?

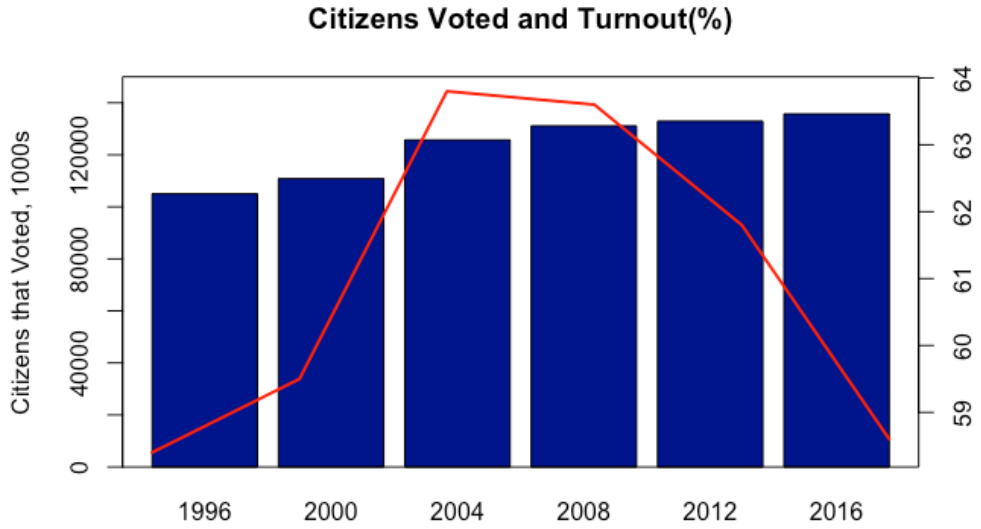
Elections and the American Electorate: 2012

Which Factors Best Explain **Voter Turnout**?

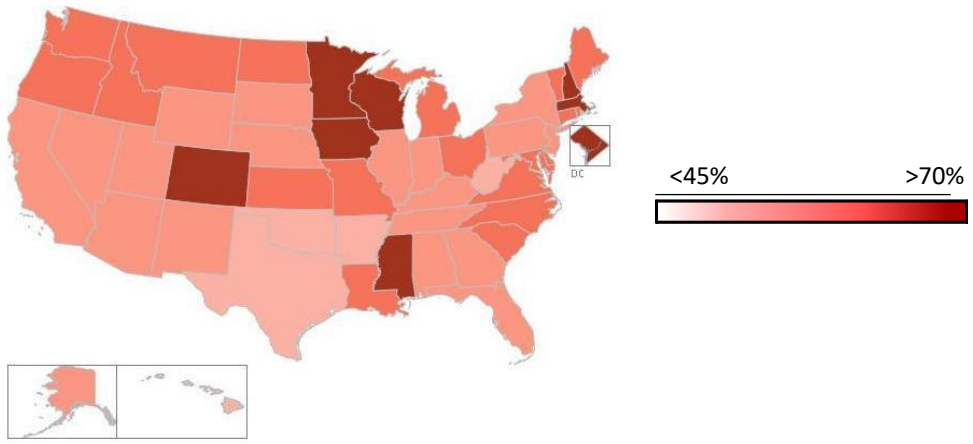
Group Member Names:
Henry Lankin
Gui Larangeira

Goal and Motivation

- To analyze the voter turnout in the 2012 elections and determine which factors best explain differences between states.
- To determine which and how categorical and numerical variables such as age bracket, race and Hispanic origin, education, income and its political “color” (**red**, **blue**, or **swing**) influence the voter turnout of the state.
- To determine if the proportion of eligible voters by race and Hispanic origin or other factors match the proportion of voter turnout by factors in the United States.



