

assi2babur2.R

jbl3

2022-02-22

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

```
elections = read.csv("C:/Users/jbl3/Downloads/electiondata3.csv")
summary(elections)
```

```
##      Vote      Year      Race      Age
## Min.   :0.0000 Min.   :1992 Min.   :1.000 Min.   :18.00
## 1st Qu.:0.0000 1st Qu.:1992 1st Qu.:1.000 1st Qu.:35.00
## Median :0.0000 Median :1996 Median :1.000 Median :47.00
## Mean   :0.4489 Mean   :1997 Mean   :1.446 Mean   :49.02
## 3rd Qu.:1.0000 3rd Qu.:2000 3rd Qu.:1.000 3rd Qu.:62.00
## Max.   :1.0000 Max.   :2004 Max.   :5.000 Max.   :93.00
##
## NA's   :39      NA's   :5
## Marital.Status Family.Income Home.Ownership Contacted.by.Republicans
## Min.   :1.000 Min.   :1.000 Min.   :1.000 Min.   :1.000
## 1st Qu.:1.000 1st Qu.:2.000 1st Qu.:1.000 1st Qu.:2.000
## Median :1.000 Median :3.000 Median :1.000 Median :2.000
## Mean   :1.979 Mean   :3.034 Mean   :1.264 Mean   :1.759
## 3rd Qu.:3.000 3rd Qu.:4.000 3rd Qu.:2.000 3rd Qu.:2.000
## Max.   :7.000 Max.   :5.000 Max.   :2.000 Max.   :2.000
## NA's   :7      NA's   :441 NA's   :88      NA's   :56
## Contacted.by.Democrats
## Min.   :1.000
## 1st Qu.:2.000
## Median :2.000
## Mean   :1.761
## 3rd Qu.:2.000
## Max.   :2.000
## NA's   :56
```

```
#year categories
```

```
year = factor(elections$Year)
levels(year)=c('1992','1996','2000','2004')
years = relevel(year, ref = '2000')
summary(lm(elections$Vote~year))
```

```
##
## Call:
## lm(formula = elections$Vote ~ year)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5080 -0.4197 -0.4156  0.5268  0.5844
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.415623   0.013471  30.853 < 2e-16 ***
## year1996     0.004107   0.020485   0.200  0.84112
## year2000     0.057592   0.020033   2.875  0.00406 **
## year2004     0.092392   0.022025   4.195 2.79e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4962 on 4318 degrees of freedom
## Multiple R-squared:  0.005498, Adjusted R-squared:  0.004807
## F-statistic: 7.958 on 3 and 4318 DF, p-value: 2.737e-05
```

```
summary('1992')
```

```
##      Length      Class      Mode
##           1 character character
```

```
dat = years
summary(dat)
```

```
## 2000 1992 1996 2004
## 1120 1357 1034  811
```

```
#race categories: omitted is white, race 2 black, asian, native american, hispanic
```

```
race = factor(elections$Race)
levels(race)=c('White','Black','Asian','Native American','Hispanic')
summary(lm(elections$Vote~race))
```

```
##
## Call:
## lm(formula = elections$Vote ~ race)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5357 -0.5126 -0.0631  0.4874  0.9369
##
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.512567   0.008208  62.445 < 2e-16 ***
## raceBlack      -0.449470   0.022314 -20.143 < 2e-16 ***
## raceAsian       0.023147   0.052422   0.442  0.659
## raceNative American -0.003947  0.044817  -0.088  0.930
## raceHispanic   -0.191466   0.033171  -5.772 8.38e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4745 on 4278 degrees of freedom
## (39 observations deleted due to missingness)
## Multiple R-squared:  0.09062,    Adjusted R-squared:  0.08977
## F-statistic: 106.6 on 4 and 4278 DF,  p-value: < 2.2e-16
```

#marital status. Omitted is married, then never married, divorced, separated, widowed, exc. 1986?

```
married = factor(elections$Marital.Status)
levels(married)=c('Married','Never Married','Divorced','Separated','Widowed','Partners')
summary(lm(elections$Vote~married))
```

```
##
## Call:
## lm(formula = elections$Vote ~ married)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5152 -0.5152 -0.3295  0.4848  0.7564
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.515212   0.009816  52.486 < 2e-16 ***
## marriedNever Married -0.185732   0.021076  -8.813 < 2e-16 ***
## marriedDivorced     -0.109387   0.023743  -4.607 4.20e-06 ***
## marriedSeparated    -0.254343   0.052083  -4.883 1.08e-06 ***
## marriedWidowed      -0.131121   0.025365  -5.169 2.46e-07 ***
## marriedPartners     -0.271622   0.056412  -4.815 1.52e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4906 on 4309 degrees of freedom
## (7 observations deleted due to missingness)
## Multiple R-squared:  0.02829,    Adjusted R-squared:  0.02716
## F-statistic: 25.09 on 5 and 4309 DF,  p-value: < 2.2e-16
```

