

Results and Operationalization

[Code ▼](#)

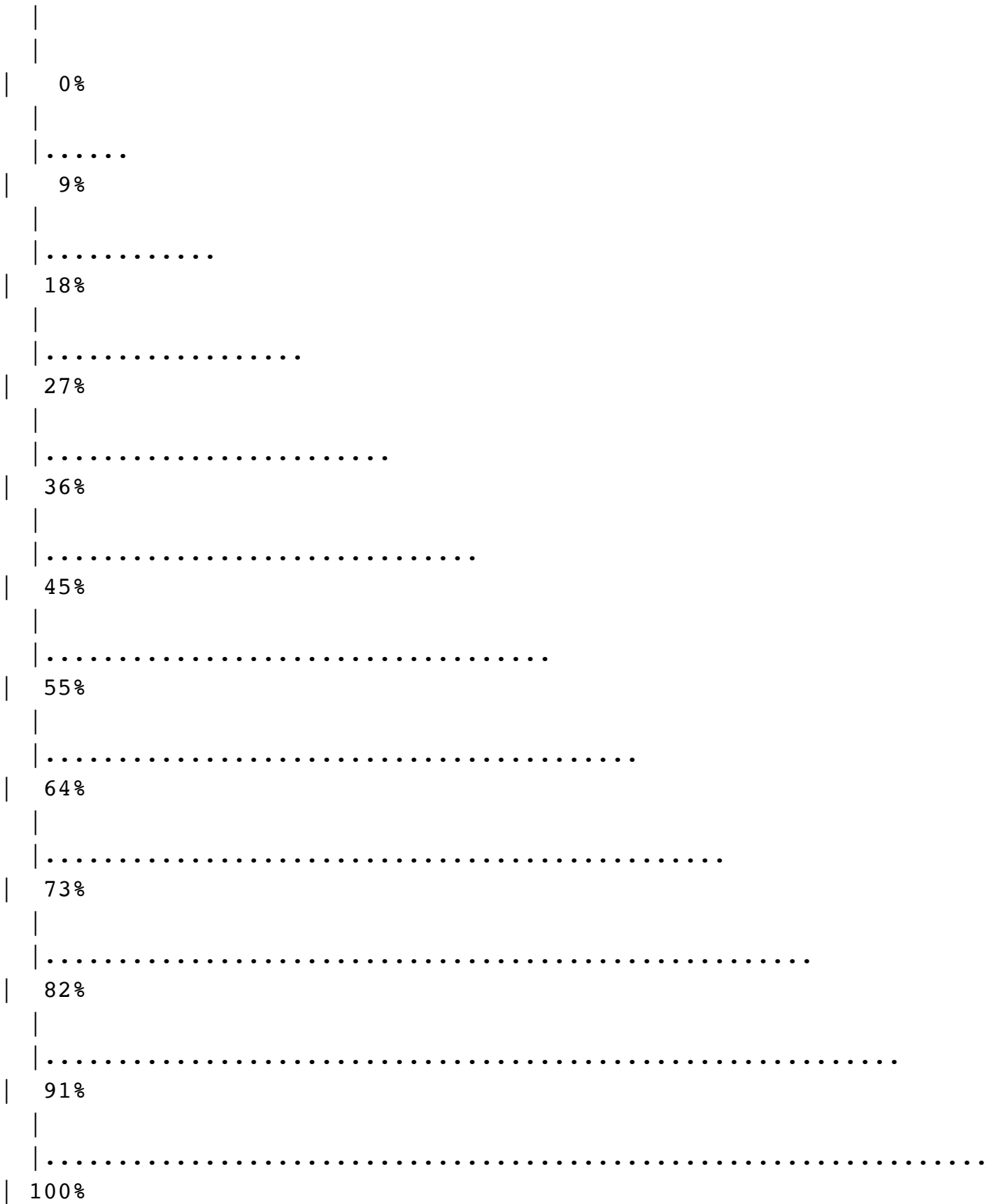
Albert Swingler

[Hide](#)[Hide](#)

```
library("tidyverse")  
library("rvest")  
library("dplyr")  
library("knitr")  
library("caret")
```

```
purl("deliverable1.Rmd", output = "part1.r") # produces r source from rmd
```

```
processing file: deliverable1.Rmd
```



output file: part1.r

```
[1] "part1.r"
```

Hide

Hide

```
purl("phase2.Rmd", output = "part2.r")
```

```
processing file: phase2.Rmd
```

```
|
|
| 0%
|
| .....
| 8%
|
| .....
| 15%
|
| .....
| 23%
|
| .....
| 31%
|
| .....
| 38%
|
| .....
| 46%
|
| .....
| 54%
|
| .....
| 62%
```



```
output file: part2.r
```

```
[1] "part2.r"
```