Percent Voted 2012



Our Dataset

Source: US CPS Current Population Survey

- Voting and Registration in the Election of November 2012

Response variable:

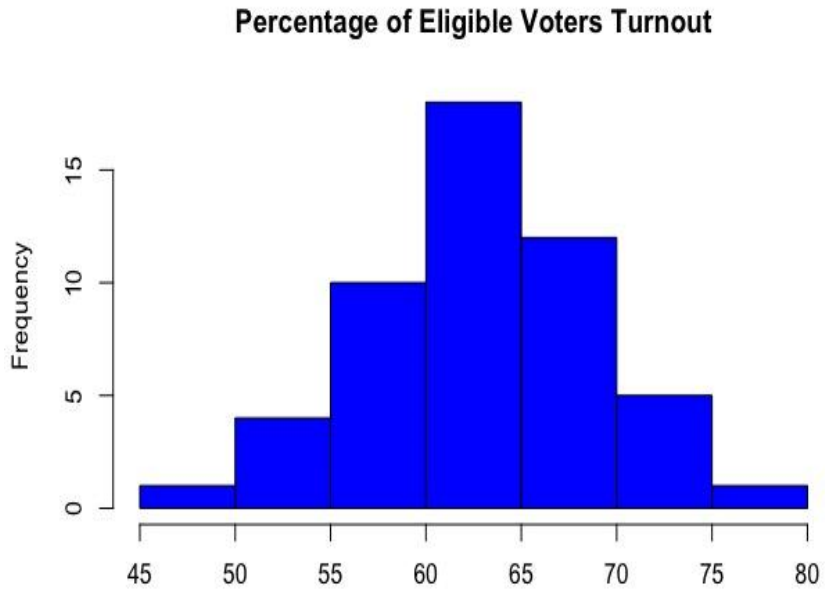
- Percent of eligible voters reported to have voted

Factors:

- State color categorized as **Red**, **Blue**, or **Swing**
- Education, Income, Age, and Race as compared to the national average (above and under)

Numerical variables:

- Race: Percent of state that is Black, White, Asian and of Hispanic origin
- Age: Percent of state by the age brackets of 18-44, 45-64, and 65 and over
- Income: Median household income and poverty rate of the state
- Education: Percent of state that is college educated (Bachelor's degree or higher)




```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  47.80   59.10   62.70   62.91   66.85   75.90
```

```
        Shapiro-Wilk normality test

data:  turnout
W = 0.99173, p-value = 0.9764
```

Race Composition of Voter Turnout

Methodology: Chi-Square Goodness of Fit Test

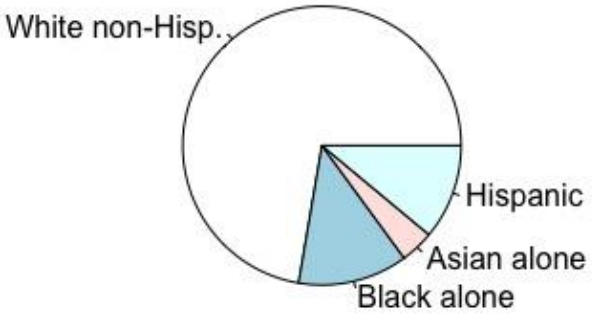
H_0: Proportion of voters by race matches eligible voters in the population

H_a: At least one proportion does not match.

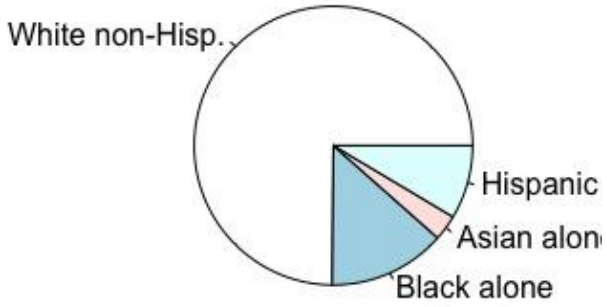
Conclusion: The proportions are not equal; thus, voting proportion by race does not accurately reflect the population ethnic groups.

- Asian and Hispanic voters were under represented
- White and Black voters were over represented.

Population Composition



Voter Composition




```
        Chi-squared test for given probabilities

data:  voters
X-squared = 1219.1, df = 3, p-value < 2.2e-16
```

