Chapter 13: Putting It Altogether

Exercises - with Updated 2016 ANES

Brian Fogarty

27 August 2018

Contents

Exercise I

Using the 2016 ANES, replicate the analysis with feelings towards Trump using feelings towards Hillary Clinton. The variable for Clinton is V161086 and use the same 8 predictor variables.

1. Rename V161086 `clinton` and recode to get rid of missing values and non-responses.
2. Write 8 empirical hypotheses.
3. Run and discuss the univariate and descriptive statistics of the `clinton` variable.
4. Create a histogram for `clinton` and each predictor variable. Discuss what they show you. Also, create two graphs with 4 histograms in each one.
5. Perform a difference of means test between `clinton` and `gender`, and discuss what you find.
6. Perform a Wilcoxon rank-sum test between `clinton` and `gender`, and discuss what you find.
7. Run a multiple regression with `clinton` and 4 predictor variables - `gender`, `education`, `partisan identification`, and `political ideology`. Evaluate the overall model and identify any statistically significant relationships.
8. Run another multiple regression with the previous 4 predictors and the 4 other predictors - `hc.law`, `economy`, `wall`, and `isis`. Evaluate the overall model and identify any statistically significant relationships.
9. Run full regression diagnostics on the second regression model. These include diagnostics for functional form, heteroscedasticity, normality, multicollinearity, and influential data points. Attempt to make corrections for any violations.
10. After diagnostics provide an interpretation of any statistically significant coefficients and discuss any significant relationships using plain language.

ANSWERS TO EXERCISE I

Question 1

```r
setwd("C:/QSSD/Exercises/Chapter 13 - Exercises/")
getwd()
```

```
[1] "C:/QSSD/Exercises/Chapter 13 - Exercises"
```

```r
library(foreign)
nes <- read.dta("anes_timeseries_2016_Stata12.dta")
```

```
Warning in read.dta("anes_timeseries_2016_Stata12.dta"): value labels
('V161029b') for 'V161029b' are missing
```

Hillary Clinton feeling thermometer

```r
class(nes$V161086)
```

```
[1] "numeric"
```

```r
table(nes$V161086)
```

```

-99 -88   0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15
 36   2 947  81  69  24  17  20   4   1   9   2  16   2   2   1   4 332
 16  17  18  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33
 33  14   6   2  16   1   1   2   2   5   1   3   3   3 257  24  10   7
 34  35  36  38  39  40  41  42  43  44  45  46  49  50  51  52  53  55
  1   6   3   1   6 228  20   6   4   2   6   1   2 195  15   4   2   8
 56  57  58  59  60  61  62  63  64  65  66  67  68  69  70  71  72  73
  1   2   1   1 373  33  10   3   3  10   5   2   1  10 414  31  13   4
 74  75  76  77  78  79  80  81  82  83  84  85  86  87  88  89  90  91
  4  22   6   4   5   3  25   4   3   1  11 434  13  13   3   3  33   3
 92  93  94  95  96  97  98  99 100
  4   4   1  18   4   8  13  11 232
```

```r
library(car)
```

```
Loading required package: carData
```

```r
nes$clinton <- recode(nes$V161086, "-99:-88=NA")
table(nes$clinton)
```

```

  0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17
947  81  69  24  17  20   4   1   9   2  16   2   2   1   4 332  33  14
 18  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35
  6   2  16   1   1   2   2   5   1   3   3   3 257  24  10   7   1   6
 36  38  39  40  41  42  43  44  45  46  49  50  51  52  53  55  56  57
  3   1   6 228  20   6   4   2   6   1   2 195  15   4   2   8   1   2
 58  59  60  61  62  63  64  65  66  67  68  69  70  71  72  73  74  75
  1   1 373  33  10   3   3  10   5   2   1  10 414  31  13   4   4  22
 76  77  78  79  80  81  82  83  84  85  86  87  88  89  90  91  92  93
  6   4   5   3  25   4   3   1  11 434  13  13   3   3  33   3   4   4
 94  95  96  97  98  99 100
  1  18   4   8  13  11 232
```

Question 2

`Gender`:

H_1: *Women are expected to have warmer feelings towards Clinton than men.*

And the null hypothesis is:

H_0: *There is no relationship between gender and feelings towards Clinton.*

`Education`:

H_2: *As education level increases, respondents are expected to have warmer feelings towards Clinton.*

And the null hypothesis is:

H_0: *There is no relationship between education and feelings towards Clinton.*

`Partisan Identification`:

H_3: *As partisan identification increases, respondents are expected to have cooler feelings towards Clinton.*

An alternative version might be:

H_3: *Republicans are expected to have cooler feelings towards Clinton than Democrats.*

And the null hypothesis is:

H_0: *There is no relationship between partisan identification and feelings towards Clinton.*

`Political Ideology`:

H_4: *As political ideology increases, respondents are expected to have cooler feelings towards Clinton.*

An alternative version might be:

H_4: *Conservatives are expected to have cooler feelings towards Clinton than liberals.*

And the null hypothesis is:

H_0: *There is no relationship between partisan ideology and feelings towards Clinton.*

`Health Care Law`:

H_5: *As support for Obamacare increases, respondents are expected to have warmer feelings towards Clinton.*

And the null hypothesis is:

H_0: *There is no relationship between support for Obamacare and feelings towards Clinton.*

`Economy`:

H_6: *As opinions on the state of the economy increase, respondents are expected to have warmer feelings towards Clinton.*

An alternative version might be:

H_6: *Respondents who believe the economy has gotten better are expected to have warmer feelings towards Clinton than respondents who believe the economy has gotten worse.*

And the null hypothesis is:

H_0: *There is no relationship between support for Obamacare and feelings towards Clinton.*

`Wall with Mexico`:

H_7: As support for a wall with Mexico increases, respondents are expected to have cooler feelings towards Clinton.

And the null hypothesis is:

H_0: There is no relationship between support for a wall with Mexico and feelings towards Clinton.

`Fighting ISIS`:

H_8: As support for sending US troops to fight ISIS increases, respondents are expected to have cooler feelings towards Clinton.

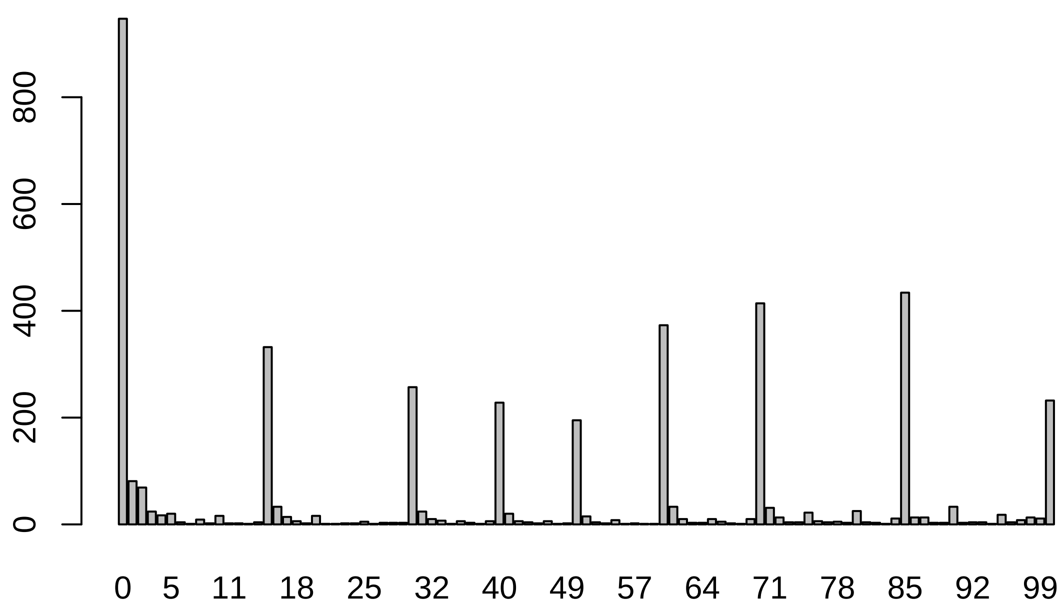
And the null hypothesis is:

H_0: There is no relationship between support for sending US troops to fight ISIS and feelings towards Clinton.

Question 3

`Feelings Towards Clinton`:

```
library(descr)
freq(ordered(nes$clinton))
```




```
ordered(nes$clinton)
      Frequency   Percent Valid Percent Cum Percent
0           947  22.17279      22.37184       22.37
1            81   1.89651       1.91354       24.29
```

2	69	1.61555	1.63005	25.92
3	24	0.56193	0.56697	26.48
4	17	0.39803	0.40161	26.88
5	20	0.46827	0.47248	27.36
6	4	0.09365	0.09450	27.45
7	1	0.02341	0.02362	27.47
8	9	0.21072	0.21262	27.69
9	2	0.04683	0.04725	27.73
10	16	0.37462	0.37798	28.11
11	2	0.04683	0.04725	28.16
12	2	0.04683	0.04725	28.21
13	1	0.02341	0.02362	28.23
14	4	0.09365	0.09450	28.33
15	332	7.77336	7.84314	36.17
16	33	0.77265	0.77959	36.95
17	14	0.32779	0.33073	37.28
18	6	0.14048	0.14174	37.42
19	2	0.04683	0.04725	37.47
20	16	0.37462	0.37798	37.85
21	1	0.02341	0.02362	37.87
22	1	0.02341	0.02362	37.89
23	2	0.04683	0.04725	37.94
24	2	0.04683	0.04725	37.99
25	5	0.11707	0.11812	38.11
26	1	0.02341	0.02362	38.13
27	3	0.07024	0.07087	38.20
28	3	0.07024	0.07087	38.27
29	3	0.07024	0.07087	38.34
30	257	6.01733	6.07134	44.41
31	24	0.56193	0.56697	44.98
32	10	0.23414	0.23624	45.22
33	7	0.16390	0.16537	45.38
34	1	0.02341	0.02362	45.41
35	6	0.14048	0.14174	45.55
36	3	0.07024	0.07087	45.62
38	1	0.02341	0.02362	45.64
39	6	0.14048	0.14174	45.78
40	228	5.33833	5.38625	51.17
41	20	0.46827	0.47248	51.64
42	6	0.14048	0.14174	51.78
43	4	0.09365	0.09450	51.88
44	2	0.04683	0.04725	51.93
45	6	0.14048	0.14174	52.07
46	1	0.02341	0.02362	52.09
49	2	0.04683	0.04725	52.14
50	195	4.56568	4.60666	56.74
51	15	0.35121	0.35436	57.10
52	4	0.09365	0.09450	57.19
53	2	0.04683	0.04725	57.24
55	8	0.18731	0.18899	57.43
56	1	0.02341	0.02362	57.45
57	2	0.04683	0.04725	57.50
58	1	0.02341	0.02362	57.52
59	1	0.02341	0.02362	57.55

```
60          373   8.73332       8.81172       66.36
61           33   0.77265       0.77959       67.14
62           10   0.23414       0.23624       67.38
63            3   0.07024       0.07087       67.45
64            3   0.07024       0.07087       67.52
65           10   0.23414       0.23624       67.75
66            5   0.11707       0.11812       67.87
67            2   0.04683       0.04725       67.92
68            1   0.02341       0.02362       67.94
69           10   0.23414       0.23624       68.18
70          414   9.69328       9.78030       77.96
71           31   0.72583       0.73234       78.69
72           13   0.30438       0.30711       79.00
73            4   0.09365       0.09450       79.09
74            4   0.09365       0.09450       79.19
75           22   0.51510       0.51973       79.71
76            6   0.14048       0.14174       79.85
77            4   0.09365       0.09450       79.94
78            5   0.11707       0.11812       80.06
79            3   0.07024       0.07087       80.13
80           25   0.58534       0.59060       80.72
81            4   0.09365       0.09450       80.82
82            3   0.07024       0.07087       80.89
83            1   0.02341       0.02362       80.91
84           11   0.25755       0.25986       81.17
85          434  10.16155      10.25278       91.42
86           13   0.30438       0.30711       91.73
87           13   0.30438       0.30711       92.04
88            3   0.07024       0.07087       92.11
89            3   0.07024       0.07087       92.18
90           33   0.77265       0.77959       92.96
91            3   0.07024       0.07087       93.03
92            4   0.09365       0.09450       93.13
93            4   0.09365       0.09450       93.22
94            1   0.02341       0.02362       93.24
95           18   0.42145       0.42523       93.67
96            4   0.09365       0.09450       93.76
97            8   0.18731       0.18899       93.95
98           13   0.30438       0.30711       94.26
99           11   0.25755       0.25986       94.52
100         232   5.43198       5.48075      100.00
NA's         38   0.88972
Total      4271 100.00000     100.00000
```

