Kuwait Data Check

We have discussed our concerns with the Kuwait dataset, particularly those surrounding high government approval numbers, which we suspected may be fabricated to a certain extent.My overall recommendation, upon which I will elaborate, is that there is nothing within the trend lines and data itself to suggest strongly that the data is fabricated.

The analysis surrounds a few assumptions. If we believe that the data was fabricated to any extent, we have to decide whether the entire data set was corrupted by falsification, or whether only a subset of variables were manipulated, in which case we could discuss ways in which the data might be salvageable.

Trends between waves 3 and 5.

I narrowed down the data frame to include only those variables for which we have observations for wave 3 and 5, so I didnt look at anything for which we dont have observations for both years. Then I took the mean for each year and anywhere where I saw a more than 10 percent jump I produced trend lines illustrating the shifts.

There are a couple points to keep in mind. The first is that most charts contain aggregates of responses for those who responded to a set of respones greater than 1. For example, for question 2011, we aggregate the reponses for those who say they trust government to a great and medium extent. This affects the analysis because its not necessarily unreasonable to expect a large number of people in a relatively prosperous country to say they trust government to a medium extent. The second point is that the comparisons are based on the presumed reliability of the wave 5 data. I.e. we begin with the assumption that the wave 5 data is wholly accurate, and that we can deduce the presence of falsified data from comparing the two data sets.

Trust in Government

Between waves 3 and 5, there is the greatest appreciable decrease in the number of people saying they trust government to a great extent, and a somewhat smaller decrease in the the number of people saying they trust government to a medium extent. In the case of 2011,(overall trust in government), 2013 (trust in parliament), 2014 (trust in regional government), in wave 3 there are either a relatively equal number of people saying they trust government to great or medium extent, or in the case of 2014, a much larger number of people saying they trust Regional Government to a great extent than those saying they trust it to only a medium extent.Tables are included with each graph to illustrate this point.

Question 2014

Certain observations are not consistent with the idea that the data was outright fabricated, although there remains the possibility that it was inflated. For example if we take question 2014, we see a decrease in the number of people saying they trust regional government to a great extent between waves 3 and 5, but an increase among those who say they trust government to a medium extent. If we believe that the wave 5 data is reliable, this could suggest that there is fabrication from shifting the entire distribution, by putting people who responded 2(medium extent) to 3(great extent), as otherwise the shift would not be so dramatic.

Question 2017

We see the same pattern in question 2017, where there is a sizeable decrease in the number of people saying they trust civil society organizations to a great extent, but an increase in the number of people saying they trust them to a medium extent.

Question 201

This pattern is repeated in question 210 with a much greater increase in the number of people sayiung that corruption is present in state institutions to a medium extent but a decrease in the number of people saying that corrruption is present to a great extent. If we assume that the data was manipulated in some way, then we have to assume that it only affects a subset of the variables, as if the data was manipulated to present a favorable picture of government, they could not neglect to alter this question. Otherwise we have to assume that none of the data was altered.

In questions 5211, 5214, 5215, 7001, 7002, 7003, we see the same pattern as observed above. The largest source of the shift is from those responding 1 to those responding 2. (greater-medium extent) This pattern is consistent across all the questions in which we see a sizeable shift across waves 3 and 5.

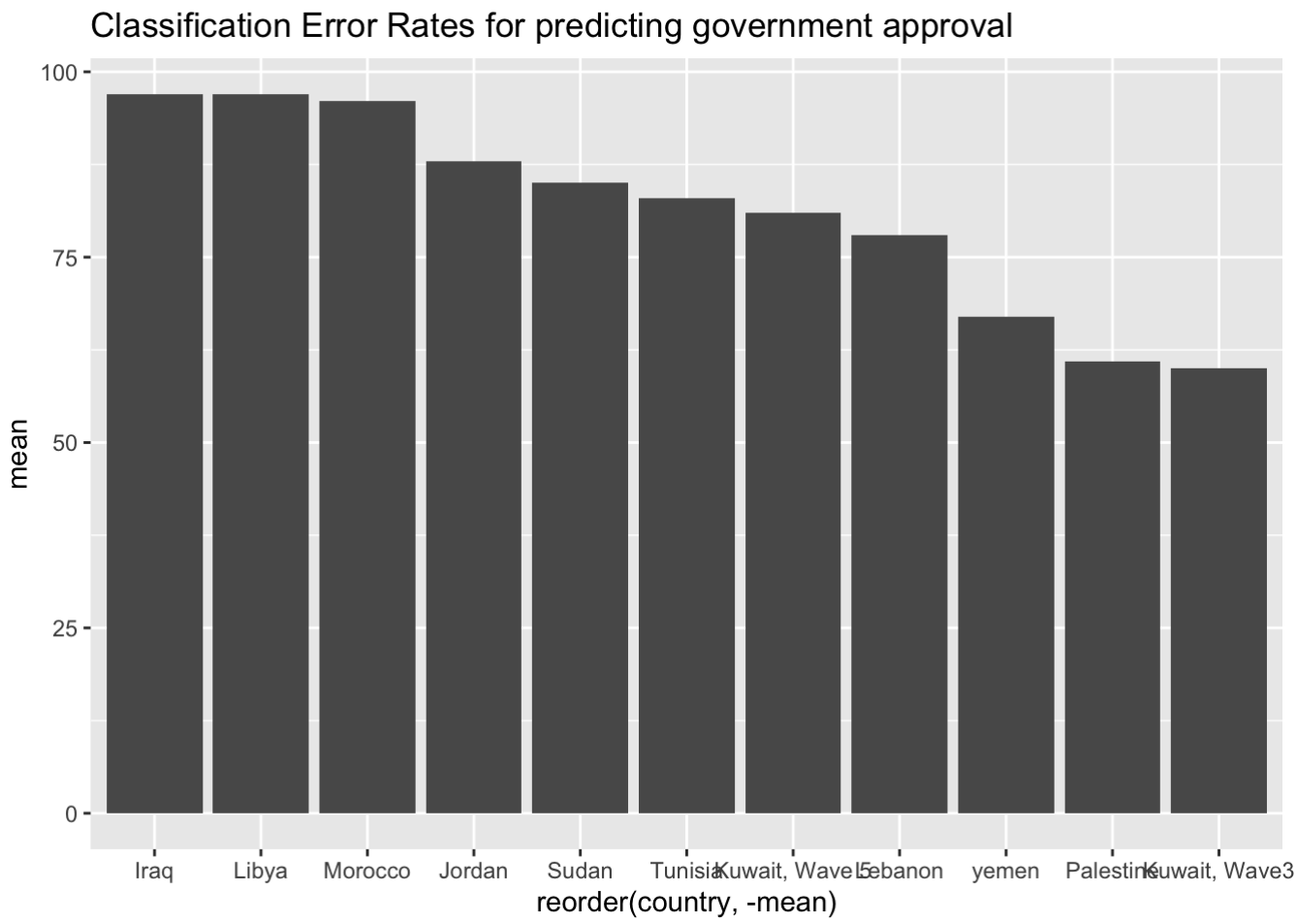
Personally, despite the high approval numbers for the government in wave 3 data, I can't find any consistent pattern in the data that would suggest that there is detectable fabrication, particularly if we are operating from the assumption that the wave 5 data is reliable. I additionally ran classification models trying to predict government approval by a smaller subset of variables ('q101','q104','q2042','q2043','q2054','q2185'). I trained the model on wave 5 data, and tested it against wave 3 data for Kuwait, and then did the same for almost every country in the merged data set. The model was least accurate for Kuwait (41% error rate). This indicates to a certain extent that the wave 3 data for Kuwait is unusual to say the least, and as a test suggests to me more strongly that the hypothesis that government approval ratings are inflated in wave3, but that other variables are not. I chose only a small subset of variables that would be highly correlated with trust in the government - if these variables are not as good predictors for levels of trust in government in Kuwait in wave 3 as they are for all other countries, it suggest that there was an attempt to inflate certain variables like trust in government without altering other underlying variables that could be correlated with them. However the case for this is not particularly strong as although Kuwait is somehwat of an outlier, we can still accurately predict government approval based on other factors within the data set.

Based on my inspection of the data, I would recommend simply removing wave 3 for certain variables that we find suspect, but leaving the overall dataset intact. The trend patterns that we see could be due to the way in which we code and report variables, combined with relatively benign trends in public approval toward the government, rather than a systematic effort to cook the books. Moreover, comparing the the trend in certain variables such as trust in Parliament to trust in Regional government, its hard to suggest that if an attempt to change data was undertaken, that it was only for certain institutions, since trust in Regional government remains high in wave 5, and we presume that to be accurate data.

The final point to keep in mind is that political events during the Arab Spring in Kuwait are also relatively consistent with the structure of Kuwaiti government, with the Parliament considered to be an opposition force to the executive, and cited by well known scholars (Nathan Brown) as being the most independent parliament in the Arab world. Protests against the government in Kuwait in 2011-2012 were largely led by stateless Bedoon rather than a large crossection of Kuwaiti citizens. The Kuwaiti National Assembly, although politically contested, is legitimately regarded as one of the most independent and democratic judiciaries in the region, and this is just as likely to be the source of its high approval as potential data manipulation or fabrication.

```
plot1_frame=tibble("country"= c("yemen","Iraq", "Jordan", "Lebanon", "Libya", "Morocco",
 "Palestine", "Tunisia", "Sudan", "Kuwait, Wave3","Kuwait, Wave 5"), mean=c(67,97,88,78,
97,96,61,83,85,60,81))
ggplot(data = plot1_frame, aes(x=reorder(country,-mean), y=mean))+
  geom_histogram(stat = "identity")+ggtitle("Classification Error Rates for predicting g
overnment approval")
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

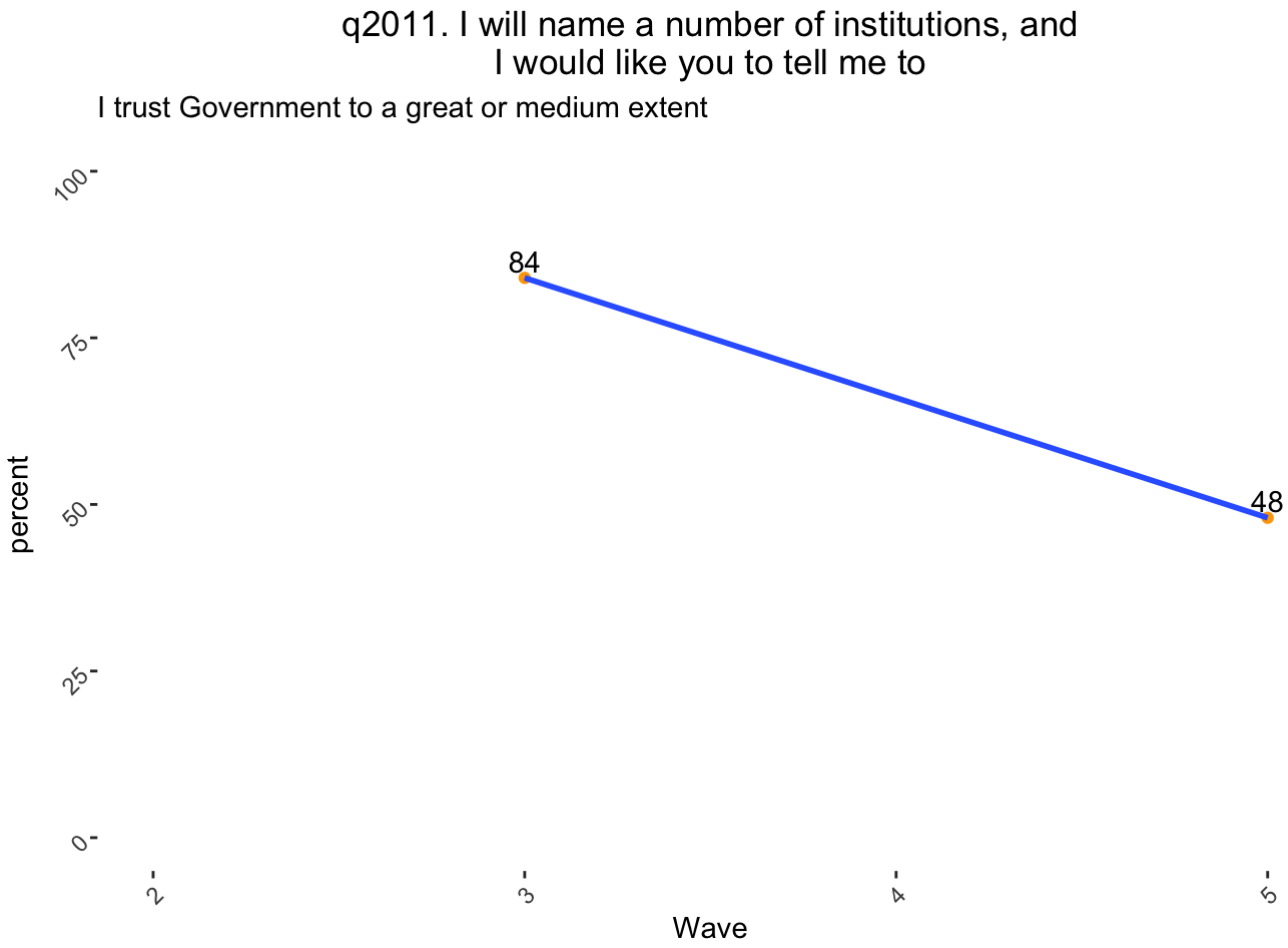



