

Modeling Michigan's 2016 Vote

Fabian McNally

Haley Smith

Imani Gooden

Maeve McCoy

Tunde Oguntimein

Will Keigler

Today's Presentation

- Background
- Objective
- Variables
- Data Analysis
- Data Limitations
- Conclusion

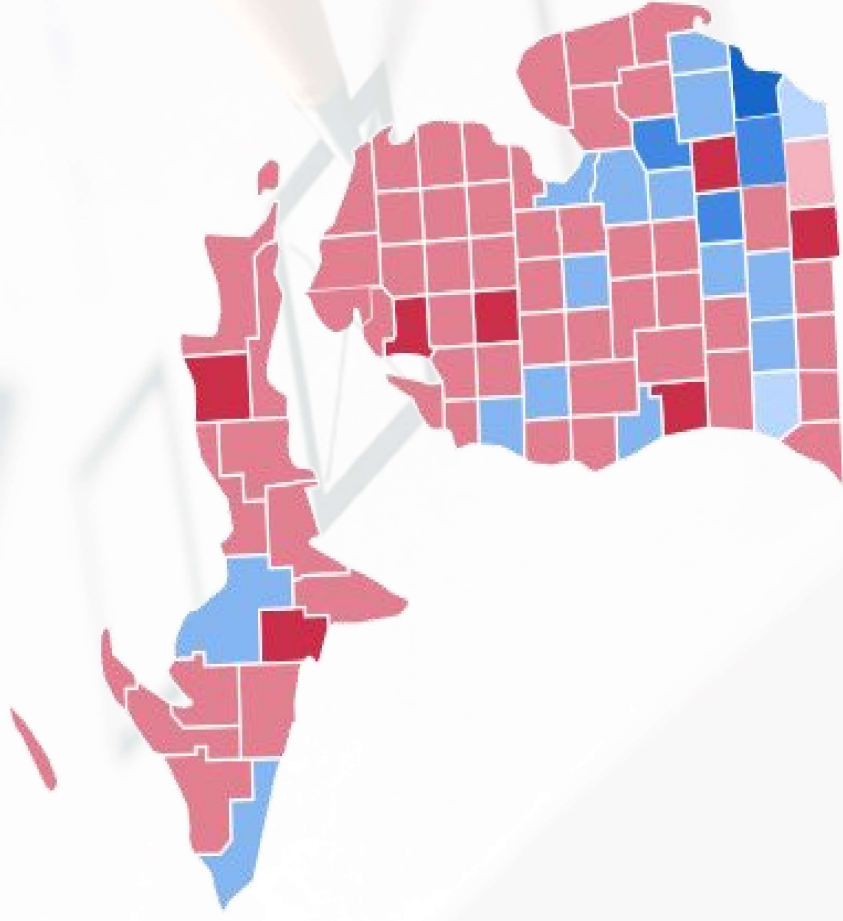


Background

- To the surprise of many, including data analysts who model election results, Donald Trump was elected President in November 2016
- The surprising results included President Trump (R) winning MI by less than 0.5% despite President Obama (D) carrying the state by double digit margins in '08 and '12
- Post election coverage has focused on the fact that there were numerous indicators that traditional publications/entities chose to ignore that would have more accurately predicted the election's outcome



2012 Election Heat Map



Obama

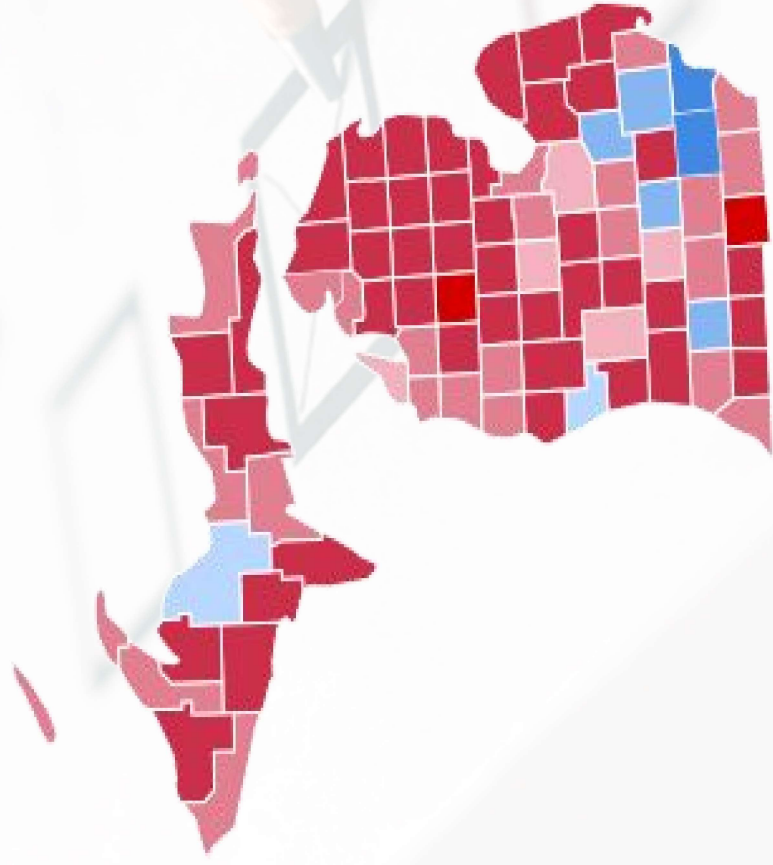
- 40-50%
- 50-60%
- 60-70%
- 70-80%

Romney

- 40-50%
- 50-60%
- 60-70%



2016 Election Heat Map



Objective

- Build a regression model using publicly available information about the Michigan electorate that would reliably predict which counties voted Republican or Democrat in the 2012 election
- Determine if that regression model could have predicted Michigan's 2016 surprising election results