We see that the mode is 0 for feelings towards Clinton.

```
median(nes$clinton,na.rm=TRUE)
```

```
[1] 40
```

```
mean(nes$clinton,na.rm=TRUE)
```

```
[1] 42.15143
```

```
sd(nes$clinton,na.rm=TRUE)
```

```
[1] 34.22733
```

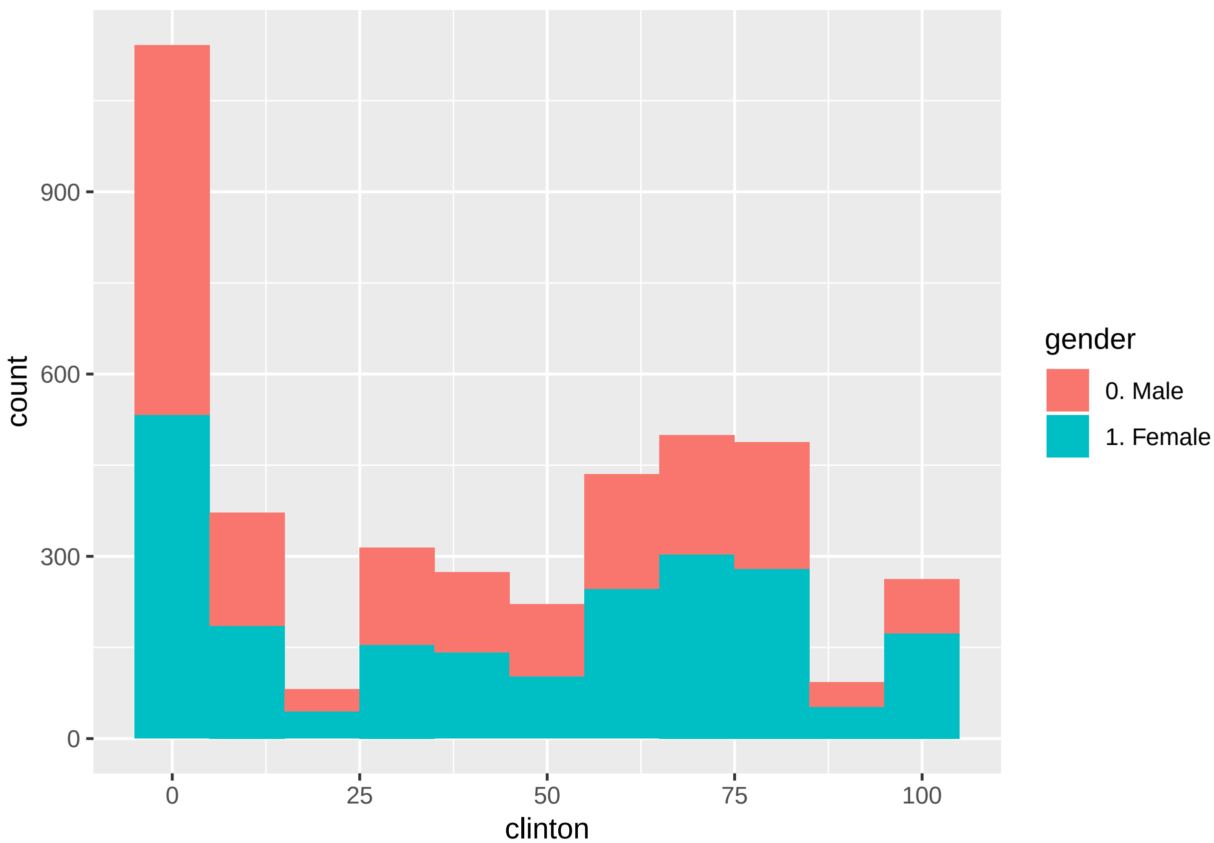
We see that the median is 40 and mean is 42.15. Based on those numbers only, we would conclude that respondents feel somewhat cool to neutral towards Clinton, but the standard deviation provides a better understanding. At 34.23, the standard deviation tells us there is a large spread of responses from the mean - that respondents are not clustered at the median or mean.

What can we conclude about feelings toward Clinton based on these descriptive statistics? While on the whole there are more respondents that feel cool than warm towards Clinton it is not overwhelmingly the case; roughly 43% of respondents have feelings above the neutral cut-point of 50. Compared to Trump, respondents have slightly warmer feelings towards Clinton.

Question 4

`Gender`

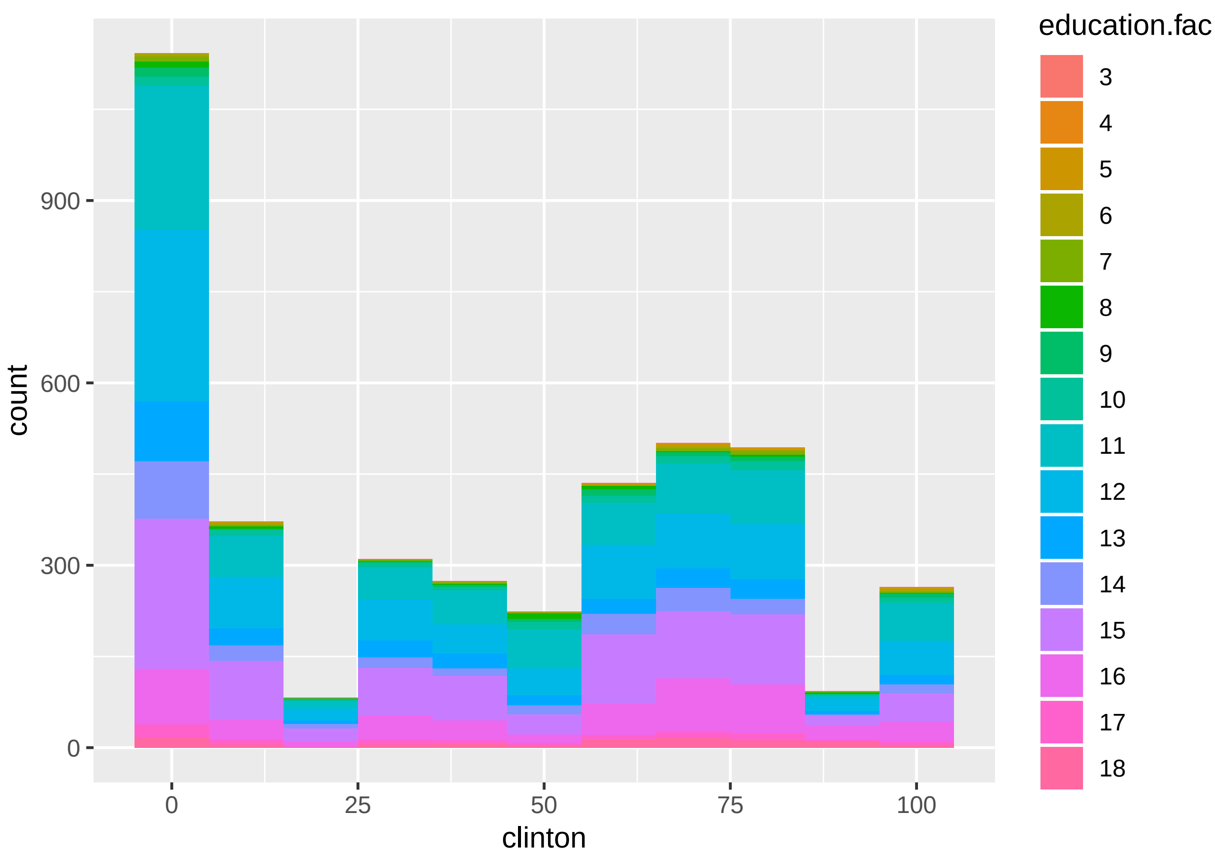
```
library(ggplot2)
nes2 <- subset(nes, !is.na(clinton) & !is.na(gender))

p1 <- ggplot(data=nes2) +
      geom_histogram(mapping=aes(clinton, fill=gender), binwidth=10)

x11()
p1
```



We see that roughly an equal number of men and women feel 0 - or very cold - towards Clinton. However, women make up larger portions of the warm and hot feelings towards Clinton than men. Therefore, it appears that women generally feel warmer to Clinton than men.

`Education`:

```r
nes$education.fac <- as.factor(nes$education)

nes2 <- subset(nes, !is.na(clinton) & !is.na(education.fac))

p2 <- ggplot(data=nes2) +
      geom_histogram(mapping=aes(clinton, fill=education.fac), binwidth=10)

x11()
p2
```




```r
table(nes$education)
```

```
  3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18
  1   3  15  22  32  40  62 107 810 899 313 288 955 499  88  93
```

```r
nes$education.fac <- cut(nes$education,
                         breaks=c(-Inf, 10, 11, 14, 15, Inf),
                         labels=c("Less than High School","High School Degree",
                                  "Some College", "Bachelor's Degree",
                                  "Graduate Degree"))
table(nes$education.fac)
```

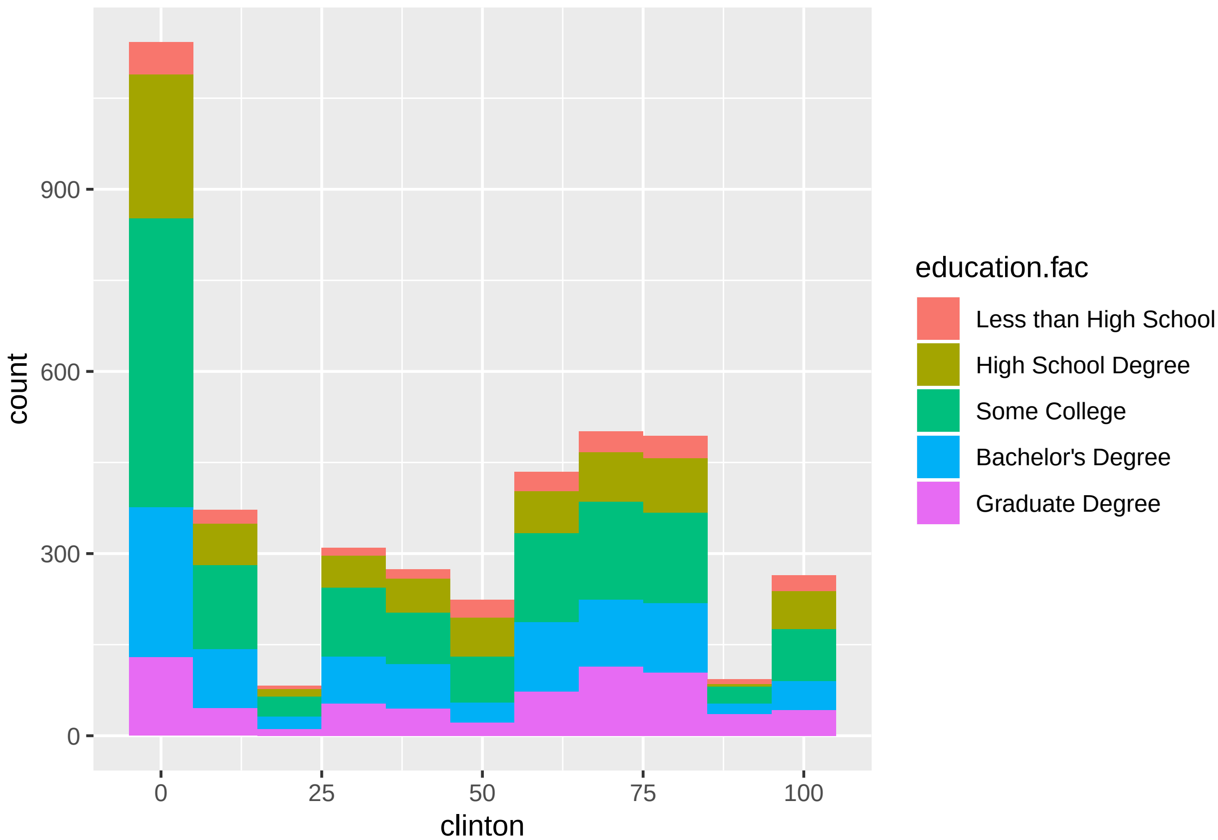
```
Less than High School    High School Degree          Some College
                  282                   810                  1500
    Bachelor's Degree       Graduate Degree
```

