 60.6
## 12 Other/Ind            1  1643 39.4
## 13 S-PBP                0  2177 87.5
## 14 S-PBP                1   312 12.5
## 15 Sinn Féin            0  5367 62.6
## 16 Sinn Féin            1  3210 37.4
## 17 Social Democrats     0  4295 87.8
## 18 Social Democrats     1   596 12.2
```

I have considered the "dat_vaa_filtered" dataframe initially since each party would specifically want to understand immig = 1 versus immig = 0 excluding the NA's to make a decision

7. Create a barplot with the `partyvote` on the x-axis and the proportion of respondents who favour more restrictive immigration policies on the y-axis. [5 points]

```
data_immig = dat_vaa_filtered %>%  group_by(partyvote, immig) %>%  summarise(cnt = n()) %>% mutate(freq
```

```
## `summarise()` has grouped output by 'partyvote'. You can override using the `.groups` argument.
```

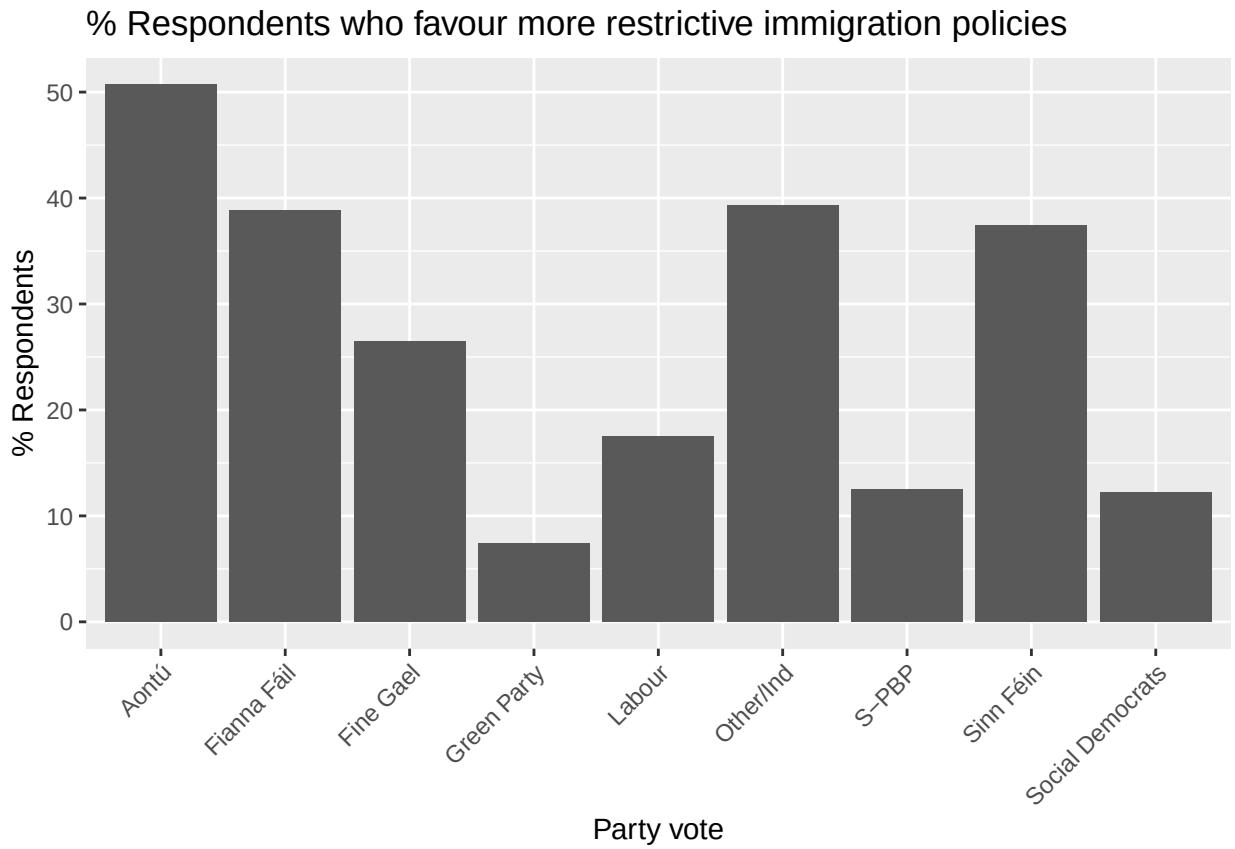
```
data_immig
```