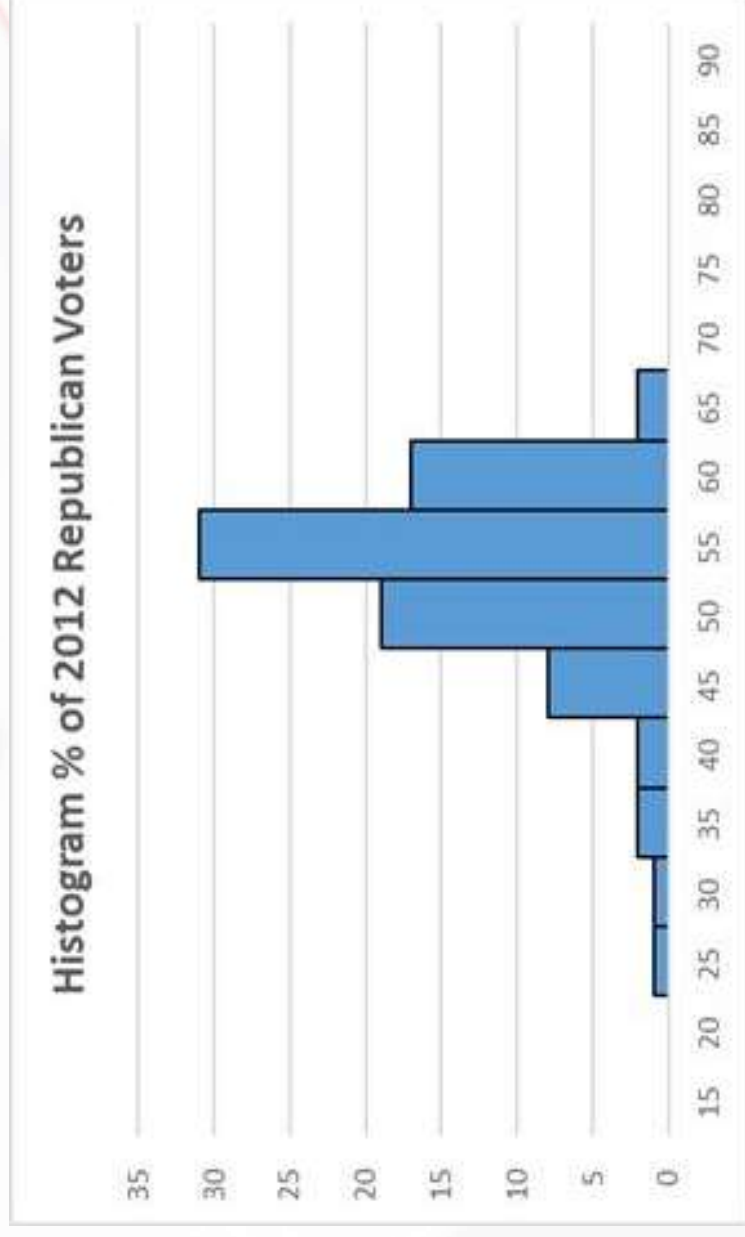


Variables

Variable	Description
County	County Name (83 total)
% Rep 2012	Percentage of County Voters who voted Republican
% Dem 2012	Percentage of County Voters who voted Democrat
% Other	Percentage of County Voters who voted Other
UIC	Urban Influence Code, 2003
Median Income	Estimate of median household income
Population	Population of County
% Poverty	Estimate of people of all ages in poverty
Labor Part	Labor participation population
Labor %	Labor participation rate
Unempl Rate	Unemployment Rate
% < HS	Percentage of population with less than a high school degree
% HS Only	Percentage of population with a high school degree only
% <= HS	Percentage of population with a high school degree or less
% Some College	Percentage of population with some college degree
% Bachelors +	Percentage of population with bachelors or more



