Case Study 3 - A Study of Demographic Differences Against Party Preference

Elliot Pickens & Dean Gladish

May 22, 2018

Introduction

Political party preference is typically thought to be associated with the demographics and geography of a populace. It is of interest to politicians, political scientists and the media alike to determine the extent of such correlation in order to understand which groups are most likely to vote for the party. Our case study, which uses data collected from U.S. adults from the 1980 and 2000 elections respectively as part of the National Election Studies project, is an investigation into the matter that allows us to model party preference using the logistic regression model. Specifically, we aim to address whether gender, regional, income, race, age, level of education, and union differences play a part in party preference over time.

Data

The dataset that we analyzed consists of a binary indicator variable indicating Democratic Party preference as well as numerous other categorical variables corresponding to factors such as year, age, gender, race, region, income, unionized and educational status. The explanatory variables that we focus on are *gender*, *race*, *income*, *age* and *union*.

What does this mean? It's unclear to the reader.

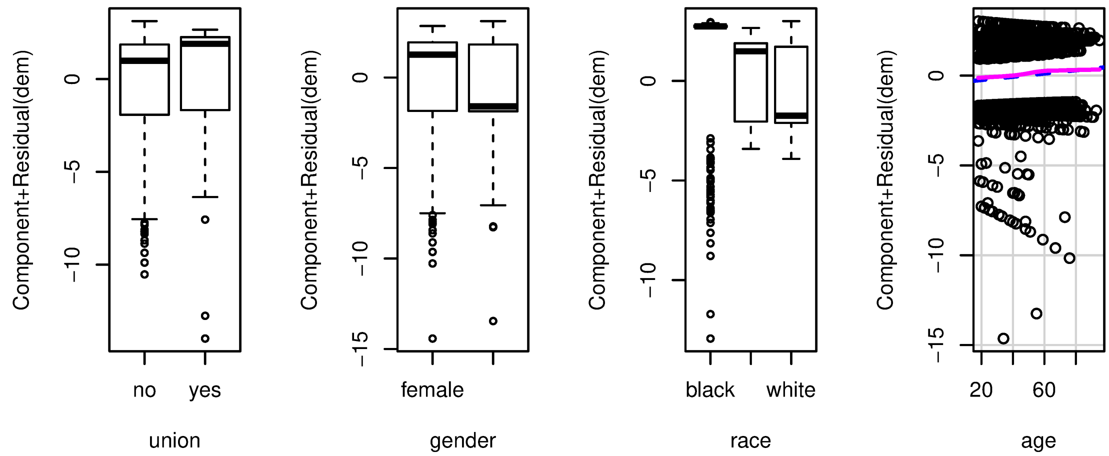
The following table gives our estimates of the important aspects (coefficients, etc.) of our model:

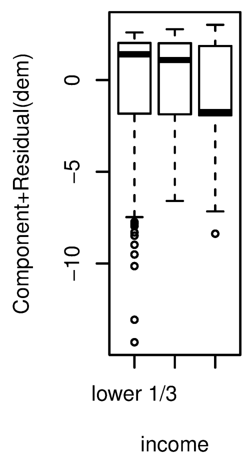
Table 1: Important Coefficients of our Logistic Regression Model

	Estimate	Standard error	z value	P-value
intercept	1.648114	0.221279	7.448	9.47e-14
raceother	-1.600862	0.227855	-7.026	2.13e-12
racewhite	-1.852649	0.182748	-10.138	2e-16
unionyes	0.692776	0.116536	5.945	2.77e-09
incomemiddle 1/3	-0.265841	0.113745	-2.337	0.01943
incomeupper 1/3	-0.491922	0.114198	-4.308	1.65e-05
age	0.007781	0.002699	2.883	0.00395
gendermale	-0.258311	0.090486	-2.855	0.00431

The following set of plots represent what is essentially the relationship between our binary model and the data for the years 1980 and 2000.

What is the reader supposed to learn from these plots? Don't just include plots, also discuss them.





Effects plots would be clearer.

This sounds like inference… clearly state what you mean

Through exploratory data analysis of significance and association, we found that interaction variables did not give us a closer fit to the data.

Results:

The description asked you not to remove terms based on significance.

Using the AIC criterion we obtained the following model:

$$\widehat{Y}_i\{dem\} = \beta_0 + \beta_1 raceother + \beta_2 racewhite + \beta_3 unionyes + \beta_4 incomemiddle\ 1/3 + \beta_5 incomeupper\ 1/3 + \beta_6 age + \beta_7 gendermale$$

Check the interaction with region and time… there is something happening in the South.

Our model was selected for its simplicity; based on the Akaike information criterion and our EDA, we know that it is reasonable to state that geographic region has little to no association with party preference. Out of our variables, racewhite and unionyes have the strongest effect on non- and pro-Democratic party preference respectively.

The model can be interpreted as follows: being in a union, for example, is associated with an $e^{0.693}$ multiplicative increase in one's odds (P(Yes)/P(No)) of preferring the Democratic Party. Being male is associated with an $e^{0.258}$ multiplicative decrease in the odds of Democratic preference.

Simplify for clarity.

Evaluating the variables of the model and their respective p-values based on an alpha level of 0.05 indicates that the race indicator is much more significant than the other indicators. While our model assumes a lack of multicollinearity, our simplified model accounts for this via the exclusion of the region variable.

How so?

A 95% confidence interval for our unionyes indicator is (0.459, 0.925), and a 95% confidence interval for gendermale is (-0.0773, -0.439).

I would recommend integrating the CIs with the interpretations.