955 680

```
class(nes$education.fac)
```

```
[1] "factor"
```

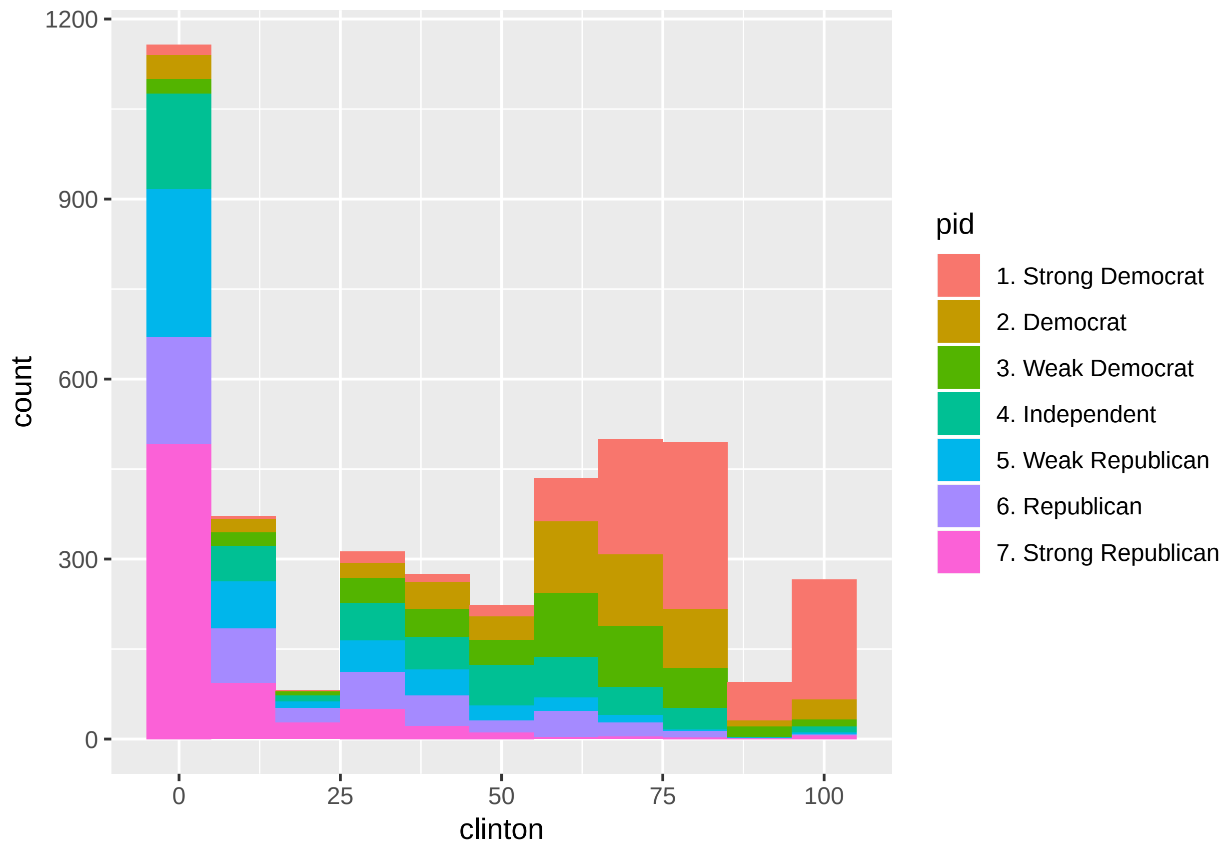
```
nes2 <- subset(nes, !is.na(clinton) & !is.na(education.fac))

p2 <- ggplot(data=nes2) +
      geom_histogram(mapping=aes(clinton, fill=education.fac), binwidth=10)

x11()
p2
```



There does not appear to be much of a difference between education levels and feelings towards Clinton. Each education level makes up a similar portion of each bar. We can therefore conclude that education does not appear to matter for explaining feelings towards Clinton.

`Partisan Identification:`

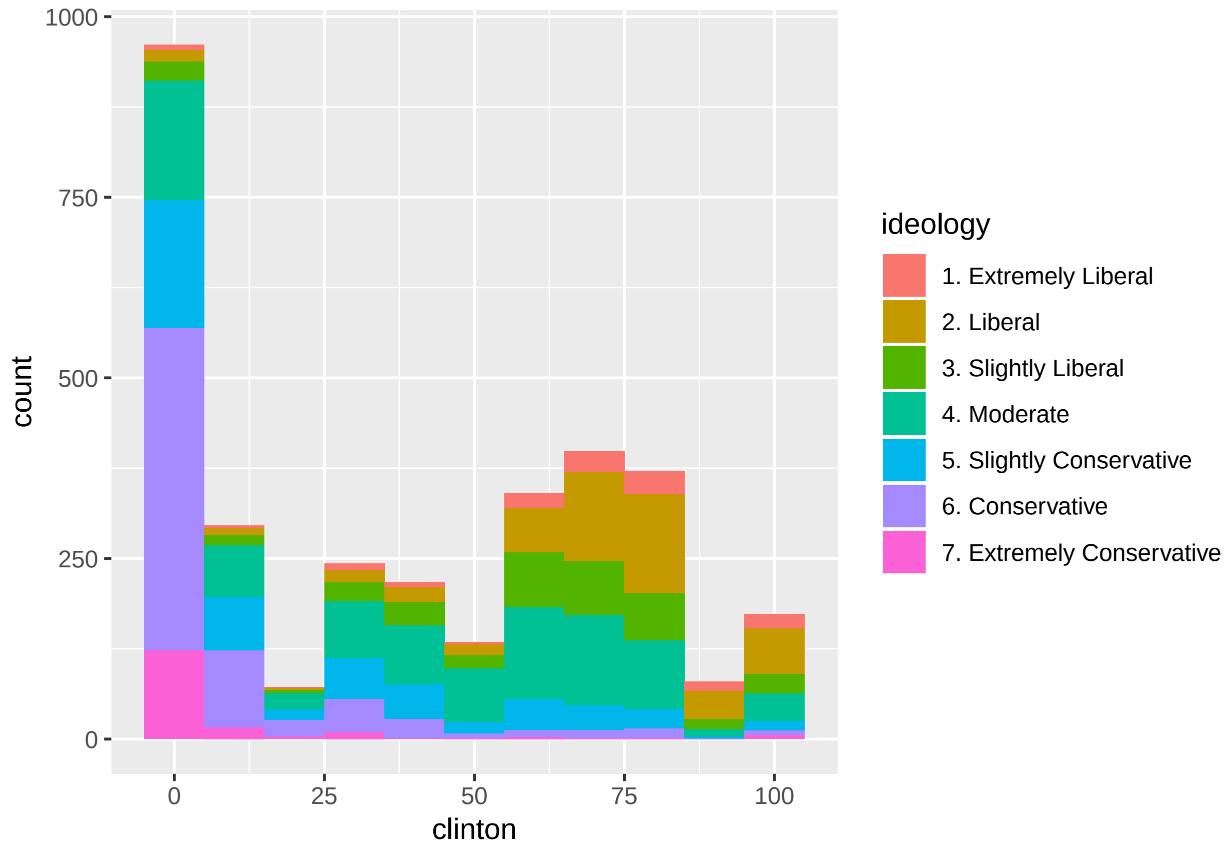
```
nes2 <- subset(nes, !is.na(clinton) & !is.na(pid))

p3 <- ggplot(data=nes2) +
      geom_histogram(mapping=aes(clinton, fill=pid), binwidth=10)

x11()
p3
```



As we would expect, strong Republicans feel the coldest towards Clinton and strong Democrats feel the warmest towards Clinton.

`Political Ideology`:

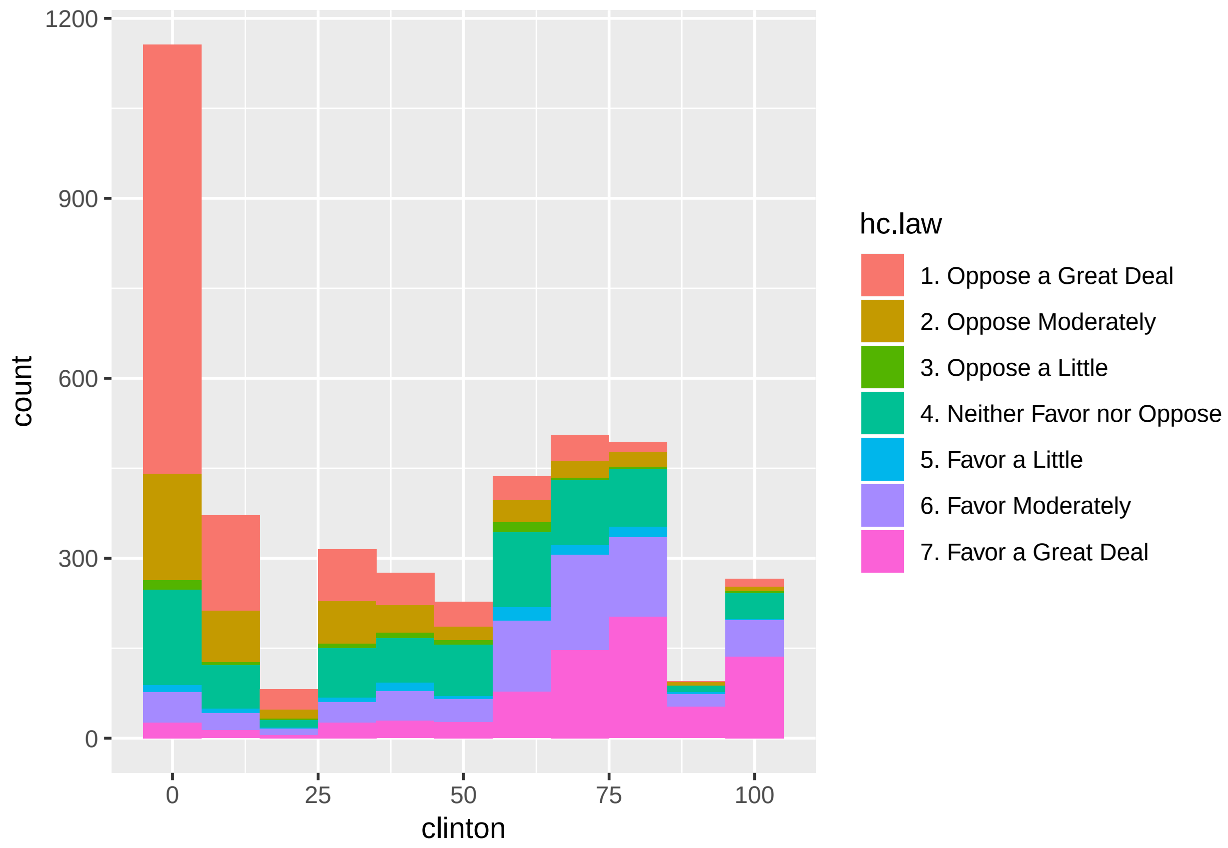
```
nes2 <- subset(nes, !is.na(clinton) & !is.na(ideology))

p4 <- ggplot(data=nes2) +
      geom_histogram(mapping=aes(clinton, fill=ideology), binwidth=10)

x11()
p4
```



Although not as strong as with partisan identification, conservatives feel the coldest towards Clinton while liberals feel the warmest towards Clinton.