```
## # A tibble: 2 x 2
##         wave filter_variable
##    <dbl+lbl>           <dbl>
## 1 3 [Wave 3]              77
## 2 5                       32
```

```
ab_graph = response_1_2(abv, q2011)
table(abv$wave,abv$q2011)
```

```
##
##       1   2   3   4   9  98  99
##   3 484 376 129  30   2   0   0
##   5 180 473 360 318   0  35   8
```

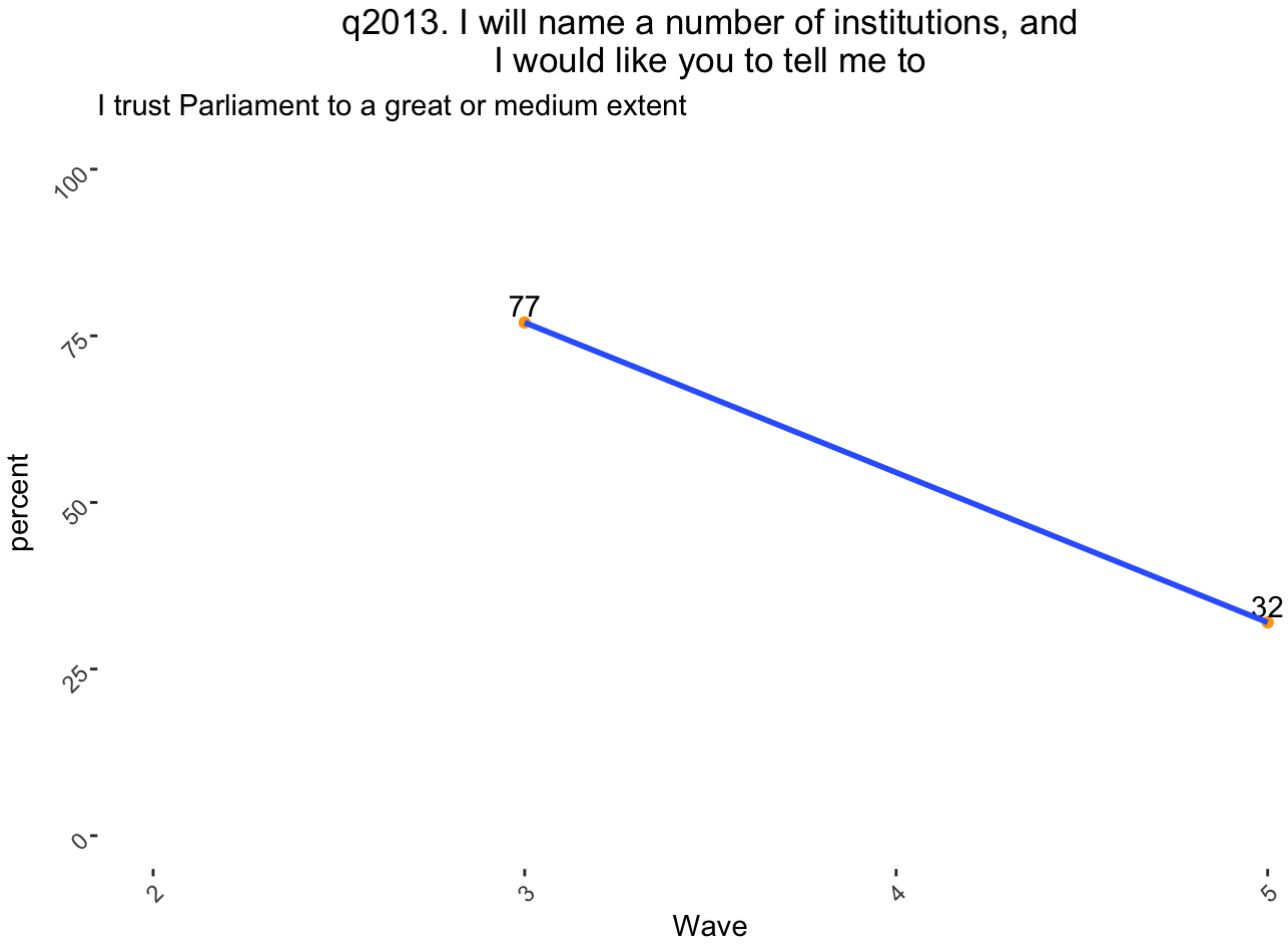
```
labelling(ab_graph$filter_variable,q2011)
plot1(ab_graph, ab_graph$filter_variable, var_lab(q2011),"I trust Government to a great
 or medium extent")
```




```
ab_graph = response_1_2(abv, q2013)
table(abv$wave,abv$q2013)
```

```
##
##       1   2   3   4   9  98  99
##   3 385 399 191  44   2   0   0
##   5  69 375 352 529   0  42   7
```

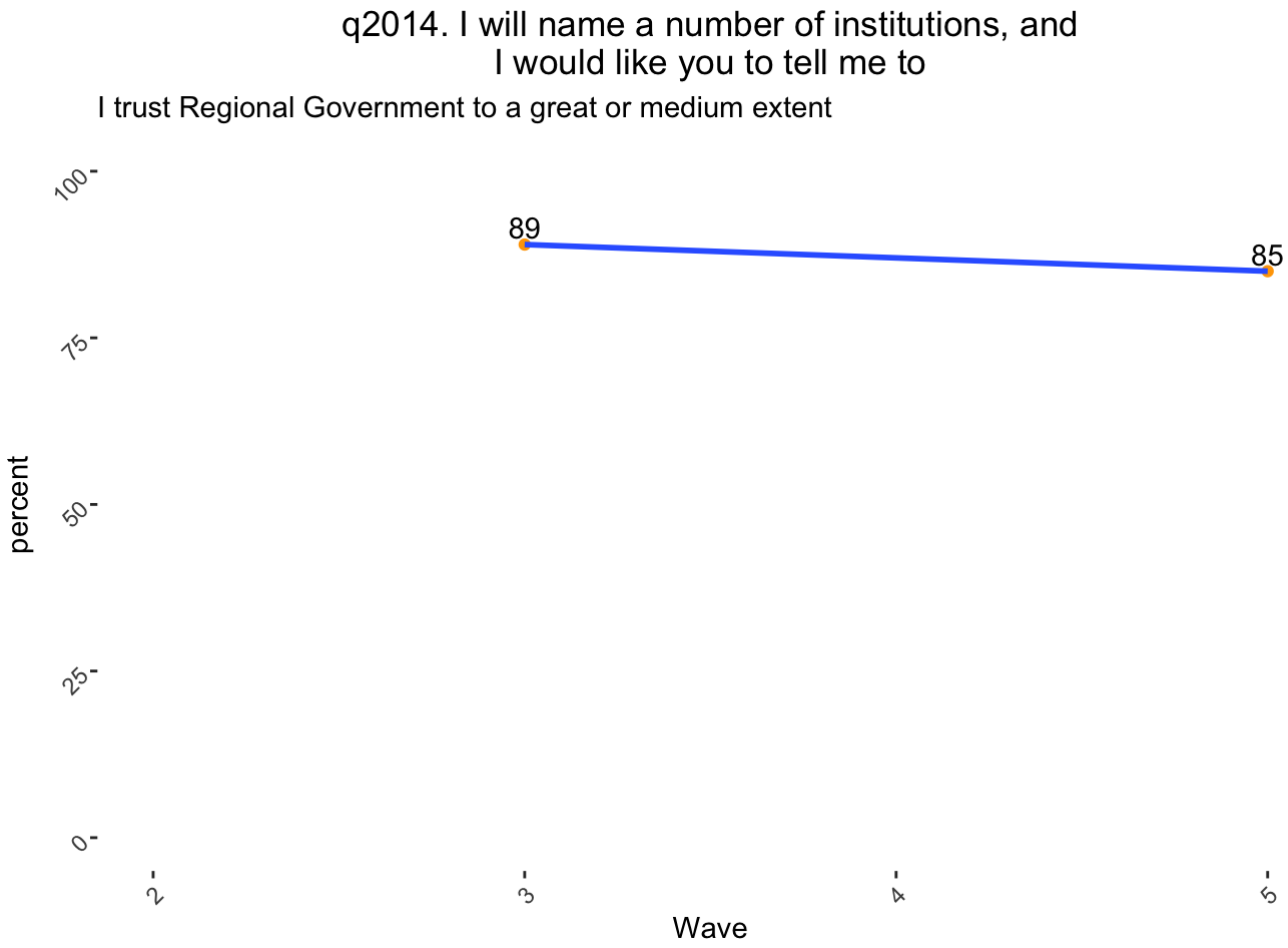
```
labelling(ab_graph$filter_variable,q2013)
plot1(ab_graph, ab_graph$filter_variable, var_lab(q2013),"I trust Parliament to a great
 or medium extent")
```




