Exploratory Data Analysis

Kevin Koehler

2025-05-03

Table of contents

1 The Tunisia survey **3**
1.1 Summary statistics . . . 3
1.2 Descriptive statistics . . . 3
1.2.1 Gender (`female`) . . . 4
1.2.2 Age (`age`) . . . 4
1.2.3 Education (`edu`) . . . 5
1.2.4 Income (`income`) . . . 5
1.2.5 Region (`region`) . . . 6
1.2.6 July 25 Events (`july25`) . . . 6
1.2.7 Members of Parliament (`mps`) . . . 7
1.2.8 People Over Politicians (`people`) . . . 7
1.2.9 Elected Officials (`officials`) . . . 8
1.2.10 Voted in 2019 Parliamentary Elections (`leg2019_voted`) . . . 8
1.2.11 Vote in 2019 Parliamentary Elections (`leg2019`) . . . 9
1.2.12 Voted in 2019 Presidential Elections, First Round (`pres2019_1_voted`) 9
1.2.13 Vote in 2019 Presidential Elections, First Round (`pres2019_1`) . . . 10
1.2.14 Voted in 2019 Presidential Elections, Second Round (`pres2019_2_voted`) 10
1.2.15 Vote in 2019 Presidential Elections, Second Round (`pres2019_2`) . . . 11

2 Exploratory Data Analysis **11**
2.1 July 25th and gender . . . 11
2.2 July 25th and age . . . 12
2.3 July 25th and education . . . 13
2.4 July 25th and income . . . 14
2.5 July 25th and region . . . 15
2.6 July 25th and “MPs lose touch” . . . 16
2.7 July 25th and “People should make decisions” . . . 17
2.8 July 25th and “Officials talk too much” . . . 18

3 Conclusion **18**

1 The Tunisia survey

This document describes the Tunisia survey data and reports the results of an exploratory data analysis. The survey was fielded by 1to1 Polling in March 2022 and included a representative sample of 1,000 Tunisians of age 18 or above. The survey included a range of items on political attitudes, voting behavior, as well as respondents' demographic background.

The next section describes the variables included in the data set in numerical and graphical form. In the second section, we describe the results of an Exploratory Data Analysis (EDA) on the data. The third section concludes.

1.1 Summary statistics

The data set includes 15 variables with 1,000 observations each. Table 1 gives a first overview of the variables contained in the data set along with basic sumamry statistics.

Table 1: Variables and summary statistics

Statistic	N	Mean	St. Dev.	Min	Max
female	1,000	0.508	0.500	0	1
age	996	41.462	15.023	18	84
edu	1,000	4.204	1.935	1	8
income	907	3.619	1.771	1	8
july25	877	1.274	0.446	1	2
mps	965	3.543	1.653	1	5
people	982	3.789	1.488	1	5
officials	984	4.181	1.416	1	5
leg2019_voted	977	1.536	0.499	1	2
leg2019	269	3.156	2.046	1	7
pres2019_1_voted	988	1.441	0.497	1	2
pres2019_1	488	2.123	2.562	1	11
pres2019_2_voted	982	1.483	0.500	1	2
pres2019_2	466	1.069	0.253	1	2

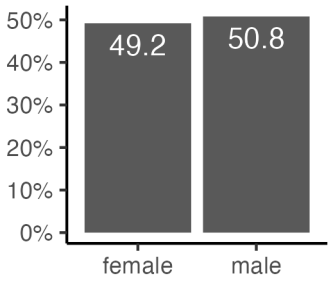
1.2 Descriptive statistics

In this section, we present descriptive statistics for all variables included in the data set.

1.2.1 Gender (`female`)

The variable is coded 0 for male and 1 for female. There are 0 missing values.