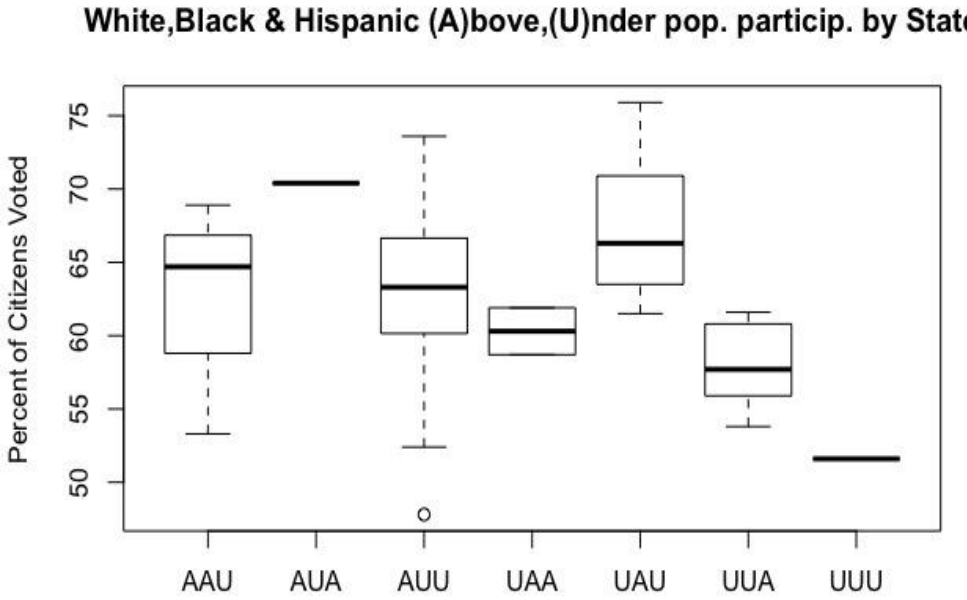
Voter turnout by Race

Methodology: Multi-factor ANOVA, “factorizing” race variables into (A)bove and (U)nder average

$\boldsymbol{H_0}$: Average Voter Turnout is same for all combinations AAU to UUA.

$\boldsymbol{H_a}$: Average Voter Turnout is not same.

Conclusion: The p-value of the overall ANOVA is significant at 0.024, indicating that race plays a role in voter turnout.



	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factorRace	6	498.5	83.09	2.735	0.024 *
Residuals	44	1336.8	30.38		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05

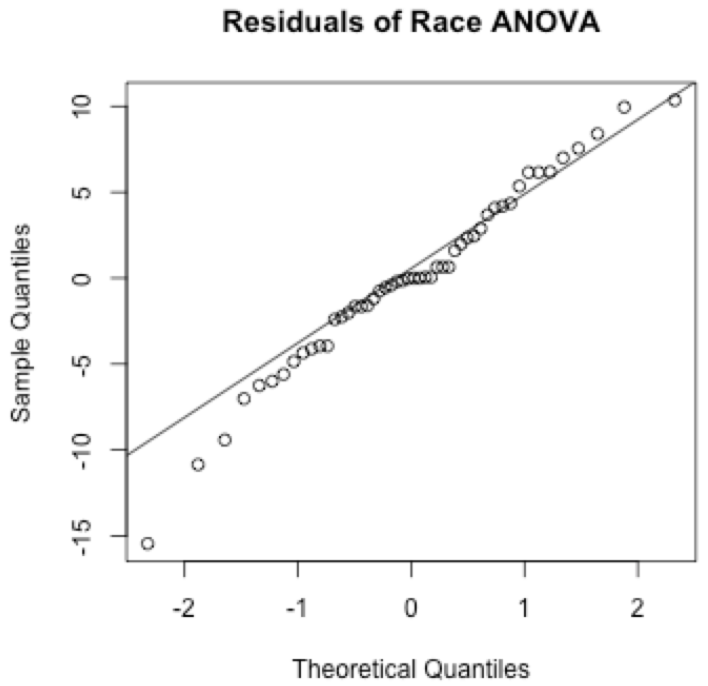
Paired Comparison for Race and Diagnostics

Conclusion of paired comparisons: Only the difference in voter turnout between the pairs UUA-UAU and UUU-UAU is significant at about 0.12.