```
ab_graph = response_1_2(abv, q2014)
table(abv$wave,abv$q2014)
```

```
##
##       1   2   3   4   9  98  99
##   3 662 251  69  38   1   0   0
##   5 549 620 141  53   0   8   3
```

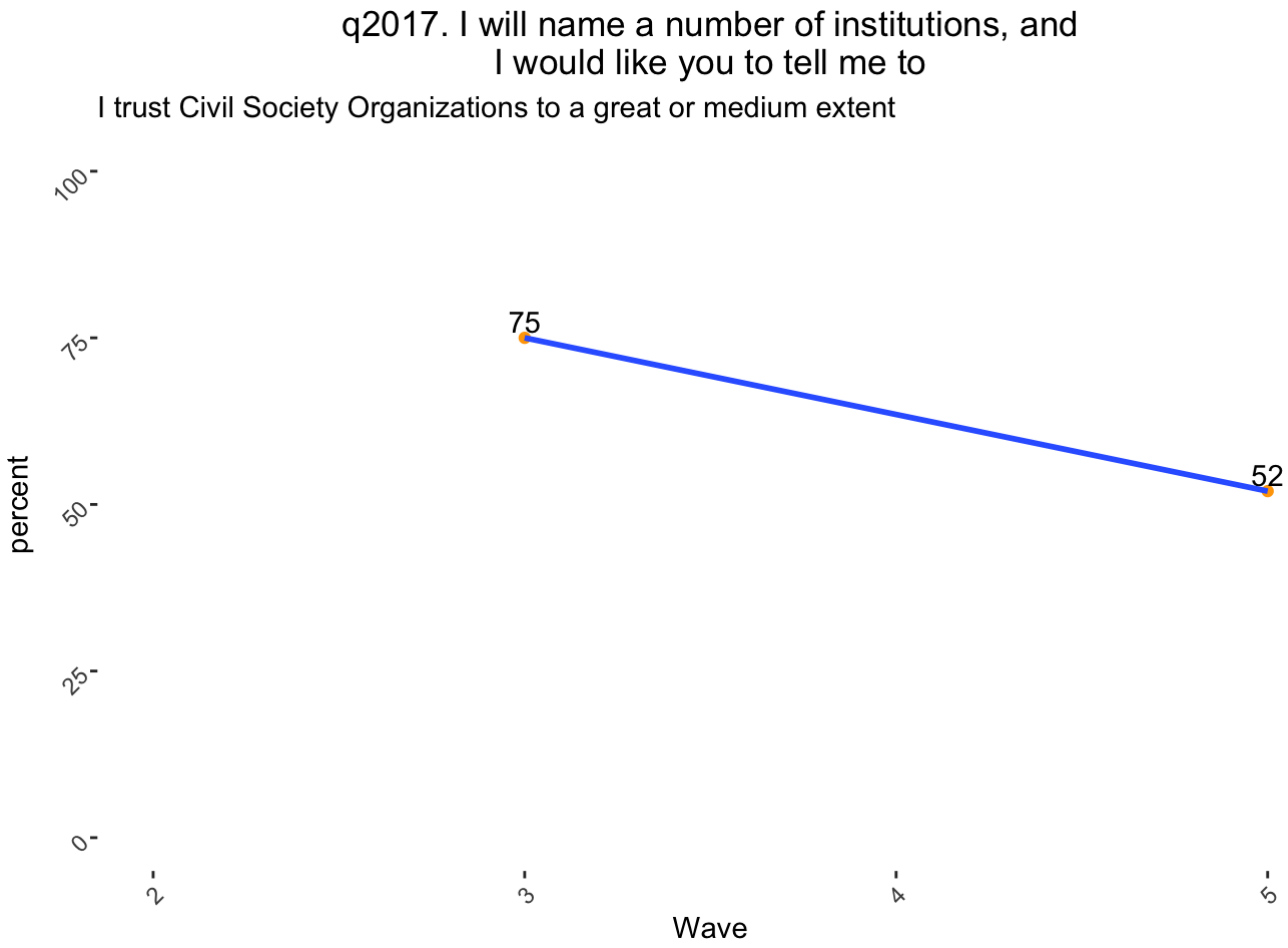
```
labelling(ab_graph$filter_variable,q2013)
plot1(ab_graph, ab_graph$filter_variable, var_lab(q2014),"I trust Regional Government to
 a great or medium extent")
```




```
ab_graph = response_1_2(abv, q2017)
table(abv$wave,abv$q2017)
```

```
##
##       1   2   3   4   9  98  99
##   3 371 395 192  45  18   0   0
##   5 120 595 377 223   0  56   3
```

```
labelling(ab_graph$filter_variable,q2017)
plot1(ab_graph, ab_graph$filter_variable, var_lab(q2017),"I trust Civil Society Organiza
tions to a great or medium extent")
```



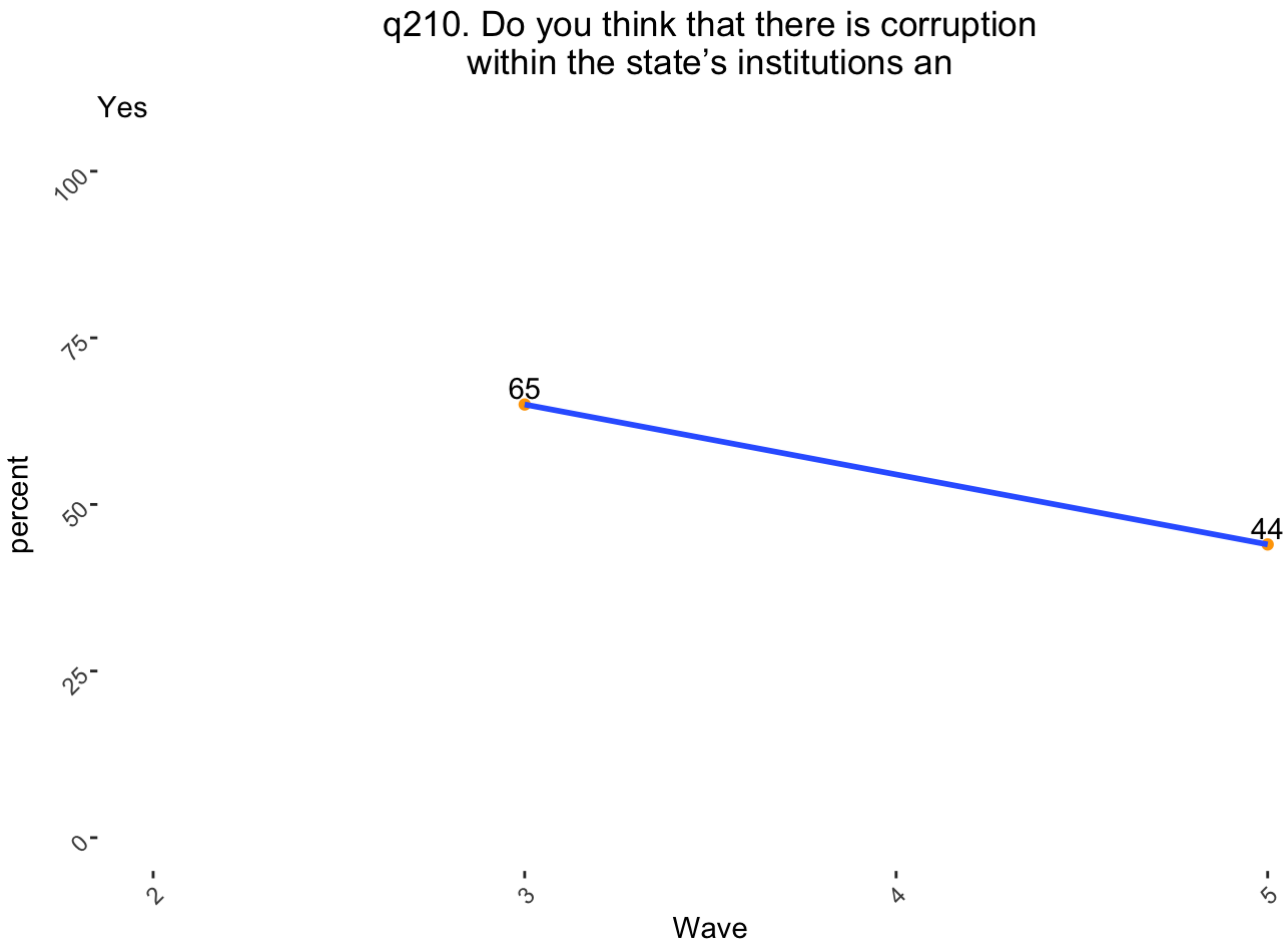

```
ab_graph = filtered_variable(abv, q210,1)
table(abv$wave,abv$q210)
```

```
##
##       1   2   3   4   9  98  99
##   3 661 307   0   0  53   0   0
##   5 599 548 161  36   0  28   2
```

```
labelling(ab_graph$filter_variable,q210)
table(abv$wave,q210)
```

```
##    q210
##       1   2   3   4   9  98  99
##   3 661 307   0   0  53   0   0
##   5 599 548 161  36   0  28   2
```

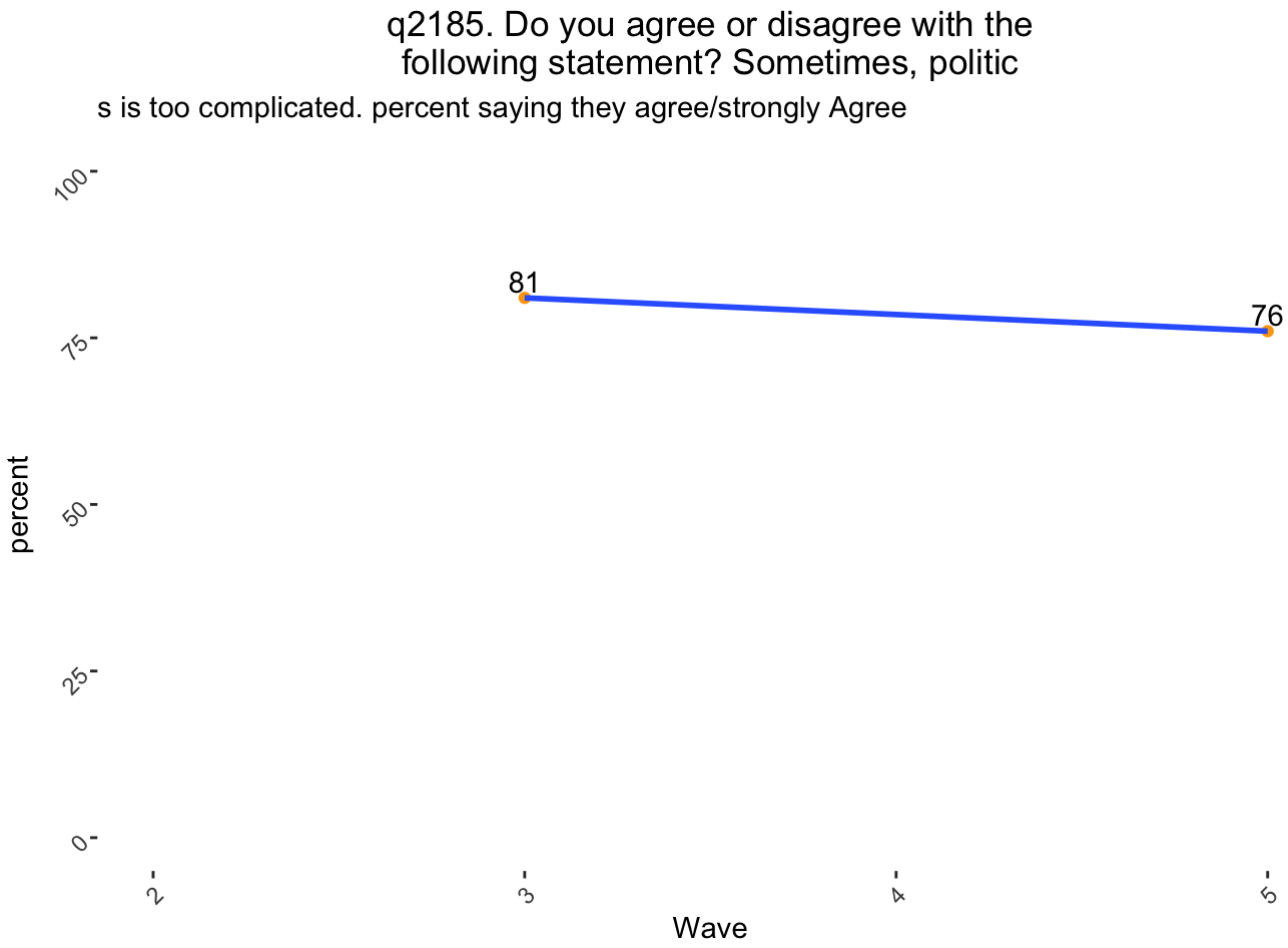
```
plot1(ab_graph, ab_graph$filter_variable, var_lab(q210),"Yes")
```




```
ab_graph = response_1_2(abv, q2185)
labelling(ab_graph$filter_variable,q2185)
table(abv$wave,abv$q2185)
```