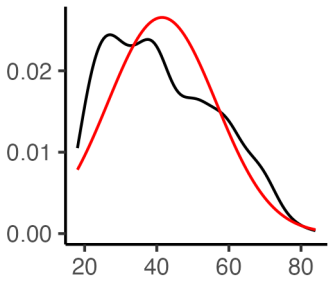
The sample is almost perfectly balanced in gender terms.



1.2.2 Age (`age`)

The variable records respondents' age in years. There are 4 missing values. The mean is 41.46 and the standard deviation 15.02, the variable varies between 18 and 84.

The density plot on the right also shows a normal distribution with identical mean and standard deviation (in red). The distribution of age does not seem to be extremely skewed.



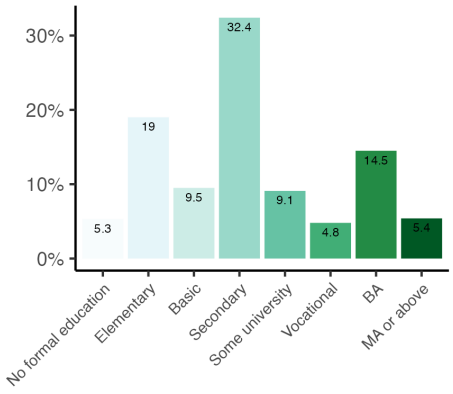
1.2.3 Education (`edu`)

The variable records respondents' education levels:

1 = Illiterate/no formal education
2 = Elementary education
3 = Basic education
4 = Secondary education
5 = Some university
6 = Vocational
7 = BA degree
8 = MA and above

There are 0 missing values.

The most frequent level of education is secondary education, about one-fifth of respondents completed university.



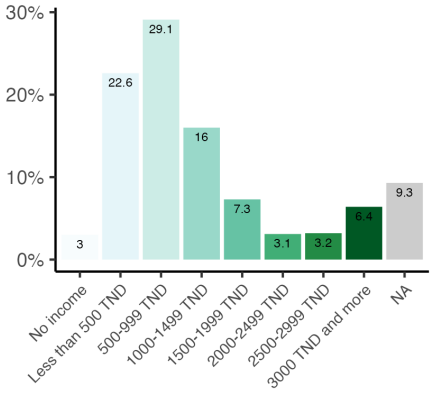
1.2.4 Income (`income`)

The variable records respondents' income in these brackets:

1 = No current income
2 = Less than 500 TND
3 = 500–999 TND
4 = 1000–1499 TND
5 = 1500–1999 TND
6 = 2000–2499 TND
7 = 2500–2999 TND
8 = 3000 TND and more

There are 93 missing values.

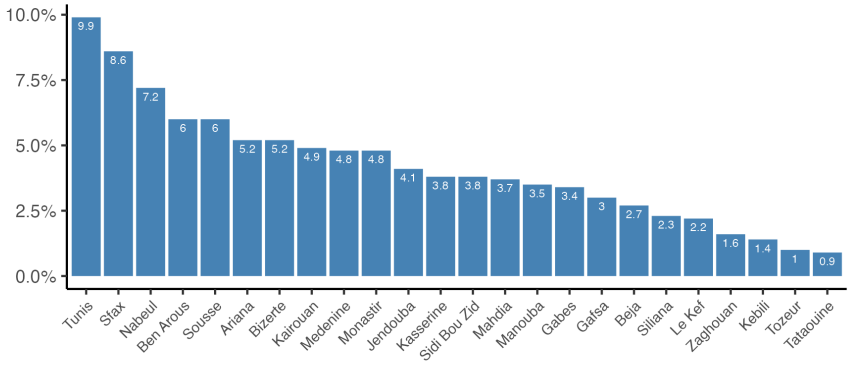
More than half report monthly incomes below 1,000 TND (approx. 308 EUR at the time of the survey).



1.2.5 Region (`region`)

Province of residence

There are 0 missing values.