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = turnout ~ factorRace)

$factorRace
                diff        lwr        upr     p adj
AUA-AAU    7.8000000 -10.400312 26.0003115 0.8371721
AUU-AAU    0.6518519  -6.569049  7.8727531 0.9999568
UAA-AAU   -2.3000000 -15.950234 11.3502336 0.9984090
UAU-AAU    4.9000000  -4.200156 14.0001558 0.6432403
UUA-AAU   -4.6833333 -14.155076  4.7884091 0.7272575
UUU-AAU  -11.0000000 -29.200312  7.2003115 0.5122371
AUU-AUA   -7.1481481 -24.485389 10.1890927 0.8601803
UAA-AUA  -10.1000000 -30.951076 10.7510763 0.7455234
UAU-AUA   -2.9000000 -21.100312 15.3003115 0.9988388
UUA-AUA  -12.4833333 -30.872254  5.9055875 0.3724345
UUU-AUA  -18.8000000 -42.876749  5.2767490 0.2181352
UAA-AUU   -2.9518519 -15.428128  9.5244240 0.9898550
UAU-AUU    4.2481481  -2.972753 11.4690494 0.5440857
UUA-AUU   -5.3351852 -13.019096  2.3487255 0.3458268
UUU-AUU  -11.6518519 -28.989093  5.6853890 0.3844027
UAU-UAA    7.2000000  -6.450234 20.8502336 0.6647246
UUA-UAA   -2.3833333 -16.284051 11.5173842 0.9982458
UUU-UAA   -8.7000000 -29.551076 12.1510763 0.8533812
UUA-UAU   -9.5833333 -19.055076 -0.1115909 0.0457115
UUU-UAU  -15.9000000 -34.100312  2.3003115 0.1227616
UUU-UUA   -6.3166667 -24.705587 12.0722542 0.9360307
```



Diagnostics: Conditions for the ANOVA model are met.

- The residuals plot and Shapiro-Wilks test show that normality in the population cannot be rejected.
- The Levene test shows that equal variances cannot be rejected.