2012 Data Analysis: Histogram

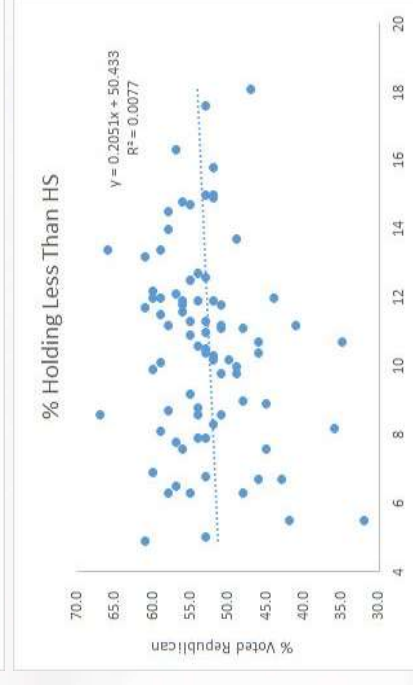
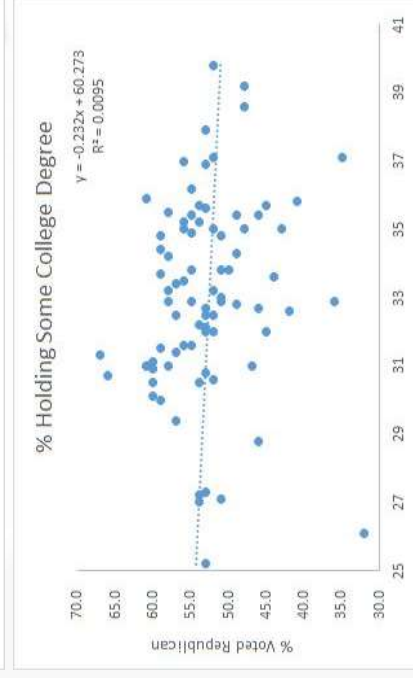
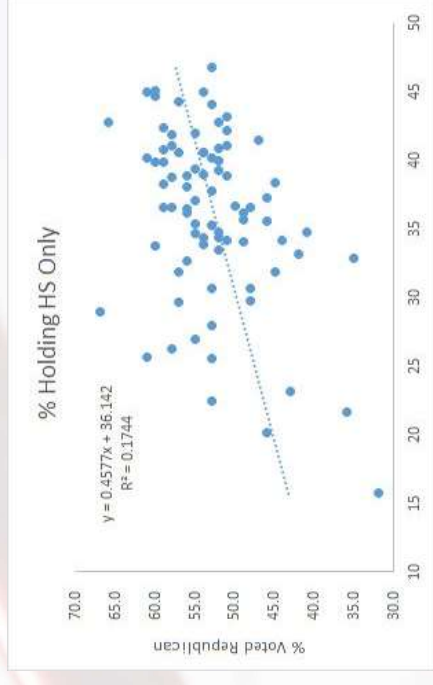
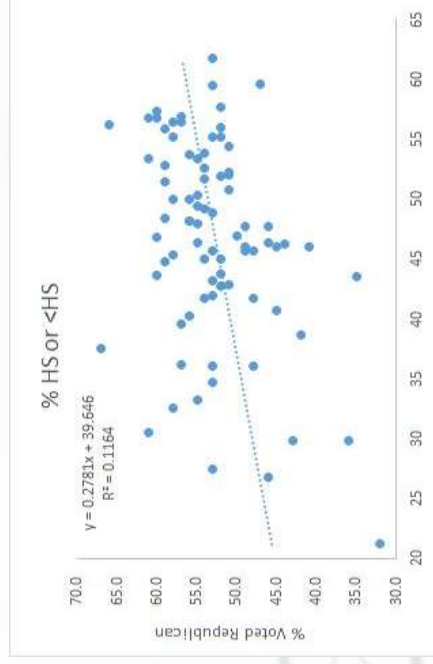
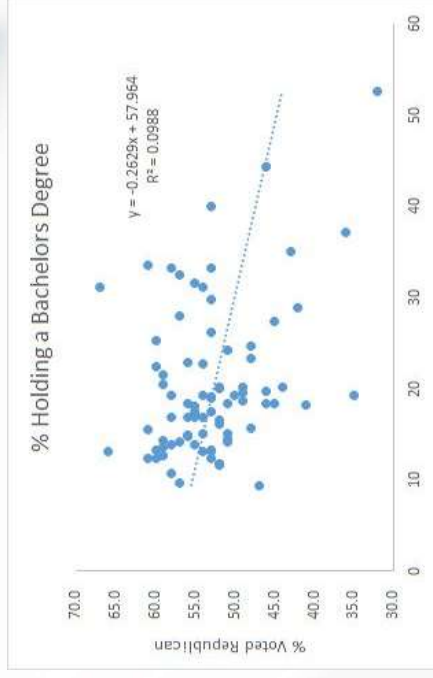