`Health Care Law:`

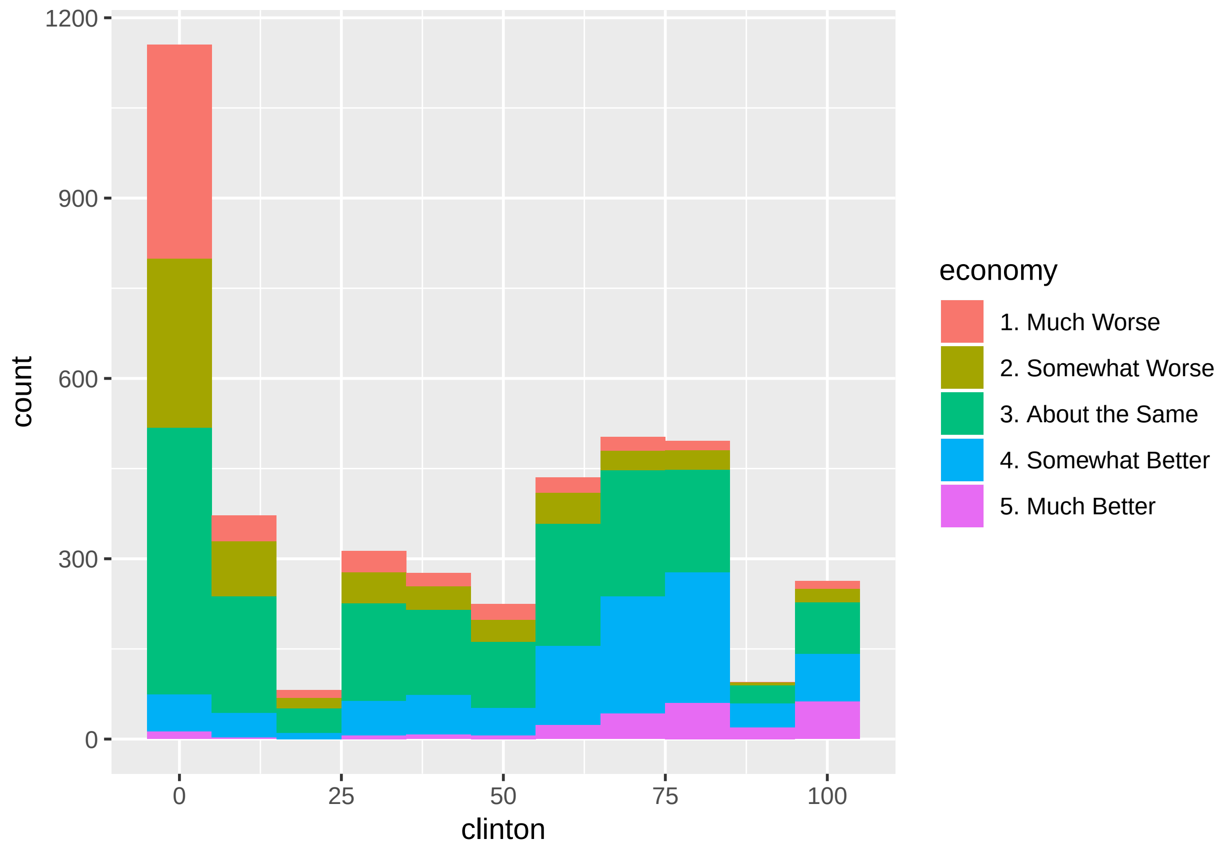
```
nes2 <- subset(nes, !is.na(clinton) & !is.na(hc.law))

p5 <- ggplot(data=nes2) +
      geom_histogram(mapping=aes(clinton, fill=hc.law), binwidth=10)

x11()
p5
```



We see that those who opposed Obamacare feel the coldest towards Clinton while those who favoured Obamacare feel the warmest towards Clinton.

`Economy`:

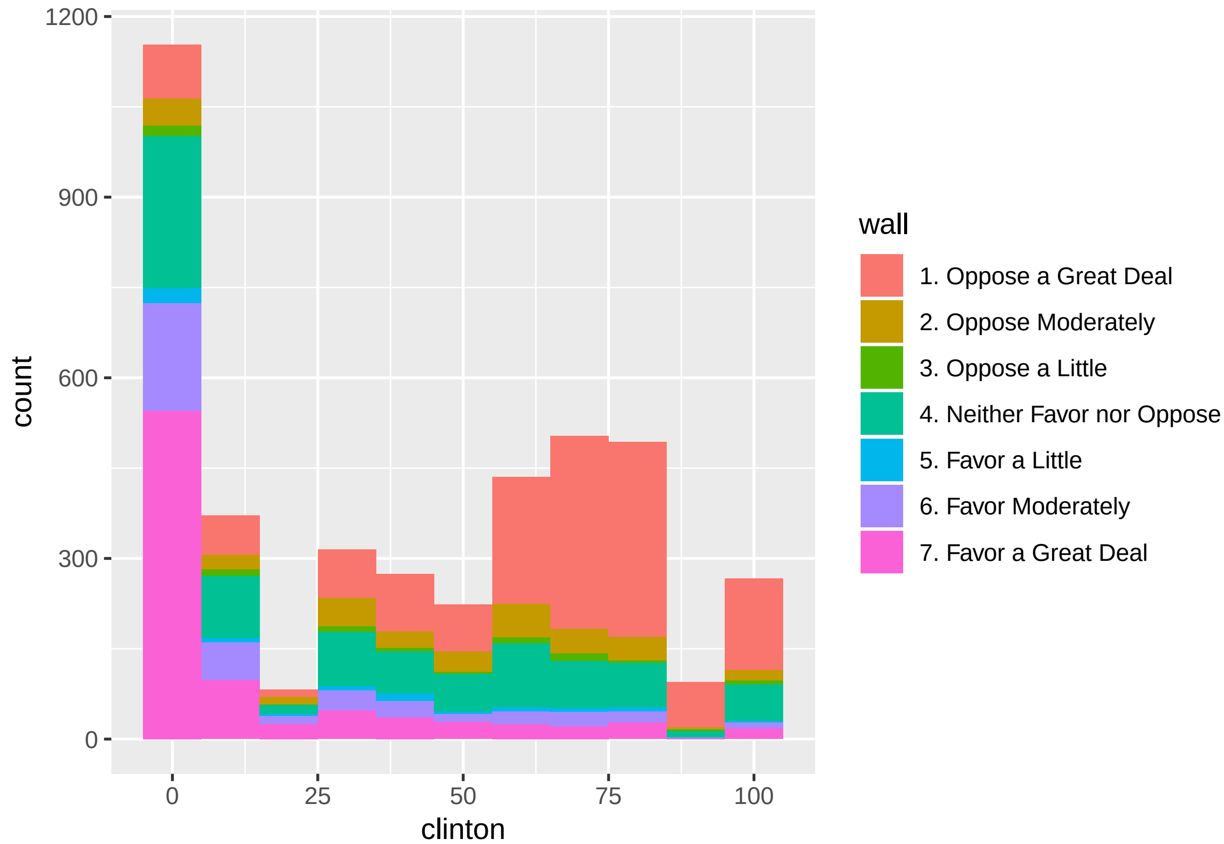
```
nes2 <- subset(nes, !is.na(clinton) & !is.na(economy))

p6 <- ggplot(data=nes2) +
      geom_histogram(mapping=aes(clinton, fill=economy), binwidth=10)

x11()
p6
```



It does not appear that views on the state of the economy matter much for explaining feelings towards Clinton. Certainly, respondents who believed the economy had gotten much worse had cold feelings towards Clinton, but the portions are similar across the bars in the histogram.

`Wall with Mexico`:

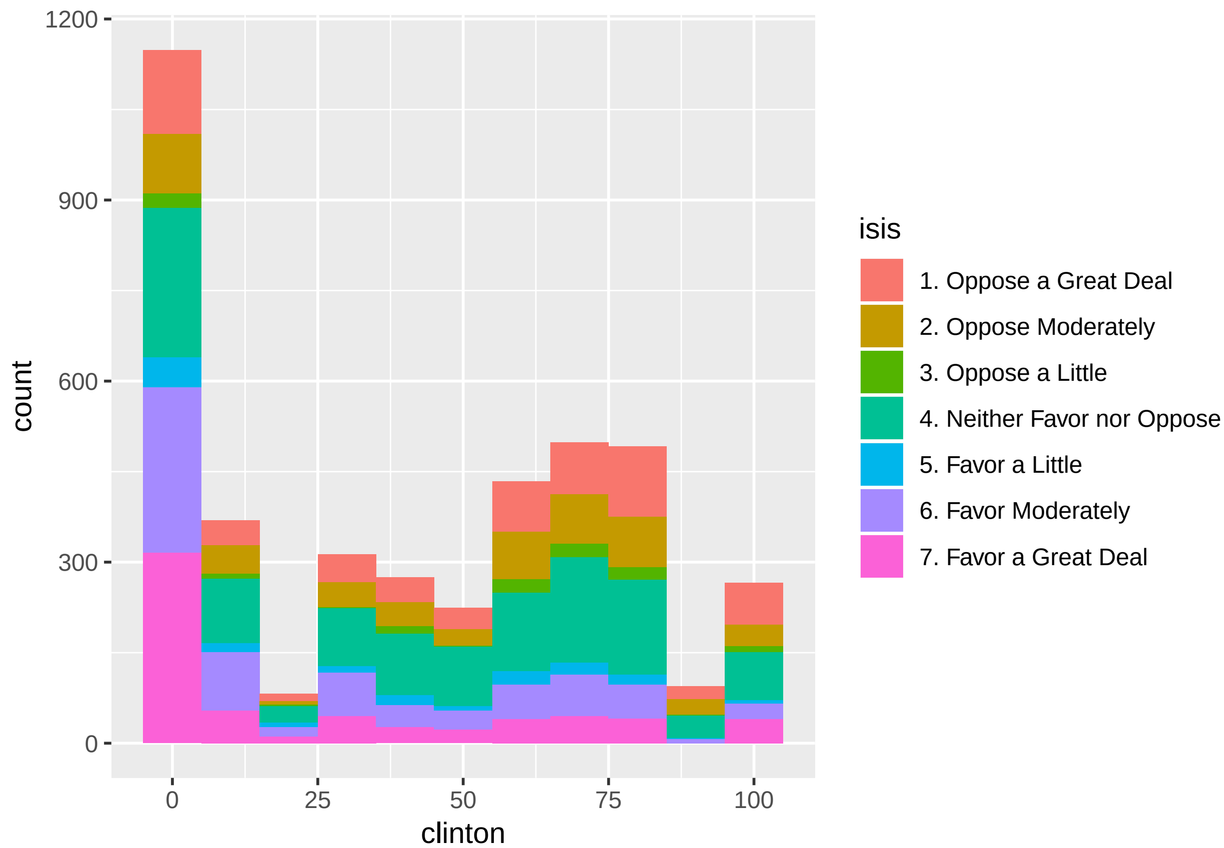
```
nes2 <- subset(nes, !is.na(clinton) & !is.na(wall))

p7 <- ggplot(data=nes2) +
      geom_histogram(mapping=aes(clinton, fill=wall), binwidth=10)

x11()
p7
```



Respondents' opinions on building a wall with Mexico do appear to matter here. Respondents who greatly favoured the wall predominately have the coldest feelings towards Clinton and vice-versa.