```
##
##       1   2   3   4   9  98  99
##   3 279 546 112  25  59   0   0
##   5 216 834 226  30   0  58  10
```

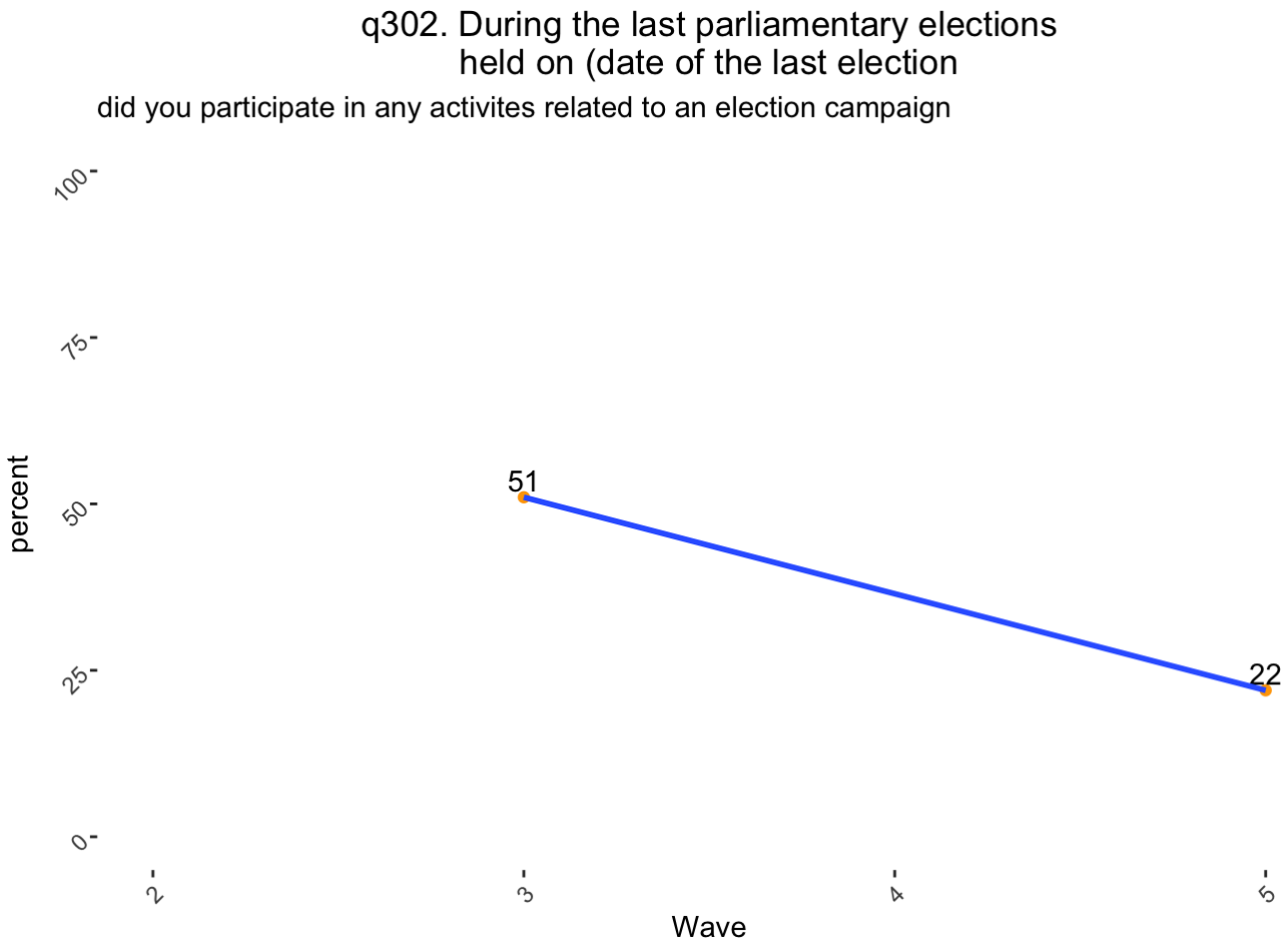
```
plot1(ab_graph, ab_graph$filter_variable, var_lab(q2185),"s is too complicated. percent
 saying they agree/strongly Agree")
```




```
ab_graph = filtered_variable(abv, q302, 1)
labelling(ab_graph$filter_variable,q302)
table(abv$wave,abv$q302)
```

```
##
##        1    2    9   98   99
##   3  523  487   11    0    0
##   5  303 1042    0   24    5
```

```
plot1(ab_graph, ab_graph$filter_variable, var_lab(q302),"did you participate in any acti
vites related to an election campaign")
```

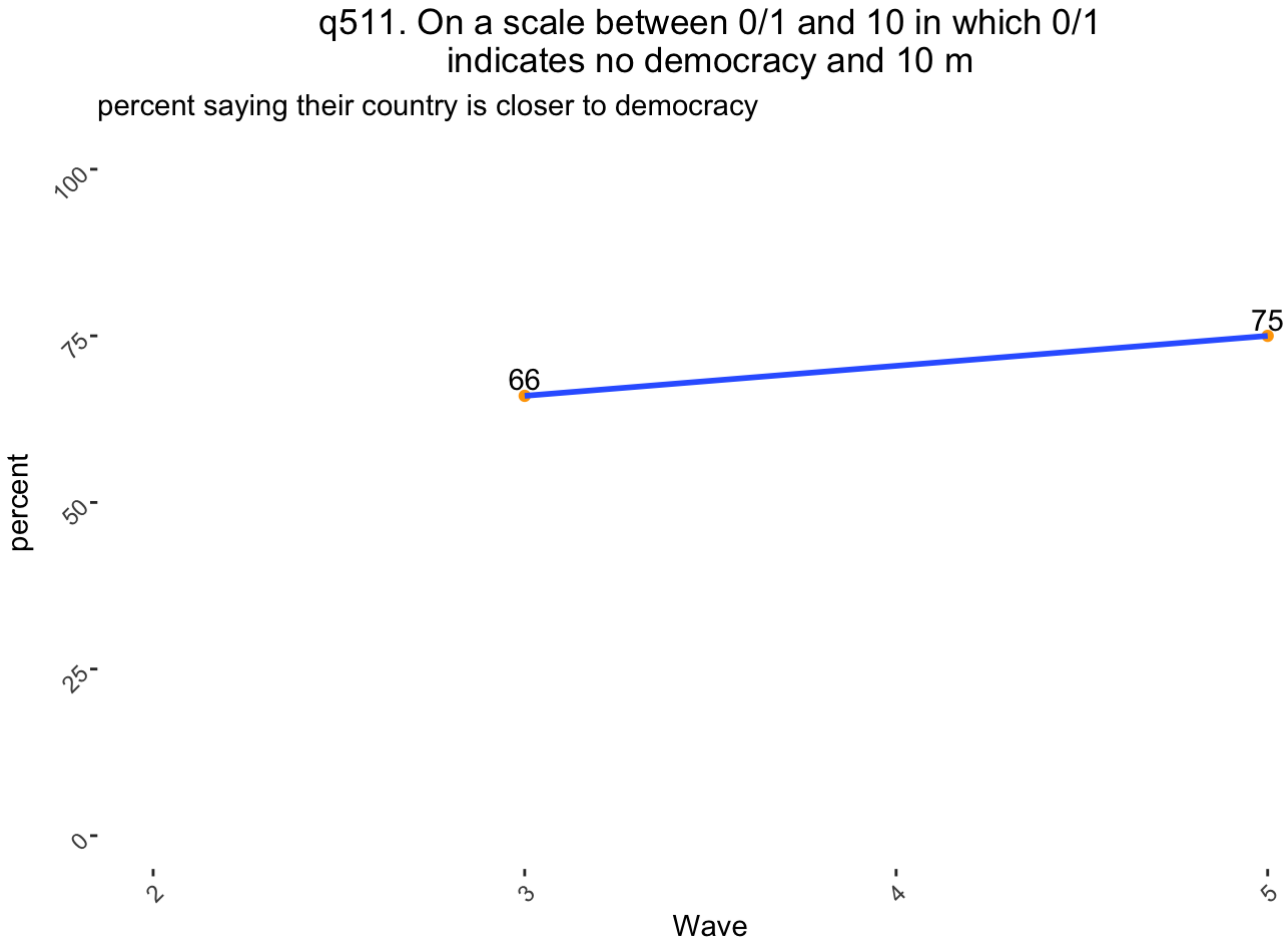



```
greater_than_six=function(data,filter_variable){
  filter_variable<-enquo(filter_variable)
  data %>%
    mutate(variable=as.numeric(ifelse(!!filter_variable==6|!!filter_variable==7|!!filter
_variable==8|!!filter_variable==9|!!filter_variable==10,1,0)))%>%
    group_by(wave)%>%
    dplyr::summarise(filter_variable=round(mean(variable,na.rm=TRUE)*100))
}


ab_graph = greater_than_six(abv, q511)
table(abv$wave,abv$q511)
```

```
##
##       0   1   2   3   4   5   6   7   8   9  10  96  98  99
##   3  18  56  42  25  50 150 127 190 174  66 113   0   0  10
##   5  16  14  22  34  46 146 192 280 241 155 169   2  47  10
```

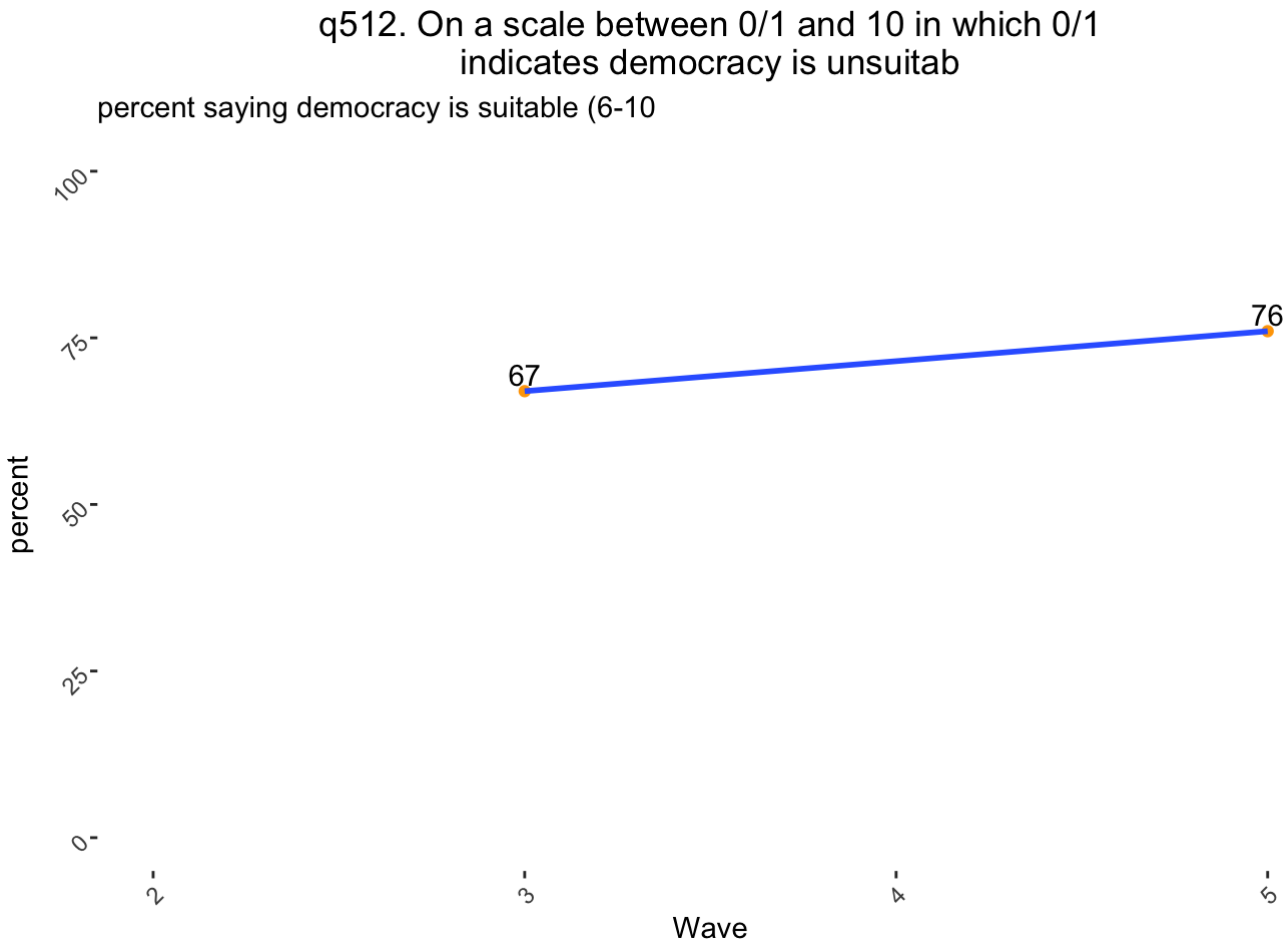
```
labelling(ab_graph$filter_variable,q511)
plot1(ab_graph, ab_graph$filter_variable, var_lab(q511),"percent saying their country is
 closer to democracy")
```




