The 2016 American presidential election saw one of the biggest upsets in recent history. Prior to voting day, virtually every major media outlet predicted Hillary Clinton in a landslide victory over Donald Trump. But on Tuesday, November 8th, 2016, the unthinkable happened: Donald Trump, the well-known long-time businessman and reality TV star, was voted into the Oval Office by a 304 to 227 victory in the electoral college. This outcome left a large portion of the general public questioning how Donald Trump, a man with no prior political experience, could be voted into the most powerful position in the world.

Due to these factors, this past election was truly one of the most unique in American history, which is why it deserves to be studied in depth to understand how and why the events transpired. This paper will attempt to explain part of Trump's voting distribution with data on the 2016 election, as well as the 2008 election. More specifically, it will attempt to explain rates of Republican voting across the United States with education levels in each county and the hypothesis for the study as follows: the expectation to see increases in Republican (Trump) voting in counties with lower education rates.

There are two main data sets for this data analysis: one corresponding to the 2016 election and one to the 2008 election. For those two elections, unemployment, education, and census data were gathered along county FIPS code. 2008 and 2016 voter turnout data at the county level were obtained from Professor Sundstrom. The 2010 county census data was gathered from the US Census Bureau, which contains data for 3083 counties, including variables for demographic, financial, education, and other characteristics. The unemployment and education data on the county level was provided by the Department of Agriculture Economic Research Service (USDA-RES).

The hypothesis is concerned with the education levels across counties in the United States. The key dependent variable in the regression is vote share for each candidate in the 2008 and 2016 election, which was calculated by taking the number of votes for one candidate divided by the total amount of votes. The key regressors for this hypothesis were determined by

the percent of people with less than a high school diploma and less than a bachelor's degree (or 4-year equivalent). The USDA-RES has data for the percent of adults according to levels of education going back to 1970. For the 2016 election, data was collected from 2011-2015. The next data set from the USDA-RES was collected in 2000, which was determined to not be a good representation of the education levels in 2008. Instead, the substitute was made to be one minus the percentages of high school and bachelors levels that were collected from the census for the 2008 election. Hillary Clinton won most of the coastal states while Donald Trump won most of the intercontinental states (states in between the mountain ranges), and there is a slight discrepancy in education levels between those two regions. Controls for regression include race, percentage above 65 years old, unemployment rate, and income per capita. By including these controls, a more valid coefficient was obtained on the education variable avoiding omitted variable bias. This study decided on including age above 65 rather than 18 because 65-year-olds are less likely of going back to school and earning a diploma than an 18-year-old. Another interesting statistic is a large increase in third party votes in the 2016 election compared to the 2008 election. This is shown in the table below as the mean percentage of third party votes. Any presidential vote that was not Republican or Democrat is considered to be a "Third Party" vote, and additionally regressed with education levels. Below is the descriptive statistics for the 2008 and 2016 election.

2016

Statistic	N	Mean	St. Dev.	Min	Max
black	3,083	9.04	14.59	0.00	85.70
hispanic	3,083	8.29	13.24	0.10	95.70
white_not_hispanic	3,083	78.19	19.87	2.70	98.70
per_capita_income	3,083	22,492.10	5,403.32	7,772	64,381
age_over_65	3,083	15.83	4.15	3.50	43.40
lessthanhs1115	3,080	14.64	6.62	1.90	53.70
lessthanbs1115	3,080	79.57	9.05	21.20	98.10
Unemployment_rate_2015	3,079	5.55	1.98	1.80	24.00
Donald.J..Trump	3,056	20,524.02	43,515.83	58	769,743
Hillary.Clinton	3,056	21,496.08	81,266.36	4	2,464,364
Total.Vote	3,056	44,706.70	127,372.80	65	3,434,308
donaldvoteshare	3,056	0.63	0.16	0.04	0.94
hillaryvoteshare	3,056	0.32	0.15	0.03	0.91
othervoteshare	3,056	0.05	0.03	0.003	0.36

2008

Table 5. 2008 Election Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Max
black	3,083	9.04	14.59	0.00	85.70
hispanic	3,083	8.29	13.24	0.10	95.70
white_not_hispanic	3,083	78.19	19.87	2.70	98.70
per_capita_income	3,083	22,492.10	5,403.32	7,772	64,381
age_over_65	3,083	15.83	4.15	3.50	43.40
lessthanhs2010	3,083	16.97	7.34	0.70	52.10
lessthanbs2010	3,083	80.95	8.69	29.00	96.30
Unemployment_rate_2008	3,076	5.85	2.05	1.30	22.60
John.S..McCain.III	3,057	19,516.94	45,201.81	67	956,425
Barack.H..Obama	3,057	22,677.74	77,776.72	8	2,295,853
Total.Vote	3,057	42,797.68	120,905.60	79	3,318,248
mccainvoteshare	3,057	0.57	0.14	0.07	0.93
obamavoteshare	3,057	0.42	0.14	0.05	0.92
othervoteshare	3,057	0.02	0.01	0.00	0.07