[Hide](#)[Hide](#)

```
source("part1.r") # executes the source
```

Parsed with column specification:

```
cols(  
  congress = [32mcol_double()][39m,  
  chamber = [31mcol_character()][39m,  
  bioguide = [31mcol_character()][39m,  
  last_name = [31mcol_character()][39m,  
  state = [31mcol_character()][39m,  
  district = [32mcol_double()][39m,  
  party = [31mcol_character()][39m,  
  votes = [32mcol_double()][39m,  
  agree_pct = [32mcol_double()][39m,  
  predicted_agree = [32mcol_double()][39m,  
  net_trump_vote = [32mcol_double()][39m  
)
```

Parsed with column specification:

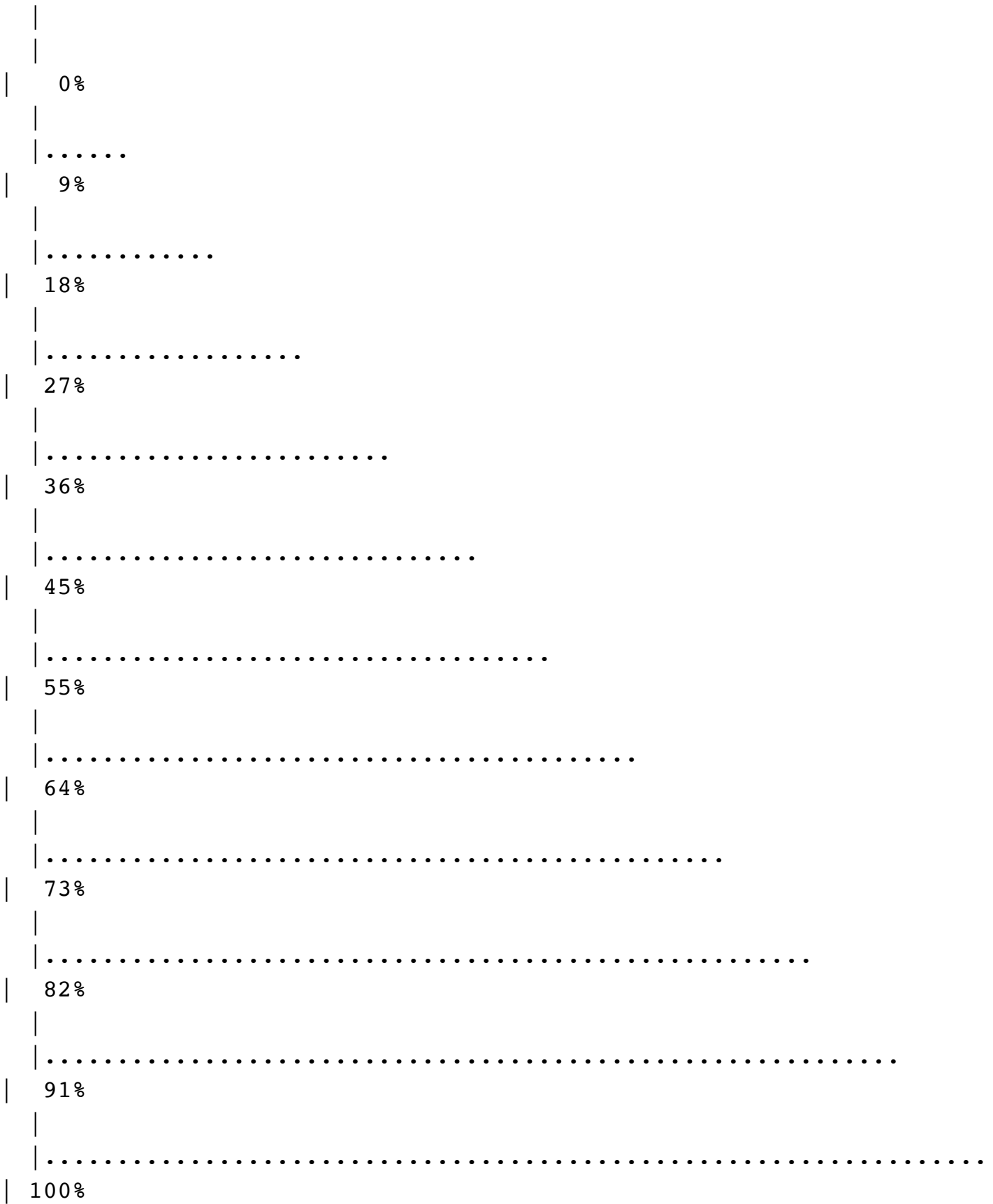
```
cols(  
  congress = [32mcol_double()][39m,  
  bill_id = [31mcol_character()][39m,  
  roll_id = [31mcol_character()][39m,  
  chamber = [31mcol_character()][39m,  
  voted_at = [31mcol_character()][39m,  
  bioguide = [31mcol_character()][39m,  
  vote = [31mcol_character()][39m,  
  trump_position = [31mcol_character()][39m,  
  last_name = [31mcol_character()][39m,  
  state = [31mcol_character()][39m,  
  district = [32mcol_double()][39m,  
  party = [31mcol_character()][39m,  
  agree = [32mcol_double()][39m,  
  yesno = [32mcol_double()][39m,  
  predicted_probability = [32mcol_double()][39m  
)
```

