Data Analysis of Roll Call Data

The summary statistics and plots of the data can be found in the plots.qmd file in the same directory.

Simple Linear Probability Model

```
df_no_change <- df %>% filter(Vote_change_dummy == 0)

# get all reps who voted only pos. or only neg. and regress with all contributions
df_no_change <- df %>% filter(Vote_change_dummy == 0)
df_no_change$"all_votes" <- 0
# create a col that indicates if the rep voted all pos or neg. votes
for (i in 1:nrow(df_no_change)) {
    if (!is.na(any(df_no_change[i, c("Vote3", "Vote4", "Vote51", "Vote52", "Vote6", "Vote7")]
        df_no_change[i, "all_votes"] <- "+"
    } else if (!is.na(any(df_no_change[i, c("Vote3", "Vote4", "Vote51", "Vote52", "Vote6", "
        df_no_change[i, "all_votes"] <- "-"
    }
}

df_no_change <- dummy_cols(df_no_change, select_columns = "all_votes")
df_no_change <- df_no_change %>%
    rename("all_votes_plus" = "all_votes_+") %>%
    rename("all_votes_minus" = "all_votes_-")

df_no_change <- df_no_change %>% select(-c(
    last_name, first_name, name, Vote_count, Vote_change,
    Vote_change, Vote_change_dummy, member_id, Vote3, Vote4, Vote51, Vote52, Vote6, Vote7
))
```

```r
view(df_no_change)

ols_1 <- lm(all_votes_minus ~ . - all_votes - all_votes_plus - state, data = df_no_change)
summary(ols_1)
```

```
Call:
lm(formula = all_votes_minus ~ . - all_votes - all_votes_plus -
    state, data = df_no_change)

Residuals:
     Min       1Q   Median       3Q      Max
-0.47119 -0.20817 -0.03031  0.07026  0.91933

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)    -7.010e-02  1.556e-01  -0.451  0.65301
partyR          3.061e-01  6.816e-02   4.491 1.53e-05 ***
District        6.406e-03  2.528e-03   2.534  0.01244 *
GeographicalMW -1.439e-02  1.637e-01  -0.088  0.93009
GeographicalNE  3.292e-02  1.732e-01   0.190  0.84957
GeographicalSO  1.616e-02  1.631e-01   0.099  0.92125
GeographicalWE -1.409e-02  1.650e-01  -0.085  0.93212
pro_env.113     1.549e-07  3.908e-06   0.040  0.96844
anti_env.113   -9.032e-07  9.140e-07  -0.988  0.32486
pro_env.114    -1.826e-06  3.852e-06  -0.474  0.63626
anti_env.114   -1.232e-06  1.027e-06  -1.200  0.23233
pro_env.115     5.669e-06  5.915e-06   0.958  0.33960
anti_env.115    1.777e-06  1.208e-06   1.471  0.14376
pro_env.116    -5.752e-06  4.786e-06  -1.202  0.23160
anti_env.116    1.892e-06  8.121e-07   2.329  0.02134 *
pro_env.117     4.799e-07  2.648e-06   0.181  0.85648
anti_env.117   -1.768e-06  5.515e-07  -3.206  0.00168 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3046 on 133 degrees of freedom
  (387 observations deleted due to missingness)
Multiple R-squared:  0.3425,    Adjusted R-squared:  0.2634
F-statistic:  4.33 on 16 and 133 DF,  p-value: 9.105e-07
```

