

Plots & Summary Statistics for Data Analysis

Table of contents

0.0.1	distribution of contributions	1
0.0.2	contributions to party plots	8
0.0.3	plot change in vote dummy vs. state, party, etc.	9
0.0.4	plot change in mind with each roll_call... (to be able to analyse the contribution better)	10
0.0.5	mind changers table -> strattman replicate	12

0.0.1 distribution of contributions

```
# summary statistics  
str(df)
```

```
'data.frame':  560 obs. of  46 variables:  
 $ last_name      : chr  "Roby" "Rogers" "Aderholt" "Brooks" ...  
 $ first_name     : chr  "Martha" "Mike D." "Robert B." "Mo" ...  
 $ name          : chr  "martha roby" "mike d rogers" "robert b aderholt" "mo brooks" ...  
 $ party         : chr  "R" "R" "R" "R" ...  
 $ state         : chr  "AL" "AL" "AL" "AL" ...  
 $ member_id     : chr  "R000591" "R000575" "A000055" "B001274" ...  
 $ Vote3         : chr  "-" "-" "-" "-" ...  
 $ Vote4         : chr  "-" "-" "-" "-" ...  
 $ Vote51        : chr  "-" "-" "-" "-" ...  
 $ Vote52        : chr  "-" "-" "-" "-" ...  
 $ Vote6         : chr  "-" "-" "-" "-" ...  
 $ Vote7         : chr  NA "-" "-" "-" ...  
 $ District      : chr  "AL-02" "AL-03" "AL-04" "AL-05" ...  
 $ Vote_count    : int  5 6 6 6 6 6 3 6 5 2 ...
```

```

$ Vote_change      : int  0 0 0 0 0 0 0 0 0 0 ...
$ Vote_change_dummy: int  0 0 0 0 0 0 0 0 0 0 ...
$ amount.oil.113   : int  7000 18750 32700 6250 NA 55450 6750 NA NA 44139 ...
$ amount.coal.113  : int  3500 15000 10500 0 NA 2250 0 NA NA 2500 ...
$ amount.mining.113: int  4500 16000 11500 1000 NA 8550 4000 NA NA 8500 ...
$ amount.gas.113   : int  1000 0 0 0 NA 7000 1500 NA NA 18639 ...
$ amount.env.113   : int  0 0 0 0 NA 4000 2500 NA NA 0 ...
$ amount.alt_en.113: int  0 0 0 0 NA 0 0 NA NA 4750 ...
$ amount.oil.114   : int  10500 7500 17500 5500 NA 43450 6700 NA NA 15400 ...
$ amount.coal.114  : int  3600 14450 13648 2500 NA 2000 0 NA NA 2000 ...
$ amount.mining.114: int  5600 14450 18397 2500 NA 11600 7500 NA NA 16250 ...
$ amount.gas.114   : int  3000 1000 0 0 NA 0 0 NA NA 4500 ...
$ amount.env.114   : int  0 0 0 0 NA 2000 7000 NA NA 0 ...
$ amount.alt_en.114: int  0 0 0 0 NA 0 1000 NA NA 13950 ...
$ amount.oil.115   : int  21950 12500 19200 NA NA 59700 NA NA NA 14150 ...
$ amount.coal.115  : int  7500 11000 8000 NA NA 2000 NA NA NA 0 ...
$ amount.mining.115: int  7500 11000 9000 NA NA 20000 NA NA NA 7000 ...
$ amount.gas.115   : int  2000 3000 0 NA NA 0 NA NA NA 2500 ...
$ amount.env.115   : int  0 0 0 NA NA 1500 NA NA NA 0 ...
$ amount.alt_en.115: int  0 0 0 NA NA 0 NA NA NA 0 ...
$ amount.oil.116   : int  28100 7000 16053 2250 NA 56250 3746 NA NA NA ...
$ amount.coal.116  : int  2000 7000 3600 2500 NA 2000 0 NA NA NA ...
$ amount.mining.116: int  2000 7000 5600 16000 NA 18450 5 NA NA NA ...
$ amount.gas.116   : int  2500 0 0 0 NA 2000 22 NA NA NA ...
$ amount.env.116   : int  0 0 0 500 NA 0 15743 NA NA NA ...
$ amount.alt_en.116: int  0 0 2500 4000 NA 0 3022 NA NA NA ...
$ amount.oil.117   : int  NA 7500 9255 2503 NA 101931 800 NA NA NA ...
$ amount.coal.117  : int  NA 3800 8760 0 NA 0 0 NA NA NA ...
$ amount.mining.117: int  NA 3800 9760 0 NA 18294 5007 NA NA NA ...
$ amount.gas.117   : int  NA 0 0 0 NA 9610 159 NA NA NA ...
$ amount.env.117   : int  NA 1000 0 0 NA 0 675 NA NA NA ...
$ amount.alt_en.117: int  NA 0 0 0 NA 6 437 NA NA NA ...