Discussion:

Out of the eight features that were contained in the original datset we ended up using only five to predict party status in our model, and this model does not include any interaction, or otherwise transformed variables. While, the simplicity of the model may suggest robustness it may may be an over simplification of the the interaction that we are ultimately trying to model. That being said (and Occam's Razor suggests) that this simplification may infact be a good thing, given that we are trying to create a model that works under a very broad range of circumstances. This may leave our model vulnerable to mis-classifying very specific groups of people, but if we wish to provide an accurate model for each and every group of people we may need the help of some political scientists that have understanding of those specific groups. Overall our model seems to predict party affiliation fairly well without completely violating its underlying assumptions or completely overfitting itself to the data.

So you can't assess changes over time

In the future, however we would suggest that another sample of the population be taken. Given that each of the data points we used in this analysis were collected in either 1980 or 2000 there may be some underlying time related pattern that is subtly influencing our data. Economies, societies, and political parties all undergo transformations over time, and we may be picking up on some of those changes in our model. Alternatively, it is possible that changes have occured since 1980 & 2000 that may render this model less effective.

Interactions can be used to model these changes

Case Study 3 Code Sup

Elliot Pickens & Dean Gladish

May 20, 2018

```
library(Sleuth3)
library(dplyr)
library(ggformula)
library(pander)
library(knitr)
library(stargazer)
library(car)
library(pander)
library(gridExtra)
library(broom)
library(ggthemes)
library(MASS)
library(leaps)
library(GGally)

library(effects)
```

After loading the necessary libraries, we also need to load the dataset of interest:

```
nes <- read.csv("http://aloy.rbind.io/data/NES.csv")
head(nes)
```

```
##   year age gender  race region     income union dem       educ
## 1 1980  70   male black      S  lower 1/3    no   1 HS or less
## 2 1980  67   male white     NC middle 1/3   yes   1 HS or less
## 3 1980  47 female black      S  lower 1/3    no   1 HS or less
## 4 1980  52 female white      W  upper 1/3   yes   0    College
## 5 1980  30 female white     NC  upper 1/3    no   1 HS or less
## 6 1980  37   male black     NC  upper 1/3    no   1    College
```

```
summary(nes)
```

```
##       year           age           gender        race      region
##  Min.   :1980   Min.   :18.00   female:1232   black: 281   NC:563
##  1st Qu.:1980   1st Qu.:33.00   male  :1000   other: 192   NE:427
##  Median :2000   Median :45.00                 white:1759   S :806
##  Mean   :1991   Mean   :46.85                              W :436
##  3rd Qu.:2000   3rd Qu.:59.00
##  Max.   :2000   Max.   :95.00
##         income     union           dem                educ
##  lower 1/3 :799   no :1788   Min.   :0.0000   College   :1179
##  middle 1/3:703   yes: 444   1st Qu.:0.0000   HS or less:1053
##  upper 1/3 :730              Median :1.0000
##                              Mean   :0.5349
##                              3rd Qu.:1.0000
##                              Max.   :1.0000
```

