Models

Maggie Nead

7/13/2020

```r
d <- data %>%
  filter(!is.na(TYPE))

d %>%
group_by(TYPE) %>%
tally
```

```
## # A tibble: 44 x 2
##    TYPE                          n
##    <chr>                     <int>
##  1 0                           116
##  2 1                        144162
##  3 2                         12790
##  4 3                         18855
##  5 4                          3322
##  6 5                         15122
##  7 6                          1065
##  8 congressional               150
##  9 Congressional               241
## 10 congressional - BP Info Only    1
## # ... with 34 more rows
```

```r
d %<>%
  filter(TYPE == c(1,2,4))

d %<>%
  mutate(interest = ifelse(TYPE %in% c(1), "constituent", NA)) %>%
  mutate(interest = ifelse(TYPE %in% c(2,4), "corporate", interest))# %>%
  #mutate(interest = ifelse(! str_detect(TYPE, "1|2|4"), "neither", interest))

d %<>%
  mutate(extreme = abs(nominate.dim1))

d %>%
  group_by(interest) %>%
  tally()
```
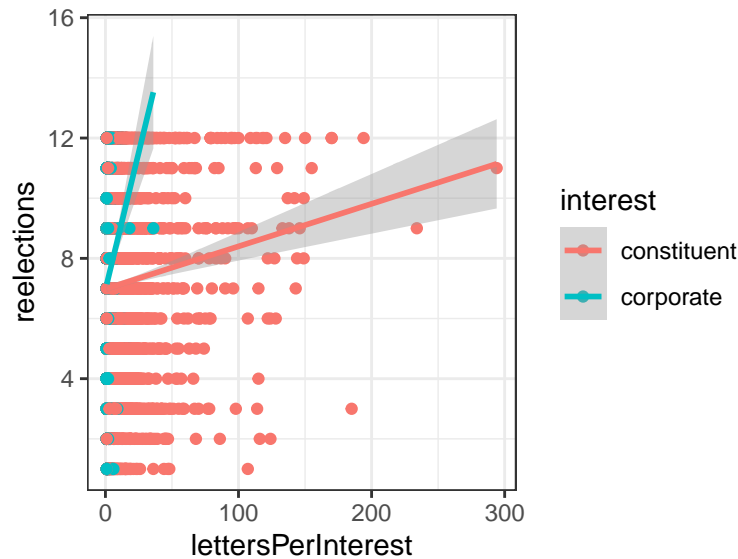
```
## # A tibble: 2 x 2
##   interest        n
##   <chr>       <int>
## 1 constituent 48052
## 2 corporate    5348
```

```
d %<>%
  count(bioname, icpsr, congress, reelections, party_name, member_reelected, interest, extreme, name =
```

```
qplot(x = lettersPerInterest, y = reelections, data = d, color = interest) +
  geom_smooth(method = "glm")
```



```
d %<>%
  spread(interest, lettersPerInterest)
```

```
d %<>%
  mutate(corporate = replace_na(corporate, 0)) %>%
  mutate(constituent = replace_na(constituent, 0))
```

Logistic Regression

This logistic regression tests for a relationship between member_reelected and letter interests (corporate or constituent).

Member Reelection and Letter Interest

```
##
## Call:
## glm(formula = member_reelected ~ corporate + constituent, family = binomial(link = "logit"),
##     data = d)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3160   0.5556   0.5746   0.5820   0.5847
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.678456   0.056797  29.552   <2e-16 ***
## corporate   0.025245   0.026050   0.969    0.332
## constituent 0.001399   0.002867   0.488    0.626
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3262.9  on 3847  degrees of freedom
## Residual deviance: 3260.7  on 3845  degrees of freedom
##   (290 observations deleted due to missingness)
## AIC: 3266.7
##
## Number of Fisher Scoring iterations: 4
```

This logit model shows no relationship between corporate or constituent interest correspodence and likelihood of being elected in the following congress. This is not unexpected as this is a simplistic model for a generally complex response variable. It can be difficult to predict the outcome of one election given only the correspodence from the previous congress as it may take many years for a trend to present itself. Using this as a jumping off point the data now lends itself to more complex models.