```
## # A tibble: 18 x 4
## # Groups:   partyvote [9]
##    partyvote        immig   cnt  freq
##    <chr>            <dbl> <int> <dbl>
##  1 Aontú                0   209 49.3
##  2 Aontú                1   215 50.7
##  3 Fianna Fáil          0  3162 61.2
##  4 Fianna Fáil          1  2007 38.8
##  5 Fine Gael            0  4888 73.5
##  6 Fine Gael            1  1759 26.5
##  7 Green Party          0  9739 92.6
##  8 Green Party          1   775  7.37
##  9 Labour               0  2469 82.5
## 10 Labour               1   525 17.5
## 11 Other/Ind            0  2532 60.6
## 12 Other/Ind            1  1643 39.4
## 13 S-PBP                0  2177 87.5
## 14 S-PBP                1   312 12.5
## 15 Sinn Féin            0  5367 62.6
## 16 Sinn Féin            1  3210 37.4
## 17 Social Democrats     0  4295 87.8
## 18 Social Democrats     1   596 12.2
```

```
data_immig1 = subset(data_immig, data_immig$immig == '1')
data_immig1
```

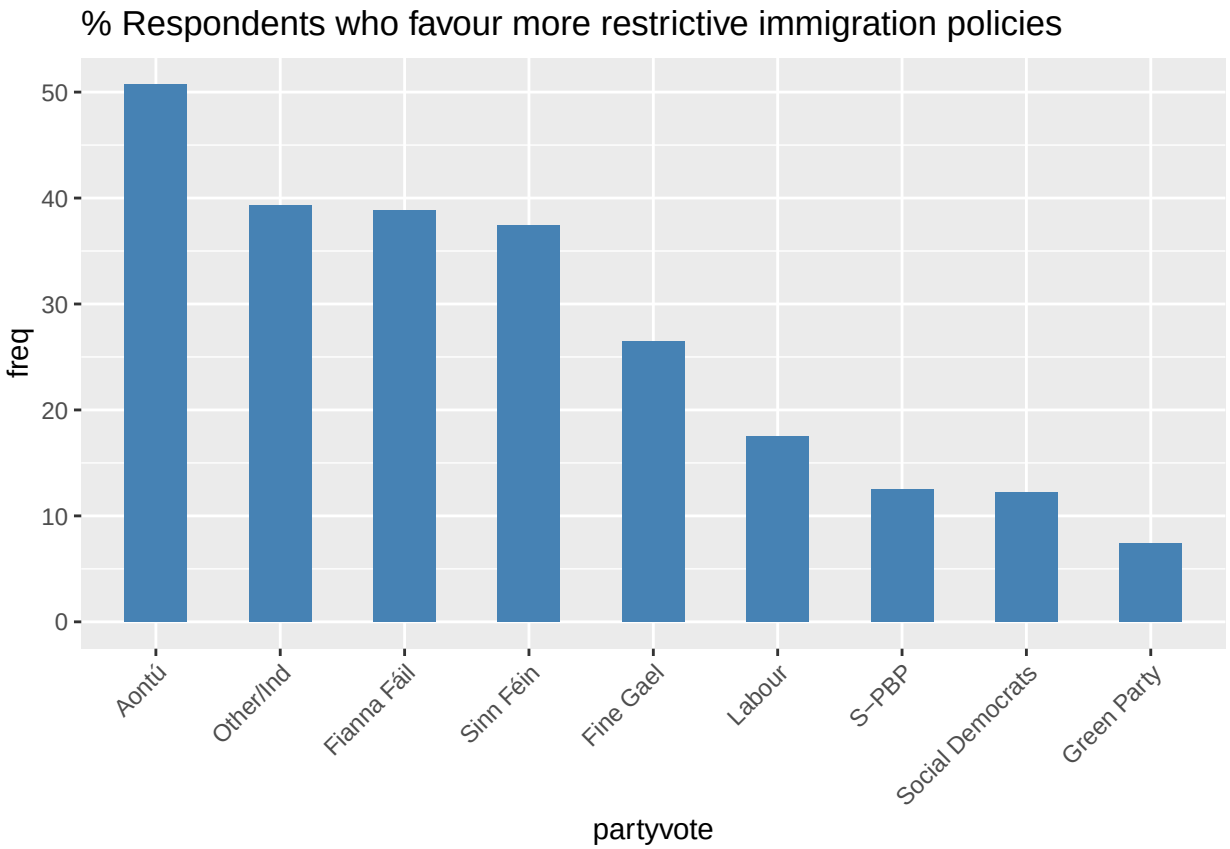
```
## # A tibble: 9 x 4
## # Groups:   partyvote [9]
##   partyvote        immig   cnt  freq
##   <chr>            <dbl> <int> <dbl>
## 1 Aontú                1   215 50.7
## 2 Fianna Fáil          1  2007 38.8
## 3 Fine Gael            1  1759 26.5
## 4 Green Party          1   775  7.37
## 5 Labour               1   525 17.5
## 6 Other/Ind            1  1643 39.4
## 7 S-PBP                1   312 12.5
## 8 Sinn Féin            1  3210 37.4
## 9 Social Democrats     1   596 12.2
```