The following explains our derivation of a model:

```r
# The following code generates our baseline model for dem.
glm.base <- glm(dem ~ gender + region + union + income + educ + year + race + age, data = nes, family =
summary(glm.base)
```

```
##
## Call:
## glm(formula = dem ~ gender + region + union + income + educ +
##     year + race + age, family = binomial, data = nes)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2970  -1.1065   0.4957   1.1314   1.5533
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -1.576601   9.492598  -0.166  0.86809
## gendermale       -0.259690   0.090771  -2.861  0.00422 **
## regionNE          0.104580   0.134548   0.777  0.43700
## regionS           0.008454   0.119120   0.071  0.94342
## regionW           0.175367   0.134395   1.305  0.19194
## unionyes          0.682129   0.120305   5.670 1.43e-08 ***
## incomemiddle 1/3 -0.258906   0.115867  -2.235  0.02545 *
## incomeupper 1/3  -0.484958   0.120274  -4.032 5.53e-05 ***
## educHS or less    0.040828   0.100289   0.407  0.68393
## year              0.001592   0.004758   0.335  0.73793
## raceother        -1.640317   0.230511  -7.116 1.11e-12 ***
## racewhite        -1.868719   0.184966 -10.103  < 2e-16 ***
## age               0.007640   0.002743   2.785  0.00535 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3083.3  on 2231  degrees of freedom
## Residual deviance: 2856.9  on 2219  degrees of freedom
## AIC: 2882.9
##
## Number of Fisher Scoring iterations: 4
```

```r
# Regressing on a constant allows us to hold everything except for dem (party
# identification) constant.
glm.basic <- glm(dem ~ 1, data = nes, family = binomial)


stpFwd <- stepAIC(glm.basic, scope = list(lower = ~1, upper = ~ year + region + union + income + educ +
```

```
## Start:  AIC=3085.3
## dem ~ 1
##
##          Df Deviance    AIC
## + race    2   2928.3 2934.3
## + income  2   3044.9 3050.9
## + union   1   3061.7 3065.7
## + educ    1   3067.5 3071.5
```

```
## + gender  1   3072.2 3076.2
## + age     1   3077.7 3081.7
## <none>        3083.3 3085.3
## + year    1   3083.3 3087.3
## + region  3   3081.9 3089.9
##
## Step:  AIC=2934.28
## dem ~ race
##
##          Df Deviance    AIC
## + union   1   2905.6 2913.6
## + income  2   2909.3 2919.3
## + age     1   2915.9 2923.9
## + gender  1   2918.6 2926.6
## + educ    1   2919.9 2927.9
## <none>        2928.3 2934.3
## + year    1   2928.2 2936.2
## + region  3   2925.0 2937.0
## - race    2   3083.3 3085.3
##
## Step:  AIC=2913.6
## dem ~ race + union
##
##          Df Deviance    AIC
## + income  2   2875.9 2887.9
## + age     1   2889.7 2899.7
## + gender  1   2893.4 2903.4
## + educ    1   2899.1 2909.1
## <none>        2905.6 2913.6
## + year    1   2905.4 2915.4
## + region  3   2903.9 2917.9
## - union   1   2928.3 2934.3
## - race    2   3061.7 3065.7
##
## Step:  AIC=2887.94
## dem ~ race + union + income
##
##          Df Deviance    AIC
## + age     1   2867.7 2881.7
## + gender  1   2867.9 2881.9
## <none>        2875.9 2887.9
## + educ    1   2875.4 2889.4
## + year    1   2875.8 2889.8
## + region  3   2873.7 2891.7
## - income  2   2905.6 2913.6
## - union   1   2909.3 2919.3
## - race    2   3007.3 3015.3
##
## Step:  AIC=2881.69
## dem ~ race + union + income + age
##
##          Df Deviance    AIC
## + gender  1   2859.5 2875.5
## <none>        2867.7 2881.7
```

```
## + educ    1   2867.6 2883.6
## + year    1   2867.7 2883.7
## + region  3   2865.4 2885.4
## - age     1   2875.9 2887.9
## - income  2   2889.7 2899.7
## - union   1   2902.4 2914.4
## - race    2   3005.5 3015.5
##
## Step:  AIC=2875.53
## dem ~ race + union + income + age + gender
##
##          Df Deviance    AIC
## <none>        2859.5 2875.5
## + year    1   2859.5 2877.5
## + educ    1   2859.5 2877.5
## + region  3   2857.2 2879.2
## - gender  1   2867.7 2881.7
## - age     1   2867.9 2881.9
## - income  2   2878.2 2890.2
## - union   1   2895.9 2909.9
## - race    2   2997.7 3009.7
```

```r
summary(stpFwd)
```

```
##
## Call:
## glm(formula = dem ~ race + union + income + age + gender, family = binomial,
##     data = nes)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3138  -1.1044   0.4946   1.1298   1.5371
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       1.648114   0.221279   7.448 9.47e-14 ***
## raceother        -1.600862   0.227855  -7.026 2.13e-12 ***
## racewhite        -1.852649   0.182748 -10.138  < 2e-16 ***
## unionyes          0.692776   0.116536   5.945 2.77e-09 ***
## incomemiddle 1/3 -0.265841   0.113745  -2.337  0.01943 *
## incomeupper 1/3  -0.491922   0.114198  -4.308 1.65e-05 ***
## age               0.007781   0.002699   2.883  0.00395 **
## gendermale       -0.258311   0.090486  -2.855  0.00431 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3083.3  on 2231  degrees of freedom
## Residual deviance: 2859.5  on 2224  degrees of freedom
## AIC: 2875.5
##
## Number of Fisher Scoring iterations: 4
```

```r
stpBk <- stepAIC(glm.base, scope = list(lower = ~1, upper = ~ year + region + union + income + educ + ge
```

```
## Start:  AIC=2882.94
## dem ~ gender + region + union + income + educ + year + race +
##     age
##
##          Df Deviance    AIC
## - region  3   2859.4 2879.4
## - year    1   2857.1 2881.1
## - educ    1   2857.1 2881.1
## <none>        2856.9 2882.9
## - age     1   2864.7 2888.7
## - gender  1   2865.1 2889.1
## - income  2   2873.3 2895.3
## - union   1   2889.9 2913.9
## - race    2   2992.9 3014.9
##
## Step:  AIC=2879.36
## dem ~ gender + union + income + educ + year + race + age
##
##          Df Deviance    AIC
## - educ    1   2859.5 2877.5
## - year    1   2859.5 2877.5
## <none>        2859.4 2879.4
## + region  3   2856.9 2882.9
## - age     1   2867.1 2885.1
## - gender  1   2867.5 2885.5
## - income  2   2875.4 2891.4
## - union   1   2894.4 2912.4
## - race    2   2996.1 3012.1
##
## Step:  AIC=2877.46
## dem ~ gender + union + income + year + race + age
##
##          Df Deviance    AIC
## - year    1   2859.5 2875.5
## <none>        2859.5 2877.5
## + educ    1   2859.4 2879.4
## + region  3   2857.1 2881.1
## - gender  1   2867.7 2883.7
## - age     1   2867.7 2883.7
## - income  2   2878.1 2892.1
## - union   1   2895.6 2911.6
## - race    2   2997.4 3011.4
##
## Step:  AIC=2875.53
## dem ~ gender + union + income + race + age
##
##          Df Deviance    AIC
## <none>        2859.5 2875.5
## + year    1   2859.5 2877.5
## + educ    1   2859.5 2877.5
## + region  3   2857.2 2879.2
## - gender  1   2867.7 2881.7
```

```
## - age      1   2867.9 2881.9
## - income   2   2878.2 2890.2
## - union    1   2895.9 2909.9
## - race     2   2997.7 3009.7
```

```r
summary(stpBk)
```

```
##
## Call:
## glm(formula = dem ~ gender + union + income + race + age, family = binomial,
##     data = nes)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3138  -1.1044   0.4946   1.1298   1.5371
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)        1.648114   0.221279   7.448 9.47e-14 ***
## gendermale        -0.258311   0.090486  -2.855  0.00431 **
## unionyes           0.692776   0.116536   5.945 2.77e-09 ***
## incomemiddle 1/3 -0.265841   0.113745  -2.337  0.01943 *
## incomeupper 1/3  -0.491922   0.114198  -4.308 1.65e-05 ***
## raceother         -1.600862   0.227855  -7.026 2.13e-12 ***
## racewhite         -1.852649   0.182748 -10.138  < 2e-16 ***
## age                0.007781   0.002699   2.883  0.00395 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3083.3  on 2231  degrees of freedom
## Residual deviance: 2859.5  on 2224  degrees of freedom
## AIC: 2875.5
##
## Number of Fisher Scoring iterations: 4
```