```

```
summary(df)
```

last_name	first_name	name	party
Length:560	Length:560	Length:560	Length:560
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

state	member_id	Vote3	Vote4
Length:560	Length:560	Length:560	Length:560
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

Vote51	Vote52	Vote6	Vote7
Length:560	Length:560	Length:560	Length:560
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

District	Vote_count	Vote_change	Vote_change_dummy
Length:560	Min. :2.000	Min. :0.00000	Min. :0.00000
Class :character	1st Qu.:3.000	1st Qu.:0.00000	1st Qu.:0.00000
Mode :character	Median :4.000	Median :0.00000	Median :0.00000
	Mean :4.237	Mean :0.05536	Mean :0.04107
	3rd Qu.:6.000	3rd Qu.:0.00000	3rd Qu.:0.00000
	Max. :6.000	Max. :4.00000	Max. :1.00000

amount.oil.113	amount.coal.113	amount.mining.113	amount.gas.113
Min. : 0	Min. : 0	Min. : -250	Min. : 0
1st Qu.: 3500	1st Qu.: 0	1st Qu.: 0	1st Qu.: 0
Median : 14475	Median : 0	Median : 625	Median : 1000
Mean : 36289	Mean : 3766	Mean : 7368	Mean : 4261
3rd Qu.: 42250	3rd Qu.: 1500	3rd Qu.: 5625	3rd Qu.: 4538
Max. :362933	Max. :89500	Max. :130250	Max. :50500
NA's :320	NA's :320	NA's :320	NA's :320

amount.env.113	amount.alt_en.113	amount.oil.114	amount.coal.114
Min. : 0	Min. : 0	Min. : 0	Min. : -250
1st Qu.: 0	1st Qu.: 0	1st Qu.: 5500	1st Qu.: 0
Median : 0	Median : 0	Median : 16750	Median : 0
Mean : 3988	Mean : 1722	Mean : 41403	Mean : 3415
3rd Qu.: 2500	3rd Qu.: 1500	3rd Qu.: 51300	3rd Qu.: 1000
Max. :56606	Max. :26602	Max. :376850	Max. :78856
NA's :320	NA's :320	NA's :291	NA's :291

amount.mining.114	amount.gas.114	amount.env.114	amount.alt_en.114
Min. : -1000	Min. : 0	Min. : 0	Min. : 0