```
data_immig1 %>% ggplot(aes(x = partyvote, y = freq)) + geom_col()+
  theme(axis.text.x = element_text(angle = 45, hjust=1)) + labs(x = "Party vote", y = "% Respondents",
```



8. Reorder the parties on the x-axis in descending order (the party with the highest proportions should be the first party in the graph). [5 points]

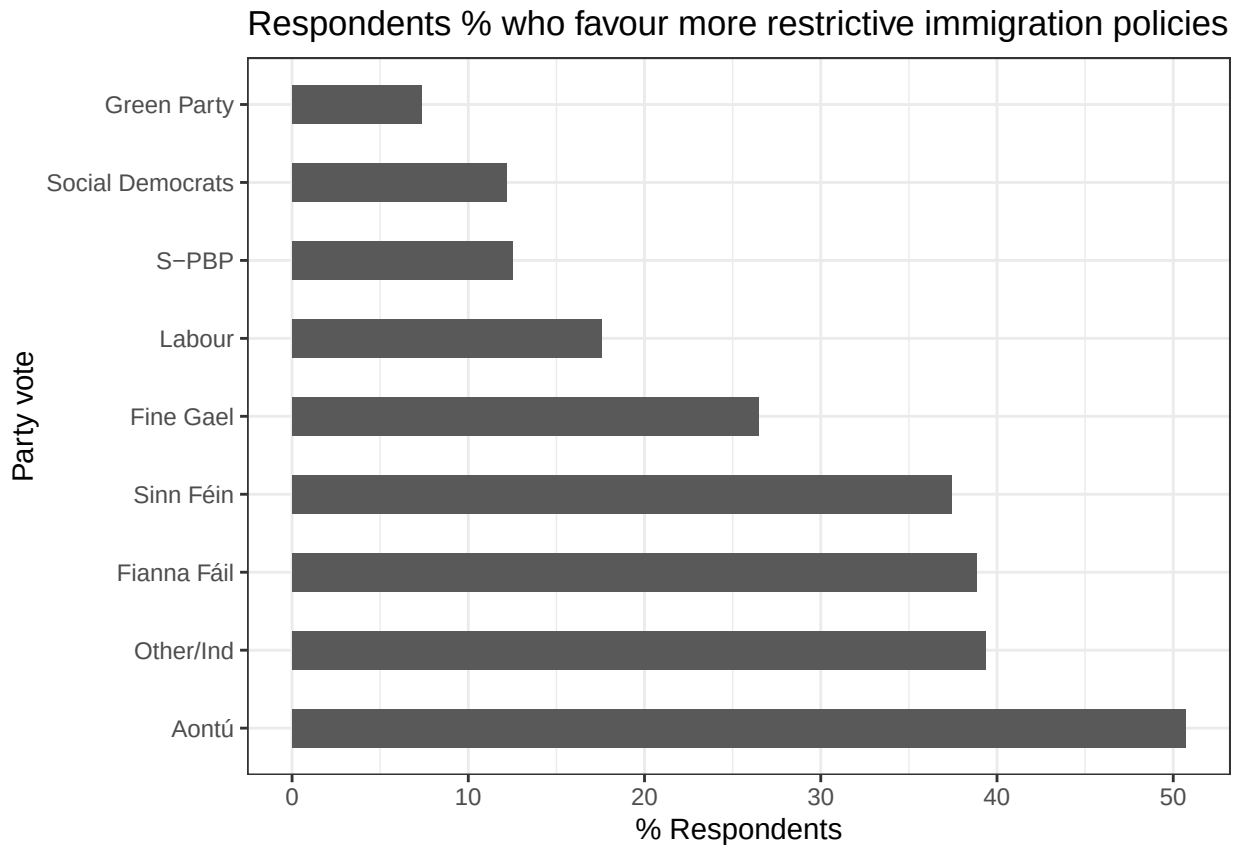
```
data_immig1$partyvote = factor(data_immig1$partyvote, levels = data_immig1$partyvote[order(data_immig1$

data_immig1 %>% ggplot(aes(x = partyvote, y = freq)) + geom_col(fill="steelblue", width=0.5)+
  theme(axis.text.x = element_text(angle = 45, hjust=1)) + labs(title="% Respondents who favour more res
```



Aontú has highest respondents who favour more restrictions

9. Use `coord_flip()` to flip the x- and y-axes. Create nicer axis labels and use `theme_bw()` to change the theme to a black-and-white theme. [5 points]