```r
glm.square <- glm(dem ~ year + region + union + income + educ + gender + race + age +
                    I(year)^2 + I(region)^2 + I(union)^2 + I(income)^2 + I(educ)^2 + I(gender)^2 + I(ra
                  data = nes, family = binomial)

glm.inter <- glm(dem ~ year + region + union + income + educ + gender + race + age +
                   age * year + age * region + age * union + age * income + age * educ + age * gender +
summary(glm.inter)
```

```
##
## Call:
## glm(formula = dem ~ year + region + union + income + educ + gender +
##     race + age + age * year + age * region + age * union + age *
##     income + age * educ + age * gender + age * race, family = binomial,
##     data = nes)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4168  -1.0933   0.4528   1.1202   1.6582
```

```
##
## Coefficients:
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -6.595e+00  2.733e+01  -0.241  0.80931
## year                  3.858e-03  1.370e-02   0.282  0.77818
## regionNE              1.102e+00  3.984e-01   2.766  0.00568 **
## regionS               1.730e-01  3.495e-01   0.495  0.62066
## regionW               2.172e-01  3.968e-01   0.547  0.58408
## unionyes              2.211e-01  3.802e-01   0.581  0.56094
## incomemiddle 1/3      2.062e-01  3.317e-01   0.622  0.53423
## incomeupper 1/3       8.013e-02  3.700e-01   0.217  0.82855
## educHS or less       -5.212e-01  2.891e-01  -1.803  0.07141 .
## gendermale           -6.244e-01  2.673e-01  -2.336  0.01951 *
## raceother            -7.516e-01  6.687e-01  -1.124  0.26103
## racewhite            -1.397e+00  5.319e-01  -2.627  0.00862 **
## age                   1.010e-01  5.517e-01   0.183  0.85479
## year:age             -4.085e-05  2.763e-04  -0.148  0.88246
## regionNE:age         -2.073e-02  7.795e-03  -2.660  0.00782 **
## regionS:age          -3.417e-03  7.006e-03  -0.488  0.62577
## regionW:age          -4.642e-04  8.192e-03  -0.057  0.95481
## unionyes:age          1.034e-02  8.180e-03   1.265  0.20602
## incomemiddle 1/3:age -9.270e-03  6.681e-03  -1.387  0.16531
## incomeupper 1/3:age  -1.174e-02  7.709e-03  -1.523  0.12772
## educHS or less:age    1.188e-02  5.970e-03   1.991  0.04649 *
## gendermale:age        7.814e-03  5.424e-03   1.441  0.14967
## raceother:age        -2.264e-02  1.568e-02  -1.444  0.14888
## racewhite:age        -1.229e-02  1.248e-02  -0.985  0.32471
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3083.3  on 2231  degrees of freedom
## Residual deviance: 2831.4  on 2208  degrees of freedom
## AIC: 2879.4
##
## Number of Fisher Scoring iterations: 4
```

```
stp.inter <- stepAIC(glm.inter, scope = list(lower = ~1, upper = ~ year + region + union + income + edu
                                              age * year + age * region + age * union + age * income +
                     direction = "both")
```

```
## Start:  AIC=2879.36
## dem ~ year + region + union + income + educ + gender + race +
##     age + age * year + age * region + age * union + age * income +
##     age * educ + age * gender + age * race
##
##              Df Deviance    AIC
## - year:age    1   2831.4 2877.4
## - race:age    2   2833.5 2877.5
## - income:age  2   2834.4 2878.4
## - union:age   1   2833.0 2879.0
## <none>            2831.4 2879.4
## - gender:age  1   2833.4 2879.4
## - educ:age    1   2835.3 2881.3
```

```
## - region:age  3   2840.3 2882.3
##
## Step:  AIC=2877.38
## dem ~ year + region + union + income + educ + gender + race +
##     age + region:age + union:age + income:age + educ:age + gender:age +
##     race:age
##
##              Df Deviance    AIC
## - year        1   2831.5 2875.5
## - race:age    2   2833.6 2875.6
## - income:age  2   2834.4 2876.4
## - union:age   1   2833.1 2877.1
## <none>            2831.4 2877.4
## - gender:age  1   2833.5 2877.5
## + year:age    1   2831.4 2879.4
## - educ:age    1   2835.7 2879.7
## - region:age  3   2840.3 2880.3
##
## Step:  AIC=2875.54
## dem ~ region + union + income + educ + gender + race + age +
##     region:age + union:age + income:age + educ:age + gender:age +
##     race:age
##
##              Df Deviance    AIC
## - race:age    2   2833.7 2873.7
## - income:age  2   2834.6 2874.6
## - union:age   1   2833.2 2875.2
## <none>            2831.5 2875.5
## - gender:age  1   2833.7 2875.7
## + year        1   2831.4 2877.4
## - educ:age    1   2835.8 2877.8
## - region:age  3   2840.5 2878.5
##
## Step:  AIC=2873.73
## dem ~ region + union + income + educ + gender + race + age +
##     region:age + union:age + income:age + educ:age + gender:age
##
##              Df Deviance    AIC
## - income:age  2   2836.8 2872.8
## - union:age   1   2835.5 2873.5
## <none>            2833.7 2873.7
## - gender:age  1   2835.8 2873.8
## + race:age    2   2831.5 2875.5
## + year        1   2833.6 2875.6
## - educ:age    1   2838.1 2876.1
## - region:age  3   2842.4 2876.4
## - race        2   2971.7 3007.7
##
## Step:  AIC=2872.75
## dem ~ region + union + income + educ + gender + race + age +
##     region:age + union:age + educ:age + gender:age
##
##              Df Deviance    AIC
## - union:age   1   2837.7 2871.7
```

```
## - gender:age  1   2838.4 2872.4
## <none>            2836.8 2872.8
## + income:age  2   2833.7 2873.7
## + year        1   2836.6 2874.6
## + race:age    2   2834.6 2874.6
## - region:age  3   2845.3 2875.3
## - educ:age    1   2845.0 2879.0
## - income      2   2851.8 2883.8
## - race        2   2974.0 3006.0
##
## Step:  AIC=2871.68
## dem ~ region + union + income + educ + gender + race + age +
##     region:age + educ:age + gender:age
##
##              Df Deviance    AIC
## - gender:age  1   2839.6 2871.6
## <none>            2837.7 2871.7
## + union:age   1   2836.8 2872.8
## + year        1   2837.5 2873.5
## + race:age    2   2835.5 2873.5
## + income:age  2   2835.5 2873.5
## - region:age  3   2846.0 2874.0
## - educ:age    1   2846.3 2878.3
## - income      2   2852.8 2882.8
## - union       1   2870.4 2902.4
## - race        2   2975.3 3005.3
##
## Step:  AIC=2871.6
## dem ~ region + union + income + educ + gender + race + age +
##     region:age + educ:age
##
##              Df Deviance    AIC
## <none>            2839.6 2871.6
## + gender:age  1   2837.7 2871.7
## + union:age   1   2838.4 2872.4
## + year        1   2839.4 2873.4
## + race:age    2   2837.4 2873.4
## + income:age  2   2837.9 2873.9
## - region:age  3   2848.4 2874.4
## - educ:age    1   2848.0 2878.0
## - gender      1   2848.3 2878.3
## - income      2   2854.3 2882.3
## - union       1   2871.9 2901.9
## - race        2   2977.6 3005.6
```

```
glm.inter <- glm(dem ~ year + region + union + income + educ + gender + race + age +
                  + age * union + age * income + age * educ + age * race, data = nes, family = binomial

stp.inter <- stepAIC(glm.inter, scope = list(lower = ~1, upper = ~ year + region + union + income + edu
                                                 age * year + age * region + age * union + age * income +
                     direction = "both", k = log(nrow(nes)))
```

```
## Start:  AIC=2989.49
## dem ~ year + region + union + income + educ + gender + race +
##     age + +age * union + age * income + age * educ + age * race
```

```