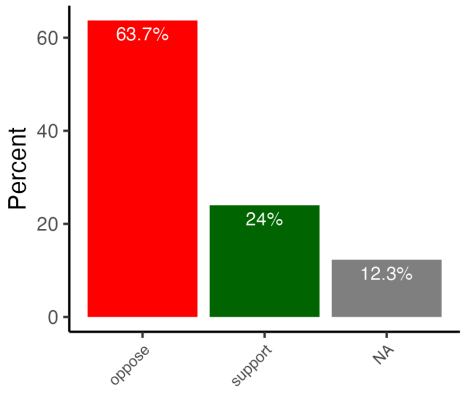


1.2.6 July 25 Events (`july25`)

What statement best characterizes the July 25 events in Tunisia?

1 = They represent a necessary correction to Tunisia's democratic transition
2 = They undermine the democratic transition
98 = Don't know
99 = Declined to answer

There are 123 missing values. More than 60% of respondents think July 25th threatens the democratic transition.

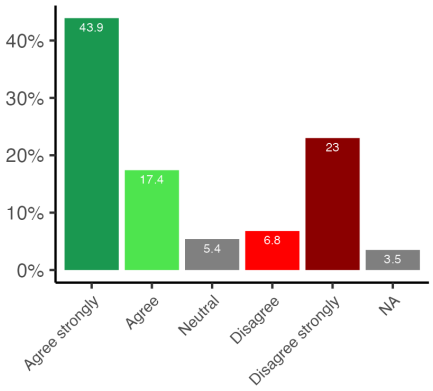


1.2.7 Members of Parliament (`mps`)

"Members of Parliament very quickly lose touch with ordinary people after they assume office."

1 = Agree Strongly
2 = Agree Somewhat
3 = Neither agree nor disagree
4 = Disagree Somewhat
5 = Disagree Strongly
98 = Don't Know
99 = Declined to answer

There are 35 missing values.

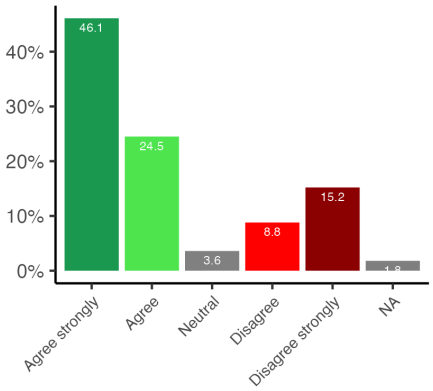


1.2.8 People Over Politicians (`people`)

"The people, not politicians, should make the most important policy decisions."

1 = Agree Strongly
2 = Agree Somewhat
3 = Neither agree nor disagree
4 = Disagree Somewhat
5 = Disagree Strongly
98 = Don't Know
99 = Declined to answer

There are 18 missing values.

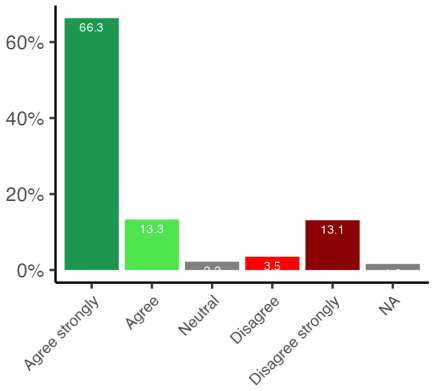


1.2.9 Elected Officials (`officials`)

"Elected officials talk too much and take too little action."

1 = Agree Strongly
2 = Agree Somewhat
3 = Neither agree nor disagree
4 = Disagree Somewhat
5 = Disagree Strongly
98 = Don't Know
99 = Declined to answer

There are 16 missing values.