Negative Binomial Regression

Corporate and Constituent Interest Letters

The most basic negative binomial model checks for a relationship between reelection counts and correspondece interest (corporate and constituent) without confounding variables.

```
##
## Call:
## glm.nb(formula = reelections ~ corporate + constituent, data = d,
##     init.theta = 14.87931665, link = log)
##
## Deviance Residuals:
##      Min        1Q     Median        3Q       Max
## -2.72700  -0.80039  -0.01964   0.70409   1.40182
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.9318043  0.0086734 222.729  < 2e-16 ***
## corporate   0.0225904  0.0036592   6.174 6.68e-10 ***
## constituent 0.0005100  0.0004171   1.223    0.222
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(14.8793) family taken to be 1)
##
##     Null deviance: 4497.4  on 4137  degrees of freedom
## Residual deviance: 4436.8  on 4135  degrees of freedom
## AIC: 21372
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  14.88
##           Std. Err.:  1.08
##
##  2 x log-likelihood:  -21363.80
```

Corporate interest correspondence is statistically significant when predicting how many times a member of congress gets reelected. For each corporate letter a member sends, the log count of times a member is reelected is expected to increase by 0.02 over a ten year period as illustrated by the figure below. There is not enough evidence to conclude that the same is true of constituent correspodence.
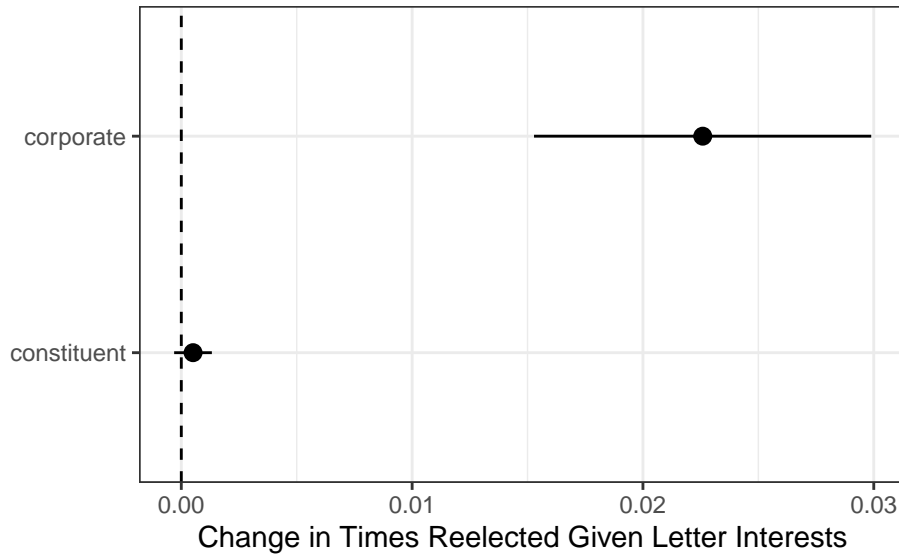
Two possibilities for this are: as stated in hypothesis #2 corporations have the money, resources and audience to help a member that advocates for them get reelected. Or it could also mean that corporations are more likely to try and build relationships with members that are more established. A member who has a position of leadership such as a commmittee chair or majority/minority leadership postition is likely to have been serving for more congresses. Corporations might see this as a relationship that is worth building as those members are more likely to continue being in congress for a long time and have more power to push through policy that could favor the corporation.

Modeling Assumptions Negative Binomial models are not assumed to have normalized residuals. The normality plot looks like it could have some skew but for the most part looks normal. There are a couple of points that stand out on the leverage plot but none that are beyond cook's distance (for diagonsitic plots see appendix model 1).