1st Qu.:	0	1st Qu.:	0	1st Qu.:	0	1st Qu.:	0
Median :	0	Median :	2000	Median :	0	Median :	0
Mean :	6553	Mean :	5461	Mean :	4784	Mean :	2148
3rd Qu.:	5500	3rd Qu.:	6500	3rd Qu.:	3600	3rd Qu.:	2000
Max. :	141519	Max. :	62100	Max. :	84869	Max. :	23900
NA's :	291	NA's :	291	NA's :	291	NA's :	291
amount.oil.115		amount.coal.115		amount.mining.115		amount.gas.115	
Min. :	-600	Min. :	0	Min. :	-1000	Min. :	0
1st Qu.:	4125	1st Qu.:	0	1st Qu.:	0	1st Qu.:	0
Median :	15500	Median :	0	Median :	0	Median :	2000
Mean :	35977	Mean :	1929	Mean :	3906	Mean :	5174
3rd Qu.:	44450	3rd Qu.:	240	3rd Qu.:	3662	3rd Qu.:	5488
Max. :	355050	Max. :	43850	Max. :	59850	Max. :	64500
NA's :	262	NA's :	262	NA's :	262	NA's :	262
amount.env.115		amount.alt_en.115		amount.oil.116		amount.coal.116	
Min. :	0	Min. :	-2500	Min. :	-2359	Min. :	0
1st Qu.:	0	1st Qu.:	0	1st Qu.:	3065	1st Qu.:	0
Median :	0	Median :	0	Median :	15100	Median :	0
Mean :	4228	Mean :	2498	Mean :	38791	Mean :	2414
3rd Qu.:	3908	3rd Qu.:	2238	3rd Qu.:	48199	3rd Qu.:	1000
Max. :	57650	Max. :	44800	Max. :	484650	Max. :	68600
NA's :	262	NA's :	262	NA's :	232	NA's :	232
amount.mining.116		amount.gas.116		amount.env.116		amount.alt_en.116	
Min. :	0	Min. :	-1500	Min. :	-500	Min. :	0.0
1st Qu.:	0	1st Qu.:	0	1st Qu.:	0	1st Qu.:	0.0
Median :	4	Median :	1010	Median :	0	Median :	919.5
Mean :	5753	Mean :	5130	Mean :	4492	Mean :	3736.4
3rd Qu.:	4962	3rd Qu.:	5600	3rd Qu.:	4431	3rd Qu.:	3500.0
Max. :	184500	Max. :	69300	Max. :	66533	Max. :	54750.0
NA's :	232	NA's :	232	NA's :	232	NA's :	232
amount.oil.117		amount.coal.117		amount.mining.117		amount.gas.117	
Min. :	-3500	Min. :	0	Min. :	-1000	Min. :	-3500
1st Qu.:	2766	1st Qu.:	0	1st Qu.:	0	1st Qu.:	0
Median :	16000	Median :	0	Median :	5	Median :	1000
Mean :	38362	Mean :	1520	Mean :	3077	Mean :	5988
3rd Qu.:	48535	3rd Qu.:	0	3rd Qu.:	2500	3rd Qu.:	7100
Max. :	556734	Max. :	56301	Max. :	66420	Max. :	92070
NA's :	279	NA's :	279	NA's :	279	NA's :	279
amount.env.117		amount.alt_en.117					
Min. :	-900	Min. :	-2500				
1st Qu.:	0	1st Qu.:	0				
Median :	500	Median :	520				
Mean :	6227	Mean :	4040				

```
3rd Qu.: 4863    3rd Qu.: 3800
Max.    :112546  Max.    :68175
NA's    :279     NA's    :279
```

```
df_long <- df %>%
  pivot_longer(cols = starts_with("amount"), names_to = "contribution_type", values_to = "contribution_amount")

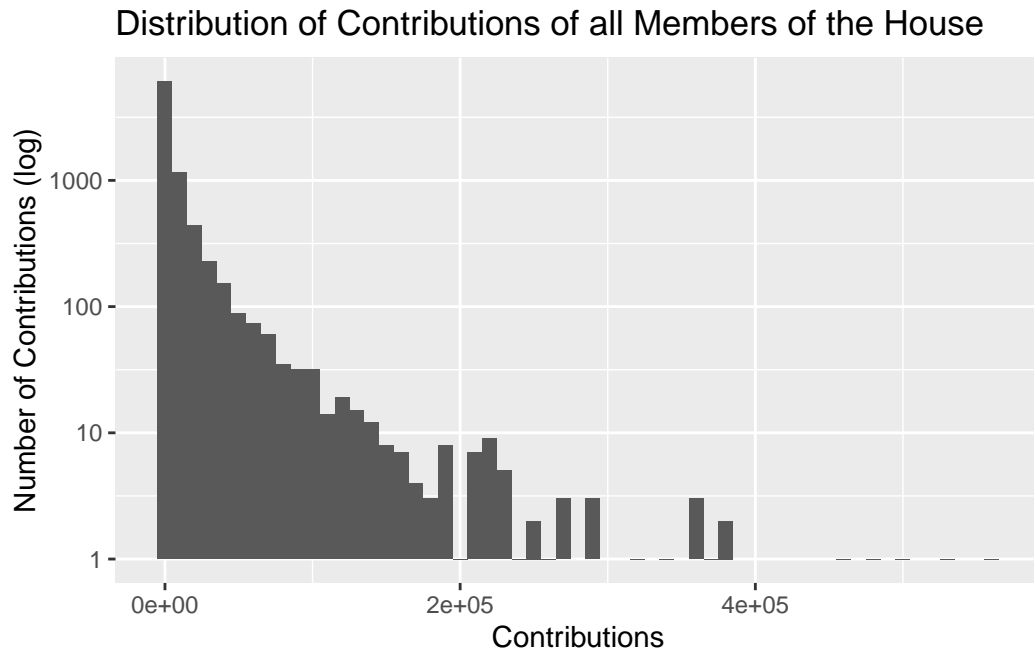
# view(df_long)
# plotting the distribution
plot_1 <- ggplot(df_long, aes(x = contribution_amount)) +
  geom_histogram(binwidth = 10000) +
  labs(title = "Distribution of Contributions of all Members of the House", x = "Contribution Amount") +
  scale_y_continuous(trans='log10') +
  theme_gray()

plot_1
```

Warning: Removed 8304 rows containing non-finite outside the scale range (``stat_bin()``).