#marry = married

```
#family income, omitted is NA, then 0-16th percentile, 17-33, 34-67, 68-95, 96-100
faminc = factor(elections$Marital.Status)
levels(faminc) = c('NA','0-16%','17-33%','34-67%','68-95%','96-100%')
summary(lm(elections$Vote~faminc))
```

```
##
## Call:
```

```
## lm(formula = elections$Vote ~ faminc)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5152 -0.5152 -0.3295  0.4848  0.7564
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.515212   0.009816  52.486 < 2e-16 ***
## faminc0-16%  -0.185732   0.021076  -8.813 < 2e-16 ***
## faminc17-33% -0.109387   0.023743  -4.607 4.20e-06 ***
## faminc34-67% -0.254343   0.052083  -4.883 1.08e-06 ***
## faminc68-95% -0.131121   0.025365  -5.169 2.46e-07 ***
## faminc96-100% -0.271622   0.056412  -4.815 1.52e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4906 on 4309 degrees of freedom
## (7 observations deleted due to missingness)
## Multiple R-squared:  0.02829,    Adjusted R-squared:  0.02716
## F-statistic: 25.09 on 5 and 4309 DF,  p-value: < 2.2e-16
```

```
#home ownership. omitted is owning, then not owning
home = factor(elections$Home.Ownership)
levels(home) = c('Owns', 'Does Not Own')
summary(lm(elections$Vote~home))
```

```
##
## Call:
## lm(formula = elections$Vote ~ home)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.4849 -0.4849 -0.3477  0.5151  0.6523
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.484926   0.008843  54.836 < 2e-16 ***
## homeDoes Not Own -0.137256   0.017225  -7.969 2.05e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4938 on 4232 degrees of freedom
## (88 observations deleted due to missingness)
## Multiple R-squared:  0.01478,    Adjusted R-squared:  0.01455
## F-statistic: 63.5 on 1 and 4232 DF,  p-value: 2.047e-15
```

```
#did the democratic party contact R. omitted is yes, then no
dem = factor(elections$Contacted.by.Democrats)
levels(dem)=c('Contacted', 'Not Contacted')
summary(lm(elections$Vote~dem))
```

```
##
```

```
## Call:
## lm(formula = elections$Vote ~ dem)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.4663 -0.4663 -0.3906  0.5337  0.6094
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.39058    0.01555  25.118 < 2e-16 ***
## demNot Contacted 0.07570    0.01782   4.247 2.21e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4964 on 4264 degrees of freedom
## (56 observations deleted due to missingness)
## Multiple R-squared:  0.004212, Adjusted R-squared:  0.003979
## F-statistic: 18.04 on 1 and 4264 DF, p-value: 2.212e-05
```

```
#did the republican party contact R. omitted is yes, then no
rep = factor(elections$Contacted.by.Republicans)
levels(rep)=c('Contacted','Not Contacted')
summary(lm(elections$Vote~rep))
```