```
#how to use x label and y label and give it colours
data_immig1 %>% ggplot(aes(x = partyvote, y = freq)) + geom_col(width=0.5)+ coord_flip() + theme_bw() +
```



10. Now we turn to left-right positions. The variable `lr_selfplacement` measures the left-right position of a respondent. How many respondents did *not* specify their left-right self-placement? [5 points]

```
sum(is.na(dat_vaa$lr_selfplacement))
```

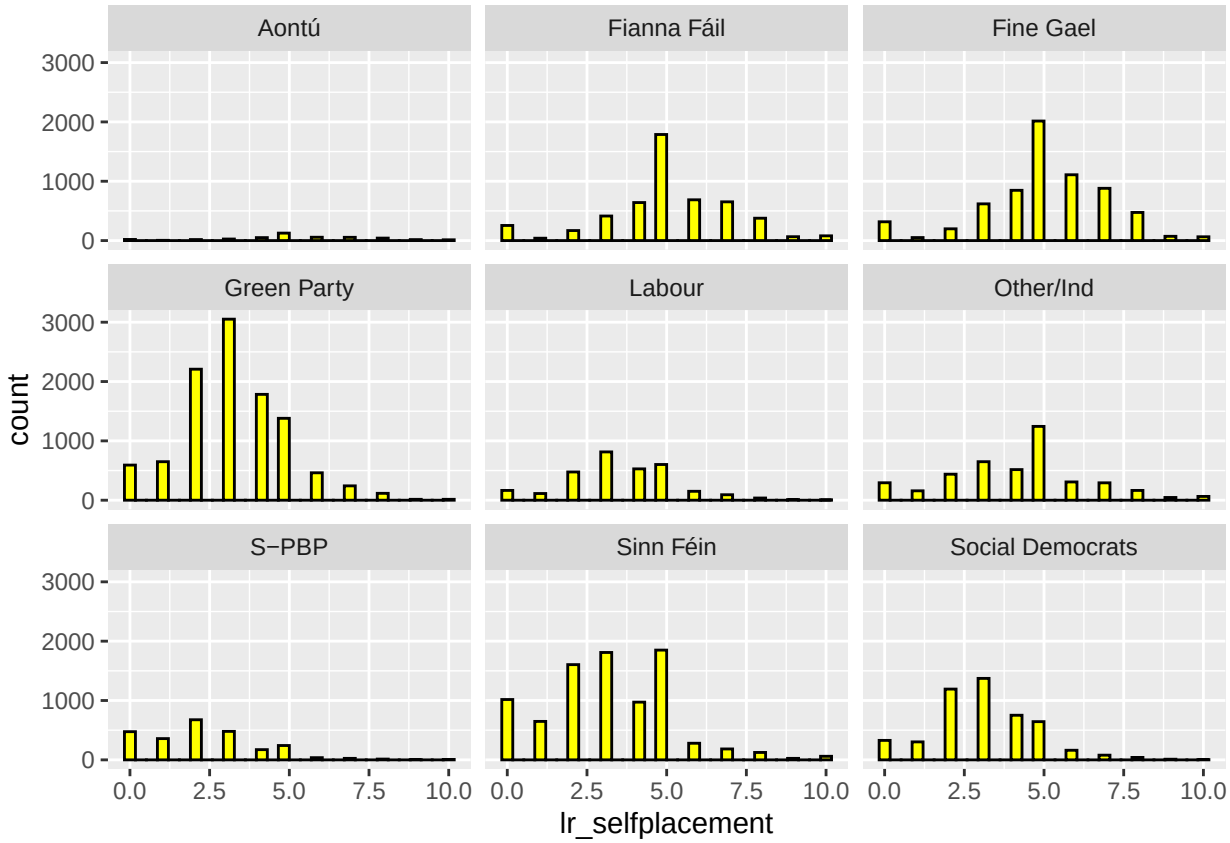
```
## [1] 61591
```

61591 respondents did not specify their left-right self-placement I have considered the original dataframe "dat_vaa" containing all NA values

11. Create a histogram of left-right positions with "small multiples" for each group of party supporters (check out `facet_wrap()`)? [5 points]

```
#Need to do facetwrap
ggplot(dat_vaa_filtered, aes(x=lr_selfplacement), main="Histogram for left-right positions", xlab="LR p
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