Mean	52.61
Median	53.00
Stdev	6.88

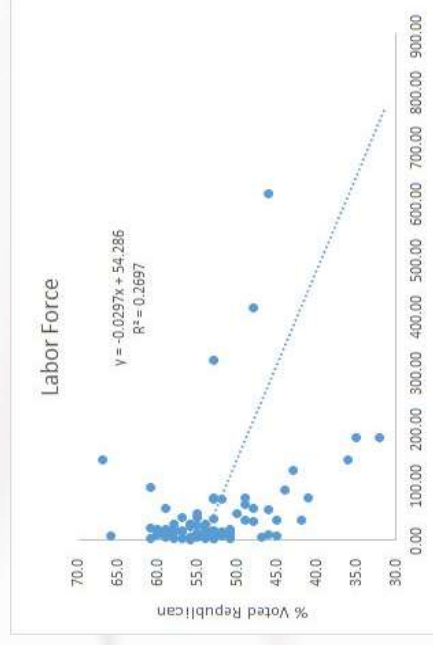
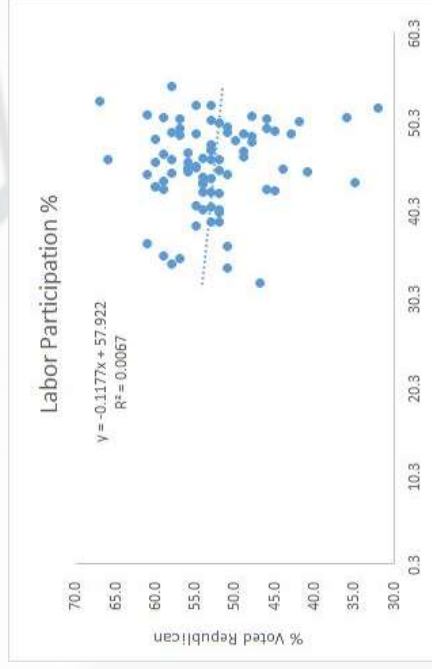
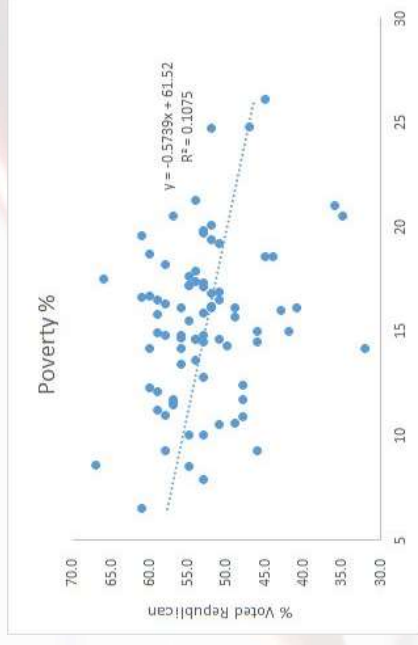
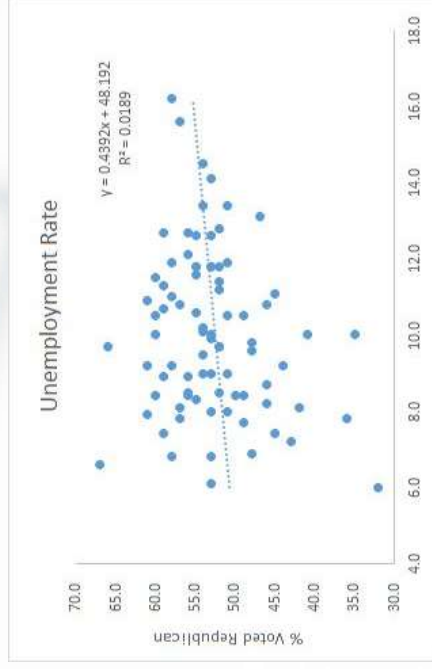
2012 Data: Correlation

	% Rep 2016	UIC	Median Income	% ≤ HS	% Bachelors +	Population	% Poverty	Econ Type	Labor Pop	Unempl Rate	% < HS	% HS Only	% Some College
% Rep 2016	1.000												
UIC	0.458 (0.325)	1.000											
Median Income		0.425 (0.587)	1.000										
% ≤ HS	0.707 (0.712)	0.425 (0.351)	0.722 (0.722)	1.000									
% Bachelors +			0.698	0.940 (0.940)	1.000								
Population		0.484 (0.592)	0.324	0.363 (0.363)	0.390	1.000							
% Poverty		0.194 (0.031)	0.754 (0.754)	0.464 (0.423)	0.423 (0.423)	0.024	1.000						
Econ Type	0.102 (0.593)	0.650 (0.496)	0.414 (0.414)	0.125 (0.397)	0.035 (0.035)	0.321 (0.321)	0.215 (0.014)	1.000					
Labor Pop			0.365 (0.684)	0.548 (0.548)	0.429 (0.511)	0.996 (0.255)	0.371 (0.371)	0.521 (0.521)	1.000 (0.284)				
Unempl Rate	0.362 (0.411)	0.655 (0.199)	0.625 (0.625)	0.816 (0.816)	0.750 (0.750)	0.042 (0.042)	0.621 (0.621)	0.082 (0.082)	0.439 (0.072)	1.000 (0.439)			
% < HS		0.478 (0.757)	0.679 (0.679)	0.963 (0.963)	0.913 (0.913)	0.469 (0.469)	0.333 (0.333)	0.130 (0.130)	0.500 (0.500)	0.531 (0.531)	0.629 (0.629)	1.000 (1.000)	
% HS Only													1.000 (0.208)
% Some College	0.032 (0.032)	0.242 (0.242)	0.121 (0.121)	0.240 (0.240)	0.104 (0.104)	0.050 (0.050)	0.153 (0.153)	0.266 (0.266)	0.062 (0.062)	0.145 (0.145)	0.247 (0.247)	0.208 (0.208)	1.000 (1.000)