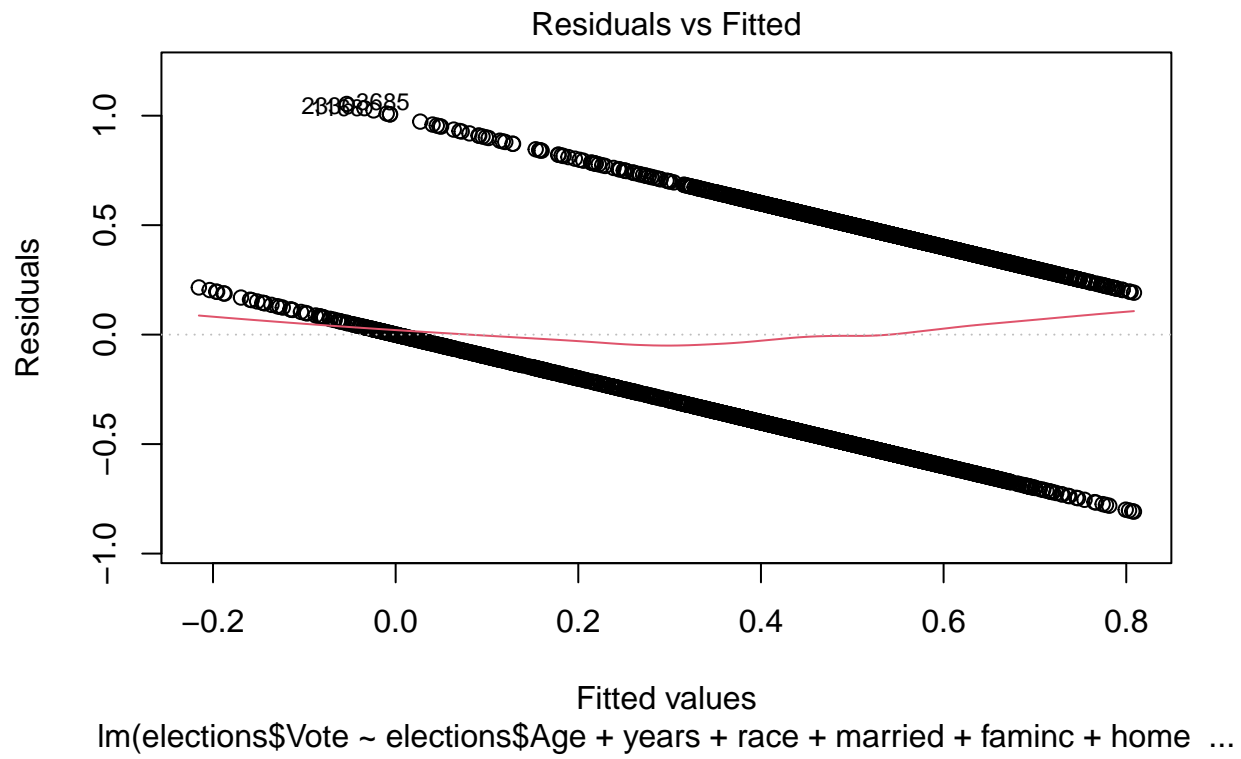
```
##
## Call:
## lm(formula = elections$Vote ~ rep)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5544 -0.4144 -0.4144  0.5856  0.5856
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.55437    0.01539  36.030 < 2e-16 ***
## repNot Contacted -0.13997    0.01767  -7.923 2.94e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4938 on 4264 degrees of freedom
## (56 observations deleted due to missingness)
## Multiple R-squared:  0.01451, Adjusted R-squared:  0.01428
## F-statistic: 62.77 on 1 and 4264 DF, p-value: 2.936e-15
```