```
ab_graph = greater_than_six(abv, q512)
labelling(ab_graph$filter_variable,q512)
table(abv$wave,abv$q512)
```

```
##
##       0   1   2   3   4   5   6   7   8   9  10  96  98  99
##   3  14  36  31  36  52 156 104 176 147  84 173   0   0  12
##   5  22  11  14  22  35 157 187 222 196 154 289   2  55   8
```

```
plot1(ab_graph, ab_graph$filter_variable, var_lab(q512),"percent saying democracy is sui
table (6-10")
```

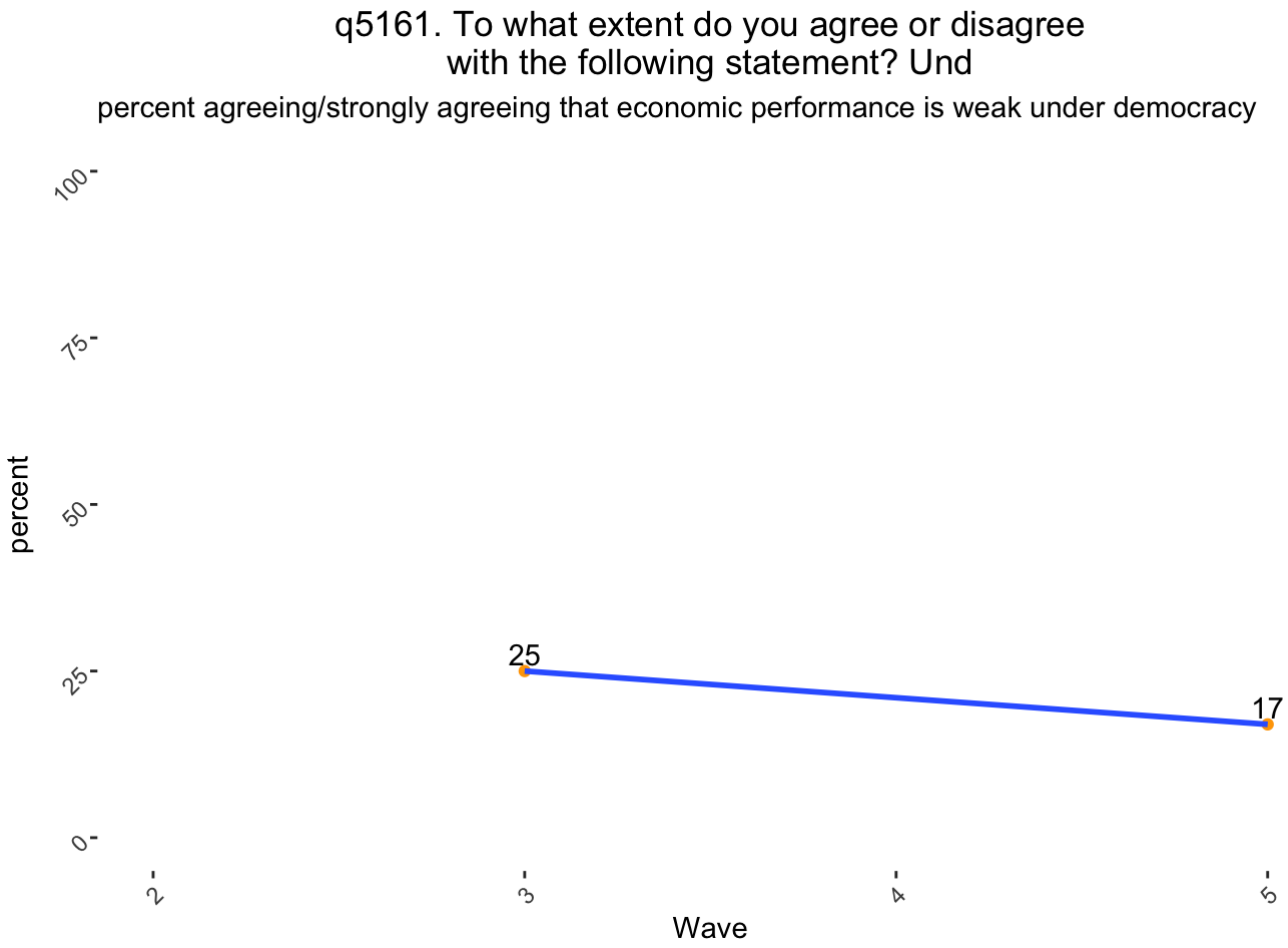



```
ab_graph = response_1_2(abv, q5161)
labelling(ab_graph$filter_variable,q5161)

table(abv$wave, abv$q5161)
```

```
##
##       1   2   3   4   9  98  99
##   3  16 239 585 154  27   0   0
##   5  39 198 738 128   0 247  24
```

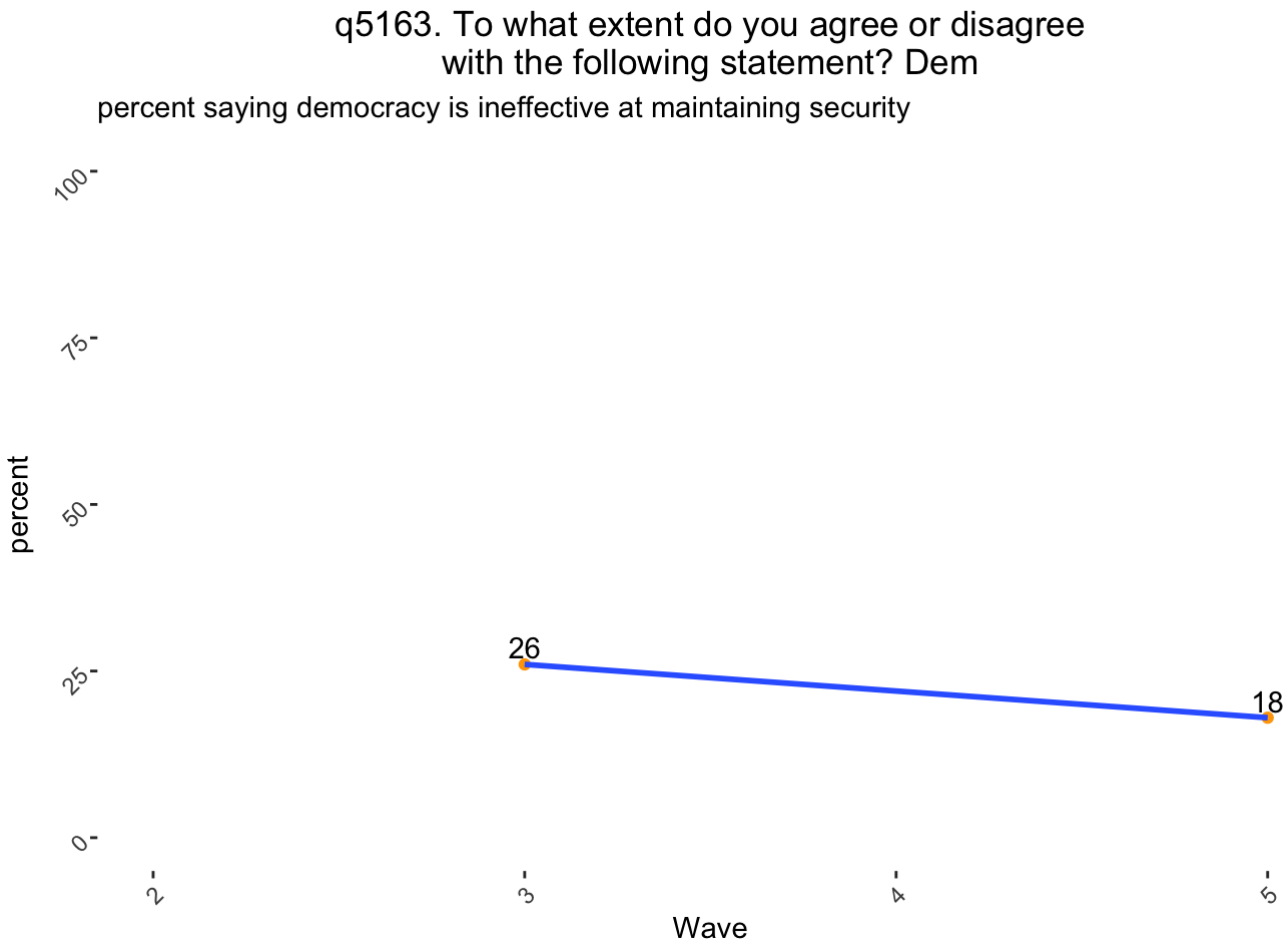
```
plot1(ab_graph, ab_graph$filter_variable, var_lab(q5161),"percent agreeing/strongly agre
eing that economic performance is weak under democracy")
```




```
table(abv$wave, abv$q5163)
```

```
##
##       1   2   3   4   9  98  99
##   3  24 244 578 142  33   0   0
##   5  37 211 698 182   0 218  28
```

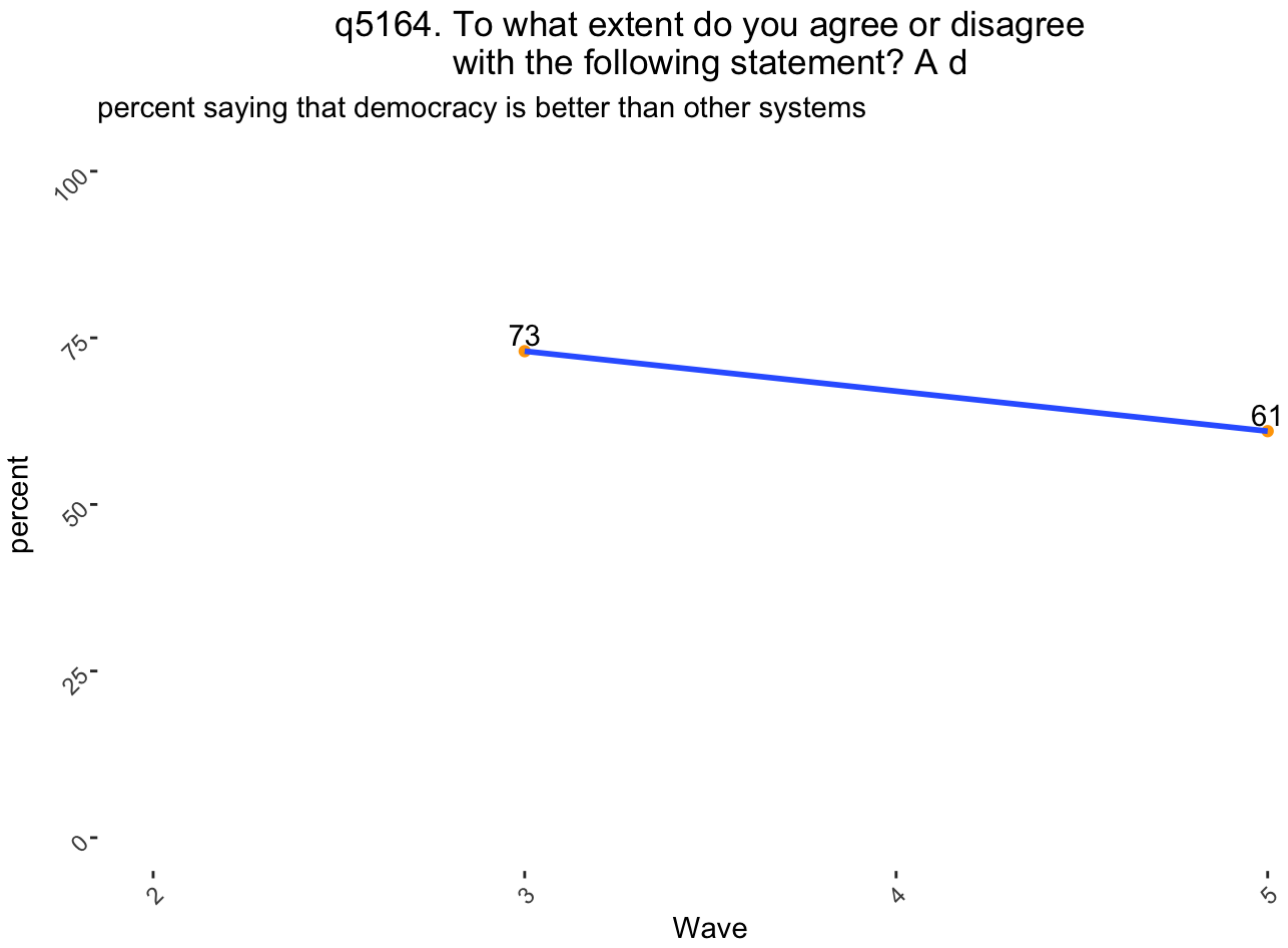
```
ab_graph = response_1_2(abv, q5163)
labelling(ab_graph$filter_variable,q5163)
plot1(ab_graph, ab_graph$filter_variable, var_lab(q5163),"percent saying democracy is in
effective at maintaining security")
```