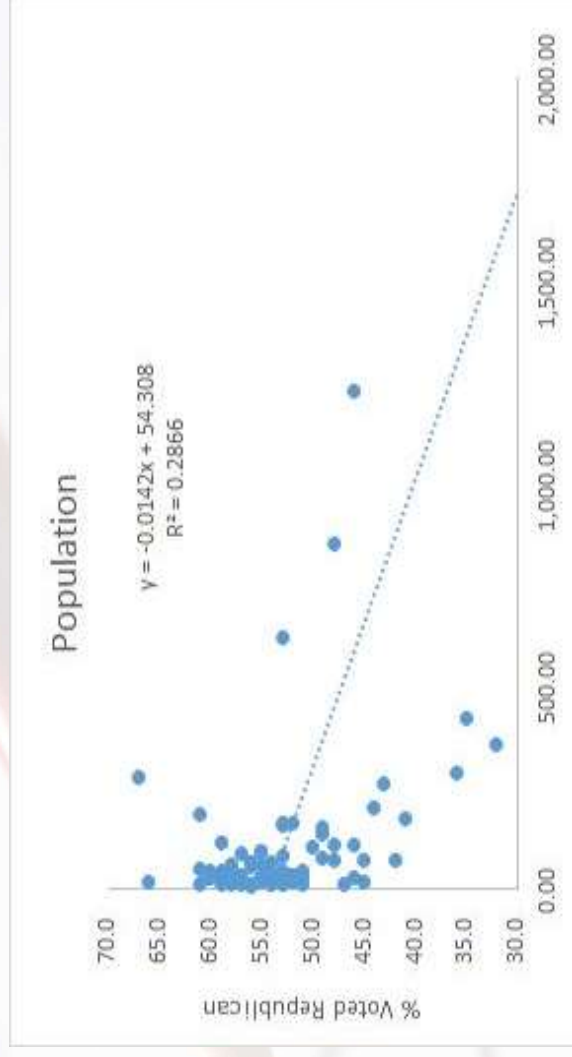
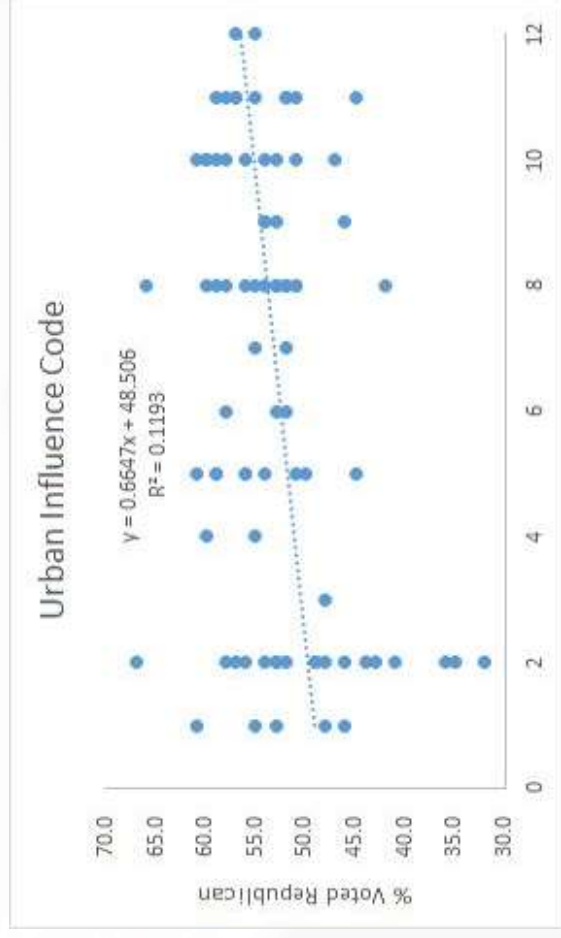
2012 Data Analysis: Education



2012 Data Analysis: Economy



2012 Data Analysis: Other



2012 Data Analysis: Full Model

Regression Statistics	
Multiple R	0.779
R Square	0.607
Adjusted R Square	0.540
Standard Error	4.669
Observations	83

ANOVA					
	df	SS	MS	F	Significance F
Regression	12	2,357.40	196.45	9.01	0.0000
Residual	70	1,526.27	21.80		
Total	82	3,883.66			

	Coefficients	Standard Error	t Stat	P-value
Intercept	(1,265.593)	878.395	(1.441)	0.154
UIC	0.604	0.266	2.267	0.026
% Poverty	(0.249)	0.309	(0.805)	0.423
Median Income	0.000	0.000	2.596	0.011
Econ Type	(0.084)	0.375	(0.223)	0.824
Population	(0.000)	0.000	(1.854)	0.068
Labor Part	0.000	0.000	1.497	0.139
Labor %	22.635	20.418	1.109	0.271
Unempl Rate	0.378	0.478	0.791	0.432
% < HS	13.186	8.760	1.505	0.137
% HS Only	13.059	8.733	1.495	0.139
% Some College	12.633	8.737	1.446	0.153
% Bachelors +	12.561	8.741	1.437	0.155