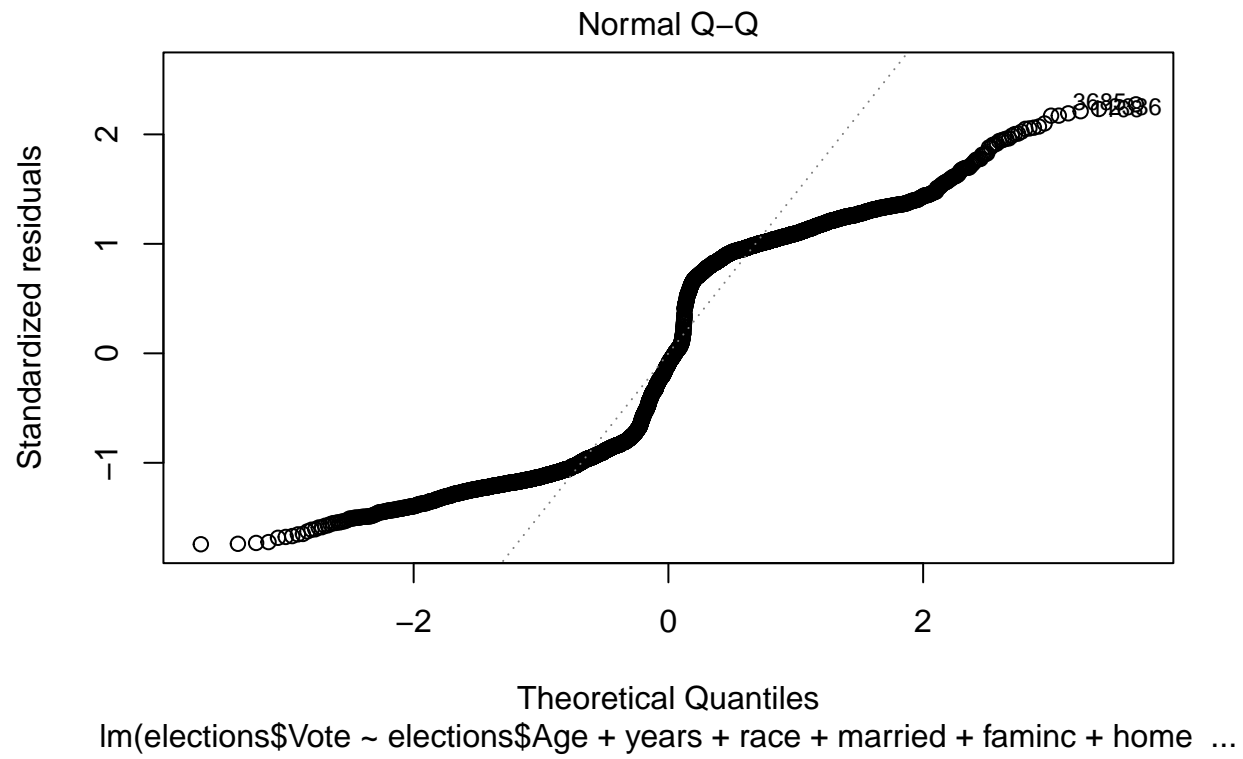
```
#the big regression
summary(lm(elections$Vote~elections$Age + years + race + married + faminc
          + home + dem + rep))
```

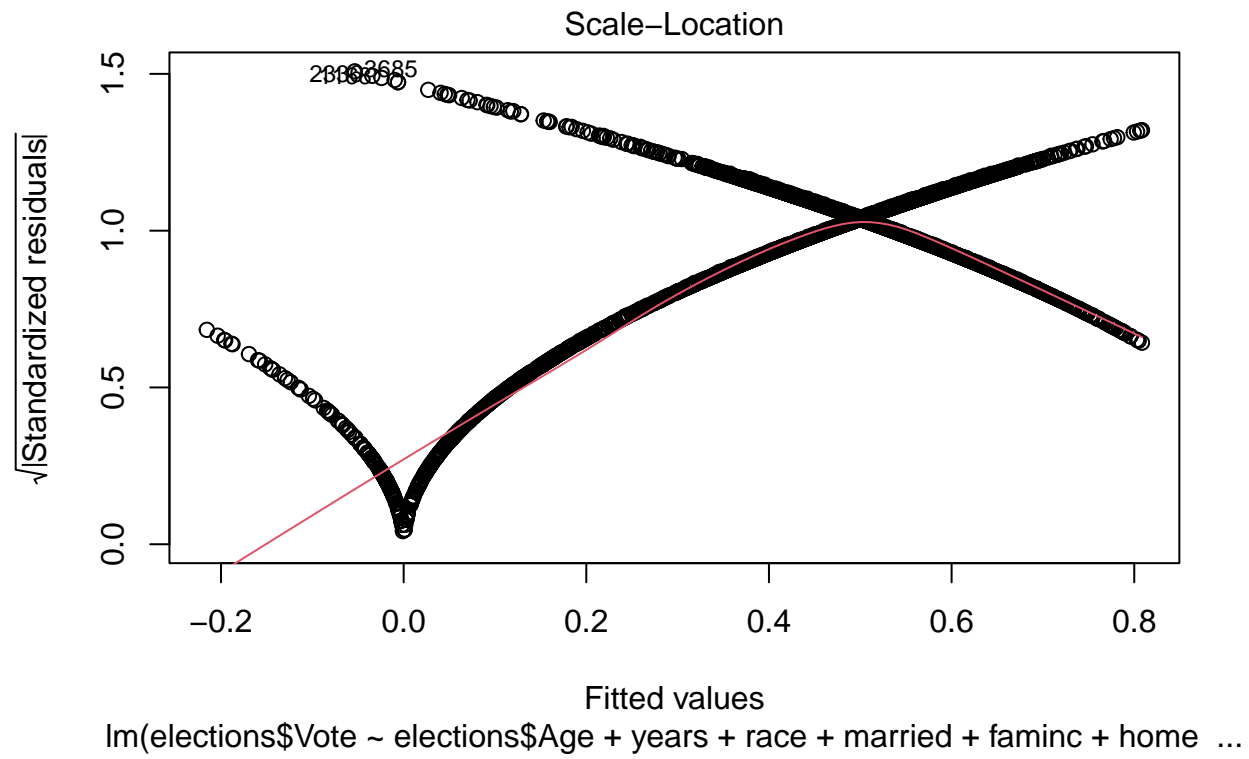
```
##
## Call:
## lm(formula = elections$Vote ~ elections$Age + years + race +
##      married + faminc + home + dem + rep)
##
```