Party and Reelections

Another variable that could have an effect on reelection is candidate political party. Parties have different priorities and voting members of those parties likely have different interests. For this reason party may play a large role in who gets reelected. For this model independents were dropped as they tend to have a significant amount of variation between them and it can be hard to determine trends in their reelection patterns.

```
##
## Call:
## glm.nb(formula = reelections ~ party_name, data = noI, init.theta = 15.16265379,
##     link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6488  -0.8398  -0.1817   0.6775   1.4665
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)          2.028533   0.009982 203.223   <2e-16 ***
## party_nameRepublican -0.120340   0.014131  -8.516   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(15.1627) family taken to be 1)
##
```

```
##     Null deviance: 4494.9  on 4123  degrees of freedom
## Residual deviance: 4422.3  on 4122  degrees of freedom
## AIC: 21272
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  15.16
##           Std. Err.:  1.11
##
##  2 x log-likelihood:  -21266.20
```

Party is a significant predictor of reelection. Looking solely at democrats and republicans, if the member is republican the log count of the number of times they are reelected is expected to be 0.12 less than that of a Democrat from the 106th to the 116th congress. This does not necessarily mean that republicans are less likely to be elected or that there was a seat turn over. It is possible that the member is replaced by someone of the same party. This model also does not account for things like retirement. This just shows that generally republicans do not hang on to their seats for as long as their democrat counterparts over a ten year period. This is visualized in the figure below.
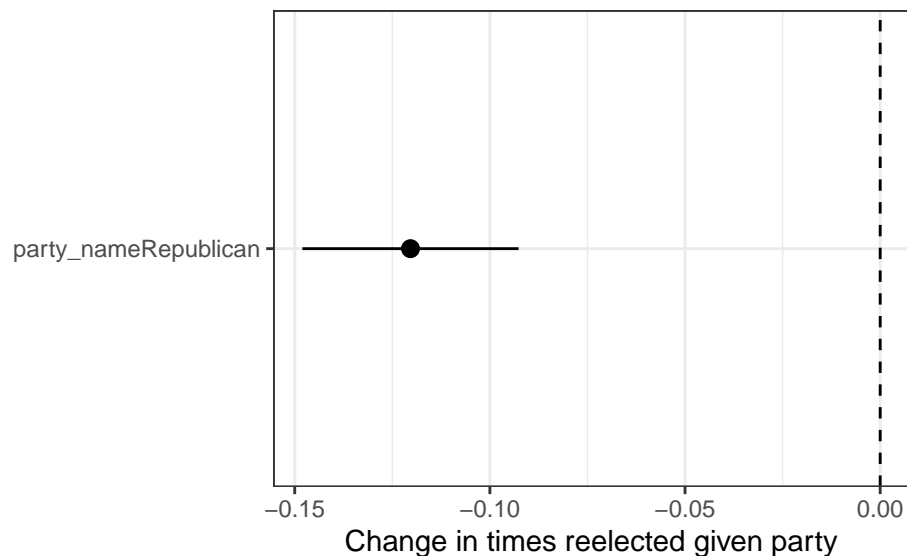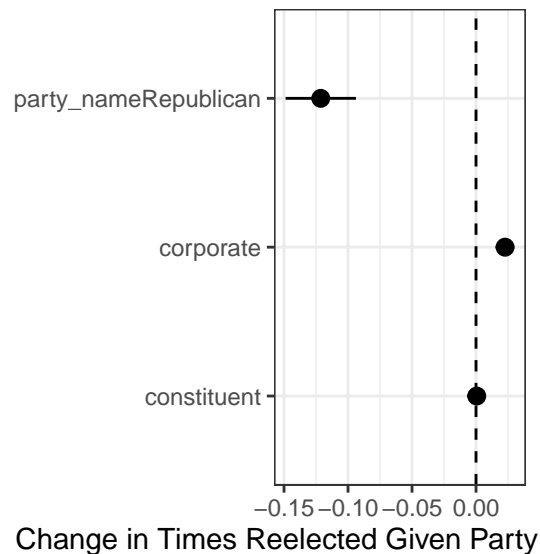

Modeling Assumptions Negative Binomial models are not assumed to have normalized residuals. The normality plot looks like it could have some skew but for the most part looks normal. There are a couple of points that stand out on the leverage plot but none that are beyond cook's distance (for diagonsitic plots see appendix model 2).