##
##              Df Deviance    AIC
## - region      3   2845.6 2969.0
## - race:age    2   2844.9 2976.0
## - income:age  2   2845.3 2976.3
## - year        1   2843.2 2982.1
## - union:age   1   2844.7 2983.4
## - educ:age    1   2847.5 2986.3
## <none>            2843.0 2989.5
## - gender      1   2851.3 2990.1
## + gender:age  1   2840.3 2994.5
## + year:age    1   2843.0 2997.2
## + region:age  3   2833.5 3003.1
##
## Step:  AIC=2968.99
## dem ~ year + union + income + educ + gender + race + age + union:age +
##     income:age + educ:age + race:age
##
##              Df Deviance    AIC
## - income:age  2   2847.7 2955.7
## - race:age    2   2847.7 2955.7
## - year        1   2845.9 2961.6
## - union:age   1   2847.2 2962.8
## - educ:age    1   2850.2 2965.8
## <none>            2845.6 2969.0
## - gender      1   2853.8 2969.4
## + gender:age  1   2842.7 2973.8
## + year:age    1   2845.6 2976.7
## + region      3   2843.0 2989.5
##
## Step:  AIC=2955.65
## dem ~ year + union + income + educ + gender + race + age + union:age +
##     educ:age + race:age
##
##              Df Deviance    AIC
## - race:age    2   2849.8 2942.3
## - year        1   2848.0 2948.2
## - union:age   1   2848.6 2948.8
## - income      2   2861.0 2953.5
## <none>            2847.7 2955.7
## - educ:age    1   2855.7 2955.9
## - gender      1   2856.3 2956.5
## + gender:age  1   2845.4 2961.0
## + year:age    1   2847.7 2963.4
## + income:age  2   2845.6 2969.0
## + region      3   2845.3 2976.3
##
## Step:  AIC=2942.35
## dem ~ year + union + income + educ + gender + race + age + union:age +
##     educ:age
##
##              Df Deviance    AIC
## - year        1   2850.1 2934.9
## - union:age   1   2850.7 2935.5
```

```
## - income      2   2863.3 2940.4
## <none>            2849.8 2942.3
## - educ:age    1   2858.2 2943.0
## - gender      1   2858.4 2943.3
## + gender:age  1   2847.5 2947.7
## + year:age    1   2849.8 2950.1
## + race:age    2   2847.7 2955.7
## + income:age  2   2847.7 2955.7
## + region      3   2847.2 2962.9
## - race        2   2988.5 3065.6
##
## Step:  AIC=2934.93
## dem ~ union + income + educ + gender + race + age + union:age +
##     educ:age
##
##              Df Deviance    AIC
## - union:age   1   2851.0 2928.1
## - income      2   2864.0 2933.4
## <none>            2850.1 2934.9
## - educ:age    1   2858.3 2935.4
## - gender      1   2858.7 2935.8
## + gender:age  1   2847.8 2940.3
## + year        1   2849.8 2942.3
## + race:age    2   2848.0 2948.2
## + income:age  2   2848.0 2948.2
## + region      3   2847.5 2955.4
## - race        2   2989.5 3058.9
##
## Step:  AIC=2928.1
## dem ~ union + income + educ + gender + race + age + educ:age
##
##              Df Deviance    AIC
## - income      2   2864.9 2926.6
## <none>            2851.0 2928.1
## - educ:age    1   2859.5 2928.9
## - gender      1   2859.6 2929.0
## + gender:age  1   2848.4 2933.2
## + union:age   1   2850.1 2934.9
## + year        1   2850.7 2935.5
## + race:age    2   2848.9 2941.4
## + income:age  2   2849.6 2942.1
## + region      3   2848.4 2948.7
## - union       1   2885.4 2954.8
## - race        2   2990.9 3052.6
##
## Step:  AIC=2926.6
## dem ~ union + educ + gender + race + age + educ:age
##
##              Df Deviance    AIC
## <none>            2864.9 2926.6
## + income      2   2851.0 2928.1
## - educ:age    1   2875.9 2929.8
## - gender      1   2875.9 2929.9
## + gender:age  1   2862.8 2932.2
```

```
## + union:age   1   2864.0 2933.4
## + year        1   2864.2 2933.6
## + race:age    2   2862.5 2939.6
## - union       1   2892.0 2945.9
## + region      3   2862.6 2947.4
## - race        2   3023.1 3069.4
```

```r
summary(stp.inter)
```

```
##
## Call:
## glm(formula = dem ~ union + educ + gender + race + age + educ:age,
##     family = binomial, data = nes)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2510  -1.1132   0.4908   1.1315   1.5399
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)          1.800e+00  2.507e-01   7.178 7.07e-13 ***
## unionyes             5.909e-01  1.149e-01   5.141 2.74e-07 ***
## educHS or less      -6.823e-01  2.653e-01  -2.572 0.010113 *
## gendermale          -2.986e-01  9.003e-02  -3.316 0.000912 ***
## raceother           -1.672e+00  2.279e-01  -7.335 2.22e-13 ***
## racewhite           -1.956e+00  1.824e-01 -10.720  < 2e-16 ***
## age                  9.047e-05  3.881e-03   0.023 0.981404
## educHS or less:age   1.745e-02  5.292e-03   3.298 0.000973 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3083.3  on 2231  degrees of freedom
## Residual deviance: 2864.9  on 2224  degrees of freedom
## AIC: 2880.9
##
## Number of Fisher Scoring iterations: 4
```

```r
stp.inter.fwd <- stepAIC(glm.basic, scope = list(lower = ~1, upper = ~ year + region + union + income +
                                                  age * year + age * region + age * union + age * incom
                         direction = "both", k = log(nrow(nes)))
```

```
## Start:  AIC=3091.01
## dem ~ 1
##
##          Df Deviance    AIC
## + race    2   2928.3 2951.4
## + income  2   3044.9 3068.0
## + union   1   3061.7 3077.1
## + educ    1   3067.5 3082.9
## + gender  1   3072.2 3087.7
## <none>        3083.3 3091.0
## + age     1   3077.7 3093.1
## + year    1   3083.3 3098.7
```

```
## + region  3   3081.9 3112.7
##
## Step:  AIC=2951.41
## dem ~ race
##
##          Df Deviance    AIC
## + union   1   2905.6 2936.4
## + age     1   2915.9 2946.7
## + income  2   2909.3 2947.8
## + gender  1   2918.6 2949.4
## + educ    1   2919.9 2950.7
## <none>        2928.3 2951.4
## + year    1   2928.2 2959.1
## + region  3   2925.0 2971.3
## - race    2   3083.3 3091.0
##
## Step:  AIC=2936.44
## dem ~ race + union
##
##          Df Deviance    AIC
## + income  2   2875.9 2922.2
## + age     1   2889.7 2928.3
## + gender  1   2893.4 2932.0
## <none>        2905.6 2936.4
## + educ    1   2899.1 2937.7
## + year    1   2905.4 2944.0
## - union   1   2928.3 2951.4
## + region  3   2903.9 2957.9
## - race    2   3061.7 3077.1
##
## Step:  AIC=2922.21
## dem ~ race + union + income
##
##          Df Deviance    AIC
## + age     1   2867.7 2921.7
## + gender  1   2867.9 2921.8
## <none>        2875.9 2922.2
## + educ    1   2875.4 2929.3
## + year    1   2875.8 2929.8
## - income  2   2905.6 2936.4
## + region  3   2873.7 2943.1
## - union   1   2909.3 2947.8
## - race    2   3007.3 3038.2
##
## Step:  AIC=2921.67
## dem ~ race + union + income + age
##
##              Df Deviance    AIC
## + gender      1   2859.5 2921.2
## <none>            2867.7 2921.7
## - age         1   2875.9 2922.2
## + union:age   1   2866.5 2928.2
## - income      2   2889.7 2928.3
## + educ        1   2867.6 2929.3
```

```
## + year        1   2867.7 2929.3
## + income:age  2   2862.8 2932.2
## + race:age    2   2865.1 2934.5
## + region      3   2865.4 2942.5
## - union       1   2902.4 2948.7
## - race        2   3005.5 3044.0
##
## Step:  AIC=2921.22
## dem ~ race + union + income + age + gender
##
##              Df Deviance    AIC
## <none>            2859.5 2921.2
## - gender      1   2867.7 2921.7
## - age         1   2867.9 2921.8
## - income      2   2878.2 2924.4
## + gender:age  1   2857.1 2926.5
## + union:age   1   2858.4 2927.8
## + year        1   2859.5 2928.9
## + educ        1   2859.5 2928.9
## + income:age  2   2855.1 2932.2
## + race:age    2   2857.0 2934.1
## + region      3   2857.2 2942.0
## - union       1   2895.9 2949.8
## - race        2   2997.7 3044.0
```