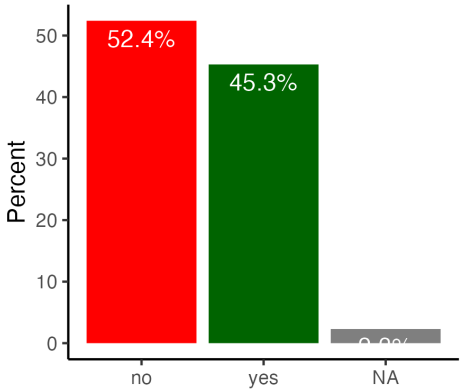


1.2.10 Voted in 2019 Parliamentary Elections (`leg2019_voted`)

Did you vote in the 2019 parliamentary elections?

1 = Yes
2 = No
98 = Don't remember
99 = Declined to answer

There are 23 missing values.

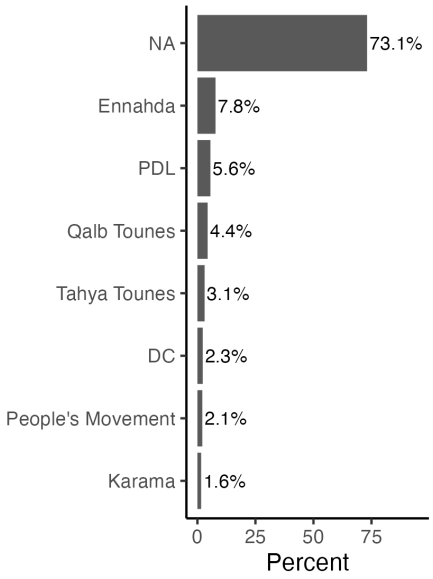


1.2.11 Vote in 2019 Parliamentary Elections (`leg2019`)

If yes, for which political party did you vote?

1 = Ennahda
2 = Qalb Tunis
3 = Free Destourian Party
4 = Democratic Current
5 = Karama
6 = People's Movement
7 = Tahya Tunis
96 = Other
97 = Blank or invalid ballot
98 = Don't know / don't remember
99 = Declined to answer

There are 731 missing values.

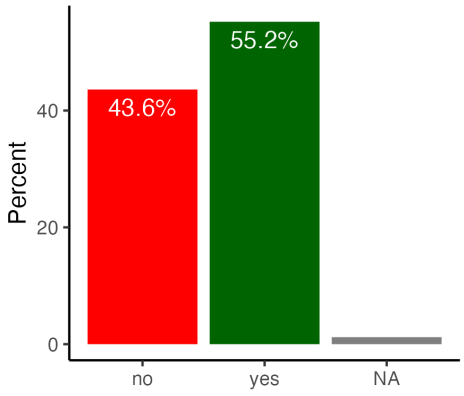


1.2.12 Voted in 2019 Presidential Elections, First Round (`pres2019_1_voted`)

Did you vote in the 2019 presidential elections (first round)?

1 = Yes
2 = No
98 = Don't know
99 = Declined to answer

There are 12 missing values.

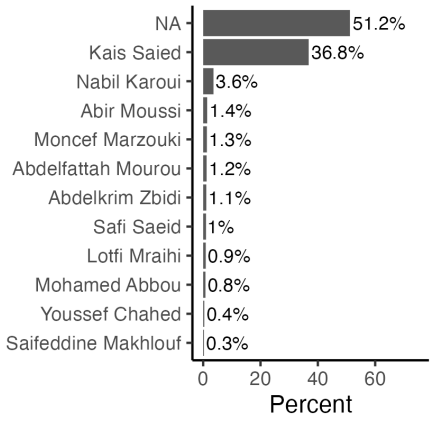


1.2.13 Vote in 2019 Presidential Elections, First Round (`pres2019_1`)

If yes, for which candidate did you vote?

1 = Kais Saied
2 = Nabil Karoui
3 = Abdelfattah Mourou
4 = Abdelkrim Zbidi
5 = Youssef Chahed
6 = Safi Saied
7 = Lotfi Mraihi
8 = Saifeddine Makhlouf
9 = Abir Moussi
10 = Mohammed Abbou
11 = Moncef Marzouki
96 = Other
97 = Blank or invalid ballot
98 = Don't know / don't remember
99 = Declined to answer

There are 512 missing values.