2012 Data Analysis: Best Model

Regression Statistics			
Multiple R	0.779		
R Square	0.607		
Adjusted R Square	0.540		
Standard Error	4.669		
Observations	83		

ANOVA					
	df	SS	MS	F	Significance F
Regression	12	2,357.40	196.45	9.01	0.0000
Residual	70	1,526.27	21.80		
Total	82	3,883.66			

	Coefficients	Standard Error	t Stat	P-value
Intercept	(1,265.593)	878.395	(1.441)	0.154
UIC	0.604	0.266	2.267	0.026
% Poverty	(0.249)	0.309	(0.805)	0.423
Median Income	0.000	0.000	2.596	0.011
Econ Type	(0.084)	0.375	(0.223)	0.824
Population	(0.000)	0.000	(1.854)	0.068
Labor Part	0.000	0.000	1.497	0.139
Labor %	22.635	20.418	1.109	0.271
Unempl Rate	0.378	0.478	0.791	0.432
% < HS	13.186	8.760	1.505	0.137
% HS Only	13.059	8.733	1.495	0.139
% Some College	12.633	8.737	1.446	0.153
% Bachelors +	12.561	8.741	1.437	0.155

Regression Statistics			
Multiple R	0.759		
R Square	0.576		
Adjusted R Square	0.548		
Standard Error	4.627		
Observations	83		

ANOVA					
	df	SS	MS	F	Significance F
Regression	5	2,235.44	447.09	20.89	0.0000
Residual	77	1,648.23	21.41		
Total	82	3,883.66			

	Coefficients	Standard Error	t Stat	P-value
Intercept	12.818	13.507	0.949	0.346
UIC	0.557	0.207	2.691	0.009
Median Income	0.457	0.160	2.863	0.005
% <= HS	0.489	0.090	5.401	0.000
Population	(0.009)	0.002	(3.824)	0.000
% Poverty	(0.424)	0.238	(1.783)	0.079

Correlation Matrix

	UIC	Median InC	% <= HS	Pop %	% Poverty
UIC	1.000				
Median Income	-0.587	1.000			
% <= HS	0.342	-0.722	1.000		
Population	-0.535	0.312	-0.353	1.000	
% Poverty	-0.328	-0.754	0.464	0.035	1.000



2012 Data Analysis: Best Model

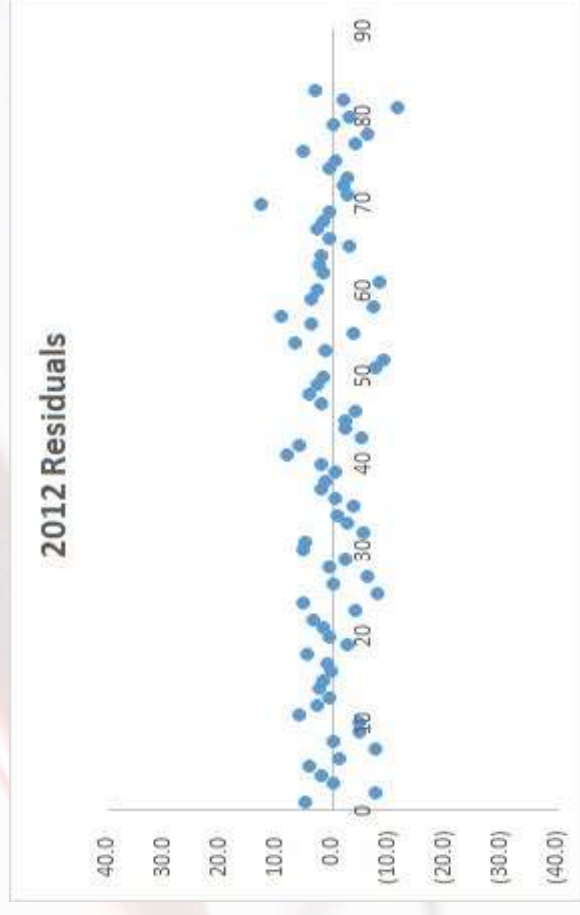
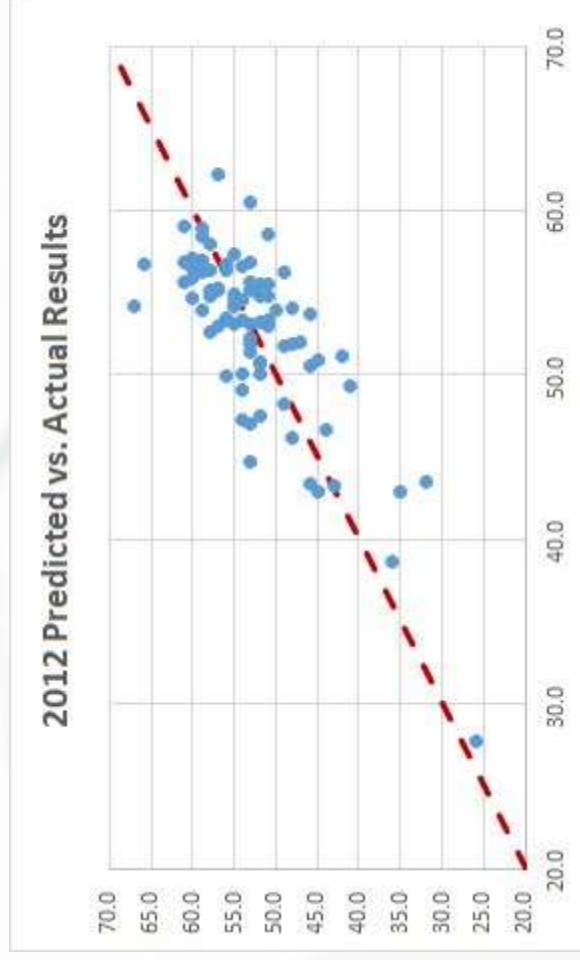
- Once we determined our best regression model (highest relative adjusted R-squared), we tested its predictive power for each individual county
- Our model accurately predicts 68 of 83 counties (82%), with a slight Republican bias