```
ab_graph = response_1_2(abv, q5164)
labelling(ab_graph$filter_variable,q5164)
table(abv$wave, abv$q5164)
```

```
##
##       1   2   3   4   9  98  99
##   3  89 654 177  71  30   0   0
##   5 206 636 242  63   0 200  27
```

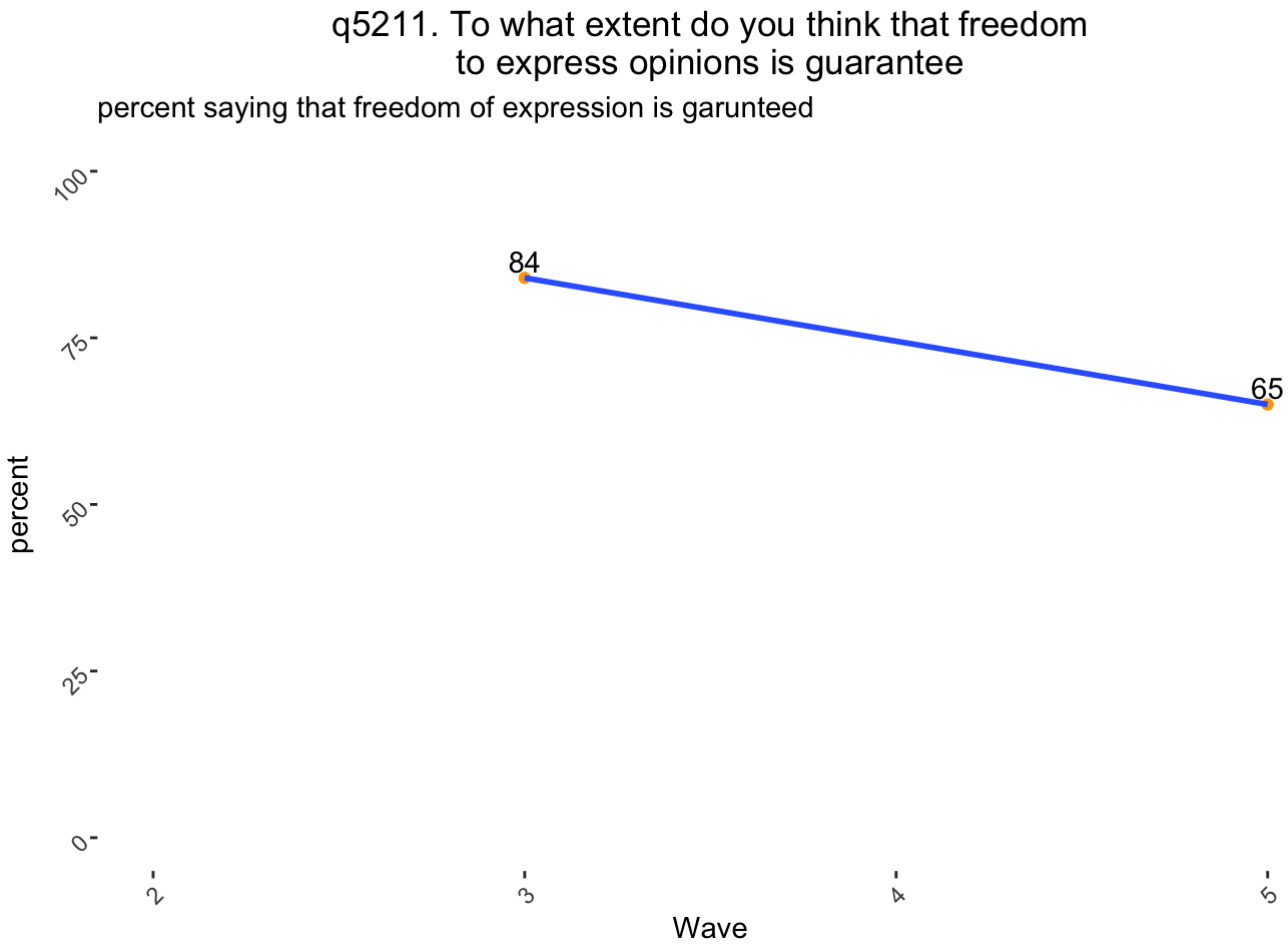
```
plot1(ab_graph, ab_graph$filter_variable, var_lab(q5164),"percent saying that democracy
 is better than other systems")
```




```
ab_graph = response_1_2(abv, q5211)
table(abv$wave,abv$q5211)
```

```
##
##       1   2   3   4   9  98  99
##   3 442 414 138  24   3   0   0
##   5 340 547 193 207   0  71  16
```

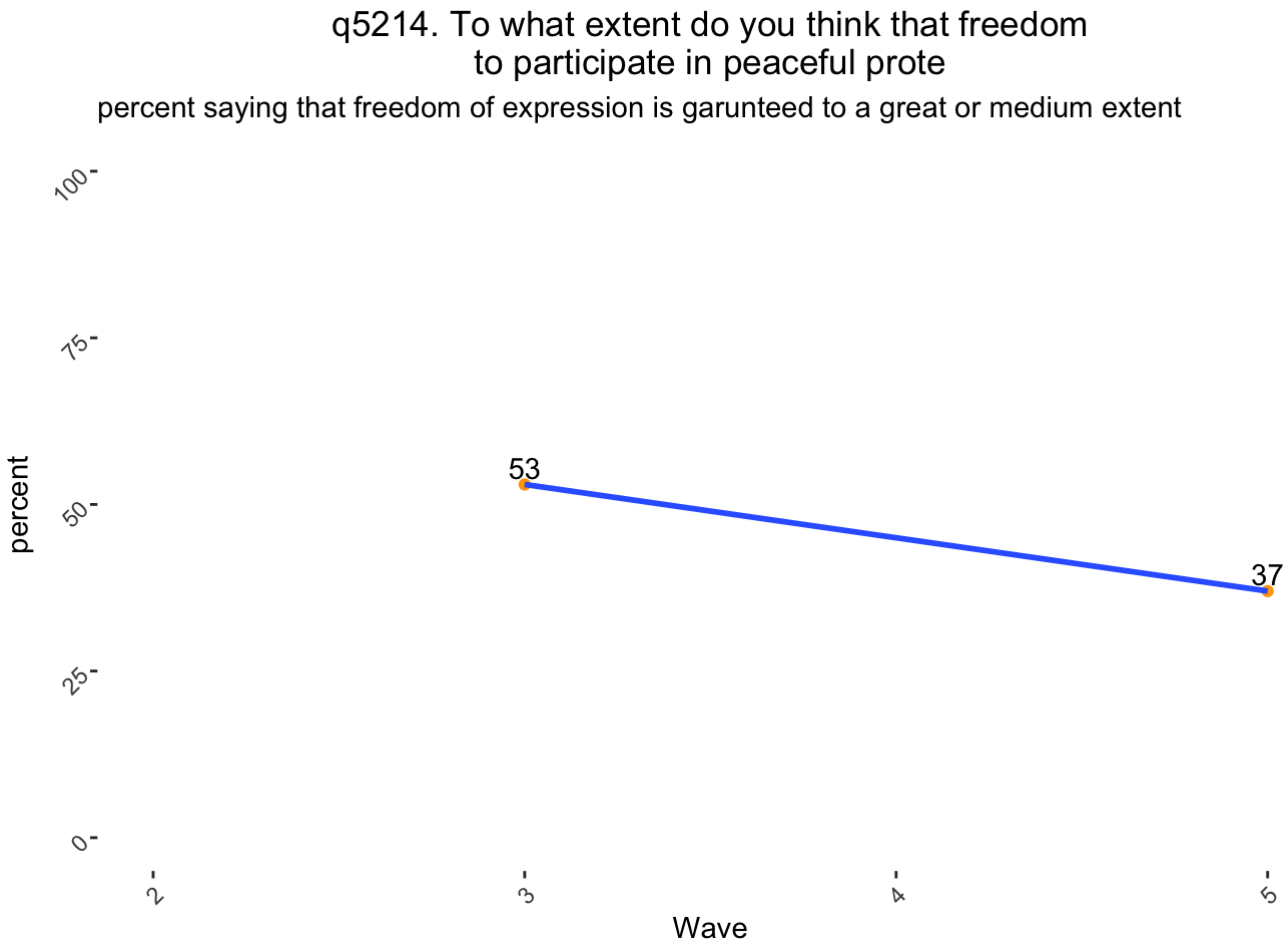
```
labelling(ab_graph$filter_variable,q5211)
plot1(ab_graph, ab_graph$filter_variable, var_lab(q5211),"percent saying that freedom of
 expression is garunteed")
```




```
ab_graph = response_1_2(abv, q5214)
table(abv$wave,abv$q5214)
```

```
##
##       1   2   3   4   9  98  99
##   3 180 365 325 122  29   0   0
##   5 128 387 264 341   0 214  40
```

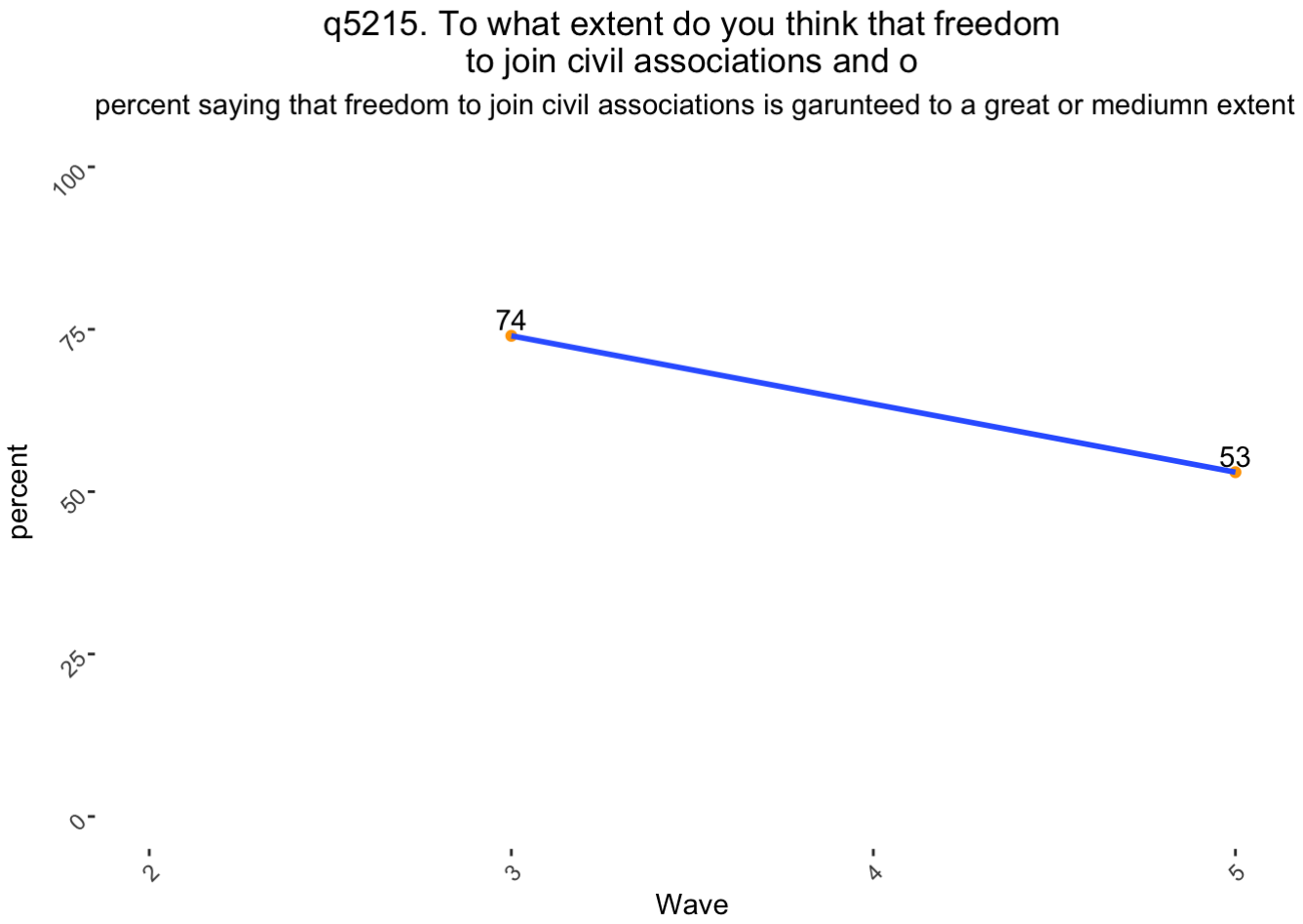
```
labelling(ab_graph$filter_variable,q5214)
plot1(ab_graph, ab_graph$filter_variable, var_lab(q5214),"percent saying that freedom of
 expression is garunteed to a great or medium extent")
```