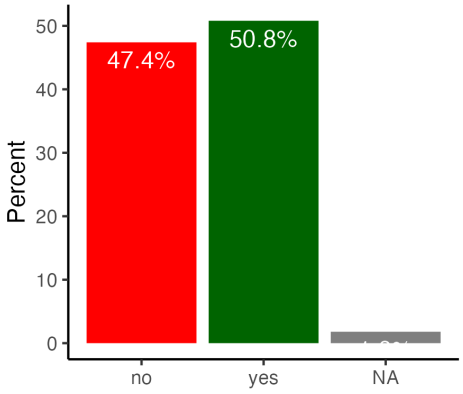


1.2.14 Voted in 2019 Presidential Elections, Second Round (`pres2019_2_voted`)

Did you vote in the 2019 presidential elections (second round)?

1 = Yes
2 = No
98 = Don't know
99 = Declined to answer

There are 18 missing values.

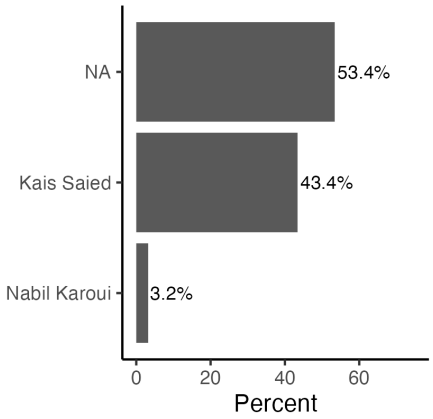


1.2.15 Vote in 2019 Presidential Elections, Second Round (`pres2019_2`)

If yes, for which candidate did you vote?

1 = Kais Saied
2 = Nabil Karoui
97 = Blank or invalid ballot
98 = Don't know / don't remember
99 = Declined to answer

There are 534 missing values.



2 Exploratory Data Analysis

In this section, we perform an exploratory data analysis and report the results. We focus in particular on potential explanations for support for (or opposition against) July 25th. We explore potential relationships between age and education, the three populism measures, and voting behavior on the one hand, and attitudes toward July 25th on the other. In each case, we first graph the relationship before we run the appropriate statistical test.

2.1 July 25th and gender

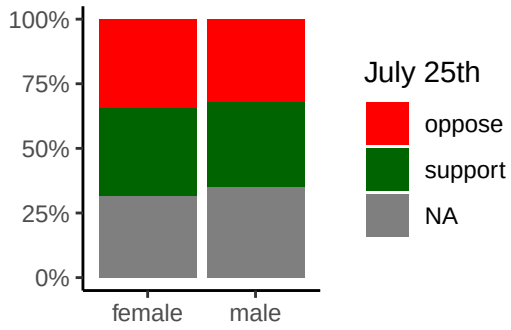


Figure 1: Support for July 25th by gender

Visually, there does not seem to be a relationship between gender and attitudes toward July 25th. We can formally test this by looking at the contingency table:

Table 2: Contingency table of gender and July 25th

female	July 25	n
0	oppose	317
1	oppose	320
0	support	119
1	support	121

Table 2 suggests that there is no relationship. This is confirmed by a χ^2 test as well. The p-value from the χ^2 test is 1, indicating no significant relationship.