Empirical Results:

Table 1. 2016 Election and County Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Max
black	3,083	9.04	14.59	0.00	85.70
hispanic	3,083	8.29	13.24	0.10	95.70
white_not_hispanic	3,083	78.19	19.8	72.70	98.70
per_capita_income	3,083	22,492.10	5,403.32	7,772	64,381
age_over_65	3,083	15.83	4.15	3.50	43.40
less_than_hs1115	3,080	14.64	6.62	1.90	53.70
less_than_bs1115	3,080	79.57	9.05	21.20	98.10
unemployment_rate_2015	3,079	5.55	1.98	1.80	24.00
Donald.J.Trump	3,056	20,524.02	43,515.83	58	769,743
Hillary.Clinton	3,056	21,496.08	81,266.36	4	2,464,364
Total.Vote	3,056	44,706.70	127,372.80	65	3,434,308
donald_vote_share	3,056	0.63	0.16	0.04	0.94
hillary_vote_share	3,056	0.32	0.15	0.03	0.91
other_vote_share	3,056	0.05	0.03	0.003	0.36

Table 2. Effect of education on Donald vote share

	Dependent variable:					
	donald_vote_share					
	(1)	(2)	(3)	(4)	(5)	(6)
Less_than_hs_1115	0.002***		0.012***		0.014***	
	(0.0004)		(0.0004)		(0.0004)	
Less_than_bs_1115		0.009***		0.009***		0.010***
		(0.0003)		(0.0002)		(0.0002)
black			-0.002***	-0.001**	-0.001***	-0.0004
			(0.0003)	(0.0003)	(0.0003)	(0.0003)
hispanic			0.00002	0.002***	-0.001**	0.001***
			(0.0004)	(0.0004)	(0.0004)	(0.0003)
white_not_hispanic			0.005***	0.005***	0.005***	0.004***

			(0.0003)	(0.0003)	(0.0003)	(0.0003)
age_over_65			0.005***	0.001	0.006***	0.001
			(0.001)	(0.001)	(0.001)	(0.0005)
Unemployment_rate_2015					-0.018***	-0.017***
					(0.001)	(0.001)
Constant	0.593***	-0.060***	-0.015	-0.452***	0.080***	-0.423***
	(0.008)	(0.022)	(0.029)	(0.030)	(0.029)	(0.029)
Observations	3,055	3,055	3,055	3,055	3,055	3,055
R2	0.011	0.253	0.503	0.570	0.534	0.601
Adjusted R2	0.010	0.252	0.502	0.569	0.533	0.600
Residual Std. Error	0.155	0.135	0.110	0.102	0.107	0.099
F Statistic	33.037***	1,032.093***	617.823***	808.921***	582.782***	766.020***
Note:						*p<0.1; **p<0.05; ***p<0.01

In table 3, regressions were ran to estimate the effect of county level education on Donald Trump's vote share in the 2016 election. As shown in simple regression model 1, a one percentage point increase in the population with less than a high school diploma is associated with a 0.2 percentage point increase in the vote share for Donald Trump, significant at the 1% level. Model 2 shows that a one percentage point increase in the population with less than a bachelor's degree is associated with a 0.9 percentage point increase in vote share for Trump, significant at the 1% level. In models 3 and 4, controls are included for race and age. Model 3 shows that a one percentage point increase in the population with less than a high school diploma is associated with a 1.2 percentage point increase in the vote share for Donald Trump, significant at the 1% level. Model 4 shows that a one percentage point increase in the population with less than a bachelor's degree is associated with a 0.9 percentage point increase in vote share for Trump, significant at the 1% level. In model 5 and 6 we include controls for unemployment rate by county. In these models, a one percentage point increase in the population with less than a high school diploma is associated with a 1.4 percentage point increase in vote share for Trump, statistically significant at the 1% level. A one percentage point increase in the population with less than a bachelor's degree is associated with a 1 percentage point increase in vote share for Trump, statistically significant at the 1% level. Interestingly, a one percentage point increase in the unemployment rate is associated with about a 1.7 percentage point decrease in the Trump's vote share, statistically significant at the 1% level.