Party and Constituent/Corporate Interest Letters

```
##
## Call:
## glm.nb(formula = reelections ~ constituent + corporate + party_name,
##     data = noI, init.theta = 16.01815096, link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.8384  -0.8689  -0.1042   0.7516   1.5729
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)             1.9919128   0.0110577 180.137   < 2e-16 ***
## constituent             0.0005396   0.0004125   1.308     0.191
## corporate               0.0226922   0.0036164   6.275   3.5e-10 ***
## party_nameRepublican   -0.1212980   0.0140142  -8.655   < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(16.0182) family taken to be 1)
##
##     Null deviance: 4565.9  on 4123   degrees of freedom
## Residual deviance: 4428.8  on 4120   degrees of freedom
## AIC: 21213
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  16.02
##           Std. Err.:  1.22
##
##  2 x log-likelihood:  -21203.38
```

As shown in the previous models above Party and corporate correspodence continue to be significant predictors of the number of times a member is reelected over a ten year period.


Change in Times Reelected Given Party

Modeling Assumptions Negative Binomial models are not assumed to have normalized residuals. The normality plot looks like it could have some skew but for the most part looks normal. There are a couple of points that stand out on the leverage plot but none that are beyond cook's distance (for diagonsitic plots see appendix model 3).

Final Model

Interaction between Party and Letter Interests As discussed previously republicans generally seem to be reelected less often than democrats over a ten year period. This is illustrated by the distribution plot below.
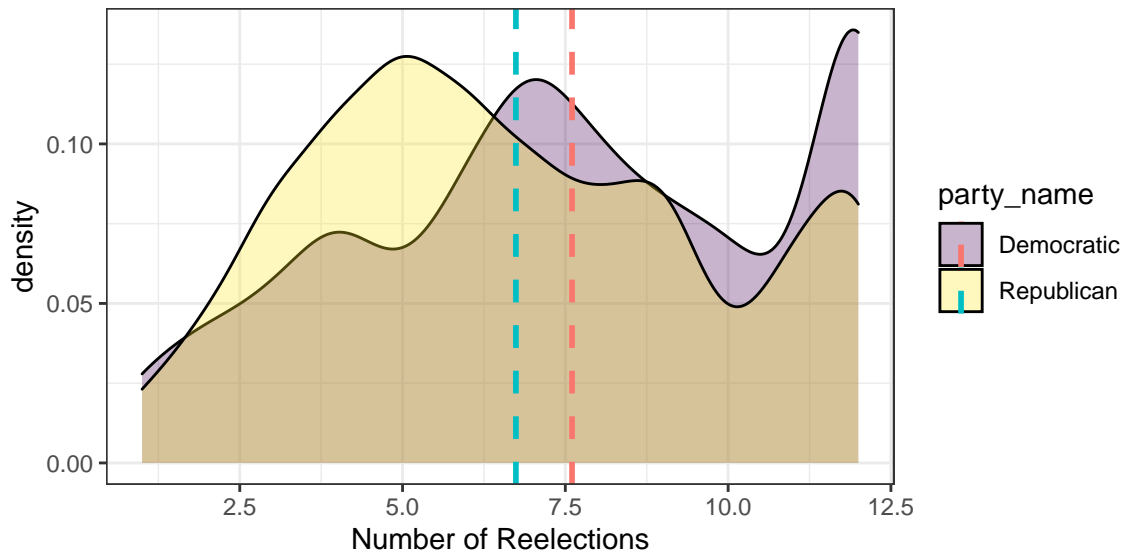