Hide

Hide

```
source("part2.r")
```

processing file: deliverable1.Rmd



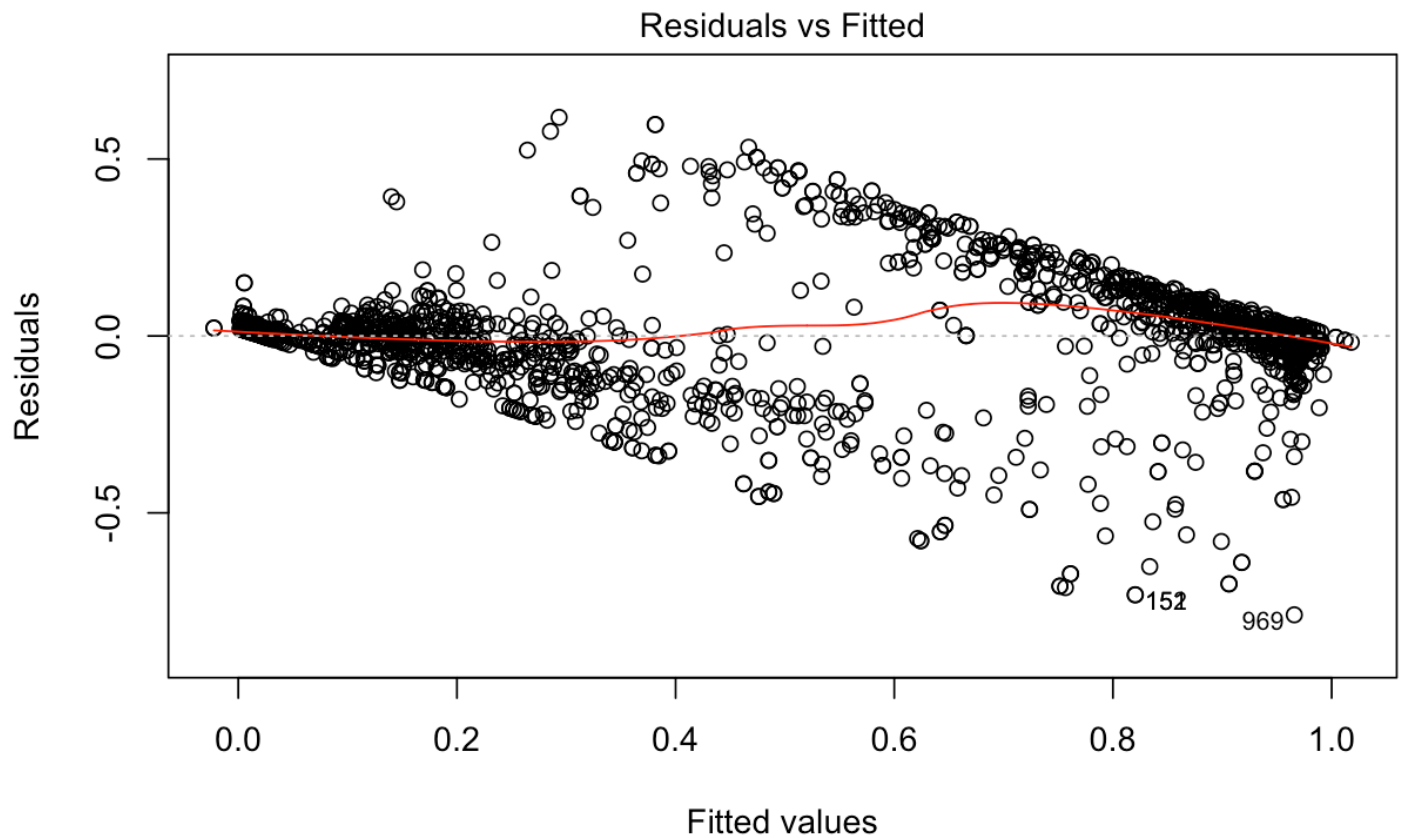
output file: part1.r

Parsed with column specification:

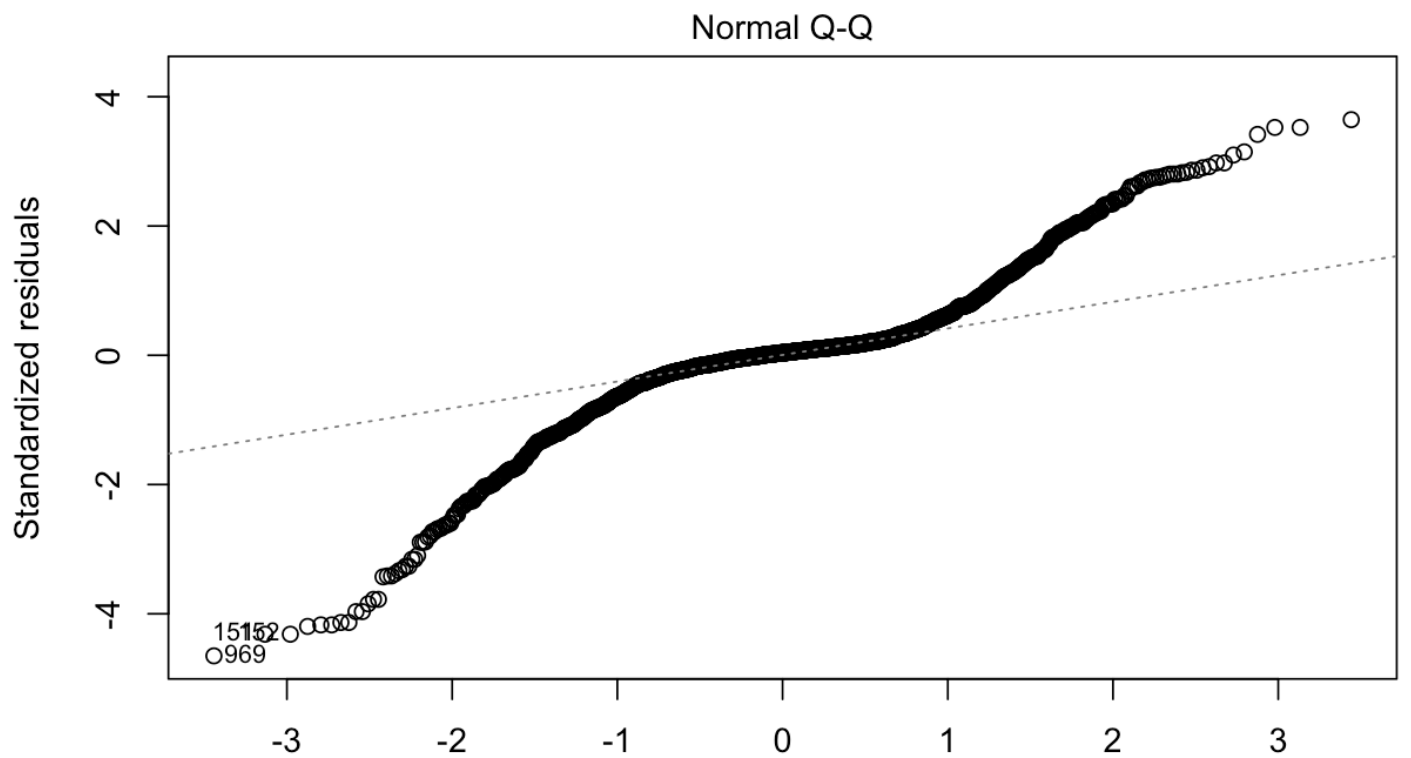
```
cols(  
  congress = [32mcol_double()][39m,  
  chamber = [31mcol_character()][39m,  
  bioguide = [31mcol_character()][39m,  
  last_name = [31mcol_character()][39m,  
  state = [31mcol_character()][39m,  
  district = [32mcol_double()][39m,  
  party = [31mcol_character()][39m,  
  votes = [32mcol_double()][39m,  
  agree_pct = [32mcol_double()][39m,  
  predicted_agree = [32mcol_double()][39m,  
  net_trump_vote = [32mcol_double()][39m  
)
```