Table 3. Effect of education on Hillary vote share

	Dependent variable:					
	hillary_vote_share					
	(1)	(2)	(3)	(4)	(5)	(6)
lessthanhs1115	-0.0001		-0.010***		-0.012***	
	(0.001)		(0.0004)		(0.0004)	
lessthanbs1115		-0.007***		-0.008***		-0.009***

		(0.0003)		(0.0002)		(0.0002)
black			0.002***	0.002***	0.002***	0.002***
			(0.0003)	(0.0003)	(0.0003)	(0.0003)
hispanic			-0.0001	-0.001***	0.001*	-0.001***
			(0.0004)	(0.0004)	(0.0004)	(0.0003)
white_not_hispanic			-0.005***	-0.004***	-0.004***	-0.004***
			(0.0003)	(0.0003)	(0.0003)	(0.0003)
age_over_65			-0.003***	0.0005	-0.004***	0.0004
			(0.001)	(0.0005)	(0.001)	(0.0005)
Unemployment_rate_2015					0.016***	0.017***
					(0.001)	(0.001)
Constant	0.319***	0.899***	0.868***	1.258***	0.781***	1.228***
	(0.008)	(0.022)	(0.027)	(0.028)	(0.027)	(0.027)
Observations	3,055	3,055	3,055	3,055	3,055	3,055
R2	0.00002	0.192	0.529	0.607	0.557	0.640
Adjusted R2	-0.0003	0.192	0.528	0.606	0.556	0.639
Residual Std. Error	0.151	0.136	0.104	0.095	0.101	0.091
F Statistic	0.051	725.729***	685.038***	941.164***	639.051***	903.656***
Note:					*p<0.1; **p<0.05; ***p<0.01	

In table 4, the same regressions are run on Clinton's vote share in the 2016 election. As expected coefficients are found in the inverse direction from those in table 2. Model 5 in table 3 suggests a one percentage point increase in the population with less than a high school diploma is associated with a 1.2 percentage point decrease in the vote share for Hillary Clinton, significant at the 1% level. Model 6 shows that a one percentage point increase in the population with less than a bachelor's degree is associated with a .9 percentage point decrease in vote share for Clinton, significant at the 1% level. The coefficients for percent less than high school diploma and percent less than bachelor's degree are smaller in magnitude than those in table 3.

Table 4. Effect of education on other parties vote share

	Dependent variable:					
	other_vote_share					
	(1)	(2)	(3)	(4)	(5)	(6)
lessthanhs1115	-0.002***		-0.002***		-0.002***	
	(0.0001)		(0.0001)		(0.0001)	
lessthanbs1115		-0.001***		-0.001***		-0.001***
		(0.0001)		(0.0001)		(0.0001)
black			-0.001***	-0.001***	-0.001***	-0.001***

			(0.0001)	(0.0001)	(0.0001)	(0.0001)
hispanic			0.0001	-0.0004***	0.0001	-0.0004***
			(0.0001)	(0.0001)	(0.0001)	(0.0001)
white_not_hispanic			-0.0004***	-0.0003***	-0.0003***	-0.0003***
			(0.0001)	(0.0001)	(0.0001)	(0.0001)
age_over_65			-0.001***	-0.001***	-0.002***	-0.001***
			(0.0002)	(0.0002)	(0.0002)	(0.0002)
Unemployment_rate_2015					0.001***	-0.00003
					(0.0003)	(0.0003)
Constant	0.088***	0.162***	0.147***	0.194***	0.139***	0.194***
	(0.001)	(0.006)	(0.009)	(0.009)	(0.009)	(0.009)
Observations	3,055	3,055	3,055	3,055	3,055	3,055
R2	0.251	0.157	0.412	0.337	0.417	0.337
Adjusted R2	0.251	0.156	0.411	0.336	0.415	0.336
Residual Std. Error	0.027	0.028	0.024	0.025	0.024	0.025
F Statistic	1,023.434***	567.115***	426.577***	310.275***	362.704***	258.480***
Note:						*p<0.1; **p<0.05; ***p<0.01

Once again, in Table 5 the same regressions were ran on the vote share for all other candidates in the 2016 election. An increase in percentage of people with lower education is negatively correlated with vote share for other candidates. These coefficients, however, are much smaller in magnitude than those in tables 3 & 4.