The sand color is the distribution of times republicans were reelected and the purple is democrat. The mean number of reelections is shown by the dashed line, blue for republicans and pink for democrats. The mean for republicans sits right around 6.5 and for democrats it is at 7.5. This plot, however, may not provide the full picture for what is happening. Parties can influence what a member advocates for and to that extent could have a significant impact on who (or what) a member is writting on behalf of. For this reason it is important to check for interaction between party and correspondence interest.
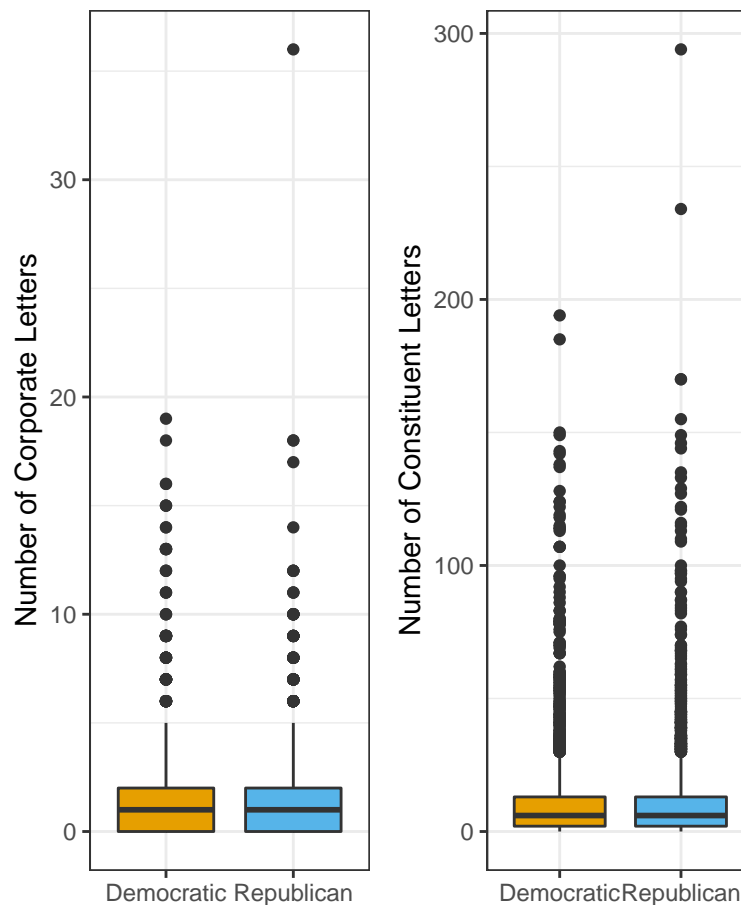
This plot looks at party and correspondence interest. The means for democrats and republicans is similar for corporate interest and constituent interest correspondence and the distributions themselves look very similar. The biggest difference is between corporate and constituent as there are more constituent interest letters.

```
##
## Call:
## glm.nb(formula = reelections ~ corporate * party_name + constituent *
##     party_name, data = noI, init.theta = 16.31536987, link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.7820  -0.8771  -0.1298   0.7388   1.8420
##
## Coefficients:
##                                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)                      2.0106949  0.0120725 166.552  < 2e-16 ***
## corporate                        0.0244032  0.0054212   4.501 6.75e-06 ***
## party_nameRepublican            -0.1585456  0.0171200  -9.261  < 2e-16 ***
## constituent                     -0.0012688  0.0006390  -1.986 0.047069 *
## corporate:party_nameRepublican  -0.0009557  0.0072721  -0.131 0.895440
## party_nameRepublican:constituent 0.0032088  0.0008309   3.862 0.000112 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(16.3154) family taken to be 1)
##
##     Null deviance: 4589.4  on 4123  degrees of freedom
## Residual deviance: 4431.9  on 4118  degrees of freedom
## AIC: 21198
##
## Number of Fisher Scoring iterations: 1
##
##
##              Theta:  16.32
##          Std. Err.:  1.26
##
##  2 x log-likelihood:  -21183.85
```