Accurate	68
Missed	15
Total	83
Pred R, D	11
Pred D, R	4

2012 ANALYSIS											
*12 Coefficients		0.56	0.46	0.49	(0.01)	(0.42)	12.82				
UIC	Median In: % <= HS	Population	% Poverty	% d	*12 %	Predicted	*12 % Rep Error	Pred Rep	*12 Act Rep Error	Accurate Missed	68 15
10	37.6	514	10.6	15.8	53.9	5.1	53.0	1	1	1	0
10	43.3	54.4	9.5	14.6	58.5	51.0	(7.5)	1	1	1	0
5	55.6	48.4	11.9	11.2	58.9	59.0	0.1	1	1	1	0
8	40.7	43.7	29.2	16.2	50.1	52.0	1.9	1	1	1	0
10	49.1	43.6	23.3	14.2	55.9	60.0	4.1	1	1	1	0
7	37.1	57.7	15.5	20.1	53.2	52.0	(1.2)	1	1	1	0
10	42.4	61.7	8.7	17.3	60.5	53.0	(7.5)	1	1	1	0
2	57.7	45.3	59.1	9.3	57.9	58.0	0.1	1	1	1	0
2	46.3	46.3	107.1	14.5	50.6	46.0	(4.6)	1	1	1	0
8	48.1	42.8	17.4	10.5	55.6	51.0	(4.6)	1	1	1	0
2	46.6	41.9	156.1	17.1	47.0	53.0	6.0	0	1	1	1
5	47.1	52.8	43.8	14.9	56.2	59.0	2.8	1	1	1	0
2	43.9	45.7	134.8	16.1	48.2	49.0	0.8	1	1	1	0
11	49.9	48.1	52.1	14.8	53.5	56.0	2.5	1	1	1	0
11	48.3	39.6	26.1	11.6	55.2	57.0	1.8	1	1	1	0
11	41.0	50.3	25.8	17.6	54.6	55.0	0.4	1	1	1	0
8	42.5	48.8	39.0	19.7	51.8	53.0	1.2	1	1	1	0
6	33.3	55.1	30.8	24.7	47.5	52.0	4.5	1	1	1	1
2	65.1	34.7	76.5	10.0	55.7	53.0	(2.7)	1	1	1	0
12	40.9	47.9	14.0	17.2	54.2	55.0	0.8	1	1	1	0
8	43.0	43.2	36.8	14.8	51.4	53.0	1.6	1	1	1	0
8	47.6	46.8	26.2	12.3	56.4	60.0	3.6	1	1	1	0
2	56.8	36.0	108.0	10.9	51.8	48.0	(3.8)	1	1	1	0
11	49.8	32.5	32.9	11.0	52.6	58.0	5.4	1	1	1	0
2	44.2	43.5	418.0	20.5	42.9	35.0	(7.9)	0	0	0	0
6	40.7	55.2	25.5	19.8	53.1	53.0	(0.1)	1	1	1	0
11	38.4	46.0	16.0	18.6	50.9	45.0	(5.9)	1	1	1	0
8	54.9	33.2	89.0	8.5	54.1	55.0	0.9	1	1	1	0
5	42.8	52.2	42.0	16.5	53.3	51.0	(2.3)	1	1	1	0
5	46.8	53.4	46.3	16.6	55.6	61.0	5.4	1	1	1	0
8	41.8	41.7	36.7	17.4	49.0	54.0	5.0	1	1	1	0
11	45.9	56.4	32.5	11.7	62.2	57.0	(5.2)	1	1	1	0
2	47.5	23.8	262.4	21.0	38.7	36.0	(2.7)	0	0	0	0
5	51.0	49.9	63.9	14.2	56.7	56.0	(0.7)	1	1	1	0
11	39.9	51.9	25.4	16.1	55.5	52.0	(3.5)	1	1	1	0
9	36.6	53.8	11.6	14.6	54.5	54.0	(0.5)	1	1	1	0
5	42.0	40.7	70.7	26.1	42.9	45.0	2.1	0	0	0	0
2	50.7	45.0	160.1	16.2	50.7	52.0	1.3	1	1	1	0
2	52.3	29.8	255.3	16.0	43.2	43.0	(0.2)	0	0	0	0