Table 5. 2008 Election and County Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Max
black	3,083	9.04	14.59	0.00	85.70
hispanic	3,083	8.29	13.24	0.10	95.70
white_not_hispanic	3,083	78.19	19.87	2.70	98.70
per_capita_income	3,083	22,492.10	5,403.32	7,772	64,381
age_over_65	3,083	15.83	4.15	3.50	43.40
less_than_hs2010	3,083	16.97	7.34	0.70	52.10
lessthanbs2010	3,083	80.95	8.69	29.00	96.30
Unemployment_rate_2008	3,076	5.85	2.05	1.30	22.60
John.S.McCain.III	3,057	19,516.94	45,201.81	67	956,425
Barack.H.Obama	3,057	22,677.74	77,776.72	8	2,295,853
Total.Vote	3,057	42,797.68	120,905.60	79	3,318,248
mccain_vote_share	3,057	0.57	0.14	0.07	0.93
obama_vote_share	3,057	0.42	0.14	0.05	0.92
other_vote_share	3,057	0.02	0.01	0.00	0.07

Table 5. 2008 Election and County Descriptive Statistics
====

TRUE

Table 6. Effect of education on McCain vote share 2008

	Dependent variable:					
	mccain_vote_share					
	(1)	(2)	(3)	(4)	(5)	(6)
lessthanhs2010	0.002***		0.006***		0.008***	
	(0.0004)		(0.0004)		(0.0004)	
lessthanbs2010		0.005***		0.005***		0.007***
		(0.0003)		(0.0003)		(0.0003)
black			-0.0001	0.0005	0.0002	0.001***
			(0.0003)	(0.0003)	(0.0003)	(0.0003)
hispanic			0.002***	0.003***	0.001***	0.002***
			(0.0004)	(0.0004)	(0.0004)	(0.0004)
white_not_hispanic			0.004***	0.004***	0.004***	0.003***
			(0.0003)	(0.0003)	(0.0003)	(0.0003)
age_over_65			0.001	-0.002***	0.001*	-0.002***
			(0.001)	(0.001)	(0.001)	(0.001)
Unemployment_rate_2008					-0.022***	-0.024***
					(0.001)	(0.001)
Constant	0.530***	0.168***	0.110***	-0.159***	0.229***	-0.131***
	(0.007)	(0.022)	(0.028)	(0.033)	(0.028)	(0.031)
Observations	3,057	3,057	3,057	3,057	3,055	3,055
R2	0.012	0.097	0.234	0.254	0.319	0.349
Adjusted R2	0.012	0.097	0.232	0.253	0.318	0.348
Residual Std. Error	0.136	0.130	0.120	0.118	0.113	0.110
F Statistic	38.131***	329.102***	185.974***	207.520***	238.467***	272.326***
Note:				*p<0.1; **p<0.05; ***p<0.01		

In table 6, the same regressions were run to estimate the effect of county level education on John McCain’s vote share in the 2008 election to see if there were any major changes in direction or magnitude. Model 1 shows that a one percentage point increase in the population with less than a high school diploma is associated with a 0.2 percentage point increase in the vote share for McCain, significant at the 1% level. Model 2 shows that a one percentage point increase in the population with less than a bachelor's degree is associated with a 0.5 percentage point increase in vote share for McCain, significant at the 1% level. In models 3 and 4, we include controls for race and age. Model 3 shows that a one percentage point increase in the population with less than a high school diploma is associated with a 0.6 percentage point increase in the vote share for McCain, significant at the 1% level. Model 4

shows that a one percentage point increase in the population with less than a bachelor's degree is associated with a 0.5 percentage point increase in vote share for McCain, significant at the 1% level. When including controls for unemployment, models 5 and 6, a one percentage point increase in the population with less than a high school diploma is associated with a 0.8 percentage point increase in vote share for McCain, statistically significant at the 1% level. A one percentage point increase in the population with less than a bachelor’s degree is associated with a 0.7 percentage point increase in vote share for McCain, statistically significant at the 1% level. A one percentage point increase in the unemployment rate is associated with roughly a 2.3 percentage point decrease in the McCain’s vote share, statistically significant at the 1% level.