As shown in the previous model the log of the expected number of times a member is reelected decreases if the member is Republican. What is interesting, however, is that the interaction between Republican and constituent correspodence is statistically significant. This suggests that while compared to democrates, republicans are not as likely to serve as long, the log expected number of times a republican was reelected increases when they send letters on behalf of constituents.
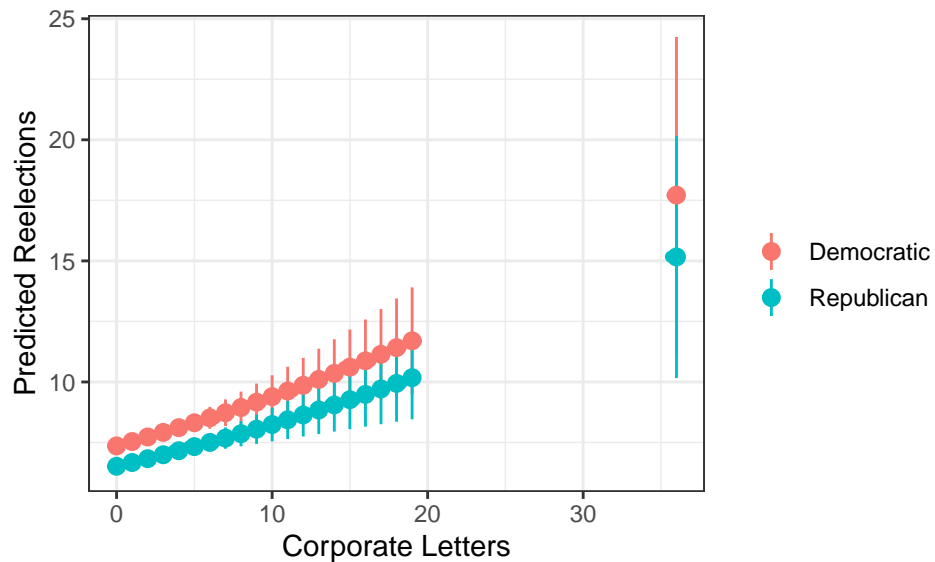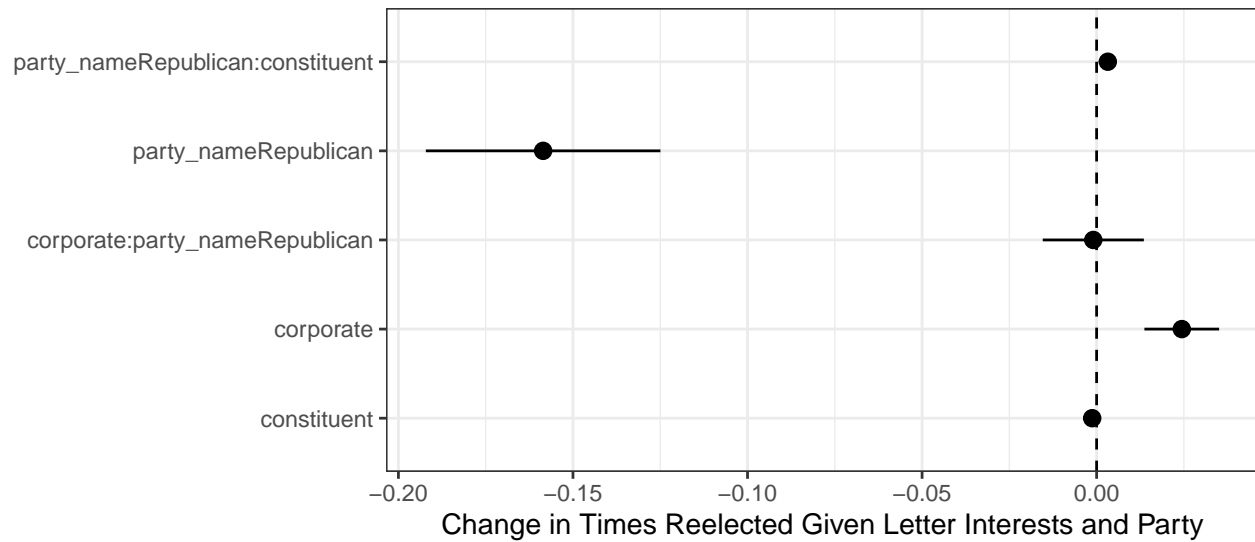
Modeled in this plot is the predicted number of reelections for a member based on their party and the number of corporate correspodence they have sent in a congress while holding constituent correspodence at its mean. The lines through each point