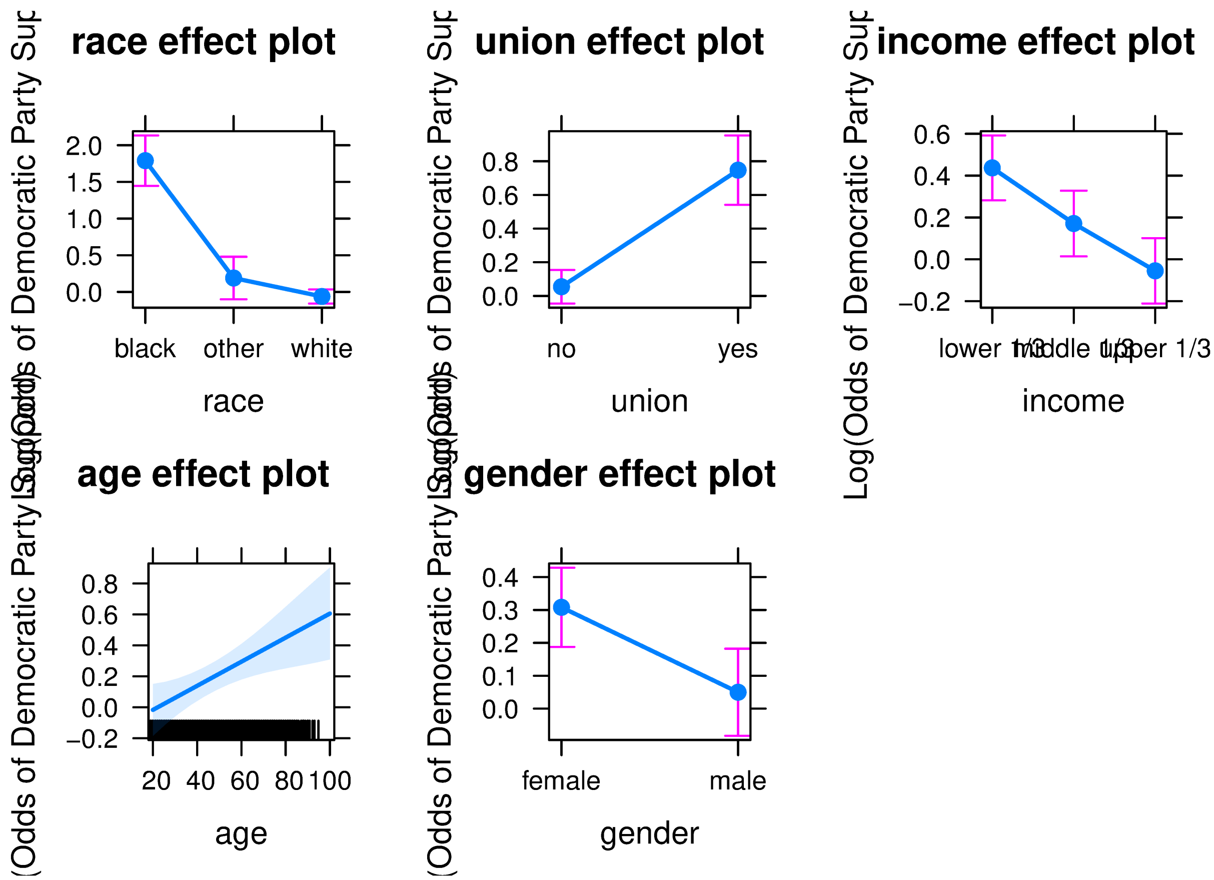
```r
summary(stp.inter.fwd)
```

```
##
## Call:
## glm(formula = dem ~ race + union + income + age + gender, family = binomial,
##     data = nes)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3138  -1.1044   0.4946   1.1298   1.5371
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       1.648114   0.221279   7.448 9.47e-14 ***
## raceother        -1.600862   0.227855  -7.026 2.13e-12 ***
## racewhite        -1.852649   0.182748 -10.138  < 2e-16 ***
## unionyes          0.692776   0.116536   5.945 2.77e-09 ***
## incomemiddle 1/3 -0.265841   0.113745  -2.337  0.01943 *
## incomeupper 1/3  -0.491922   0.114198  -4.308 1.65e-05 ***
## age               0.007781   0.002699   2.883  0.00395 **
## gendermale       -0.258311   0.090486  -2.855  0.00431 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3083.3  on 2231  degrees of freedom
## Residual deviance: 2859.5  on 2224  degrees of freedom
## AIC: 2875.5
##
```

```
## Number of Fisher Scoring iterations: 4
```

```r
# In order to do some preliminary investigation into whether there are
# associations between gender and party preference,
# between region and party preference,
# and between unionized status and party preference,
# I have created some plots of gender, region, and union.

plot(allEffects(stp.inter.fwd), rows = 2, cols = 3, type = "link",
     ylab = "Log(Odds of Democratic Party Support)")
```