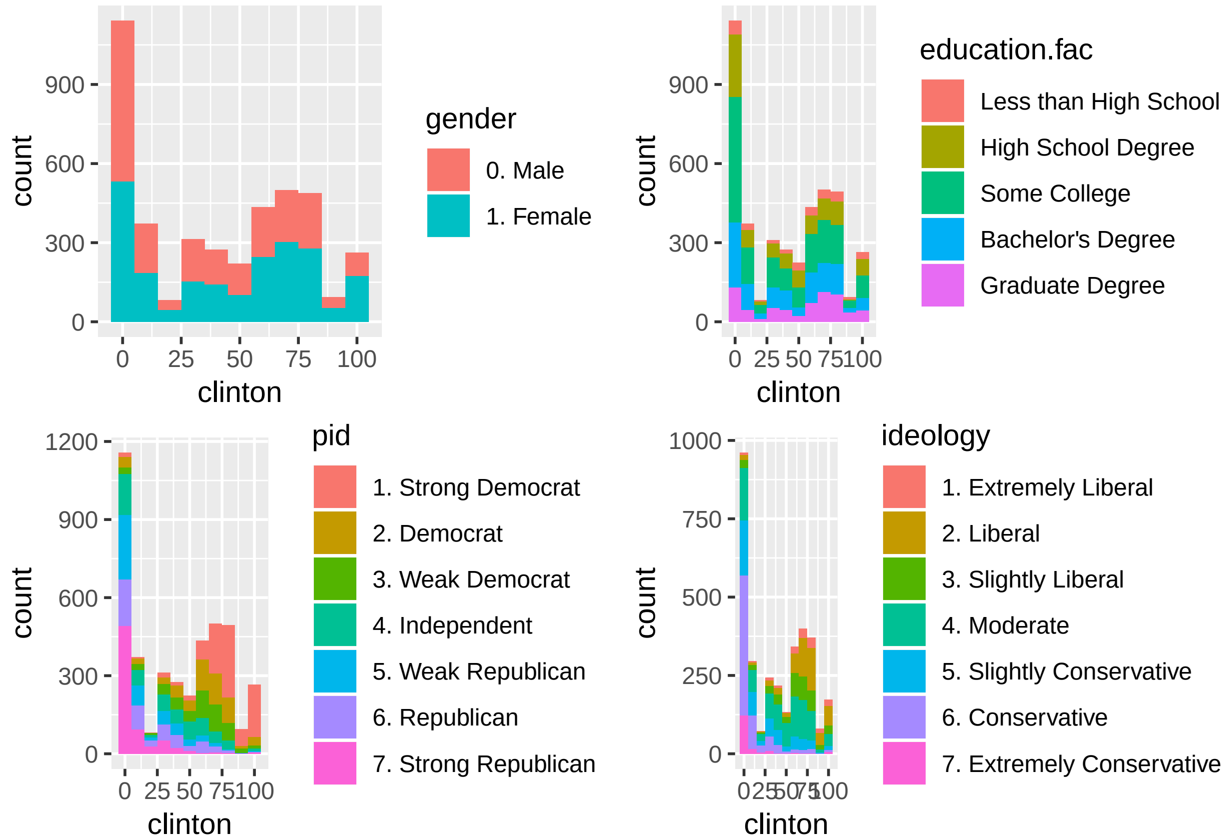
`Fighting ISIS`:

```
nes2 <- subset(nes, !is.na(clinton) & !is.na(isis))

p8 <- ggplot(data=nes2) +
      geom_histogram(mapping=aes(clinton, fill=isis), binwidth=10)

x11()
p8
```

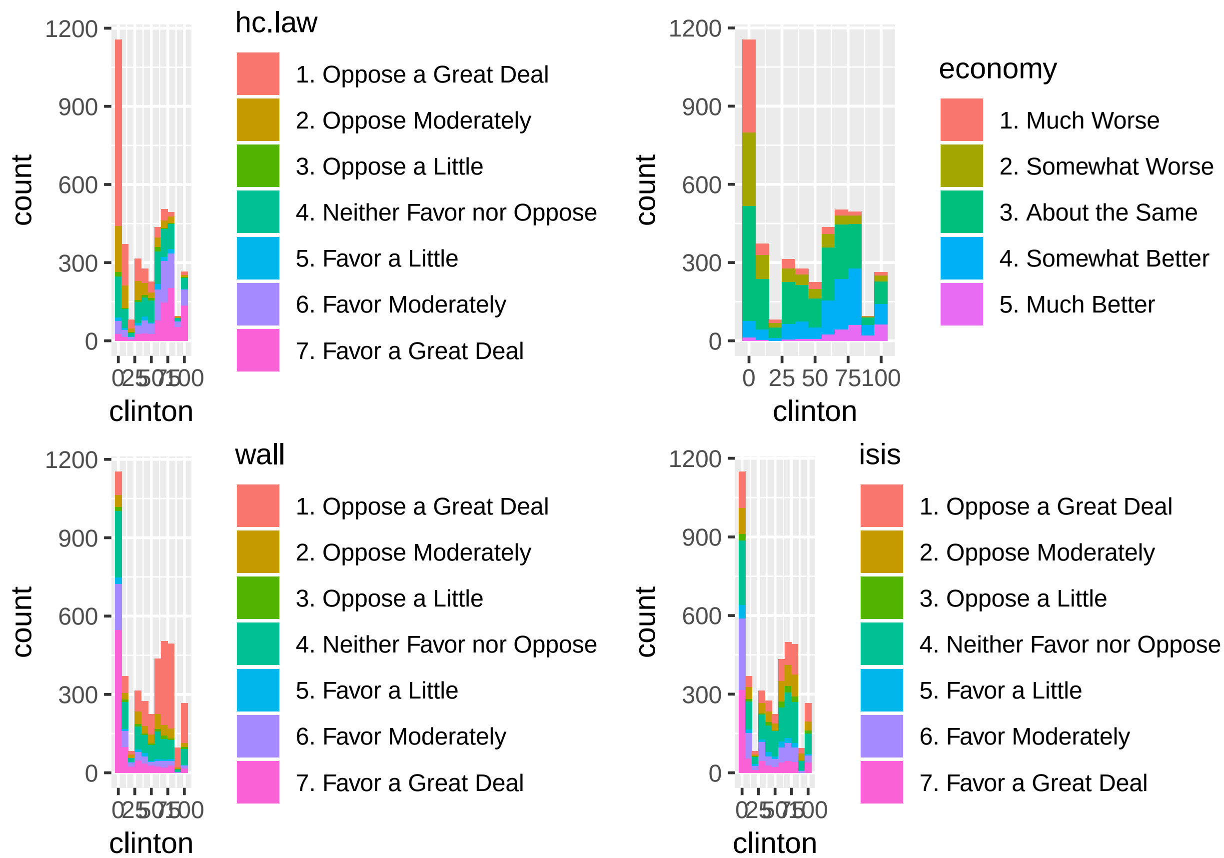


As with Trump, this histogram is somewhat of a mixed picture. However, it does appear that respondents who strongly favoured sending US troops to fight ISIS generally have cold feelings towards Clinton and vice-versa.

```
library(gridExtra)
x11()
grid.arrange(p1, p2, p3, p4, ncol=2, nrow=2)
```




```
x11()
grid.arrange(p5, p6, p7, p8, ncol=2, nrow=2)
```



Question 5

```r
t.test(clinton~gender, data=nes)
```

```

	Welch Two Sample t-test

data:  clinton by gender
t = -6.9894, df = 4143.3, p-value = 3.203e-12
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -9.409579 -5.287143
sample estimates:
  mean in group 0. Male mean in group 1. Female 
               38.17895               45.52731 
```

Since $p \leq .05$, we conclude that there is a statistically significant difference between men and women's feelings towards Clinton. Further, we see that men's mean rating of Clinton is 38.18, while women's mean rating is 45.53. Let's see whether men's feelings toward Clinton are significantly smaller than women's feelings toward Clinton by adding the option `alternative="less"`.

```r
t.test(clinton~gender, alternative="less", data=nes)
```

```

	Welch Two Sample t-test
```

```
data:  clinton by gender
t = -6.9894, df = 4143.3, p-value = 1.601e-12
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
      -Inf -5.618651
sample estimates:
  mean in group 0. Male mean in group 1. Female
               38.17895                45.52731
```

Based on the p-value, we do find that men's ratings of Clinton are significantly smaller than women's ratings. Next, let's see whether these differences hold in a non-parametric environment.

Question 6

```r
wilcox.test(clinton~gender, data=nes)
```

```
    Wilcoxon rank sum test with continuity correction

data:  clinton by gender
W = 1904200, p-value = 1.343e-12
alternative hypothesis: true location shift is not equal to 0
```

Since $p \leq .05$, we do find a significant difference between men and women's feelings towards Clinton. Let's see if these results are the same when we consider the directional test.

```r
wilcox.test(clinton~gender, alternative="less", data=nes)
```

```
    Wilcoxon rank sum test with continuity correction

data:  clinton by gender
W = 1904200, p-value = 6.714e-13
alternative hypothesis: true location shift is less than 0
```

Yes, according to the p-value, there is a significant difference between men and women.

Question 7

```r
nes2 <- na.omit(nes)

nes2 <- subset(nes2, select=c(clinton:education.fac))
```

```r
summary(model.1 <- lm(clinton ~ gender + education + pid.num + ideology.num, data=nes2))
```

```
Call:
lm(formula = clinton ~ gender + education + pid.num + ideology.num,
    data = nes2)

Residuals:
    Min      1Q  Median      3Q     Max
-76.044 -12.818  -1.375  14.183  91.210
```

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)      90.5794     4.5749  19.799  < 2e-16 ***
gender1. Female   1.0264     1.2756   0.805    0.421    
education         0.2645     0.2859   0.925    0.355    
pid.num         -10.0759     0.4103 -24.558  < 2e-16 ***
ideology.num     -2.8864     0.5686  -5.076 4.45e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 22.12 on 1208 degrees of freedom
Multiple R-squared:  0.5681,    Adjusted R-squared:  0.5666 
F-statistic: 397.2 on 4 and 1208 DF,  p-value: < 2.2e-16
```

```r
confint(model.1, level=.95)
```

```
                      2.5 %    97.5 %
(Intercept)      81.6038047 99.555082
gender1. Female  -1.4763138  3.529139
education        -0.2965445  0.825475
pid.num         -10.8808514 -9.270932
ideology.num     -4.0018986 -1.770850
```

We see that the overall model is statistically significant and the adjusted R^2 is .57 indicating that our model explains roughly 57% of the variance in feelings towards Clinton. Gender and education are not significant but the other two predictors are statistically significant. Partisan identification and political ideology both have negative coefficients indicating that as respondents' become more Republican and more conservative their feelings towards Clinton decrease. And we see our significance results are supported by the 95% confidence intervals.

Question 8

```r
summary(model.2 <- lm(clinton ~ gender + education + pid.num + ideology.num +
                        hc.law.num + economy.num + wall.num + isis.num, data=nes2))
```

```
Call:
lm(formula = clinton ~ gender + education + pid.num + ideology.num + 
    hc.law.num + economy.num + wall.num + isis.num, data = nes2)

Residuals:
    Min      1Q  Median      3Q     Max 
-74.255 -12.017  -0.559  12.498 100.638 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)      56.3224     5.2530  10.722  < 2e-16 ***
gender1. Female   2.9070     1.1781   2.467   0.0137 *  
education        -0.3011     0.2667  -1.129   0.2591    
pid.num          -6.9608     0.4302 -16.181  < 2e-16 ***
ideology.num     -0.4538     0.5474  -0.829   0.4073    
hc.law.num        3.5386     0.3168  11.169  < 2e-16 ***
economy.num       3.8651     0.6743   5.732 1.25e-08 ***
wall.num         -1.6545     0.3109  -5.321 1.23e-07 ***
```