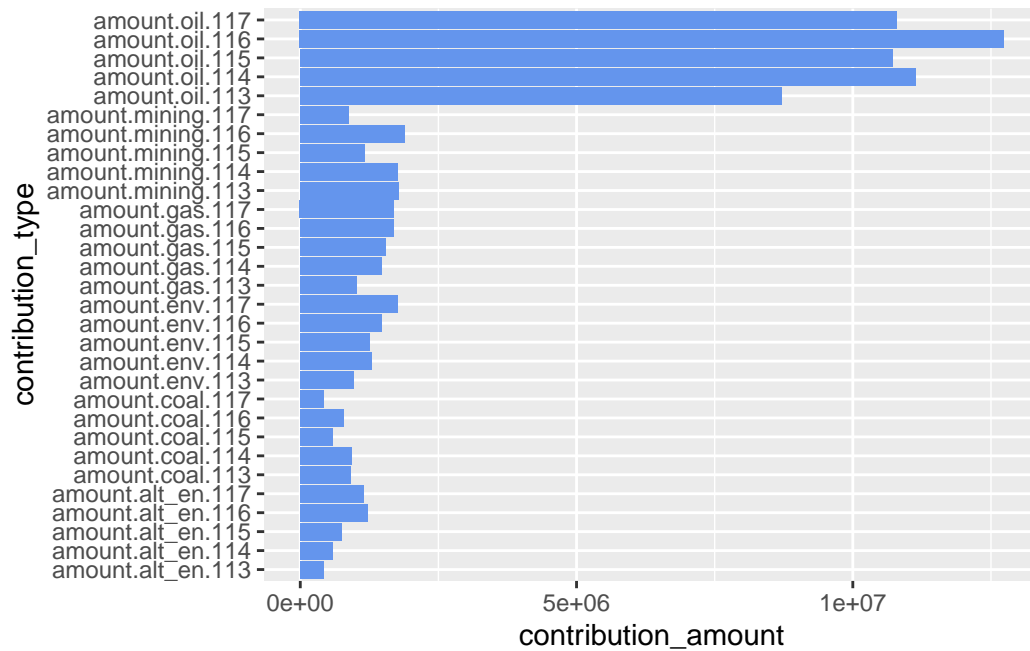
Warning in `scale_y_continuous(trans = "log10")`: log-10 transformation introduced infinite values.

Warning: Removed 18 rows containing missing values or values outside the scale range (``geom_bar()``).



```
# plot of the contributions amount to type
plot_2 <- ggplot(df_long, aes(x = contribution_amount, y = contribution_type)) + geom_col(fill = "#f0f0f0")
plot_2
```

Warning: Removed 8304 rows containing missing values or values outside the scale range (`geom_col()`).