```
        Shapiro-Wilk normality test

data:  result.race$residuals
W = 0.97936, p-value = 0.5254
```

```
Levene's Test for Homogeneity of Variance
      Df F value Pr(>F)
group  6  0.7436 0.6176
      43
```

Voter Turnout by State Color

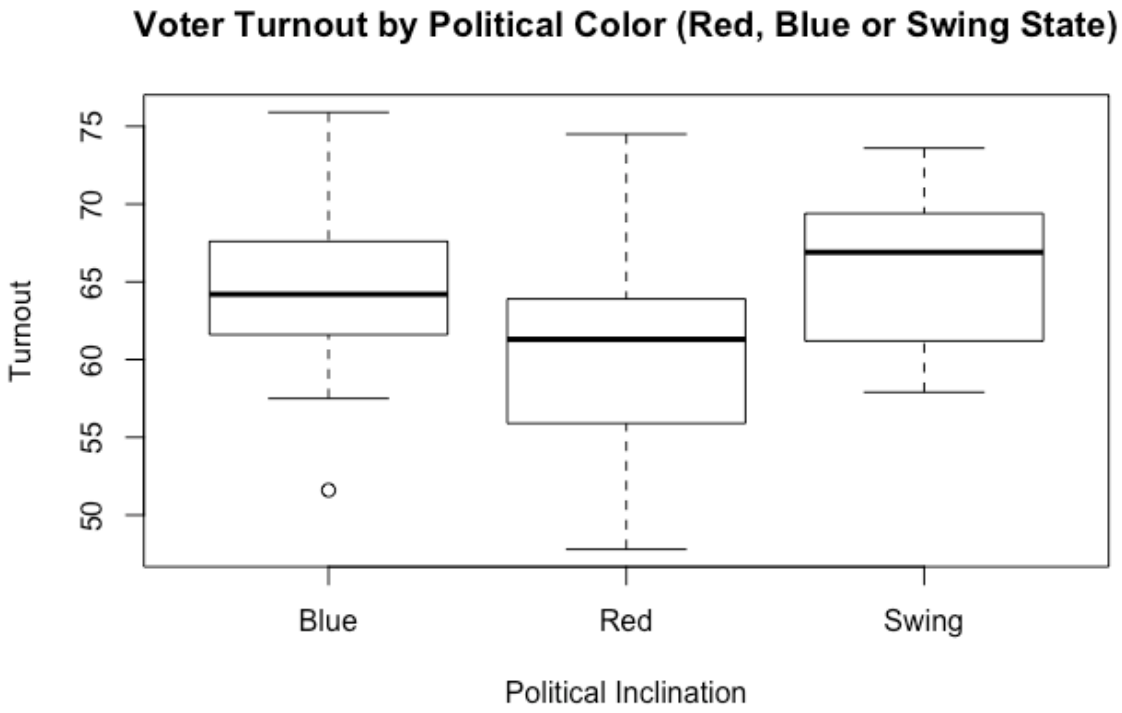
Methodology: One-way ANOVA

H_0: Proportion of voter turnout by **red**, **blue**, and **swing** states are the same

H_a: The proportions are not the same.

Conclusion: The voter turnout between political color of state is different at a significance level below 0.02.

- As expected, the Swing states have the highest average turnout.

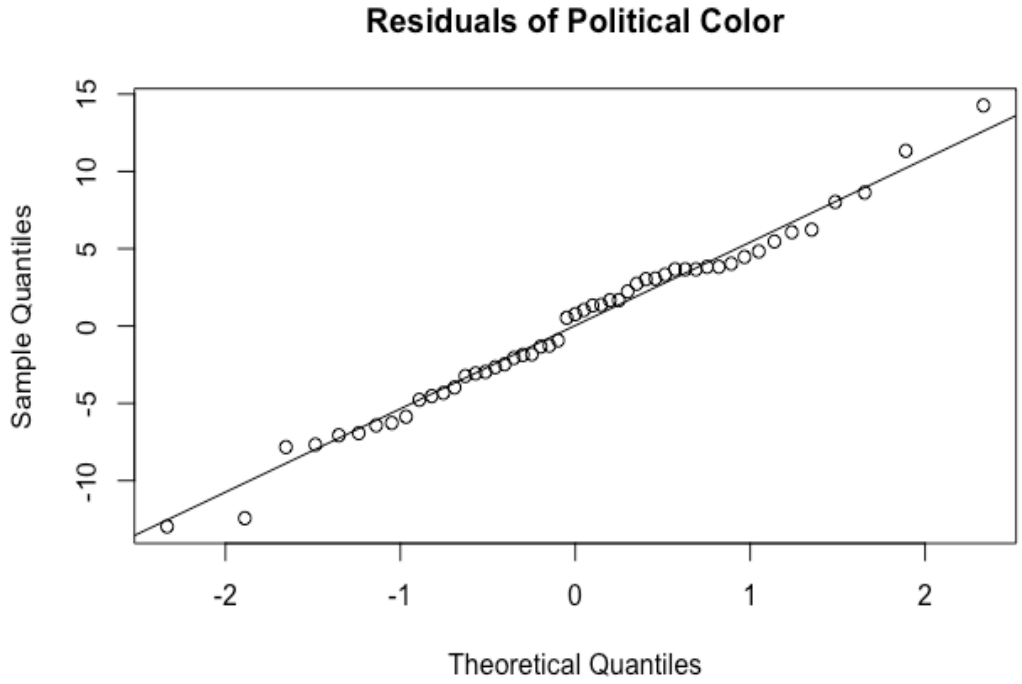



```
             Df Sum Sq Mean Sq F value Pr(>F)
polit.color   2  285.2   142.6   4.415 0.0174 *
Residuals    48 1550.2    32.3
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Diagnostics for Political Color

Conditions for the ANOVA model are met.

- The residuals plot and Shapiro-Wilks test show that normality cannot be rejected.
- The Levene test shows that equal variances cannot be rejected.