12. Use `summarise()` to calculate the

- average left-right position
- median left-right position
- standard deviation of the left-right position

and store this data frame as a new object (always exclude missing responses). [5 points]

```
summarised_object = dat_vaa %>% summarise(mean = mean(lr_selfplacement, na.rm = TRUE), median = median(
summarised_object
```

```
## # A tibble: 10 x 4
##    partyvote         mean median std_dev
##    <chr>            <dbl>  <dbl>   <dbl>
##  1 Aontú             5.35      5    2.14
##  2 Fianna Fáil       5.09      5    1.99
##  3 Fine Gael         5.10      5    1.97
##  4 Green Party       3.31      3    1.70
##  5 Labour            3.66      4    1.75
##  6 Other/Ind         4.28      5    2.14
##  7 S-PBP             2.40      2    1.80
##  8 Sinn Féin         3.27      3    2.02
##  9 Social Democrats  3.14      3    1.67
## 10 <NA>              4.07      4    1.94
```

I have considered the dataframe with NA initially and removed the NA's at column level for the calculation purpose

13. Repeat the step above, but first group the data frame by **partyvote** and then get the average, median, and standard deviation separately for supporters from each party. [5 points]

```
mean = dat_vaa %>%  group_by(partyvote) %>% summarise(mean_lr_replacement = mean(lr_selfplacement, na.r
mean
```

```
## # A tibble: 10 x 2
##    partyvote        mean_lr_replacement
##    <chr>                          <dbl>
##  1 Aontú                           5.35
##  2 Fianna Fáil                     5.09
##  3 Fine Gael                       5.10
##  4 Green Party                     3.31
##  5 Labour                          3.66
##  6 Other/Ind                       4.28
##  7 S-PBP                           2.40
##  8 Sinn Féin                       3.27
##  9 Social Democrats                3.14
## 10 <NA>                            4.07
```

```
med = dat_vaa %>%  group_by(partyvote) %>% summarise(median_lr_replacement = median(lr_selfplacement, na
med
```

```
## # A tibble: 10 x 2
##    partyvote        median_lr_replacement
##    <chr>                            <dbl>
##  1 Aontú                                5
##  2 Fianna Fáil                          5
##  3 Fine Gael                            5
##  4 Green Party                          3
##  5 Labour                               4
##  6 Other/Ind                            5
##  7 S-PBP                                2
##  8 Sinn Féin                            3
##  9 Social Democrats                     3
## 10 <NA>                                 4
```

```
sd = dat_vaa %>%  group_by(partyvote) %>% summarise(sd_lr_replacement = sd(lr_selfplacement, na.rm = TRU
sd
```

```
## # A tibble: 10 x 2
##    partyvote        sd_lr_replacement
##    <chr>                        <dbl>
##  1 Aontú                         2.14
##  2 Fianna Fáil                   1.99
##  3 Fine Gael                     1.97
##  4 Green Party                   1.70
##  5 Labour                        1.75
##  6 Other/Ind                     2.14
```

```
##  7 S-PBP                          1.80
##  8 Sinn Féin                      2.02
##  9 Social Democrats               1.67
## 10 <NA>                           1.94
```

```
new_object = data.frame(merge(merge(mean,med, by.x = "partyvote", by.y = "partyvote"), sd, by.x = "party
new_object
```

```
##           partyvote mean_lr_replacement median_lr_replacement sd_lr_replacement
## 1             Aontú            5.347059                      5          2.144839
## 2       Fianna Fáil            5.086515                      5          1.990496
## 3         Fine Gael            5.099989                      5          1.968478
## 4       Green Party            3.313337                      3          1.697322
## 5            Labour            3.659694                      4          1.754298
## 6         Other/Ind            4.275785                      5          2.142248
## 7             S-PBP            2.396239                      2          1.804933
## 8         Sinn Féin            3.272919                      3          2.017916
## 9  Social Democrats            3.143819                      3          1.665479
## 10             <NA>            4.067062                      4          1.942032
```

I have explored the 3 results using Merge function using column 'partyvote' to ensure 1:1 mapping and then stored into object as a data frame

14. Interpret the output: voters from which party are - on average - the most "left", and voters from which party are - on average - most "right"? For which party do we observe the largest standard deviation, and what does a larger standard deviation imply regarding the distribution of left-right self-placements? [5 points]

Voters from Aontú are the most right with largest average 5.34 and voters from S-PBP are most left with smallest average 2.39 Aontú has largest standard deviation 2.14 Larger standard deviation imply data is largely spreadout and its distance from mean left-right self-placements is higher than the rest