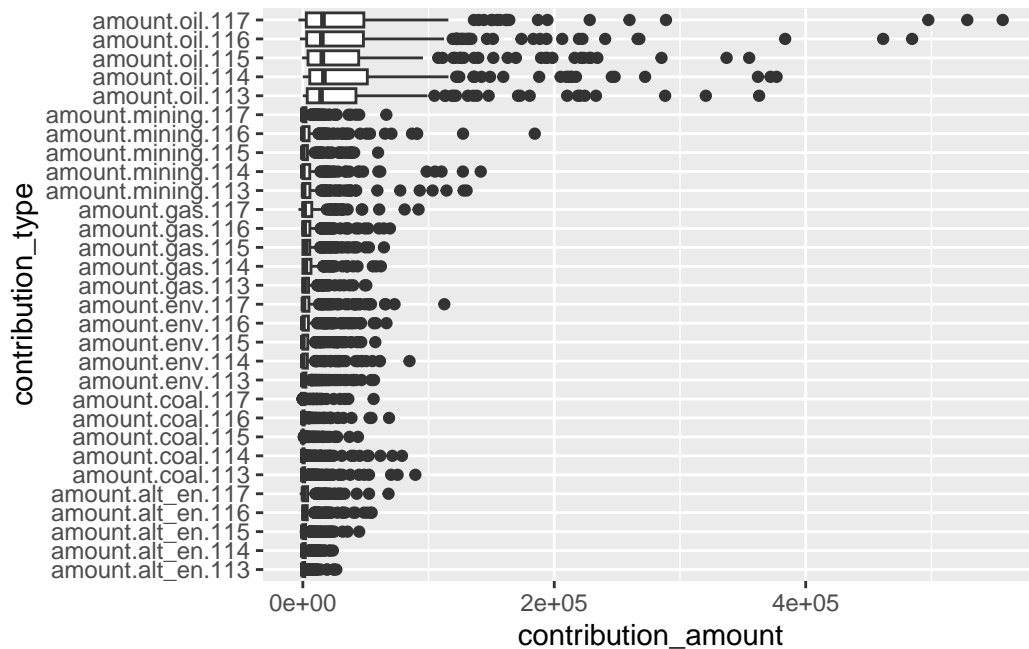


```
# boxplot of the distribution of the contribution data by amount

# Plot with limits applied on the x-axis
boxplot_1 <- ggplot(df_long, aes(x = contribution_amount, y = contribution_type)) +
  geom_boxplot()

boxplot_1
```

Warning: Removed 8304 rows containing non-finite outside the scale range (`stat_boxplot()`).



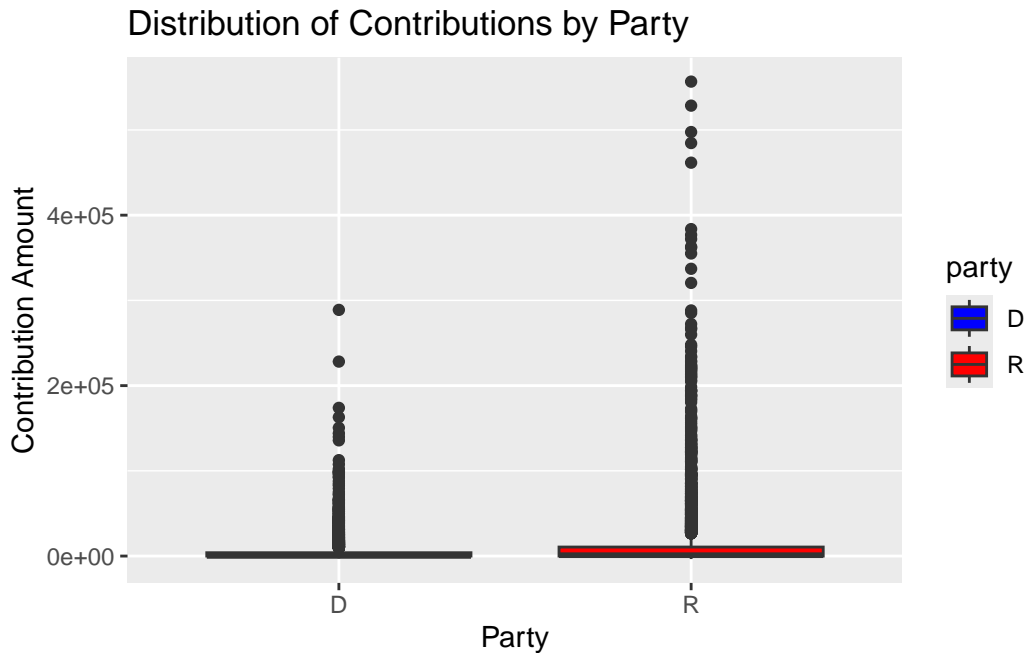
0.0.2 contributions to party plots

```
# plot of contributions by party
# Calculate upper and lower limits for outliers
upper.limit <- quantile(df_long$contribution_amount, na.rm=TRUE)[4] + 1.5*IQR(df_long$contribution_amount)
lower.limit <- quantile(df_long$contribution_amount, na.rm=TRUE)[2] - 1.5*IQR(df_long$contribution_amount)
# maybe set lower limit to 0, to eliminate neg. contributions.

# Plot with limits applied
plot_3 <- df_long %>%
  group_by(party) %>%
  ggplot(aes(y = contribution_amount, x = party, fill = party)) +
  geom_boxplot() +
  labs(title = "Distribution of Contributions by Party", x = "Party", y = "Contribution Amount") +
  # coord_cartesian(ylim=c(lower.limit, upper.limit)) +
  scale_fill_manual(values = c("R" = "red", "D" = "blue"))

plot_3
```

Warning: Removed 8304 rows containing non-finite outside the scale range (``stat_boxplot()``).