Parsed with column specification:

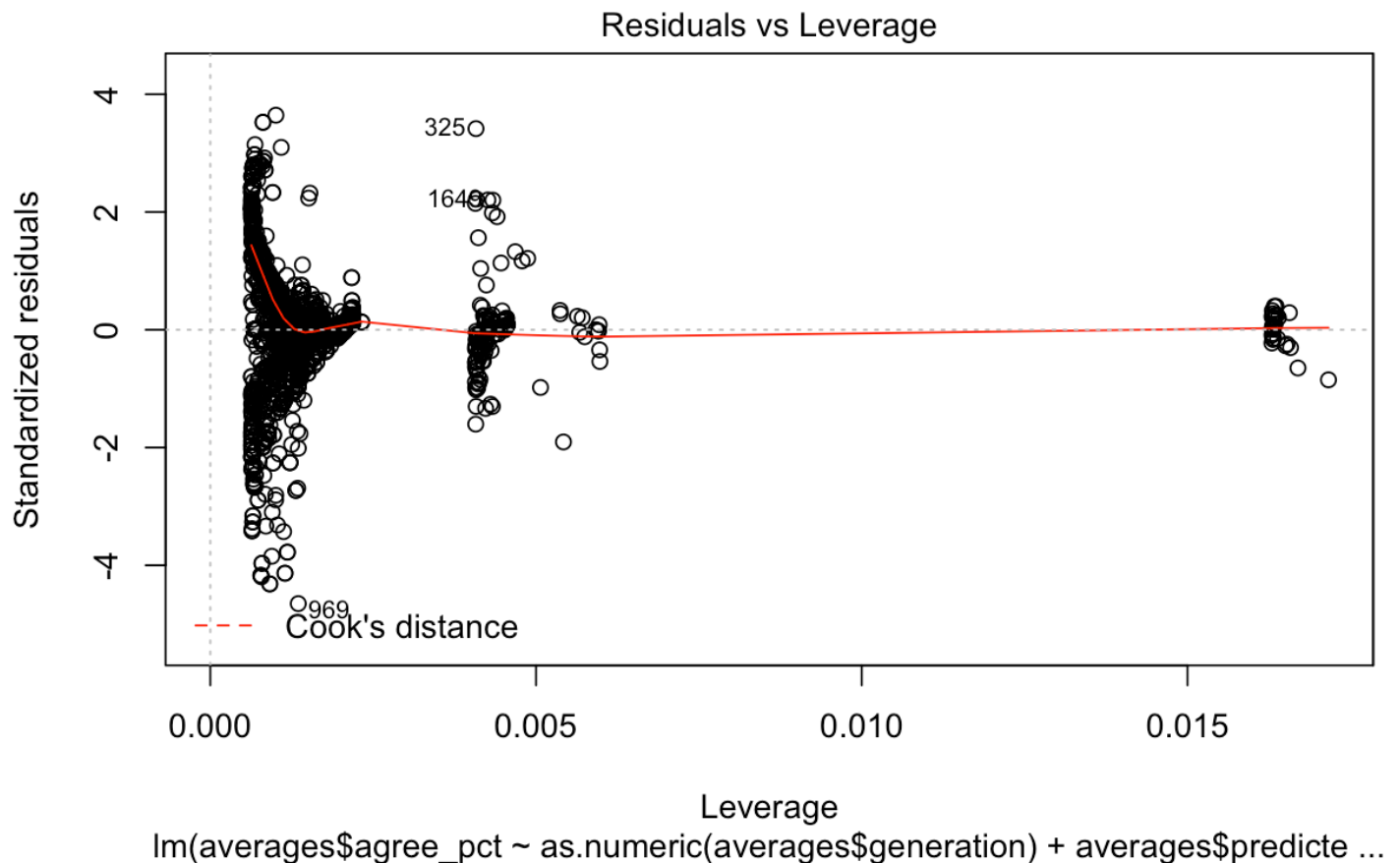
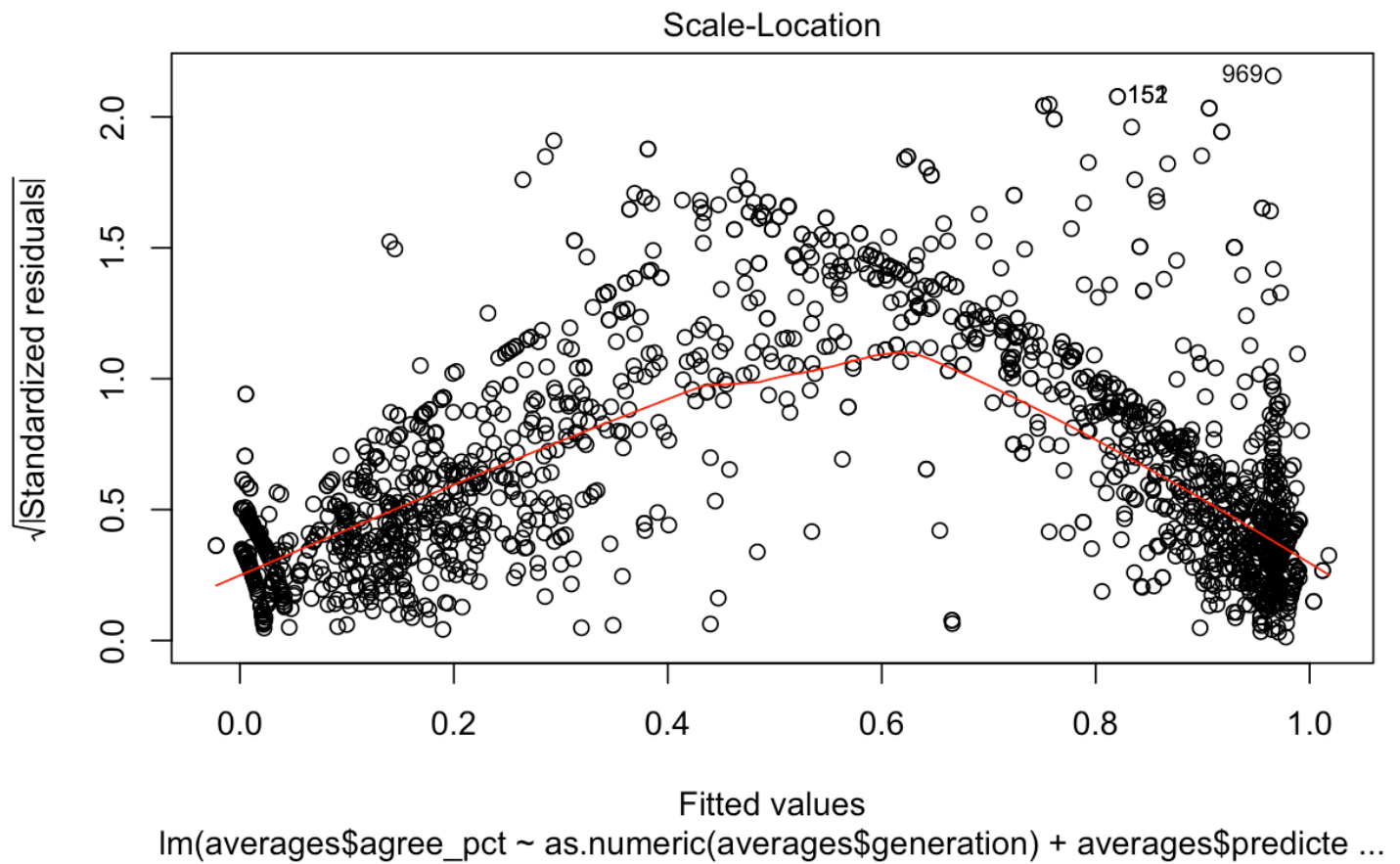
```
cols(  
  congress = [32mcol_double()][39m,  
  bill_id = [31mcol_character()][39m,  
  roll_id = [31mcol_character()][39m,  
  chamber = [31mcol_character()][39m,  
  voted_at = [31mcol_character()][39m,  
  bioguide = [31mcol_character()][39m,  
  vote = [31mcol_character()][39m,  
  trump_position = [31mcol_character()][39m,  
  last_name = [31mcol_character()][39m,  
  state = [31mcol_character()][39m,  
  district = [32mcol_double()][39m,  
  party = [31mcol_character()][39m,  
  agree = [32mcol_double()][39m,  
  yesno = [32mcol_double()][39m,  
  predicted_probability = [32mcol_double()][39m  
)
```