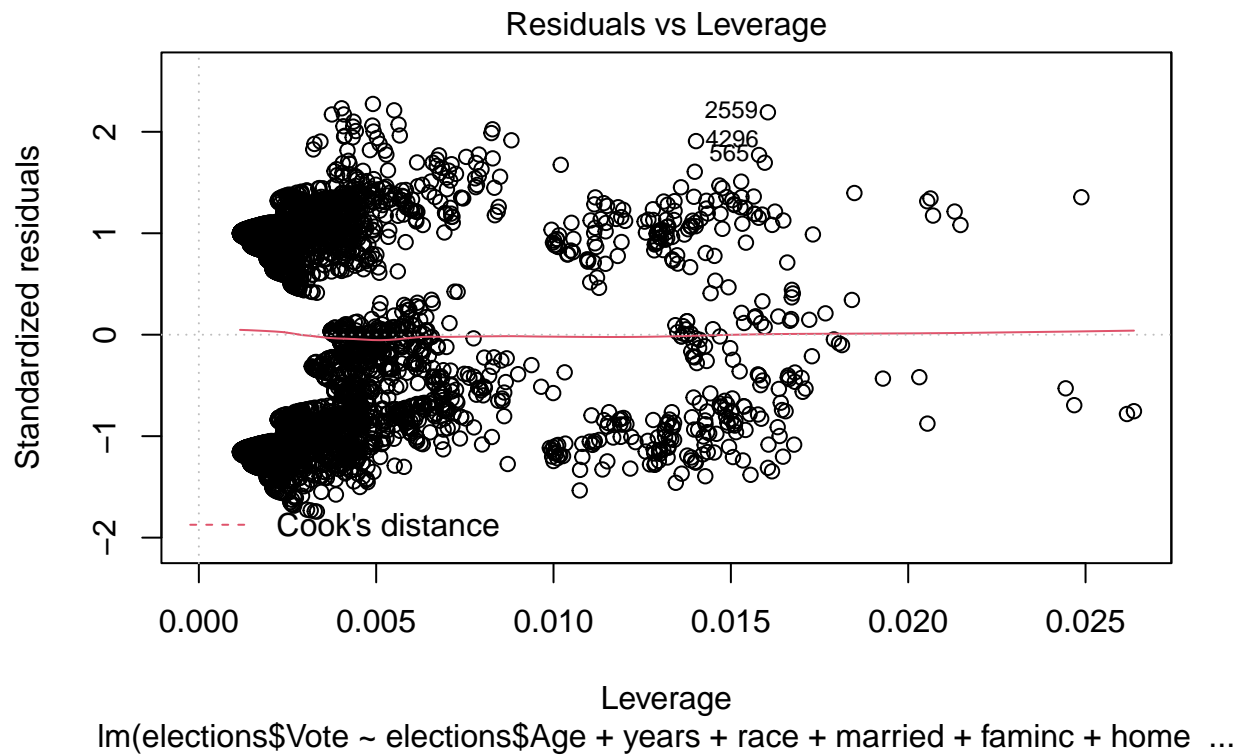
```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.80842 -0.45715 -0.04542  0.45784  1.05359
##
## Coefficients: (5 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.6449149  0.0351112  18.368 < 2e-16 ***
## elections$Age   -0.0018027  0.0005279  -3.415 0.000644 ***
## years1992      -0.0505313  0.0196603  -2.570 0.010198 *
## years1996      -0.0569170  0.0204564  -2.782 0.005421 **
## years2004       0.0670553  0.0218206   3.073 0.002133 **
## raceBlack      -0.3965810  0.0231126 -17.159 < 2e-16 ***
## raceAsian       0.0001817  0.0518472   0.004 0.997204
## raceNative American -0.0071822  0.0445477  -0.161 0.871923
## raceHispanic   -0.1824675  0.0331128  -5.510 3.80e-08 ***
## marriedNever Married -0.1426338  0.0229241  -6.222 5.40e-10 ***
## marriedDivorced -0.0858048  0.0236291  -3.631 0.000285 ***
## marriedSeparated -0.1364833  0.0514855  -2.651 0.008058 **
## marriedWidowed  -0.0596384  0.0274497  -2.173 0.029864 *
## marriedPartners -0.1967141  0.0561734  -3.502 0.000467 ***
## faminc0-16%      NA          NA        NA      NA
## faminc17-33%      NA          NA        NA      NA
## faminc34-67%      NA          NA        NA      NA
## faminc68-95%      NA          NA        NA      NA
## faminc96-100%     NA          NA        NA      NA
## homeDoes Not Own  -0.0457928  0.0182396  -2.511 0.012089 *
## demNot Contacted  0.1469300  0.0182222   8.063 9.65e-16 ***
## repNot Contacted -0.1174555  0.0184156  -6.378 1.99e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4644 on 4119 degrees of freedom
## (186 observations deleted due to missingness)
## Multiple R-squared:  0.1314, Adjusted R-squared:  0.128
## F-statistic: 38.95 on 16 and 4119 DF, p-value: < 2.2e-16
```

```
reg = (lm(elections$Vote~elections$Age + years + race + married + faminc
          + home + dem + rep))
#plotting residuals for analysis
resid = reg$residuals
plot(reg)
```