15. Group the data frame by `immig` and get the average left-right position for the two groups. Which group has a higher average left-right value? [5 points]

```
dat_vaa_filtered %>%  group_by(immig) %>% summarise(mean_lr_selfplacement = mean(lr_selfplacement, na.r
```

```
## # A tibble: 2 x 2
##   immig mean_lr_selfplacement
##   <dbl>                 <dbl>
## 1     0                  3.46
## 2     1                  4.68
```

The group with immig = 1 (support more restrictive policies) has higher average left-right value

16. We now use a categorical variable `age`. Use `count()` to get the distribution of age groups (using absolute frequencies/counts). Which age category is the modal age category? [5 points]

```
dat_vaa_filtered %>%  group_by(age) %>% summarise(cnt = n()) %>%
  mutate(freq = round(cnt / sum(cnt)*100, 3)) %>%  arrange(desc(freq))
```

```
## # A tibble: 6 x 3
##   age       cnt  freq
##   <fct>   <int> <dbl>
## 1 25-34   15214 33.2
## 2 35-44   11696 25.5
## 3 18-24    9547 20.8
## 4 45-54    5319 11.6
## 5 55-65    2899  6.32
## 6 over 65  1205  2.63
```

25-34 is the modal age category with highest count

17. Create a new binary variable that distinguishes between respondents who would vote for Sinn Féin and all other respondents. You can use `ifelse()` to recode a variable into two categories. You can name this variable `sinn_fein_binary`. [5 points]

```
head(dat_vaa_filtered,6)
```

```
## # A tibble: 6 x 4
## # Groups:   partyvote [5]
##   immig partyvote        lr_selfplacement age
##   <dbl> <chr>                       <dbl> <fct>
## 1     0 Social Democrats                2 18-24
## 2     1 Fianna Fáil                     2 25-34
## 3     0 Fine Gael                       4 35-44
## 4     0 Green Party                     3 18-24
## 5     0 S-PBP                           0 18-24
## 6     0 Social Democrats                3 55-65
```

```
w <- function(x) return(ifelse(x == "Sinn Féin", 'yes' , 'no'))
#head(w(dat_elections$party_name))

#creating the binary variable sinn_fein_binary with values 'yes' or 'no'
dat_sinn_fein_binary <- dat_vaa %>% mutate(sinn_fein_binary = w(partyvote))
head(dat_sinn_fein_binary,20)
```

```
## # A tibble: 20 x 5
## # Groups:   partyvote [7]
##    immig partyvote        lr_selfplacement age   sinn_fein_binary
##    <dbl> <chr>                       <dbl> <fct> <chr>
##  1    NA <NA>                           NA <NA>  <NA>
##  2    NA <NA>                           NA <NA>  <NA>
##  3    NA <NA>                           NA <NA>  <NA>
##  4     0 Social Democrats                2 18-24 no
##  5     0 <NA>                           NA <NA>  <NA>
##  6     0 <NA>                           NA <NA>  <NA>
##  7     1 <NA>                            3 25-34 <NA>
##  8     1 Fianna Fáil                     2 25-34 no
##  9     0 Fine Gael                       4 35-44 no
## 10    NA Green Party                     5 35-44 no
## 11     0 Labour                          1 <NA>  no
## 12     0 <NA>                           NA <NA>  <NA>
```

```
## 13    NA <NA>                            5 18-24 <NA>
## 14    NA <NA>                            6 35-44 <NA>
## 15     0 <NA>                           NA <NA>  <NA>
## 16     0 Green Party                     3 18-24 no
## 17     0 S-PBP                           0 18-24 no
## 18     0 <NA>                           NA <NA>  <NA>
## 19     0 Social Democrats                3 55-65 no
## 20     0 <NA>                           NA <NA>  <NA>
```

```
#checking if we have rows with sinn_fein_binary == "yes" and "no"
filter(dat_sinn_fein_binary, partyvote != "Sinn Féin")
```

```
## # A tibble: 51,015 x 5
## # Groups:   partyvote [8]
##    immig partyvote        lr_selfplacement age   sinn_fein_binary
##    <dbl> <chr>                       <dbl> <fct> <chr>
##  1     0 Social Democrats                2 18-24 no
##  2     1 Fianna Fáil                     2 25-34 no
##  3     0 Fine Gael                       4 35-44 no
##  4    NA Green Party                     5 35-44 no
##  5     0 Labour                          1 <NA>  no
##  6     0 Green Party                     3 18-24 no
##  7     0 S-PBP                           0 18-24 no
##  8     0 Social Democrats                3 55-65 no
##  9     0 S-PBP                           1 18-24 no
## 10     0 Green Party                     2 18-24 no
## # ... with 51,005 more rows
```

```
filter(dat_sinn_fein_binary, partyvote == "Sinn Féin")
```

```
## # A tibble: 12,132 x 5
## # Groups:   partyvote [1]
##    immig partyvote lr_selfplacement age   sinn_fein_binary
##    <dbl> <chr>                <dbl> <fct> <chr>
##  1    NA Sinn Féin                5 35-44 yes
##  2     0 Sinn Féin                2 25-34 yes
##  3     0 Sinn Féin                1 18-24 yes
##  4     0 Sinn Féin                4 18-24 yes
##  5    NA Sinn Féin                2 18-24 yes
##  6     0 Sinn Féin                2 25-34 yes
##  7     0 Sinn Féin                2 18-24 yes
##  8     0 Sinn Féin               NA 25-34 yes
##  9     0 Sinn Féin                2 25-34 yes
## 10     1 Sinn Féin               NA 25-34 yes
## # ... with 12,122 more rows
```

```
table(dat_sinn_fein_binary$sinn_fein_binary)
```