```
isis.num         -0.2914     0.3014  -0.967   0.3337
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 20.22 on 1204 degrees of freedom
Multiple R-squared:  0.6402,    Adjusted R-squared:  0.6379
F-statistic: 267.8 on 8 and 1204 DF,  p-value: < 2.2e-16
```

```
confint(model.2, level=.95)
```

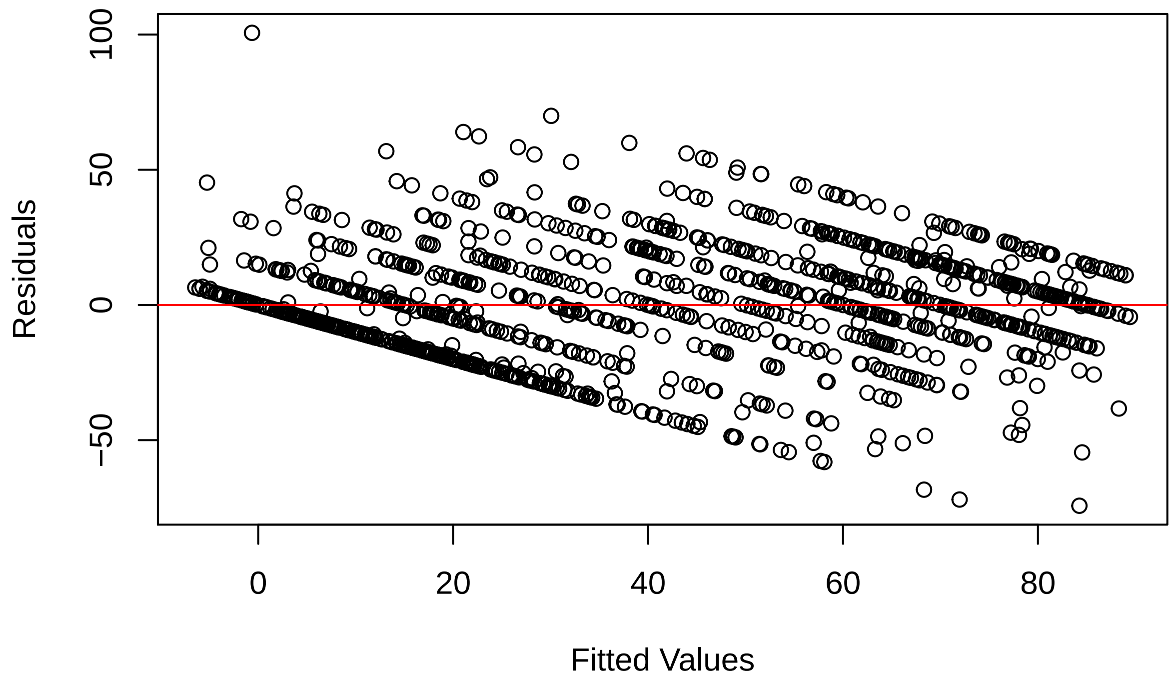
```
                     2.5 %     97.5 %
(Intercept)     46.0163685 66.6285020
gender1. Female  0.5955658  5.2183826
education       -0.8243674  0.2221068
pid.num         -7.8047580 -6.1167445
ideology.num    -1.5277532  0.6202416
hc.law.num       2.9169790  4.1601661
economy.num      2.5422703  5.1880177
wall.num        -2.2644603 -1.0444911
isis.num        -0.8827097  0.2998143
```

The overall model is again statistically significant and the adjusted R^2 is now .64 indicating that our model explains roughly 64% of the variance in feelings towards Clinton. We now find that gender is significant, but political ideology is not, when controlling for the four additional predictors. The positive coefficient for gender implies that women have warmer feelings toward Clinton than men. The coefficients for Obamacare and the economy are both positive indicating that as respondents' support for the health care law and positive evaluation of the economy increases their feelings towards Clinton increase. The negative coefficient for the wall with Mexico suggests that as respondents' support of building a wall increases their feelings for Clinton decrease.

Question 9

9.1 Functional Form

```
x11()
plot(y=model.2$residuals,x=model.2$fitted.values, xlab="Fitted Values", ylab="Residuals")
abline(h=0, col="red")
```



It does not look like the local means are 0, thus we may have violated functional form.

```
library(lmtest)
```

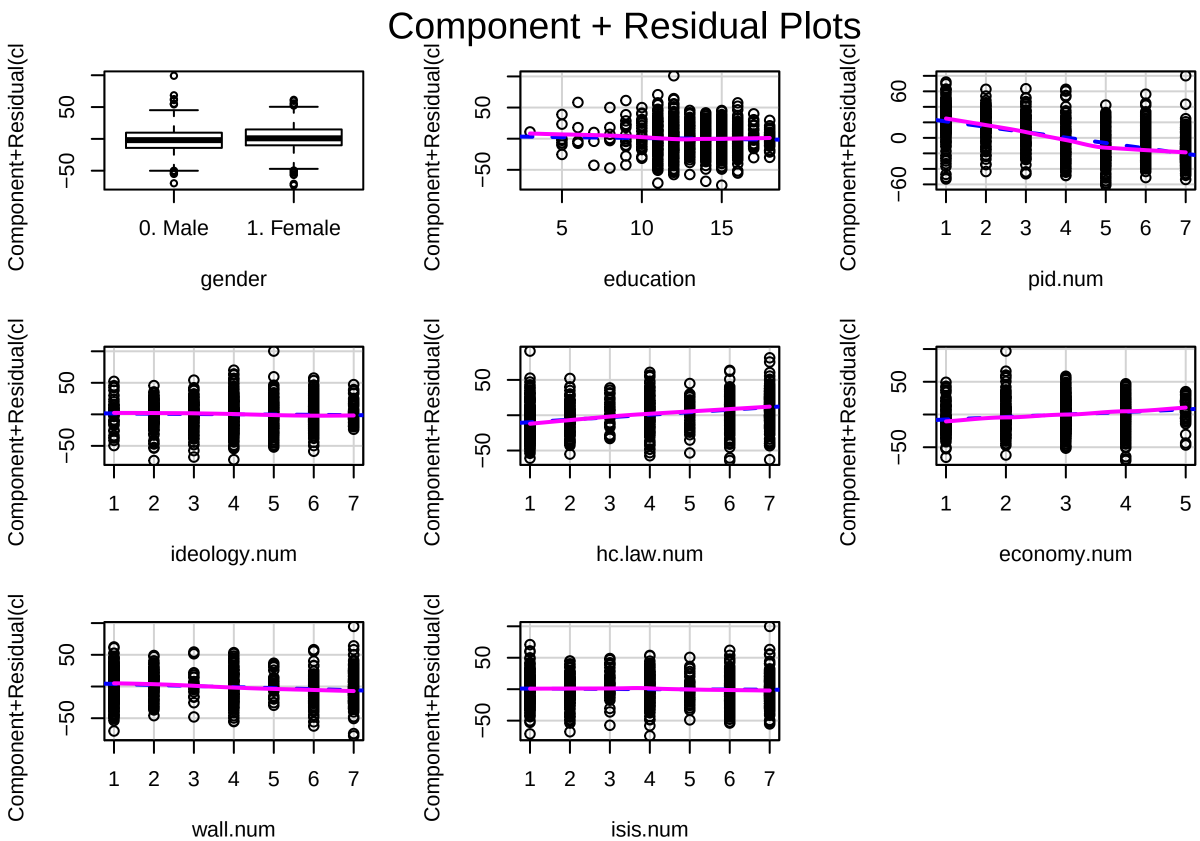
```
Loading required package: zoo


Attaching package: 'zoo'

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

    as.Date, as.Date.numeric
```

```
resettest(model.2, power=2:3, type="fitted")
```

```

    RESET test

data:  model.2
RESET = 9.5918, df1 = 2, df2 = 1202, p-value = 7.366e-05
```

We find that $p \leq .05$, thus rejecting the null and concluding we have violated the functional form assumptions.

```
x11()
crPlots(model.2)
```



It appears that education and partisan identification are the least linearly related to our outcome variable.

Let's use a Box-Tidwell test to check education and partisan identification.

We need to create a new Clinton variable that does not have 0 and we also need to include `max.iter=20000` for the test to converge.

```
nes2$clinton2 <- nes2$clinton +1

boxTidwell(clinton2 ~ education + pid.num, data=nes2, na.action=na.exclude,
           max.iter=20000)
```

```
Warning in boxTidwell.default(y, X1, X2, max.iter = max.iter, tol = tol, :
maximum iterations exceeded

          MLE of lambda Score Statistic (z) Pr(>|z|)
education        0.9584              1.8484 0.064547 .
pid.num          0.7435              2.6679 0.007633 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

iterations =  20001
```

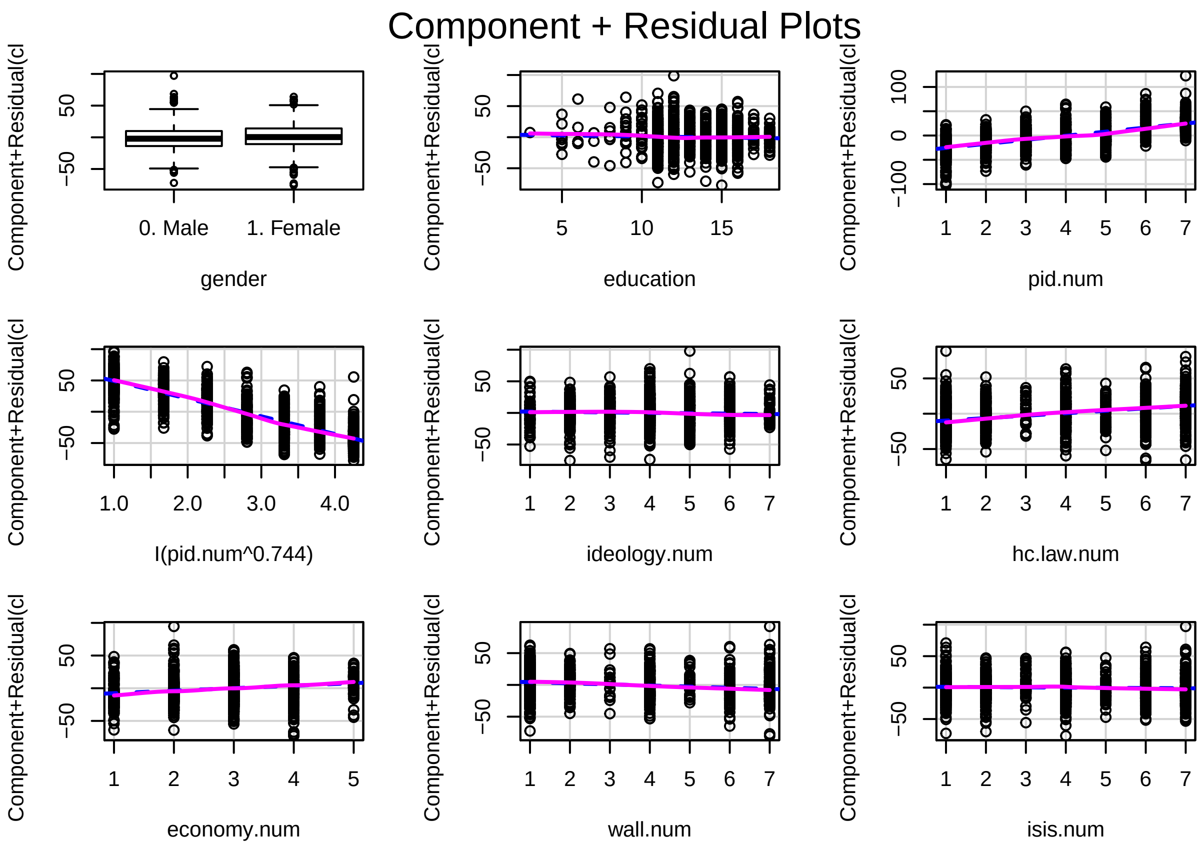
We see that we do not need to transform `education`, but we do need to transform `pid.num`. We need to raise `pid.num` by .744.