`lm(averages$agree_pct ~ as.numeric(averages$generation) + averages$predicte ...`



`lm(averages$agree_pct ~ as.numeric(averages$generation) + averages$predicte ...`



-Updated my p2 model a lot after the peer review and made it more clear what was being displayed and what it meant for my analysis. -Updated the last graph in p1 to show the party of each politician represented. This made the chart much better at

showing what I found. -Removed unnecessary commented out code and white space

More models. Switched the formula to model predicted agree from the first deliverable data source instead of the actual agree just to see if there was a difference.

Hide

Hide

```
set.seed(333)

index <- createDataPartition(averages$agree_pct, p = 0.75, list = FALSE)

train <- averages[index, ]
test <- averages[-index, ]

trained_model <- lm(train, formula = predicted_agree~agree_pct+generation)
summary(trained_model)
```

Call:

```
lm(formula = predicted_agree ~ agree_pct + generation, data = train)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.53601	-0.08818	0.00084	0.07371	0.62017

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.04182	0.02783	1.503	0.13318
agree_pct	0.79010	0.01014	77.895	< 2e-16 ***
generation2	0.02659	0.03131	0.849	0.39598
generation3	0.07783	0.02841	2.740	0.00623 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1417 on 1296 degrees of freedom

Multiple R-squared: 0.8362, Adjusted R-squared: 0.8358

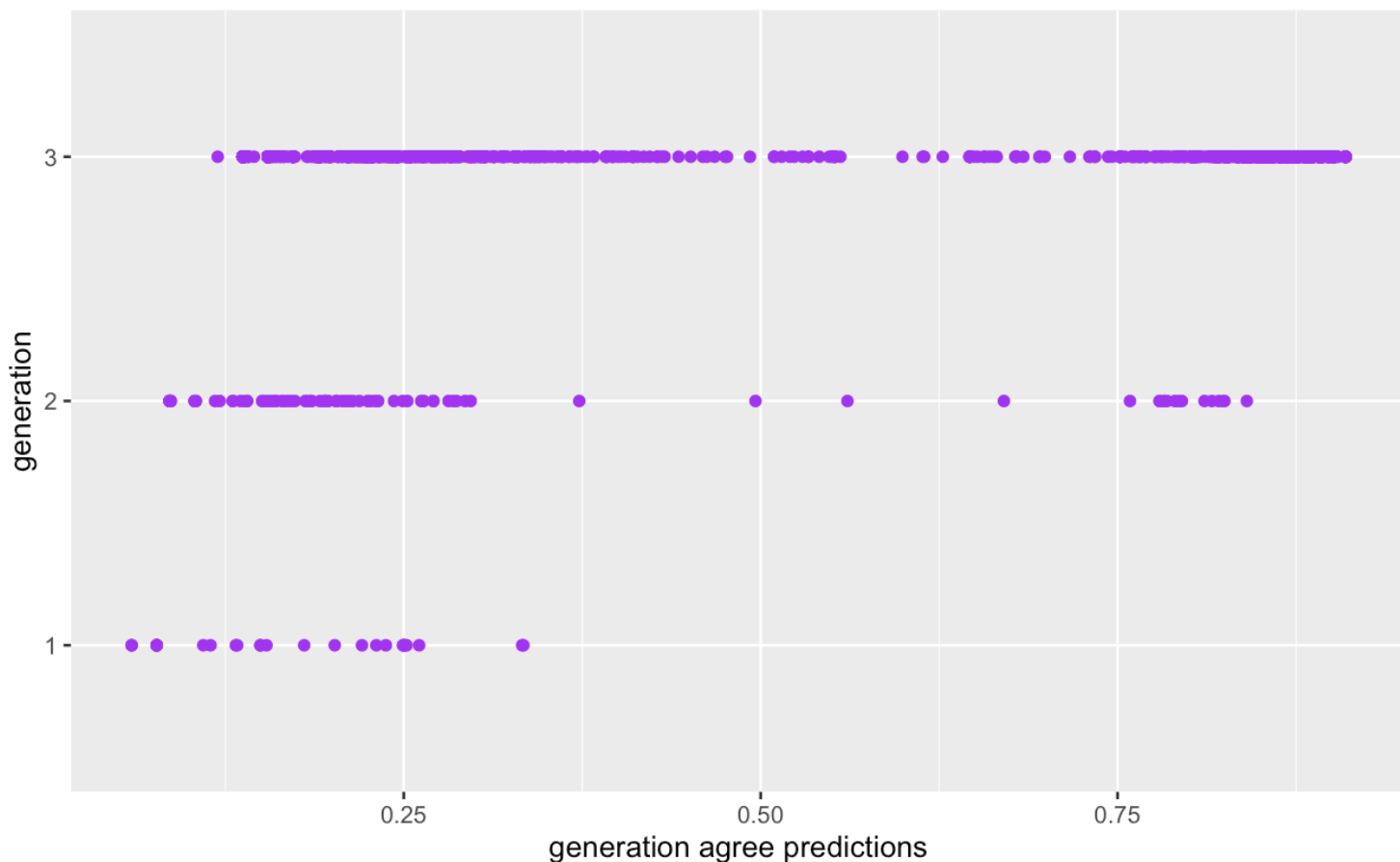
F-statistic: 2205 on 3 and 1296 DF, p-value: < 2.2e-16

Hide

Hide

Hide

```
generation_agree_predictions <- predict(trained_model, test)
ggplot(data=test, aes(generation_agree_predictions, generation)) +
  geom_point(color='purple') + xlab('generation agree predictions')
```



Hide

Hide

NA

NA

Cross-validation

Operationalization: The information learned from this project can be used to educate voters on whether their representatives in congress are voting for things they want. Although it shouldn't be used exclusively to decide who to vote for, being able to predict how a given politician will vote based on their generation may be useful information for voters to factor in to their decisions.

This data analysis has several important social and ethical implications. The information found that the more recent a congressperson and their family immigrated to the United States are much less likely to vote with Trump is good to realize. Knowing this information someone could choose to push for or advocate for one candidate over the other before even hearing what their opinion on the issues is. A media outlet could be more or less willing to urge their viewers to support a particular candidate based on this alone.

The next thing to do for this project would be to continue gathering data and trying to predict votes. Maybe data on things such as politician age or education could be used as predictors. The education idea could potentially have some really interesting results.

There isn't a ton that can be done on this either as it is. It's no surprise that immigrants are less likely to vote with Trump. No one is able to decide what generation they are like they are with their policy positions so it's possible someone could be judged one way or the other before being given a chance to express their views. People need to be aware that while one's generation is just one of the many things that contribute to one's political alignment, it is far from the only one.

operationaization -what next what do we model? data collection, data automation, unintended consequences

social impact -who it serves/impacts -social responses -ethical dilemmas