LPM per session

```r
# Vote 3
df_113 <- filter_session_data(df, "113")
ols_113 <- lm(Vote3_minus ~ . - Vote3 - Vote3_plus, data = df_113)
summary(ols_113)
```

```
Call:
lm(formula = Vote3_minus ~ . - Vote3 - Vote3_plus, data = df_113)

Residuals:
     Min       1Q   Median       3Q      Max
-0.13695 -0.04341 -0.01370  0.01183  0.94492

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     9.490e-03  6.482e-02   0.146   0.8837
partyR          9.157e-01  2.438e-02  37.556   <2e-16 ***
anti_env.113    2.658e-07  1.418e-07   1.875   0.0622 .
pro_env.113    -8.293e-07  9.413e-07  -0.881   0.3793
GeographicalMW  4.318e-02  6.761e-02   0.639   0.5237
GeographicalNE  1.109e-03  6.891e-02   0.016   0.9872
GeographicalSO  5.428e-02  6.690e-02   0.811   0.4180
GeographicalWE  4.445e-03  6.717e-02   0.066   0.9473
District        1.819e-03  9.192e-04   1.979   0.0491 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1423 on 220 degrees of freedom
  (135 observations deleted due to missingness)
Multiple R-squared:  0.9209,    Adjusted R-squared:  0.918
F-statistic: 320.1 on 8 and 220 DF,  p-value: < 2.2e-16
```

```r
# Vote 4
df_114 <- filter_session_data(df, "114")
ols_114 <- lm(Vote4_minus ~ . - Vote4 - Vote4_plus, data = df_114)
summary(ols_114)
```

```
Call:
lm(formula = Vote4_minus ~ . - Vote4 - Vote4_plus, data = df_114)

Residuals:
     Min       1Q   Median       3Q      Max
-0.97759 -0.02922 -0.00279  0.01551  0.93622

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     1.629e-02  6.656e-02   0.245   0.8068
partyR          9.056e-01  2.339e-02  38.722   <2e-16 ***
anti_env.114    2.749e-07  1.242e-07   2.213   0.0278 *
pro_env.114    -3.516e-07  7.747e-07  -0.454   0.6503
GeographicalMW  5.203e-02  6.966e-02   0.747   0.4557
GeographicalNE -1.250e-02  7.071e-02  -0.177   0.8598
GeographicalSO  6.590e-02  6.905e-02   0.954   0.3408
GeographicalWE  1.757e-02  6.955e-02   0.253   0.8007
District       -5.713e-04  8.826e-04  -0.647   0.5181
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1477 on 254 degrees of freedom
  (157 observations deleted due to missingness)
Multiple R-squared:  0.9138,    Adjusted R-squared:  0.9111
F-statistic: 336.8 on 8 and 254 DF,  p-value: < 2.2e-16
```

```r
# Vote 51
df_1151 <- filter_session_data(df, "1151")
ols_1151 <- lm(Vote51_minus ~ . - Vote51 - Vote51_plus, data = df_1151)
summary(ols_1151)
```

```
Call:
lm(formula = Vote51_minus ~ . - Vote51 - Vote51_plus, data = df_1151)

Residuals:
     Min       1Q   Median       3Q      Max
-0.97624 -0.02020  0.00563  0.02094  0.94197

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     1.028e-02  6.176e-02   0.166   0.8679
```

```
partyR          9.182e-01  2.340e-02  39.241   <2e-16 ***
anti_env.115    2.584e-07  1.563e-07   1.653   0.0995 .
pro_env.115    -1.976e-07  8.812e-07  -0.224   0.8228
GeographicalMW  5.239e-02  6.485e-02   0.808   0.4199
GeographicalNE -3.385e-02  6.543e-02  -0.517   0.6054
GeographicalSO  4.556e-02  6.444e-02   0.707   0.4802
GeographicalWE  6.325e-03  6.439e-02   0.098   0.9218
District        1.615e-04  9.015e-04   0.179   0.8579
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1491 on 251 degrees of freedom
  (155 observations deleted due to missingness)
Multiple R-squared:  0.913, Adjusted R-squared:  0.9102
F-statistic: 329.3 on 8 and 251 DF,  p-value: < 2.2e-16
```

```r
# Vote 52
df_1152 <- filter_session_data(df, "1152")
ols_1152 <- lm(Vote52_minus ~ . - Vote52 - Vote52_plus, data = df_1152)
summary(ols_1152)
```

```

Call:
lm(formula = Vote52_minus ~ . - Vote52 - Vote52_plus, data = df_1152)

Residuals:
     Min       1Q   Median       3Q      Max
-0.97369 -0.01948  0.02799  0.05939  0.96102

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     2.444e-02  8.853e-02   0.276    0.783
partyR          9.094e-01  3.033e-02  29.980   <2e-16 ***
anti_env.115   -2.552e-08  2.042e-07  -0.125    0.901
pro_env.115    -2.523e-07  1.171e-06  -0.215    0.830
GeographicalMW  1.177e-02  9.194e-02   0.128    0.898
GeographicalNE -3.073e-02  9.268e-02  -0.332    0.740
GeographicalSO  4.079e-02  9.127e-02   0.447    0.655
GeographicalWE  4.852e-03  9.167e-02   0.053    0.958
District       -7.628e-04  1.233e-03  -0.619    0.537
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.1951 on 246 degrees of freedom
  (156 observations deleted due to missingness)
Multiple R-squared:  0.8525,    Adjusted R-squared:  0.8477
F-statistic: 177.8 on 8 and 246 DF,  p-value: < 2.2e-16
```

```r
# Vote 6
df_116 <- filter_session_data(df, "116")
ols_116 <- lm(Vote6_minus ~ . - Vote6 - Vote6_plus, data = df_116)
summary(ols_116)
```

```
Call:
lm(formula = Vote6_minus ~ . - Vote6 - Vote6_plus, data = df_116)

Residuals:
     Min       1Q   Median       3Q      Max
-0.97630 -0.00988  0.01239  0.03222  0.96874

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     1.289e-02  6.191e-02   0.208    0.835
partyR          9.538e-01  2.202e-02  43.310   <2e-16 ***
anti_env.116    6.534e-08  1.188e-07   0.550    0.583
pro_env.116     2.305e-08  6.691e-07   0.034    0.973
GeographicalMW  1.775e-03  6.484e-02   0.027    0.978
GeographicalNE -2.095e-02  6.538e-02  -0.320    0.749
GeographicalSO  2.771e-03  6.426e-02   0.043    0.966
GeographicalWE  2.558e-02  6.490e-02   0.394    0.694
District       -9.834e-04  8.545e-04  -1.151    0.251
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1376 on 253 degrees of freedom
  (146 observations deleted due to missingness)
Multiple R-squared:  0.9263,    Adjusted R-squared:  0.924
F-statistic: 397.6 on 8 and 253 DF,  p-value: < 2.2e-16
```

```r
# Vote 7
df_117 <- filter_session_data(df, "117")
ols_117 <- lm(Vote7_minus ~ . - Vote7 - Vote7_plus, data = df_117)
summary(ols_117)
```

```
Call:
lm(formula = Vote7_minus ~ . - Vote7 - Vote7_plus, data = df_117)

Residuals:
     Min       1Q   Median       3Q      Max
-0.98317 -0.01405  0.02273  0.03471  0.10469

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     1.746e-02  6.426e-02   0.272    0.786
partyR          9.524e-01  2.066e-02  46.099   <2e-16 ***
anti_env.117   -4.561e-08  1.034e-07  -0.441    0.659
pro_env.117     1.868e-08  4.796e-07   0.039    0.969
GeographicalMW  3.341e-03  6.652e-02   0.050    0.960
GeographicalNE -4.592e-02  6.719e-02  -0.683    0.495
GeographicalSO  1.487e-02  6.629e-02   0.224    0.823
GeographicalWE  1.044e-02  6.654e-02   0.157    0.875
District       -1.116e-03  8.490e-04  -1.315    0.190
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1276 on 229 degrees of freedom
  (117 observations deleted due to missingness)
Multiple R-squared:  0.9361,    Adjusted R-squared:  0.9338
F-statistic:   419 on 8 and 229 DF,  p-value: < 2.2e-16
```