```
ab_graph = response_1_2(abv, q5215)
table(abv$wave,abv$q5215)
```

```
##
##       1   2   3   4   9  98  99
##   3 392 359 170  52  48   0   0
##   5 284 451 215 167   0 236  21
```

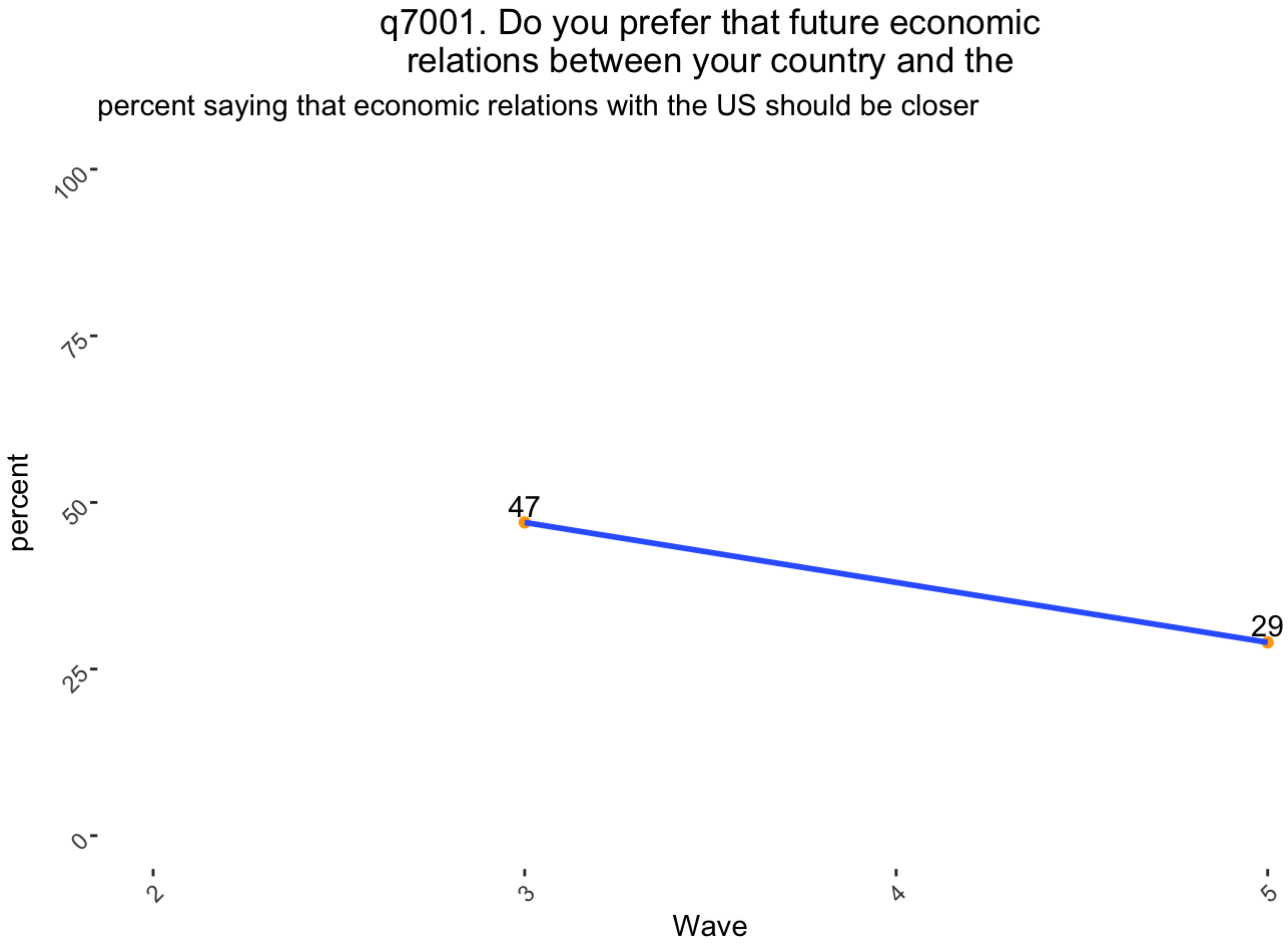
```
labelling(ab_graph$filter_variable,q5215)
plot1(ab_graph, ab_graph$filter_variable, var_lab(q5215),"percent saying that freedom to
 join civil associations is garunteed to a great or mediumn extent")
```




```
ab_graph = filtered_variable(abv, q7001, 1)
labelling(ab_graph$filter_variable,q7001)
table(abv$wave,abv$q7001)
```

```
##
##       1   2   3   9  98  99
##   3 478 361 164  18   0   0
##   5 397 582 234   0 148  13
```

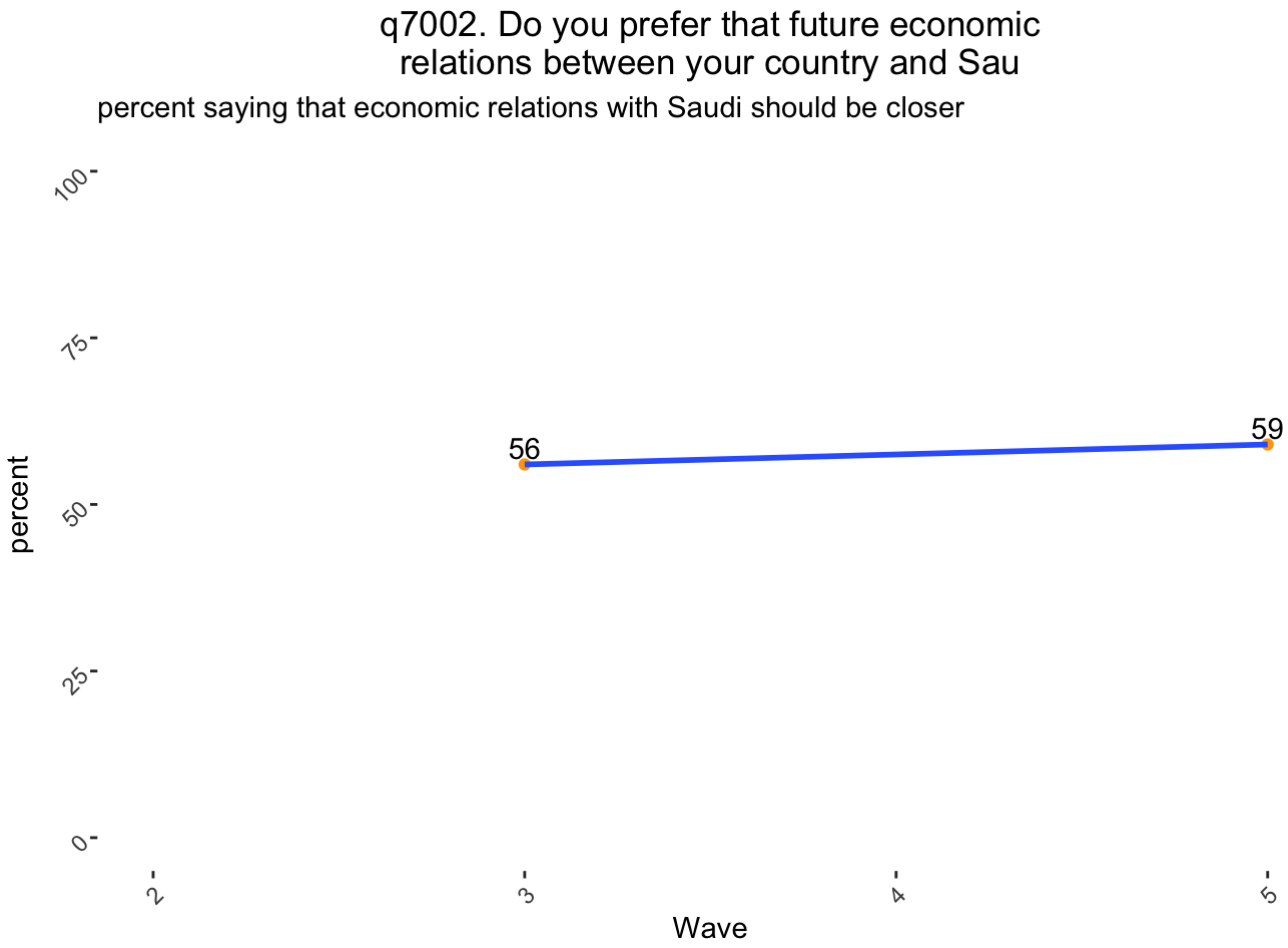
```
plot1(ab_graph, ab_graph$filter_variable, var_lab(q7001),"percent saying that economic r
elations with the US should be closer")
```




```
ab_graph = filtered_variable(abv, q7002, 1)
labelling(ab_graph$filter_variable,q7002)
table(abv$wave,abv$q7002)
```