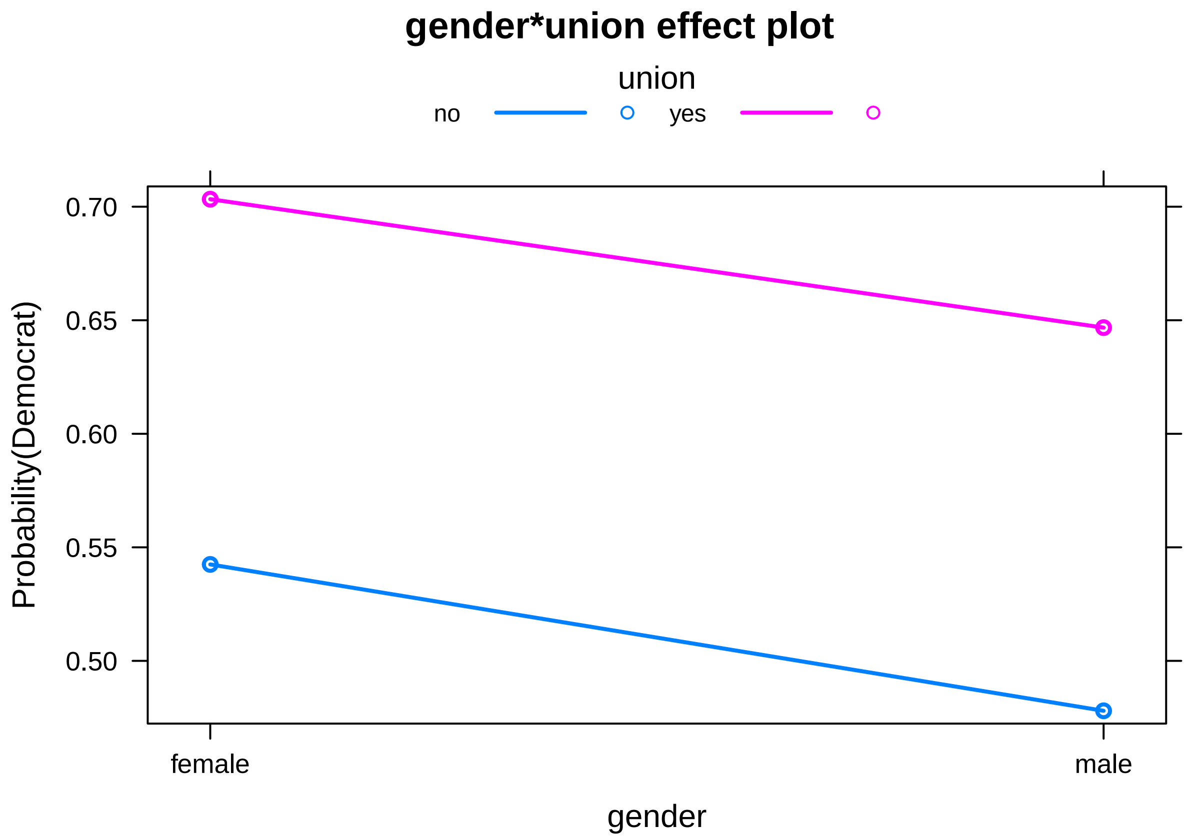
As we can see from the plots, males have lower odds of supporting the Democratic party.