2.2 July 25th and age

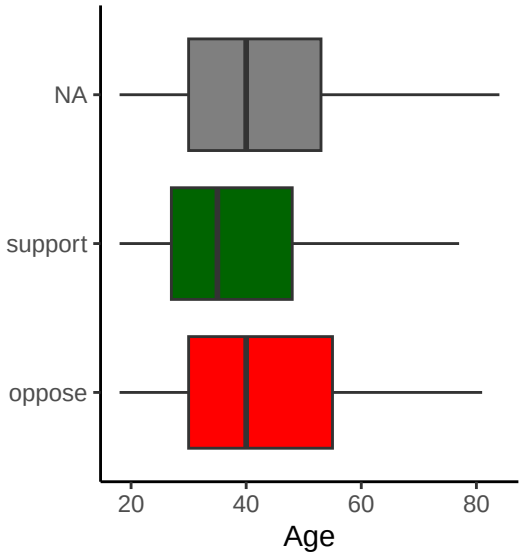


Figure 2: Support for July 25th by age

Figure 2 suggests that there indeed might be a relationship such that older respondents are more likely to see July 25th as weakening the democratic transition. This is confirmed by a formal t-test: The average age of supporters of July 25th is 37.74, while the average age of opponents is 42.71 (the difference is significant with 5×10^{-6}).

2.3 July 25th and education

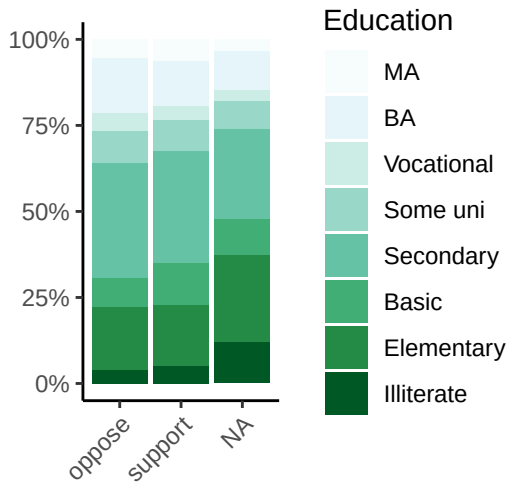


Figure 3: Support for July 25th by education

There might be a difference based on Figure 3. A Wilcoxon test reveals that there is none, however (p=0.3078).

2.4 July 25th and income

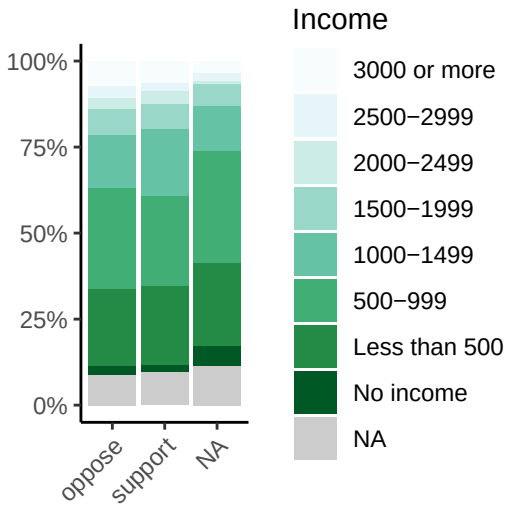


Figure 4: Support for July 25th by income

Figure 4 does not suggest a relationship, a Wilcoxon test confirms this impression (p=0.8861).