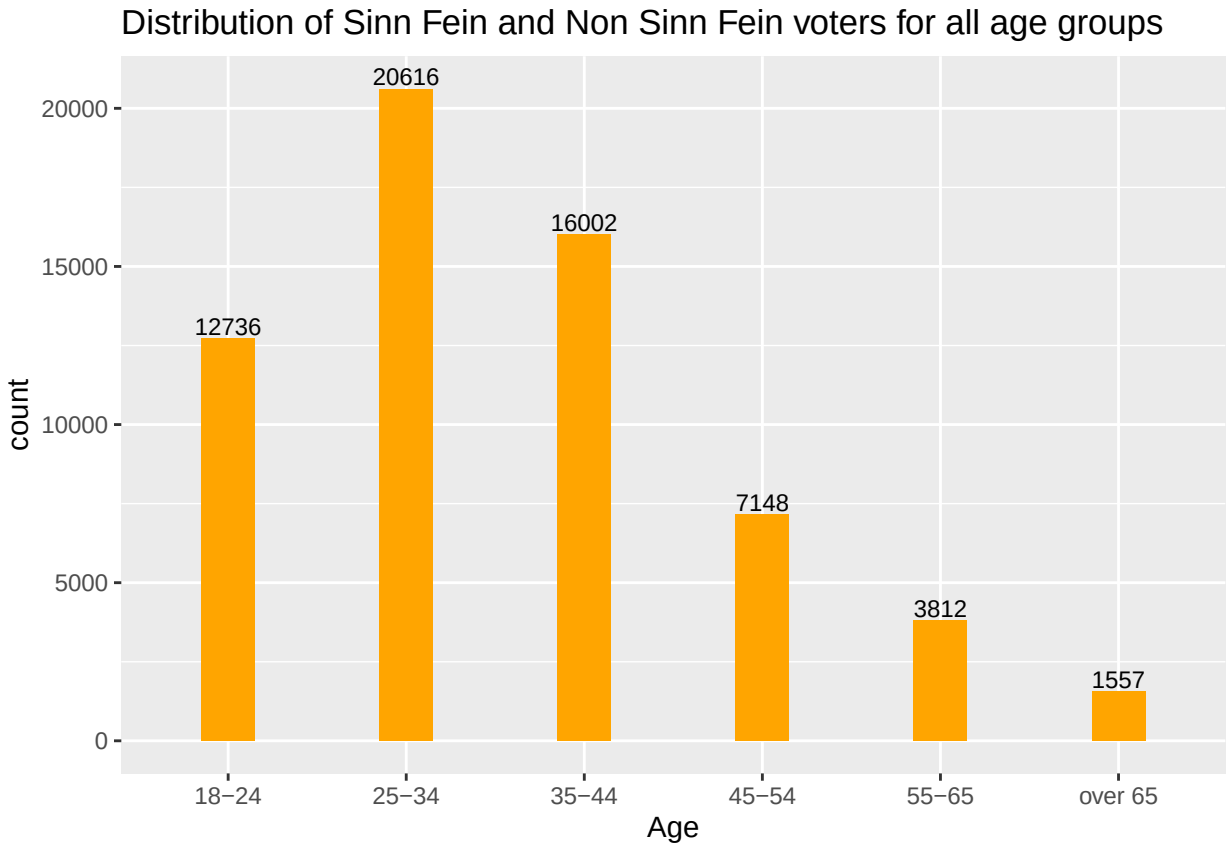
```
##
##    no   yes
## 51015 12132
```