Although people in North Carolina and the Southern region have lower odds of supporting the Democratic party, our model has been simplified such that these do not matter.

Those who are not in unions also have lower odds of supporting the Democratic party.

For further analysis of the probability that any given individual supports the Democrats, we can use the following code:

```r
plot(Effect(c("gender", "union"), stp.inter.fwd), multiline = TRUE, type = "response", ylab = "Probabili
```



This code allows us to more clearly see that Support of the Democratic Party tends to come from people who are in unions and who are female.

More specifically, Unionization seems to have the largest effect on support, followed by Gender and then Region.

NOW, we need to assess the significance of these effects regardless of time.

```
for (i in c(2, 3, 4, 5, 6, 7, 8)) {
  coefficient <- coef(stp.inter.fwd)[i]
  standardError <- sqrt(vcov(stp.inter.fwd)[i,i])
  waldStat <- (coefficient / standardError)^2
  print(1-pchisq(waldStat, df = 1))
}
```

```
##    raceother
## 2.128298e-12
## racewhite
##         0
##     unionyes
## 2.768933e-09
## incomemiddle 1/3
##       0.01943014
## incomeupper 1/3
##    1.650271e-05
##          age
## 0.003945303
## gendermale
```

```
## 0.00430756
```

Based off these p-values, we can reject the null hypothesis that the coefficients are zero. We can reject them for small p-values. Specifically, unionyes and gendermale seem to have an undeniable impact at an alpha level of 0.05.

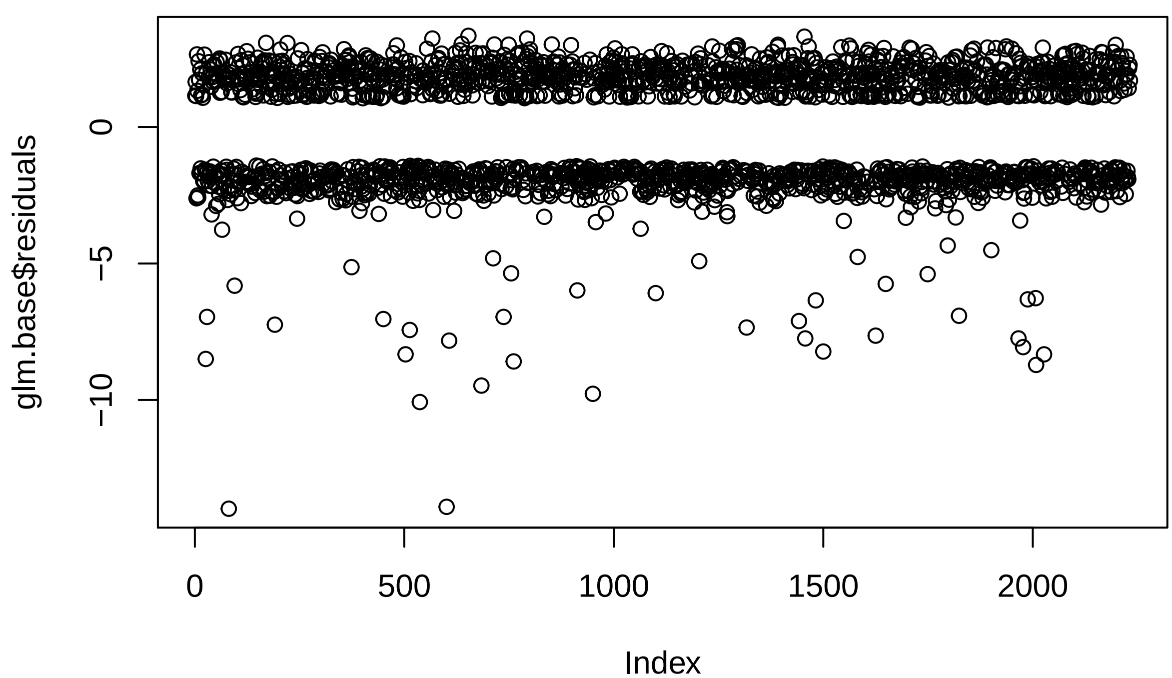
If we also want to look at the significance of the union variable,

```
gender_only <- glm(dem ~ union, family = binomial, data = nes)
anova(gender_only, glm.base, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: dem ~ union
## Model 2: dem ~ gender + region + union + income + educ + year + race +
##     age
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1      2230     3061.7
## 2      2219     2856.9 11   204.72 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

shows that we can reject the notion that the other coefficients are not necessary.

```
plot(glm.base$residuals)
```



The residuals plot shows that our model generally fits the data.