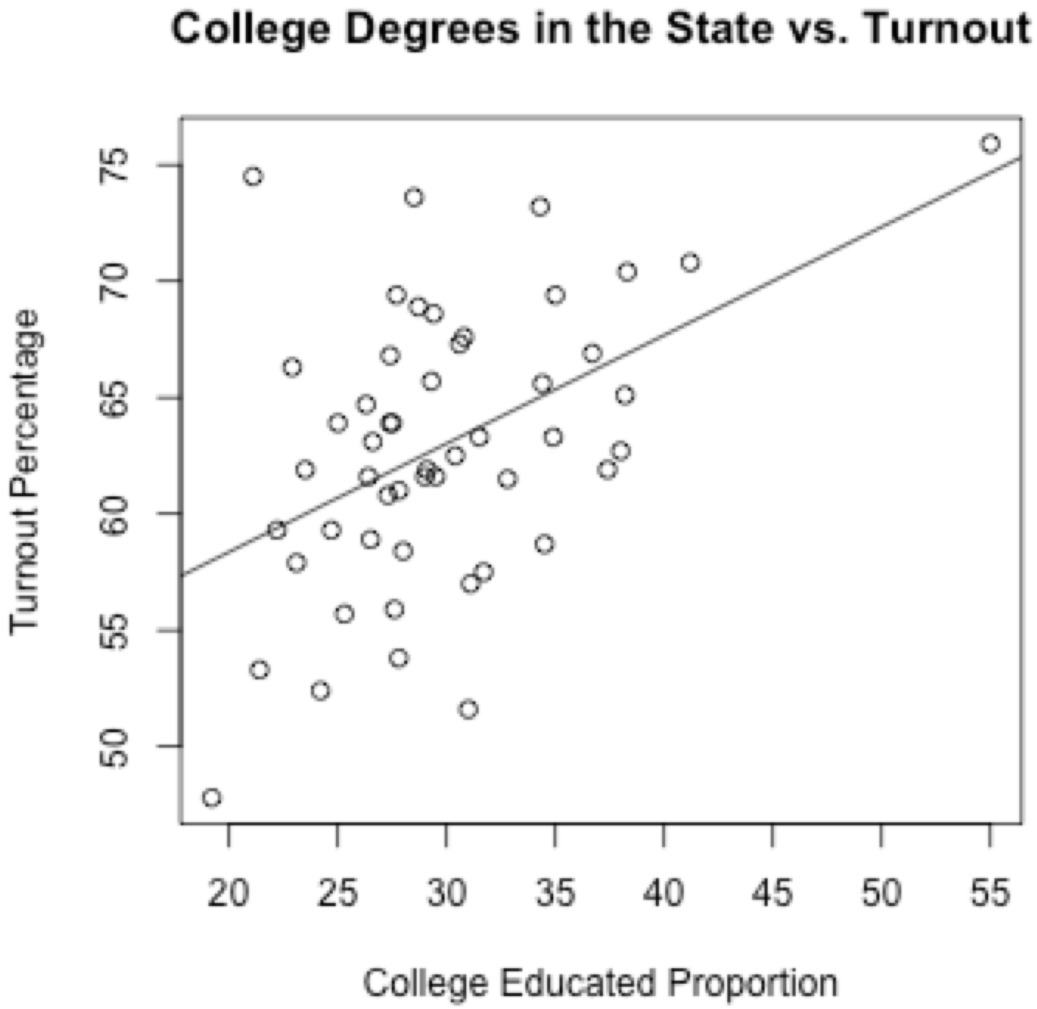



```
Levene's Test for Homogeneity of Variance
      Df F value Pr(>F)
group  2   6e-04 0.9994
      48
```

```
        Shapiro-Wilk normality test

data:  polit.color.aovResult$residuals
W = 0.98684, p-value = 0.8391
```

Percent College Educated



Methodology: Simple linear regression between percent of state that is college educated and voter turnout.

Conclusion: The regression line shows a positive correlation between percent of the state that is college educated and the voter turnout of the state.

- Percent of college educated is quite significant (less than 1%) for voter turnout.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  49.0520     3.8078  12.882  < 2e-16 ***
education     0.4657     0.1254   3.715 0.000522 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.406 on 49 degrees of freedom
Multiple R-squared:  0.2197, Adjusted R-squared:  0.2038
F-statistic:  13.8 on 1 and 49 DF,  p-value: 0.0005218
```

All Numerical Variables Combined

Methodology: Multi-linear regression

```
lm(formula = turnout ~ education + age.bracket.18.44 + age.bracket.45.64 +
    age.bracket.65 + med.income + poverty.rate + race.white +
    race.black + race.hispanic)

Residuals:
    Min      1Q  Median      3Q     Max
-9.3102 -2.6567 -0.2893  2.4306 12.4293