```
summary(model.2a <- lm(clinton ~ gender + education + pid.num + I(pid.num^.744) +
                        ideology.num + hc.law.num + economy.num + wall.num + isis.num,
                        data=nes2))
```

```
Call:
lm(formula = clinton ~ gender + education + pid.num + I(pid.num^0.744) +
    ideology.num + hc.law.num + economy.num + wall.num + isis.num,
    data = nes2)

Residuals:
    Min      1Q  Median      3Q     Max
-76.782 -12.465  -0.866  12.146  98.318

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       74.2692     7.3020  10.171  < 2e-16 ***
gender1. Female    2.4296     1.1804   2.058 0.039780 *
education         -0.3419     0.2657  -1.287 0.198390
pid.num            8.3823     4.3789   1.914 0.055826 .
I(pid.num^0.744) -28.2552     8.0254  -3.521 0.000446 ***
ideology.num      -0.6361     0.5473  -1.162 0.245350
hc.law.num         3.4650     0.3160  10.964  < 2e-16 ***
economy.num        3.8339     0.6712   5.712 1.40e-08 ***
wall.num          -1.7654     0.3110  -5.676 1.73e-08 ***
isis.num          -0.3671     0.3007  -1.221 0.222462
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 20.12 on 1203 degrees of freedom
Multiple R-squared:  0.6439,    Adjusted R-squared:  0.6412
F-statistic: 241.7 on 9 and 1203 DF,  p-value: < 2.2e-16
```

The transformed partisan identification is significant, but the coefficient is very strange. This indicates that we should be suspicious of using this model.

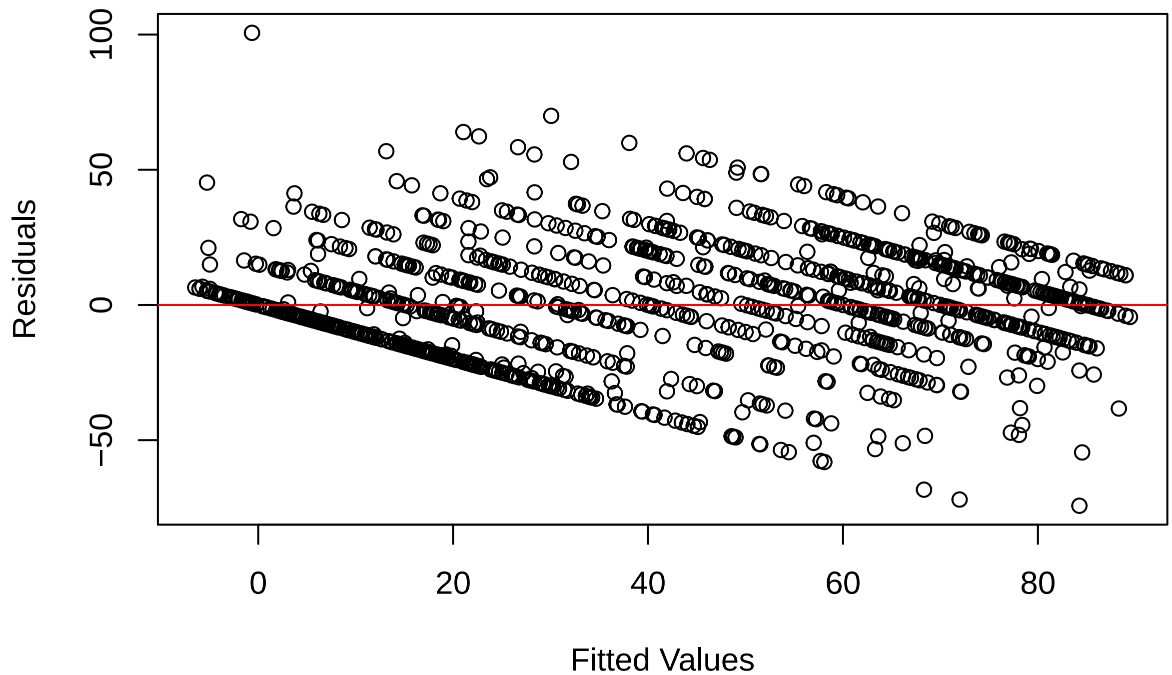
```r
x11()
crPlots(model.2a)
```



However, we do see that partisan identification and the transformed version now look close to linear.

9.2 Heteroscedasticity

```
x11()
plot(y=model.2$residuals,x=model.2$fitted.values, xlab="Fitted Values", ylab="Residuals")
abline(h=0, col="red")
```



As you may have concluded when we first created the plot, there appears to be heteroscedasticity present; the downwards-slanting residuals clearly make for a pattern.

```
bptest(model.2, studentize=FALSE)
```

```

	Breusch-Pagan test

data:  model.2
BP = 45.018, df = 8, p-value = 3.651e-07
```

Since $p \leq .05$, we reject the null of constant error variance and conclude that we have heteroscedasticity. To deal with heteroscedasticity, we will re-run our model with robust standard errors using the `coeftest()` function from the `sandwich` library.

```
library(sandwich)
coeftest(model.2, vcov = vcovHC)
```

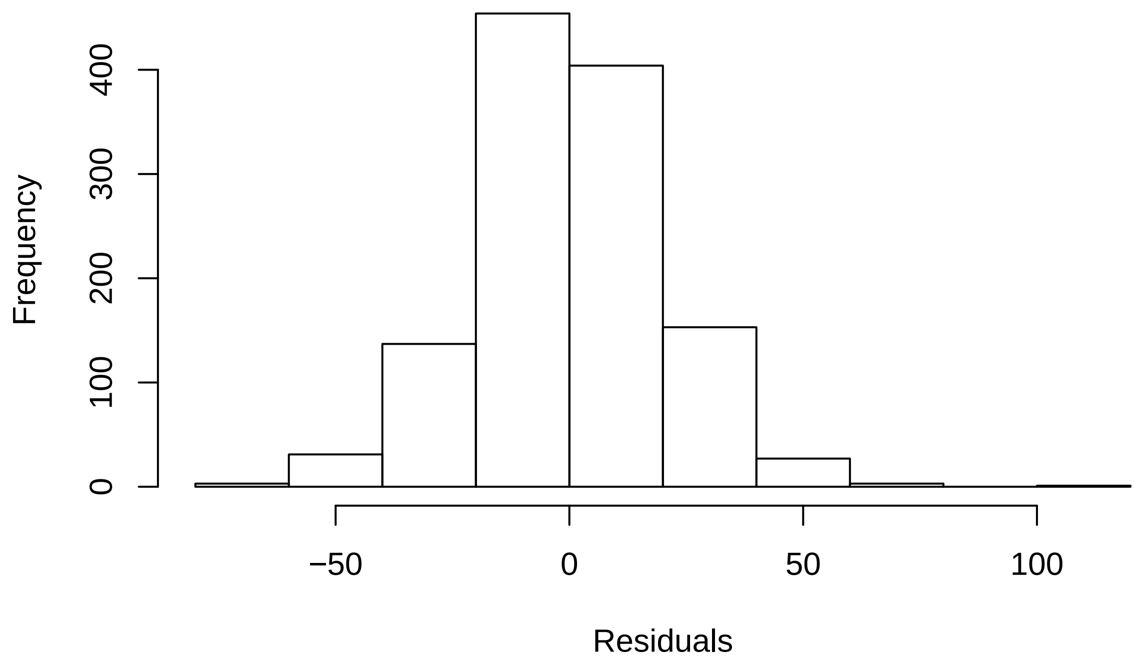
```

t test of coefficients:

                Estimate Std. Error  t value  Pr(>|t|)    
(Intercept)     56.32244    5.70546   9.8717 < 2.2e-16 ***
gender1. Female  2.90697    1.17780   2.4681   0.01372 *  
education       -0.30113    0.27013  -1.1148   0.26517    
pid.num         -6.96075    0.51496 -13.5170 < 2.2e-16 ***
ideology.num    -0.45376    0.62835  -0.7221   0.47035    
hc.law.num       3.53857    0.37419   9.4566 < 2.2e-16 ***
economy.num      3.86514    0.70384   5.4915 4.857e-08 ***
```

```
wall.num        -1.65448    0.36514  -4.5311 6.451e-06 ***
isis.num        -0.29145    0.32757  -0.8897   0.37380
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We find the same significance results as we did before.

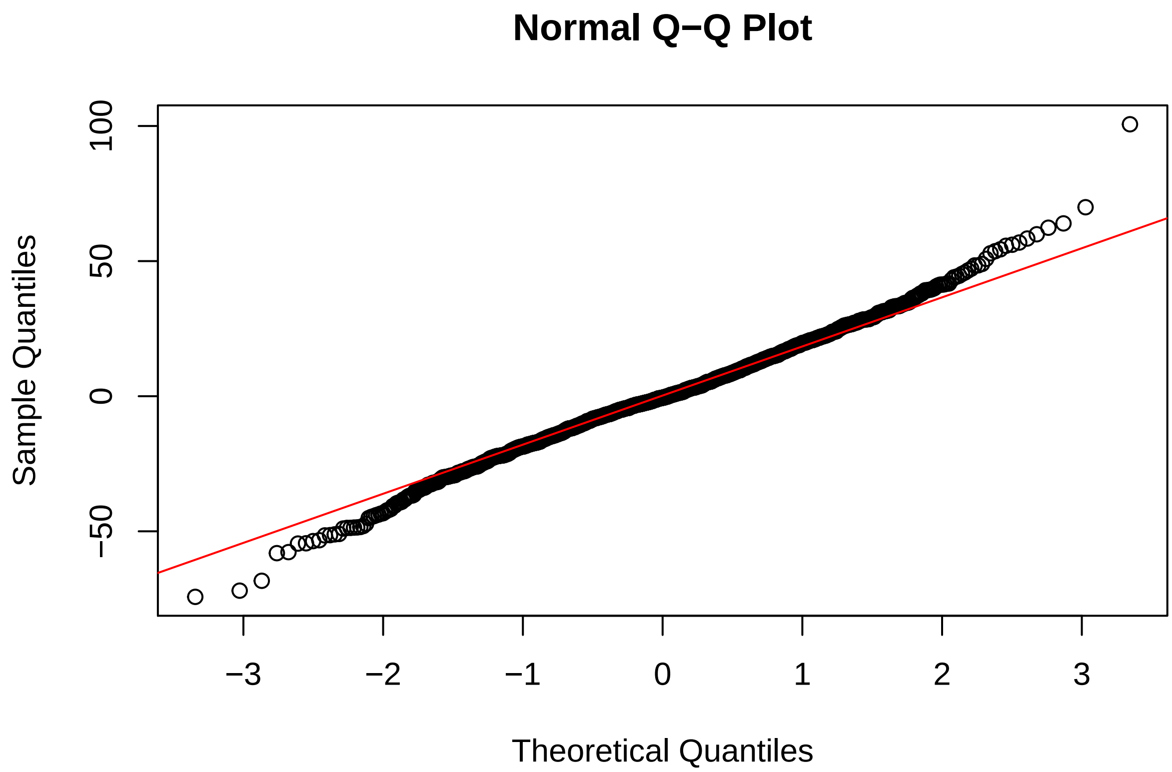
9.3 Normality

```r
x11()
hist(model.2$residuals,xlab="Residuals",main="")
```



The residuals are somewhat normally distributed, but there are long tails. This suggests we may have violated the normality assumption.

```r
x11()
qqnorm(model.2$residuals)
qqline(model.2$residuals,col="red")
```



As with the histogram, the residuals appear somewhat normally distributed, but a large portion of the tails are off the line.

```
shapiro.test(model.2$residuals)
```

```

    Shapiro-Wilk normality test

data:  model.2$residuals
W = 0.99305, p-value = 1.849e-05
```