18. **BONUS POINTS**: Create a barplot (`geom_bar()`) and plot the distribution of age groups for both groups of voters. Use `facet_wrap()` to create a plot with two "small multiples". Are respondents who would vote for Sinn Féin younger than respondents who expressed a different vote choice? [5 points]

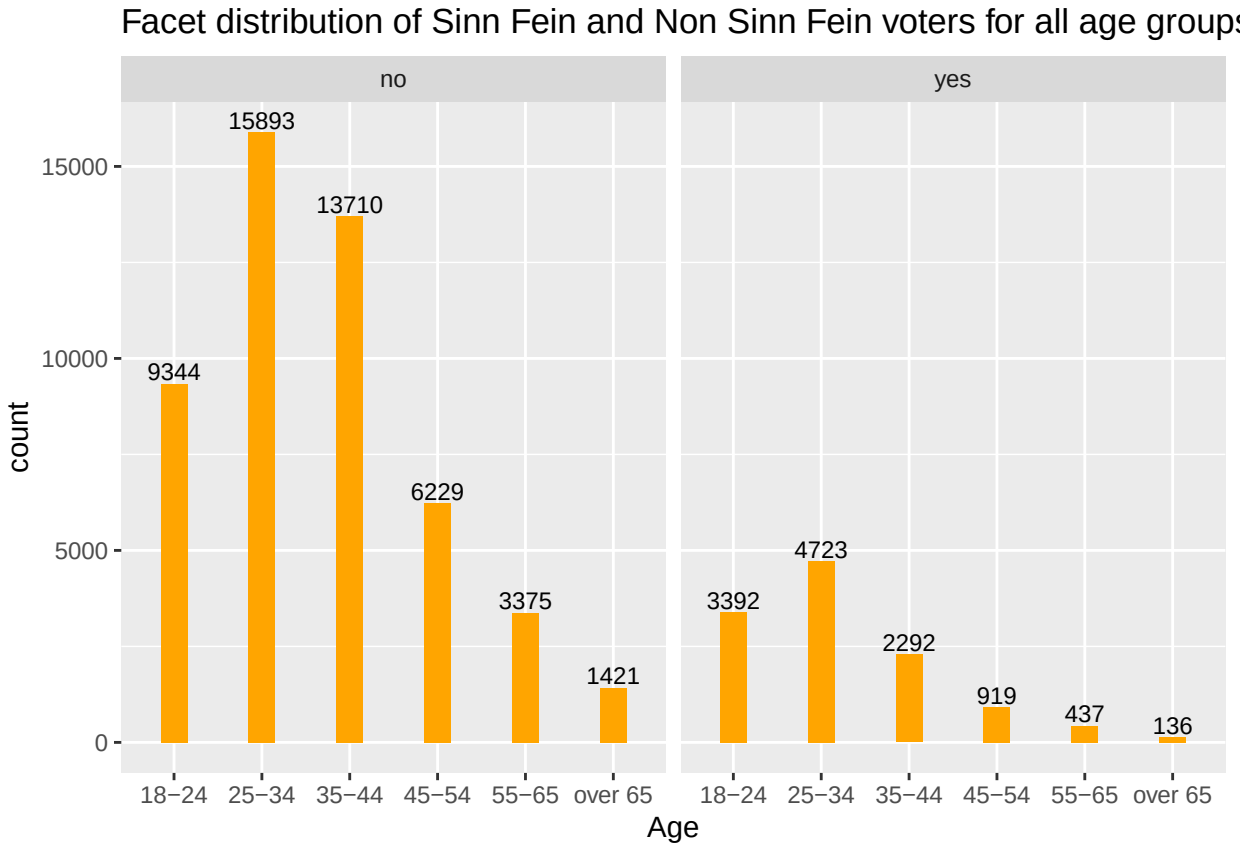
```
dim(dat_sinn_fein_binary)
```

```
## [1] 141936      5
```

```
#Bar graph
ggplot(filter(dat_sinn_fein_binary,!is.na(age)&!is.na(sinn_fein_binary)), aes(x = age)) + geom_bar(widt
```




```
#Graph with facet added
ggplot(filter(dat_sinn_fein_binary,!is.na(age)&!is.na(sinn_fein_binary)), aes(x = age)) + geom_bar(widt
```



We cannot truly identify younger respondents as 18-24 or 25-34 or both of them combined If we are to assume younger respondents as 18-24, then we still do not know the individual ages of each respondents to come to a conclusion For example : 3392 can have 90% having age 24 and 10% between age 18-23. However, 9344 can have 5% with age 18-20 and 95% with age 21-24 So there is no convincing answer to this question unless we know individual ages of all respondents