is the confidence interval for that prediction. There is a clear overlap in the confidence intervals for democrats and republicans at each predicted count. This verfies what is seen in the model that the interaction between corporate correspondence and party is not significant.
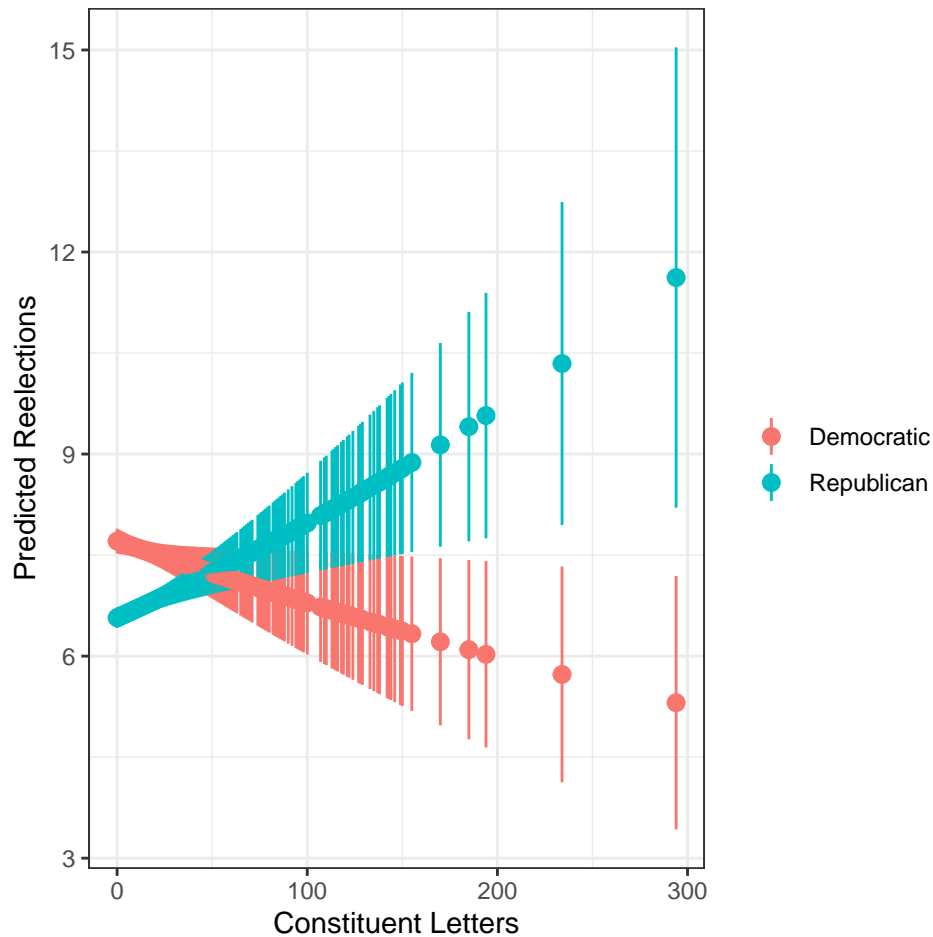
While there may not be evidence to suggest that constituent correspondence is a significant predictor of the number of times a member will be reelected, the interaction between party and constituent correspodence yeilds some interesting results. Holding the number of corporate correspondence at its mean, The number of predicted reelections for a republican member increases based on how much constituent correspondence they send where the opposite appears to be true for democrats.