We find that the p-value is below .05 and thus we reject the null that our residuals are normally distributed.

```
summary(powerTransform(nes2$clinton2))
```

```
bcPower Transformation to Normality
              Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
nes2$clinton2    0.4346        0.43       0.3893       0.4798

Likelihood ratio test that transformation parameter is equal to 0
 (log transformation)
                           LRT df       pval
LR test, lambda = (0) 382.3138  1 < 2.22e-16

Likelihood ratio test that no transformation is needed
                          LRT df       pval
LR test, lambda = (1) 502.906  1 < 2.22e-16
```

The test suggests that we should transform our outcome variable by raising the variable to .4346.

```r
nes2$clinton3 <- (nes2$clinton2)^.4346
```

```r
summary(model.2a <- lm(clinton3 ~ gender + education + pid.num + ideology.num +
                        hc.law.num + economy.num + wall.num + isis.num, data=nes2))
```

```
Call:
lm(formula = clinton3 ~ gender + education + pid.num + ideology.num +
    hc.law.num + economy.num + wall.num + isis.num, data = nes2)

Residuals:
    Min      1Q  Median      3Q     Max
-5.0935 -0.9420  0.0402  0.9284  5.8507

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      5.21546    0.37995  13.727  < 2e-16 ***
gender1. Female  0.16094    0.08521   1.889   0.0592 .
education       -0.01786    0.01929  -0.926   0.3548
pid.num         -0.38343    0.03112 -12.323  < 2e-16 ***
ideology.num    -0.06682    0.03959  -1.688   0.0917 .
hc.law.num       0.24551    0.02292  10.713  < 2e-16 ***
economy.num      0.30075    0.04877   6.167 9.51e-10 ***
wall.num        -0.15555    0.02249  -6.917 7.47e-12 ***
isis.num        -0.02291    0.02180  -1.051   0.2934
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.462 on 1204 degrees of freedom
Multiple R-squared:  0.6044,    Adjusted R-squared:  0.6018
F-statistic:   230 on 8 and 1204 DF,  p-value: < 2.2e-16
```

```r
confint(model.2a, level=.95)
```

```
                       2.5 %      97.5 %
(Intercept)      4.470016963  5.96089861
gender1. Female -0.006240204  0.32812949
education       -0.055703512  0.01998827
pid.num         -0.444476912 -0.32238240
ideology.num    -0.144505730  0.01085936
hc.law.num       0.200550807  0.29047089
economy.num      0.205061320  0.39642900
wall.num        -0.199674432 -0.11143370
isis.num        -0.065676835  0.01985547
```

In this model, we now find that gender is not a significant predictor at the .05-level. As we previously discussed, transforming the outcome variable in a non-intuitive way makes it difficult to interpret the coefficients. Therefore, we may be better off leaving the outcome variable in its original form.

9.4 Multicollinearity

```r
nes2$gender.num <- as.numeric(nes2$gender)
data <- data.frame(nes2$gender.num, nes2$education, nes2$pid.num,
                   nes2$ideology.num, nes2$hc.law.num, nes2$economy.num,
```

```r
                    nes2$wall.num, nes2$isis.num)
head(data)
```

```
  nes2.gender.num nes2.education nes2.pid.num nes2.ideology.num
1               1             15            6                 5
2               2             11            3                 4
3               1             16            5                 6
4               2             11            4                 5
5               1             11            5                 4
6               2             12            3                 3
  nes2.hc.law.num nes2.economy.num nes2.wall.num nes2.isis.num
1               4                3             7             6
2               4                3             1             1
3               1                1             7             7
4               4                2             7             6
5               1                2             7             6
6               6                3             1             7
```

```r
cor(data, use="pairwise.complete.obs")
```

```
                  nes2.gender.num nes2.education nes2.pid.num
nes2.gender.num      1.0000000000   0.0003711709  -0.08431888
nes2.education       0.0003711709   1.0000000000  -0.05555047
nes2.pid.num        -0.0843188796  -0.0555504651   1.00000000
nes2.ideology.num   -0.0871314291  -0.1583629030   0.69672019
nes2.hc.law.num      0.0089716363   0.0902526412  -0.59336452
nes2.economy.num    -0.0685063974   0.1771258154  -0.46524989
nes2.wall.num       -0.0290107912  -0.1881997817   0.57894630
nes2.isis.num       -0.0726686242  -0.0517619374   0.36426632
                  nes2.ideology.num nes2.hc.law.num nes2.economy.num
nes2.gender.num         -0.08713143     0.008971636       -0.0685064
nes2.education          -0.15836290     0.090252641        0.1771258
nes2.pid.num             0.69672019    -0.593364518       -0.4652499
nes2.ideology.num        1.00000000    -0.539686616       -0.4309842
nes2.hc.law.num         -0.53968662     1.000000000        0.4197381
nes2.economy.num        -0.43098419     0.419738055        1.0000000
nes2.wall.num            0.54045756    -0.449930821       -0.4327909
nes2.isis.num            0.35305154    -0.234867045       -0.1876537
                  nes2.wall.num nes2.isis.num
nes2.gender.num     -0.02901079   -0.07266862
nes2.education      -0.18819978   -0.05176194
nes2.pid.num         0.57894630    0.36426632
nes2.ideology.num    0.54045756    0.35305154
nes2.hc.law.num     -0.44993082   -0.23486704
nes2.economy.num    -0.43279095   -0.18765370
nes2.wall.num        1.00000000    0.30321382
nes2.isis.num        0.30321382    1.00000000
```

There are no high correlations.

```r
vif(model.2)
```

```
      gender    education      pid.num ideology.num   hc.law.num
    1.029746     1.074266     2.574591     2.221188     1.685446
 economy.num     wall.num     isis.num
    1.438744     1.712557     1.191452
```

None of the VIFs are near 10 and thus we have no multicollinearity.

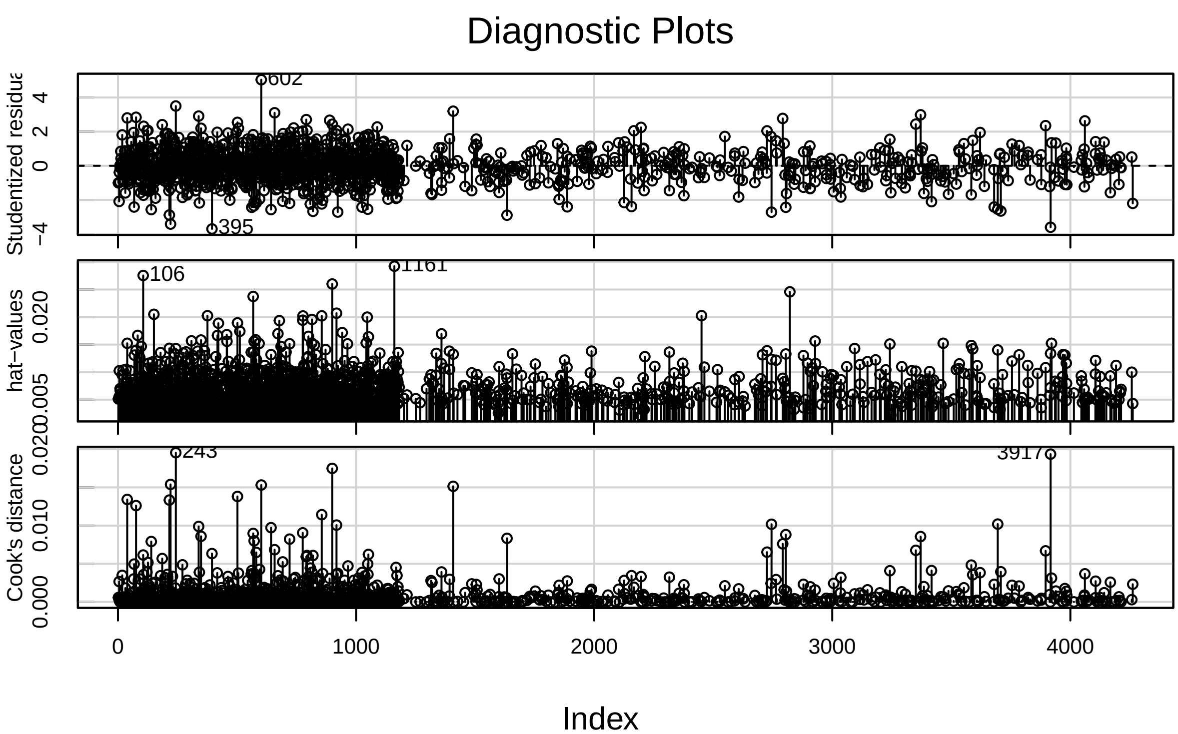
9.5 Outliers, Leverage, and Influential Data Points

```
(2*(8+1))/1213
```

```
[1] 0.01483924
```

The cut-point for high leverage is .015.

```
x11()
influenceIndexPlot(model.2,
                   vars=c("Studentized","hat","Cook"))
```



There are some outliers, points with high leverage, but no influential data points. Therefore, we do not need to make any corrections.

Question 10

For our interpretations, we will use the version of the model with robust standard errors. We run it again to have an immediate reference:

```
coeftest(model.2, vcov = vcovHC)
```

```
t test of coefficients:
```

```
                Estimate Std. Error  t value  Pr(>|t|)    
(Intercept)     56.32244    5.70546   9.8717 < 2.2e-16 ***
gender1. Female  2.90697    1.17780   2.4681   0.01372 *  
education       -0.30113    0.27013  -1.1148   0.26517    
pid.num         -6.96075    0.51496 -13.5170 < 2.2e-16 ***
ideology.num    -0.45376    0.62835  -0.7221   0.47035    
hc.law.num       3.53857    0.37419   9.4566 < 2.2e-16 ***
economy.num      3.86514    0.70384   5.4915 4.857e-08 ***
wall.num        -1.65448    0.36514  -4.5311 6.451e-06 ***
isis.num        -0.29145    0.32757  -0.8897   0.37380    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Again, we find that all the predictors are statistically significant except for education, political ideology, and support for sending troops to fight ISIS. Thus hypotheses 2, 4, and 8 are not supported, but the rest of our hypotheses have statistical merit. Let's go through each significant predictor and perform a coefficient interpretation and brief discussion.

First, *women are expected to rate Clinton 2.91 points higher than men, while controlling for other variables.* As suggested by our plotting, women have warmer feelings towards Clinton than men. Clinton was the first female major party presidential candidate which may have led to drawing more support from women.

Second, *for a one-unit increase in partisan identification, respondents are expected to rate Clinton 6.96 points lower.* This means that as respondents become more Republican, they are expected to have colder feelings towards Clinton. Partisan identification has traditionally been a stronger driver of candidate evaluation and vote choice in US elections.

Third, *for a one-unit increase in support for Obamacare, respondents are expected to rate Clinton 3.54 points higher.* Although Obamacare was not Clinton's plan, she was a supporter of the plan and proposed to continue with its implementation. Hence, respondents who favoured the plan had warmer feelings towards Clinton and vice-versa.

Fourth, *for a one-unit increase in the evaluation of economy, respondents are expected to rate Clinton 3.87 points higher.* Respondents who think the economy has gotten better over the past year are expected to rate Clinton higher than respondents who thought the economy has gotten worse. Commonly, people who think the economy is better have a higher evaluation of the incumbent or incumbent's party and vice-versa; although we are not ruling out a partisan effect on whether people think the economy is better or worse.

Fifth, *for a one-unit increase in support for building a wall with Mexico, respondents are expected to rate Clinton 1.65 points lower.* Although this effect is not as large as it was for Trump, respondents' positions on the wall still mattered in their evaluation of Clinton.

In sum, we find that partisan identification and view of the economy had the largest effect on respondents' feelings towards Clinton. Respondents' gender, support for Obamacare, and support for building a wall with Mexico also had a significant effect on the evaluation of Clinton.