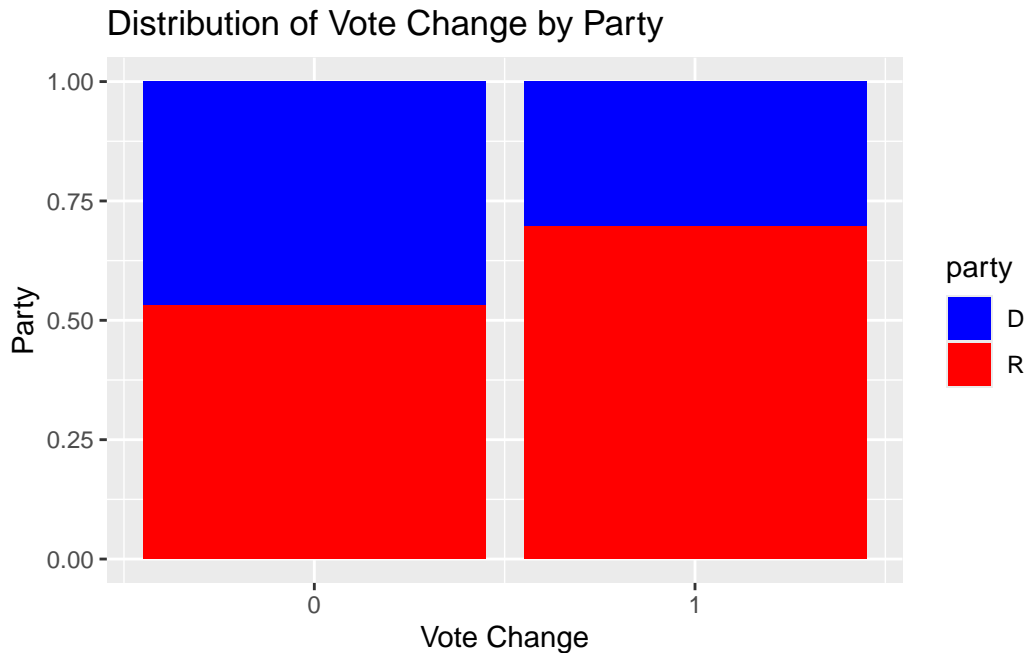
The reason for the negative contributions which some representatives received is.... Since negative contributions still is data for contributions, I will keep them in the data set.

The reason for the NAs in some contributions, e.g. in 113th congress, there are 320 NAs. This is because some of the representatives that are in thge later congresses were not in the 113th congress, e.g. and thus potential contributions they received are not important for this analysis.

0.0.3 plot change in vote dummy vs. state, party, etc.

```
# determine vote_change by party
# view(df_long)
plot_4 <- df_long %>% group_by(party) %>%
  ggplot(aes(x = Vote_change_dummy, fill = party)) +
  geom_bar(position = "fill") + labs(title = "Distribution of Vote Change by Party", x = "Vote") +
  scale_x_continuous(breaks = c(0,1)) +
  scale_fill_manual(values = c("R" = "red", "D" = "blue"))

plot_4
```



```
# count how many people of which party affiliation there are
party_count <- df %>% group_by(party) %>% count()
print(party_count)
```

```
# A tibble: 2 x 2
# Groups:   party [2]
  party     n
  <chr> <int>
1 D       259
2 R       301
```

The above plot shows, that in amount, Republicans and Democrats almost equally often do not change their votes. However, Republicans are more likely to change their votes. The reason as to why the non-vote changers are almost equal in number and still the Republicans change their votes more often, is because the amount of Republicans in the data set is higher than the amount of Democrats.