LPM of only mind changers

```r
df_subsample <- df_subsample %>% select(-c(
    last_name, first_name, name, Vote_count, Vote_change,
    Vote_change, Vote_change_dummy, votes, member_id, Vote3, Vote4, Vote51, Vote52, Vote6, Vo
))
# view(df_subsample)

ols_subsample <- lm(vote_change_to_anti ~ ., data = df_subsample)
summary(ols_subsample)
```

```
Call:
```

```
lm(formula = vote_change_to_anti ~ ., data = df_subsample)

Residuals:
    Min      1Q  Median      3Q     Max
-0.5000 -0.3333  0.0000  0.3750  0.6667

Coefficients: (8 not defined because of singularities)
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)    -7.167e-01  2.576e+00  -0.278    0.790
partyR         -2.071e-01  1.629e+00  -0.127    0.903
District       -1.756e-02  2.279e-02  -0.771    0.470
GeographicalNE  1.269e+00  2.920e+00   0.435    0.679
GeographicalSO  1.256e+00  2.544e+00   0.494    0.639
GeographicalWE  9.631e-01  1.905e+00   0.506    0.631
pro_env.113     1.276e-04  1.488e-04   0.857    0.424
anti_env.113   -4.329e-07  8.008e-06  -0.054    0.959
pro_env.114            NA         NA      NA       NA
anti_env.114           NA         NA      NA       NA
pro_env.115            NA         NA      NA       NA
anti_env.115           NA         NA      NA       NA
pro_env.116            NA         NA      NA       NA
anti_env.116           NA         NA      NA       NA
pro_env.117            NA         NA      NA       NA
anti_env.117           NA         NA      NA       NA

Residual standard error: 0.6009 on 6 degrees of freedom
  (17 observations deleted due to missingness)
Multiple R-squared:  0.2417,    Adjusted R-squared:  -0.6431
F-statistic: 0.2732 on 7 and 6 DF,  p-value: 0.9432
```

Why warning message?