Modeling Assumptions
Negative Binomial models are not assumed to have normalized residuals. The normality plot looks like it could have some skew but for the most part looks normal. There are a couple of points that stand out on the leverage plot but none that are beyond cook's distance (for diagonsitic plots see appendix model 4).

```
##
## Call:
## glm.nb(formula = reelections ~ extreme, data = noI, init.theta = 14.28508207,
##     link = log)
##
## Deviance Residuals:
##      Min       1Q    Median        3Q       Max
## -2.58519  -0.73903  -0.04601   0.78697   1.35065
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.94173    0.02101  92.400   <2e-16 ***
## extreme      0.06345    0.04801   1.322    0.186
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(14.2851) family taken to be 1)
##
##     Null deviance: 4416.0  on 4123  degrees of freedom
## Residual deviance: 4414.3  on 4122  degrees of freedom
## AIC: 21342
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  14.29
##           Std. Err.:  1.01
##
##  2 x log-likelihood:  -21336.37

##
## Call:
## glm.nb(formula = reelections ~ corporate * party_name + constituent *
##     party_name, data = noI, init.theta = 16.31536987, link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.7820  -0.8771  -0.1298   0.7388   1.8420
##
## Coefficients:
##                                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)                      2.0106949  0.0120725 166.552  < 2e-16 ***
## corporate                        0.0244032  0.0054212   4.501 6.75e-06 ***
## party_nameRepublican            -0.1585456  0.0171200  -9.261  < 2e-16 ***
## constituent                     -0.0012688  0.0006390  -1.986 0.047069 *
## corporate:party_nameRepublican  -0.0009557  0.0072721  -0.131 0.895440
## party_nameRepublican:constituent 0.0032088  0.0008309   3.862 0.000112 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(16.3154) family taken to be 1)
##
##     Null deviance: 4589.4  on 4123  degrees of freedom
## Residual deviance: 4431.9  on 4118  degrees of freedom
## AIC: 21198
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  16.32
##           Std. Err.:  1.26
##
##  2 x log-likelihood:  -21183.85

##
## Call:
## glm.nb(formula = reelections ~ corporate * party_name * extreme +
##     constituent * party_name * extreme, data = noI, init.theta = 21.55361024,
##     link = log)
```

```
## 
## Deviance Residuals: 
##     Min       1Q   Median       3Q      Max  
## -3.6182  -0.8001  -0.0425   0.7183   2.0289  
## 
## Coefficients:
##                                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)                              1.543259   0.036709  42.040  < 2e-16
## corporate                                0.053172   0.015250   3.487 0.000489
## party_nameRepublican                     0.521182   0.052535   9.921  < 2e-16
## extreme                                  1.226070   0.093545  13.107  < 2e-16
## constituent                             -0.003697   0.001871  -1.976 0.048184
## corporate:party_nameRepublican          -0.062092   0.024013  -2.586 0.009717
## corporate:extreme                       -0.076758   0.042249  -1.817 0.069248
## party_nameRepublican:extreme            -1.686980   0.123027 -13.712  < 2e-16
## party_nameRepublican:constituent         0.007387   0.002750   2.686 0.007233
## extreme:constituent                      0.009305   0.005218   1.783 0.074541
## corporate:party_nameRepublican:extreme   0.145732   0.058247   2.502 0.012350
## party_nameRepublican:extreme:constituent -0.013609   0.006995  -1.946 0.051710
## 
## (Intercept)                              ***
## corporate                                ***
## party_nameRepublican                     ***
## extreme                                  ***
## constituent                              *  
## corporate:party_nameRepublican           ** 
## corporate:extreme                        .  
## party_nameRepublican:extreme             ***
## party_nameRepublican:constituent         ** 
## extreme:constituent                      .  
## corporate:party_nameRepublican:extreme   *  
## party_nameRepublican:extreme:constituent .  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for Negative Binomial(21.5536) family taken to be 1)
## 
##     Null deviance: 4921.9  on 4123  degrees of freedom
## Residual deviance: 4444.7  on 4112  degrees of freedom
## AIC: 20914
## 
## Number of Fisher Scoring iterations: 1
## 
## 
##               Theta:  21.55 
##           Std. Err.:  2.04 
## 
##  2 x log-likelihood:  -20887.70 
```