0.0.4 plot change in mind with each roll_call... (to be able to analyse the contribution better)

```

df_113 <- filter_session_data(df, 113)
view(df_113)

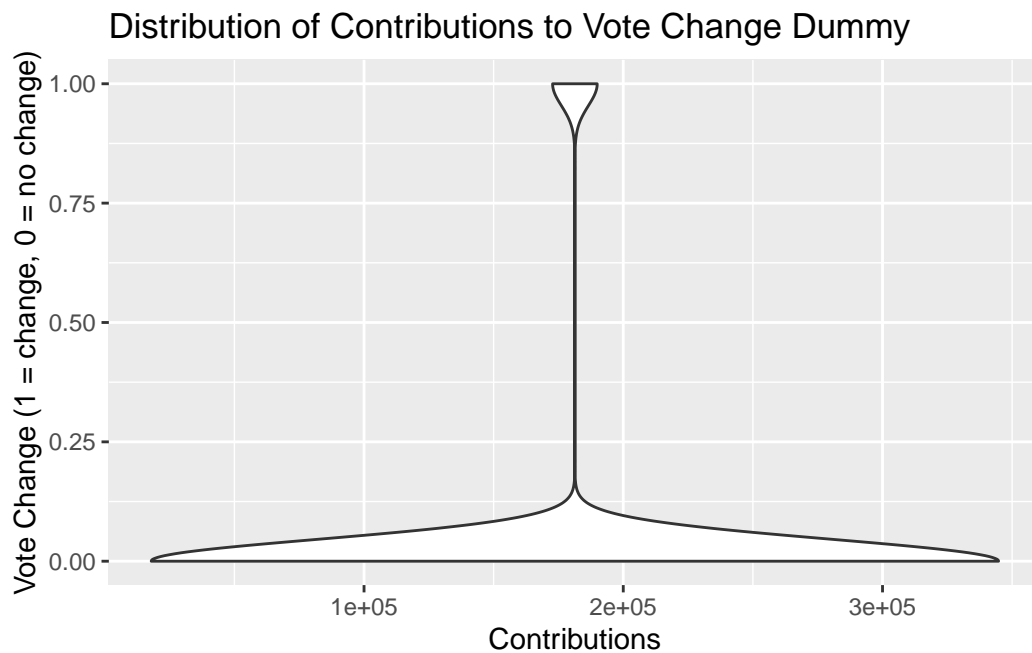
# plot vote_change_dummy with the amount cols
df_long_113 <- df_113 %>%
  pivot_longer(cols = starts_with("amount"), names_to = "contribution_type", values_to = "contribution_amount")

# view(df_long_113)

# plot change in mind with each roll_call
# create a plot which plots the contributions to the vote_change dummy per session.
plot_113 <- df_long_113 %>% ggplot(aes(x = contribution_amount, y = Vote_change_dummy)) + geom_density()
plot_113

```

Warning: Removed 1920 rows containing non-finite outside the scale range (``stat_ydensity()``).



```

# plot all other congresses too

```

0.0.5 mind changers table -> strattman replicate

```
# vote changers
vote_change <- df %>% filter(Vote_change_dummy == 1)

# non-vote changers
non_vote_change <- df %>% filter(Vote_change_dummy == 0)

# Initialize counters
pos_first <- 0
neg_first <- 0
pos <- 0
neg <- 0

for (i in 1:nrow(df)) {
  if (df$Vote_change[i] > 0) {
    # Loop through vote columns for each row
    for (vote_col in c("Vote3", "Vote4", "Vote51", "Vote52", "Vote6", "Vote7")) {
      vote <- df[[vote_col]][i]
      if (!is.na(vote)) {
        # If vote is found, increment the respective counter and break the loop
        if (vote == "-") {
          neg_first <- neg_first + 1
        } else if (vote == "+") {
          pos_first <- pos_first + 1
        }
        break
      }
    }
  }
  if (df$Vote_change[i] == 0) {
    for (vote_col in c("Vote3", "Vote4", "Vote51", "Vote52", "Vote6", "Vote7")) {
      vote <- df[[vote_col]][i]
      if (!is.na(vote)) {
        # If vote is found, increment the respective counter and break the loop
        if (vote == "-") {
          neg <- neg + 1
        } else if (vote == "+") {
          pos <- pos + 1
        }
        break
      }
    }
  }
}
```

```
    }  
  }  
}  
  
list(pos_first = pos_first, neg_first = neg_first, pos = pos, neg = neg)
```

```
$pos_first  
[1] 4
```

```
$neg_first  
[1] 19
```

```
$pos  
[1] 259
```

```
$neg  
[1] 278
```

```
# create a table for these vars with the labels (always positive, always negative, change - c
```