2.5 July 25th and region

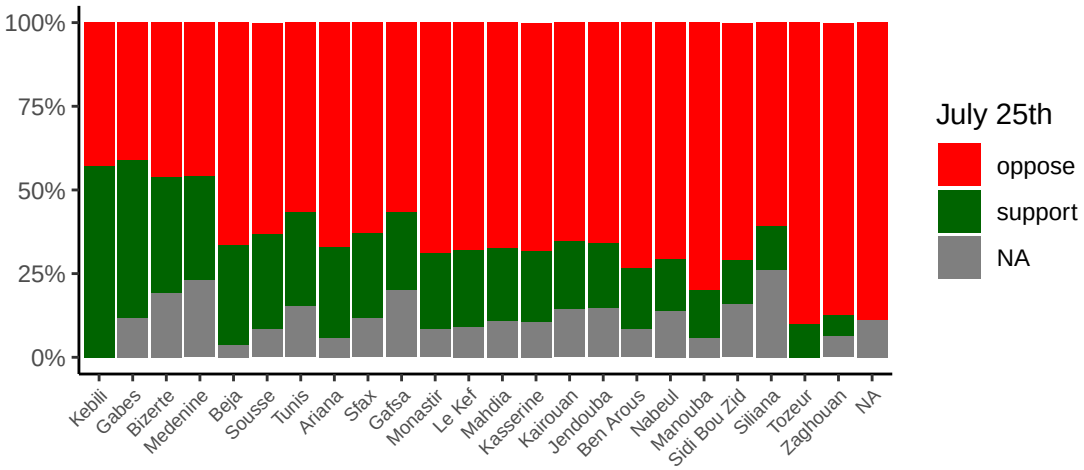


Figure 5: Support for July 25th by region

There is a relationship between region and support for July 25th (p=0.0032 in a χ^2-test).

2.6 July 25th and "MPs lose touch"

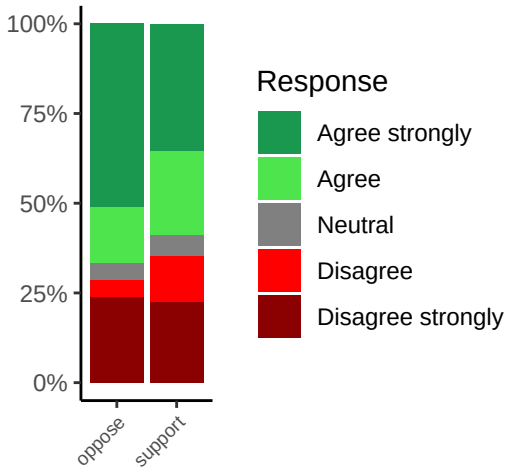


Figure 6: Support for July 25th and MPs lose touch

We can conclude that the difference in Figure 6 is significant with p = 0.0028.

2.7 July 25th and "People should make decisions"

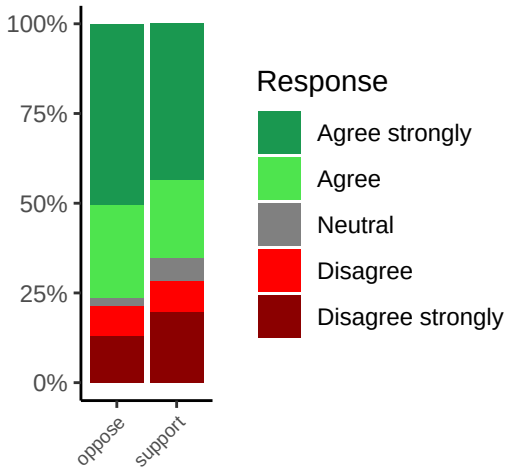


Figure 7: Support for July 25th and people should make decisions

We can conclude that the difference in Figure 7 is significant with p = 0.0102.

2.8 July 25th and "Officials talk too much"

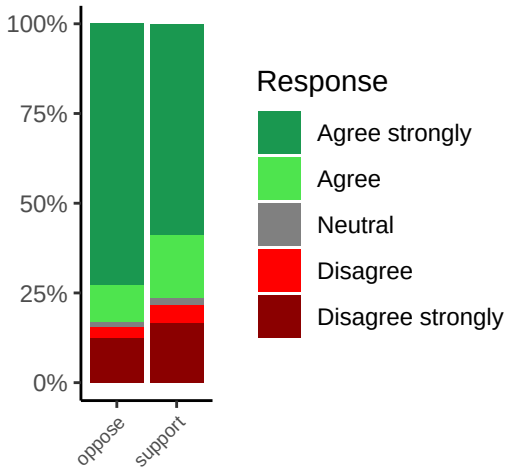


Figure 8: Support for July 25th and officials talk too much

We can conclude that the difference in Figure 8 is significant with $p = 2 \times 10^{-4}$.

3 Conclusion