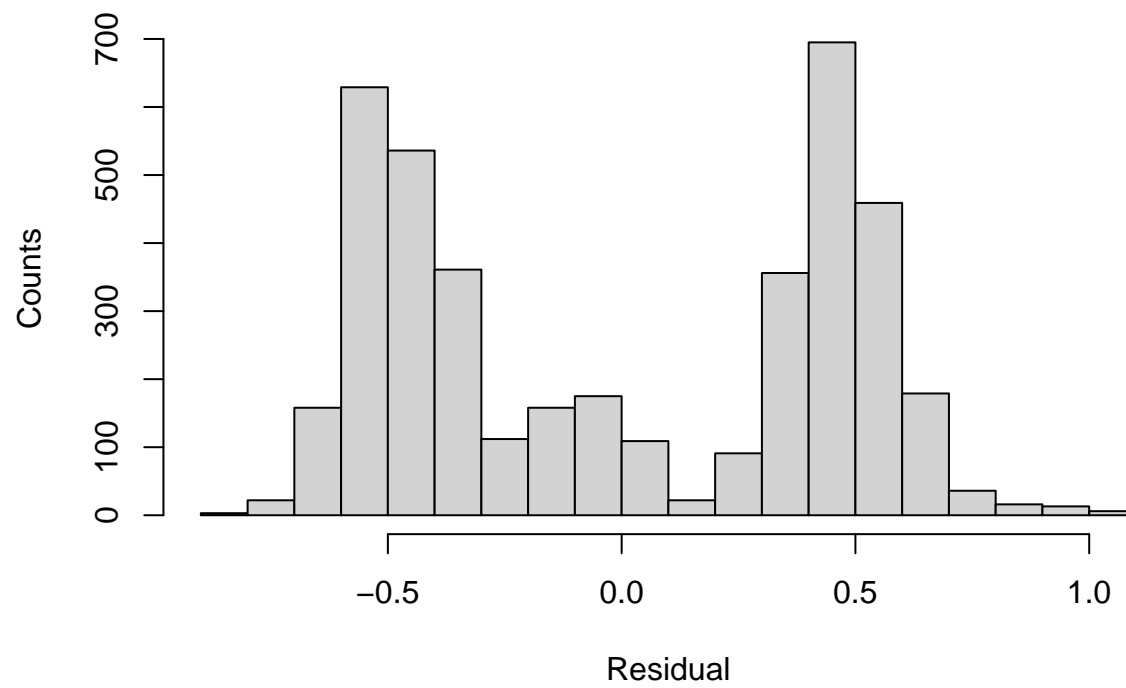






```
hist(resid,20,main=paste("Analysis of Residuals"), xlab="Residual",
     ylab="Counts")
```

Analysis of Residuals



```
summary(predict(reg))
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -0.2155  0.3850  0.4922  0.4478  0.5608  0.8084
```