Table 7. Effect of education on Obama vote share 2008

	Dependent variable:					
	obama_vote_share					
	(1)	(2)	(3)	(4)	(5)	(6)
lessthanhs2010	-0.002***		-0.006***		-0.008***	
	(0.0004)		(0.0004)		(0.0004)	
lessthanbs2010		-0.005***		-0.005***		-0.007***
		(0.0003)		(0.0003)		(0.0003)
black			0.0002	-0.0003	0.00004	-0.001**
			(0.0003)	(0.0003)	(0.0003)	(0.0003)
hispanic			-0.002***	-0.003***	-0.001***	-0.002***
			(0.0004)	(0.0004)	(0.0004)	(0.0004)
white_not_hispanic			-0.004***	-0.004***	-0.004***	-0.004***
			(0.0003)	(0.0003)	(0.0003)	(0.0003)
age_over_65			-0.001	0.001**	-0.001**	0.002***
			(0.001)	(0.001)	(0.001)	(0.001)
Unemployment_rate_2008					0.022***	0.023***
					(0.001)	(0.001)
Constant	0.448***	0.818***	0.878***	1.150***	0.763***	1.122***
	(0.007)	(0.022)	(0.028)	(0.033)	(0.027)	(0.031)
Observations	3,057	3,057	3,057	3,057	3,055	3,055
R2	0.009	0.098	0.254	0.277	0.334	0.367
Adjusted R2	0.009	0.098	0.253	0.275	0.333	0.365
Residual Std. Error	0.137	0.130	0.119	0.117	0.112	0.109
F Statistic	27.773***	331.885***	207.888***	233.302***	255.174***	294.201***
Note:					*p<0.1; **p<0.05; ***p<0.01	

In table 7, the same regressions are run on Obama's vote share in the 2008 election. Model 5 shows that a one percentage point increase in the county's population without a high school diploma is associated with a 0.8 percentage point decrease in vote share for Obama, statistically significant at the 1% level. A one percentage point increase in the people without a bachelor's degree is associated with a 0.7 percentage point decrease in Obama's vote share, statistically significant at the 1% level.

Table 8. Effect of education on other parties vote share 2008

	Dependent variable:					
	other_vote_share					
	(1)	(2)	(3)	(4)	(5)	(6)
lessthanhs2010	-0.0003***		-0.00004**		-0.0001***	
	(0.00002)		(0.00002)		(0.00002)	
lessthanbs2010		0.00003*		0.00004***		-0.00001
		(0.00001)		(0.00001)		(0.00002)
black			-0.0002***	-0.0002***	-0.0002***	-0.0002***
			(0.00002)	(0.00002)	(0.00002)	(0.00002)
hispanic			-0.0001***	-0.0001***	-0.00004*	-0.0001***
			(0.00002)	(0.00002)	(0.00003)	(0.00003)
white_not_hispanic			0.0001***	0.0001***	0.0001***	0.0001***
			(0.00002)	(0.00002)	(0.00002)	(0.00002)
age_over_65			0.0001***	0.0001**	0.0001***	0.0001**
			(0.00004)	(0.00004)	(0.00004)	(0.00004)
Unemployment_rate_2008					0.001***	0.001***
					(0.0001)	(0.0001)
Constant	0.021***	0.014***	0.012***	0.009***	0.008***	0.009***
	(0.0003)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	3,057	3,057	3,057	3,057	3,055	3,055
R2	0.067	0.001	0.261	0.261	0.283	0.277
Adjusted R2	0.067	0.0004	0.259	0.260	0.282	0.276
Residual Std. Error	0.008	0.008	0.007	0.007	0.007	0.007
F Statistic	221.127***	2.268	215.053***	215.760***	200.714***	194.980***
Note:					*p<0.1; **p<0.05; ***p<0.01	

Table 8 shows that there is a statistically significant but very small, negative relationship between the percent of people without a high school diploma and the vote share of other candidates.

Conclusion

So as one can see from the multiple regression tables above, education rates do a very good job of predicting voting rates, for both Republican and Democrat candidates, and for both the 2016 and 2008 elections. It is evident that lower rates of education (less than high school degree, less than bachelor's) are both positively correlated with Republican voting and negatively associated with Democrat voting. Furthermore, these positive/negative correlations are even stronger for the 2016 candidates as compared to the 2008 candidates, insinuating that Trump seemed to capitalize on these populations better than McCain had in 2008. Although education rates alone do not explain Trump's victory, they have proved to be fairly correlated with his voting distribution, from the regressions of this study.

Works Cited

"County-level Data Sets." *USDA ERS - County-level Data Sets*. United States Department Agriculture Economic Research Services, n.d. Web. 3 June 2017. <https://www.ers.usda.gov/data-products/county-level-data-sets/>.

"United States Census Data." *OpenIntro*. United States Census Bureau, 2010. Web. 3 June 2017. <https://www.openintro.org/stat/data/cc.php>.

Voter Data from 2008 and 2016 elections provided by Professor Sundstrom