```
##
##       1   2   3   9  98  99
##   3 570 350  93   8   0   0
##   5 815 427  57   0  67   8
```

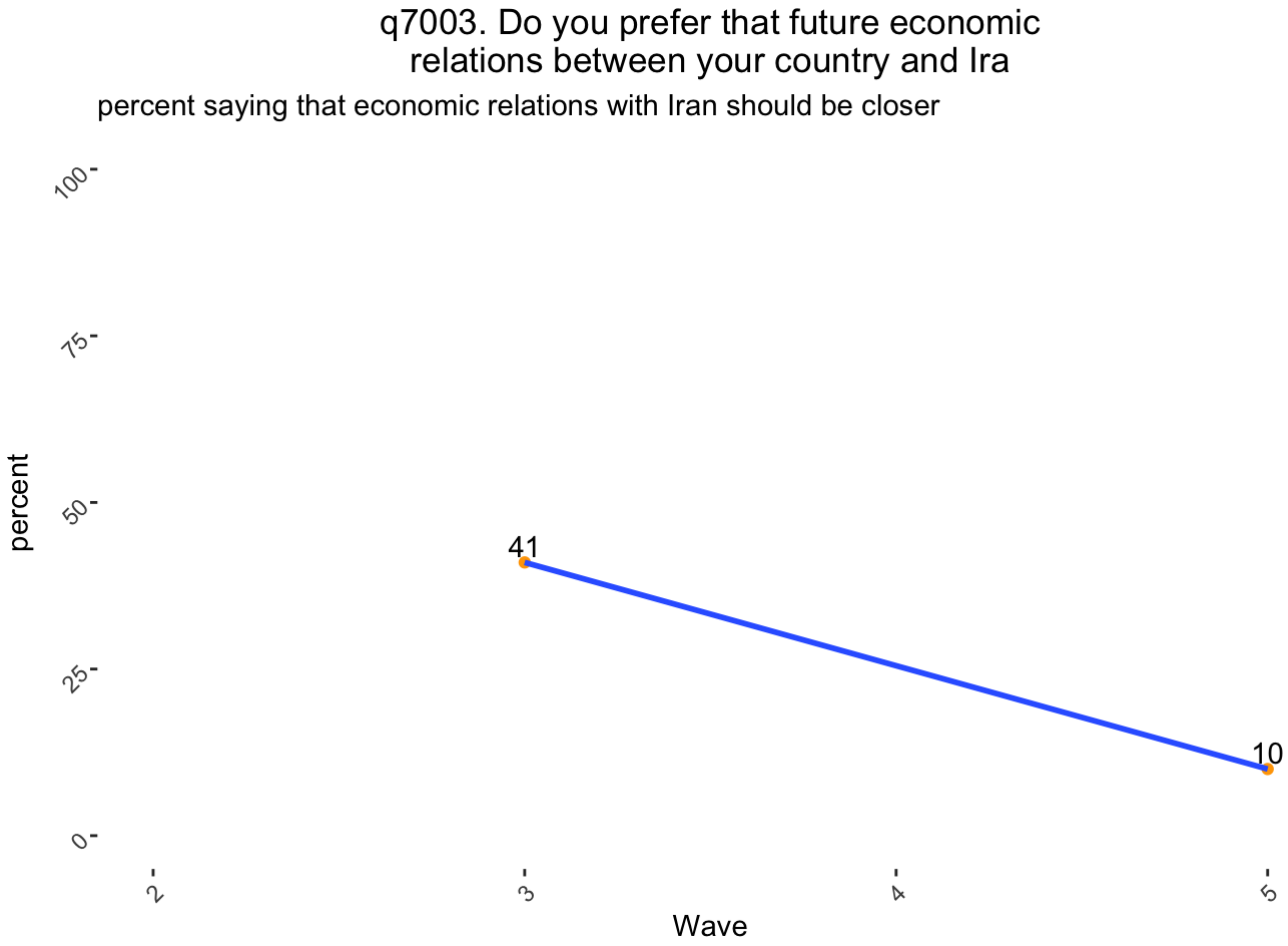
```
plot1(ab_graph, ab_graph$filter_variable, var_lab(q7002),"percent saying that economic r
elations with Saudi should be closer")
```




```
ab_graph = filtered_variable(abv, q7003, 1)
labelling(ab_graph$filter_variable,q7003)
table(abv$wave,abv$q7003)
```

```
##
##       1   2   3   9  98  99
##   3 418 353 229  21   0   0
##   5 138 375 622   0 171  68
```

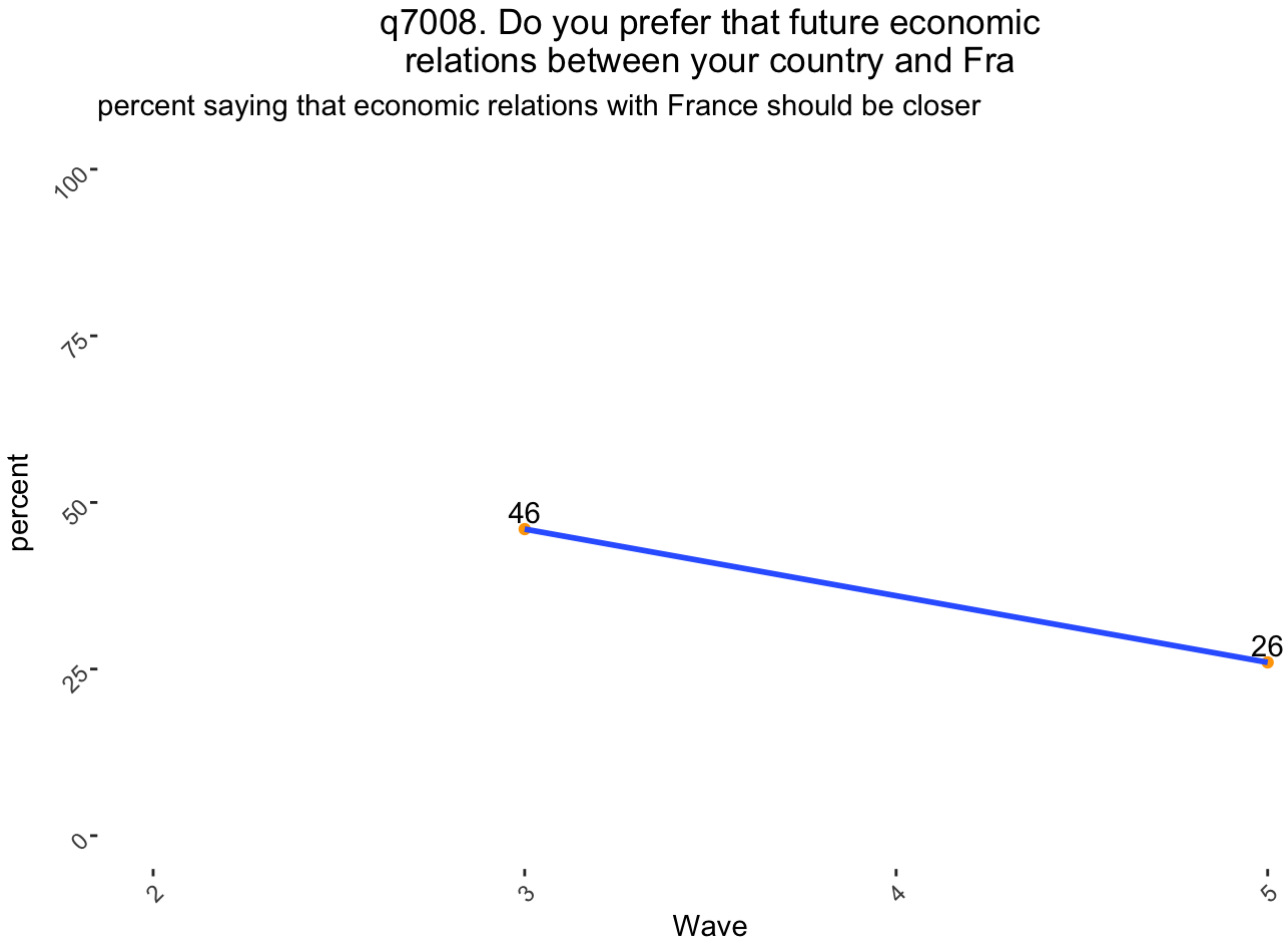
```
plot1(ab_graph, ab_graph$filter_variable, var_lab(q7003),"percent saying that economic r
elations with Iran should be closer")
```




```
ab_graph = filtered_variable(abv, q7008, 1)
labelling(ab_graph$filter_variable,q7008)
table(abv$wave,abv$q7008)
```

```
##
##       1   2   3   9  98  99
##   3 471 413 116  21   0   0
##   5 354 527 260   0 215  18
```

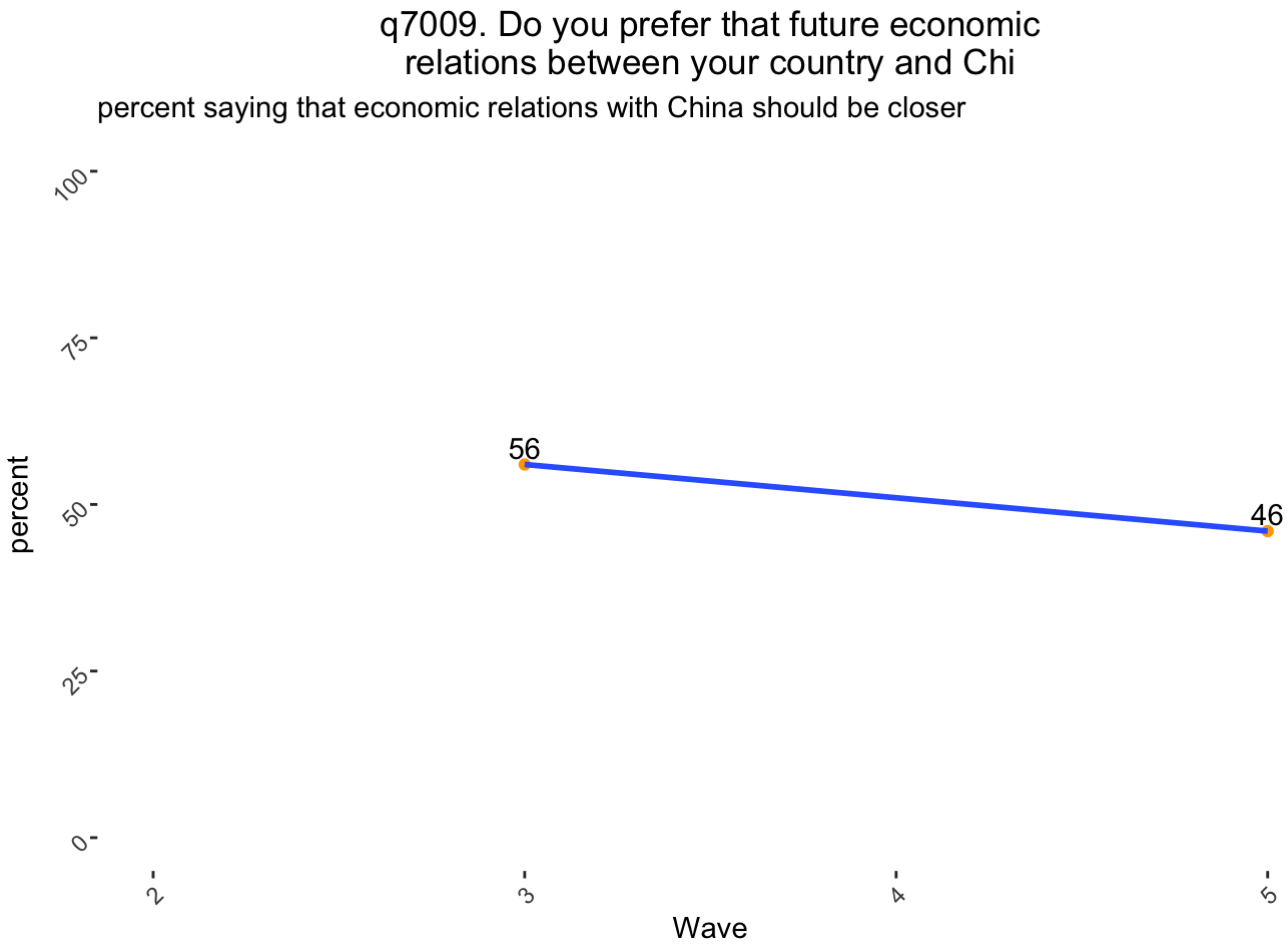
```
plot1(ab_graph, ab_graph$filter_variable, var_lab(q7008),"percent saying that economic r
elations with France should be closer")
```




```
ab_graph = filtered_variable(abv, q7009, 1)
labelling(ab_graph$filter_variable,q7009)
table(abv$wave,abv$q7009)
```

```
##
##       1   2   3   9  98  99
##   3 572 315 118  16   0   0
##   5 631 517 101   0 111  14
```

```
plot1(ab_graph, ab_graph$filter_variable, var_lab(q7009),"percent saying that economic r
elations with China should be closer")
```


q7009. Do you prefer that future economic relations between your country and Chi
percent saying that economic relations with China should be closer
100
75
50
25
0
percent
56
46
2
3
4
5
Wave