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       82.6952698 67.4933384   1.225  0.22748
education          0.7840068  0.2427676   3.229  0.00244 **
age.bracket.18.44 -0.3832911  0.6709828  -0.571  0.57096
age.bracket.45.64  0.3050352  0.9635025   0.317  0.75316
age.bracket.65    -0.6355510  0.4975402  -1.277  0.20865
med.income        -0.0004220  0.0002631  -1.604  0.11638
poverty.rate      -0.8784384  0.6055036  -1.451  0.15446
race.white         0.1032968  0.1258660   0.821  0.41657
race.black         0.2730112  0.1404027   1.944  0.05872 .
race.hispanic      0.0367869  0.1481445   0.248  0.80513
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.688 on 41 degrees of freedom
Multiple R-squared:  0.509,  Adjusted R-squared:  0.4013
F-statistic: 4.723 on 9 and 41 DF,  p-value: 0.0002469
```

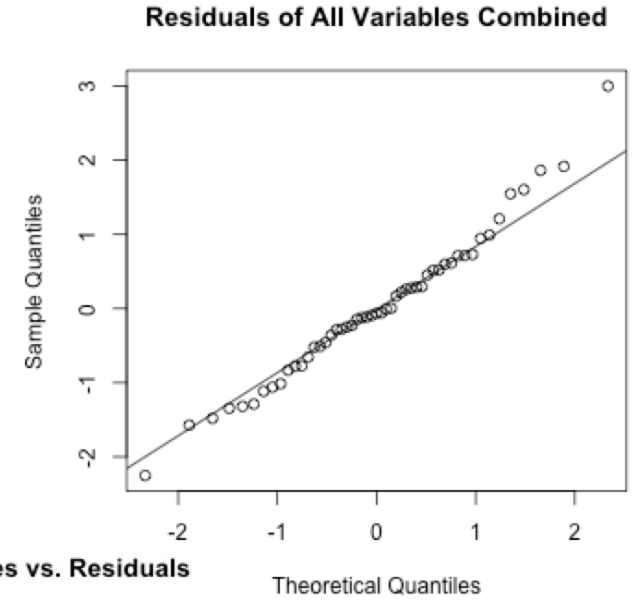
Conclusion: The correlation coefficient (multiple r^2 value) of the regression model is low and is likely affected by the noise of all the variables being included.

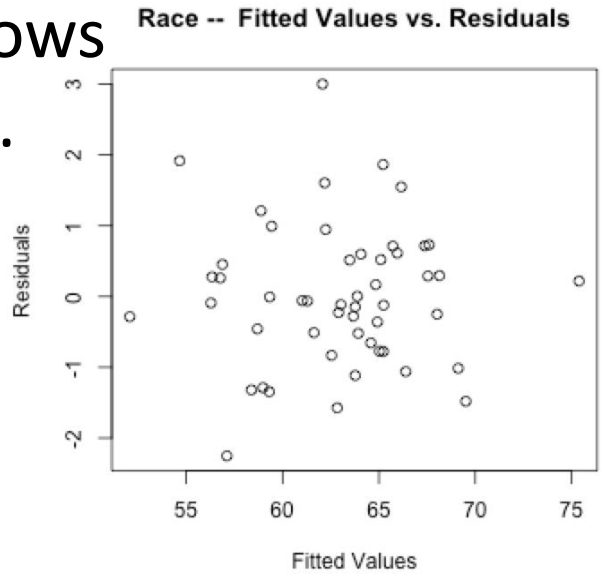
- There is a high significance of less than 1% for education and the positive estimator confirms the result from the SLR of education.
- There is also a fairly high significance of a little over 5% for Black race, confirming the results of the ANOVA and Goodness of Fit.

Diagnostics for Multi-Linear Regression

Conditions for the regression model are met.

- The residuals plot and Shapiro-Wilks test show that normality cannot be rejected.
- The residuals vs. fitted values plot shows that equal variances cannot be rejected.






```
        Shapiro-Wilk normality test

data:  rs.all
W = 0.98222, p-value = 0.6362
```

Final Conclusion

- Our dataset of over 50 states on the percent of eligible voter turnout has a **normal distribution.**
- The composition of voters by race and Hispanic origin **does not mirror** the overall population of eligible voters.
- Of the factors, **state color** has the **most influence** on voter turnout, with **swing states** clearly above average voter turnout.
- Of the numerical variables tested, **education** has the **most significant influence**.
- Policy recommendations to increase voter turnout:
  - Implement a **direct popular vote** and **abolish the Electoral College** so that we eliminate the bias introduced by different states.
  - **Make voting day a national civic holiday** so people can more easily participate in the vote.
  - **Expand registration efforts** and **implement early voting and vote-by-mail options in more counties** to increase voting participation.

Appendix

1. Political Color ANOVA: Red, Blue and Swing States (1996-2012 Presidential Elections)