2012 Data Analysis: Residuals



Low heteroskedasticity

2016 Data Analysis: 2012 Test

- Using the 2012 regression model, we sought to predict election 2016 results of Michigan counties
- Accurately predicts 74 of 83 counties (89%) with a slight Democrat bias

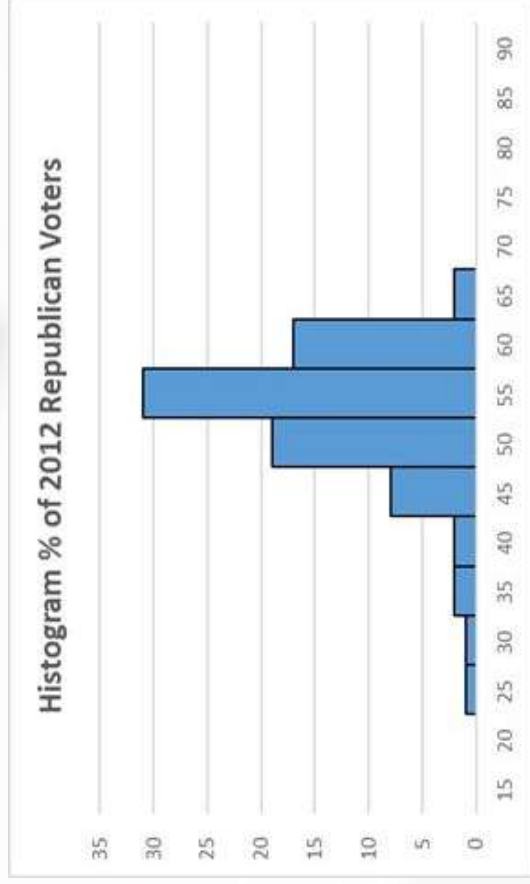
Accurate	74
Missed	9
Total	83
Pred R, D	2
Pred D, R	7

2016 ANALYSIS												0.891566	Accurate	74
*12 Coefficients													Missed	9
UIC	Median Income	% ≤ HS	0.46	0.49	0.01	(0.42)	12.82	*16 % Rep Error			Pred Rep	*16 Act Rep Error	Pred R, C	2
					Population	% Poverty	Predicted						Pred D, F	7
					n		d *16 %							
10	37,618	51.4	10,352	15.8	53.9	14.1	1	68.0	14.1	1	1	1	1	0
10	43,341	54.4	9,219	14.6	58.5	11.2	1	57.3	11.2	1	1	1	1	0
5	55,611	48.4	115,548	11.2	58.3	2.4	1	61.3	2.4	1	1	1	1	0
8	40,685	43.7	28,704	18.2	50.1	11.8	1	61.9	11.8	1	1	1	1	0
10	43,116	43.6	23,144	14.2	55.9	6.5	1	62.4	6.5	1	1	1	1	0
7	37,12	57.7	15,122	20.1	53.2	9.9	1	63.1	9.9	1	1	1	1	0
10	42,389	61.7	8,503	17.3	60.5	1.4	1	61.9	1.4	1	1	1	1	0
2	57,685	45.3	59,702	9.3	57.9	63.4	1	53.5	5.5	1	1	1	1	0
2	46,256	46.3	104,747	14.5	50.6	2.9	1	53.5	2.9	1	1	1	1	0
8	48,127	42.8	17,572	10.5	55.6	54.2	1	54.2	(1.4)	1	1	1	1	0
2	46,588	41.9	154,01	17.1	47.0	53.8	1	53.8	6.8	1	1	1	1	1
5	47,145	52.8	43,427	14.9	56.2	86.9	1	10.7	1	1	1	1	1	0
2	43,864	45.7	134,366	16.1	48.2	53.6	1	5.4	1	1	1	1	1	0
2	43,889	48.1	51,599	14.8	53.5	81.2	1	7.7	1	1	1	1	1	0
11	48,267	39.6	26,174	11.6	55.2	4.3	1	59.5	4.3	1	1	1	1	0
11	41,036	50.3	25,401	17.6	54.6	9.0	1	63.6	9.0	1	1	1	1	0
8	42,473	48.8	37,724	19.7	51.8	53.1	1	7.3	1	1	1	1	1	0
6	33.25	55.1	30,358	24.7	47.5	63.6	1	16.1	1	1	1	1	1	0
2	65,051	34.7	77,888	10	55.7	53.3	1	1	1	1	1	1	1	0
12	40,907	47.3	13,144	17.2	54.2	3.5	1	63.7	3.5	1	1	1	1	0
8	43.02	43.2	36,202	14.8	51.4	8.8	1	60.2	8.8	1	1	1	1	0
8	47,616	46.8	25,535	12.3	56.4	8.9	1	65.3	8.9	1	1	1	1	0
2	56,76	36	109,16	10.9	51.8	49.6	1	3.9	(2.2)	1	1	1	1	0
11	43,774	32.5	33,182	11	52.6	56.5	1	52.4	3.9	1	1	1	1	0
2	44,181	43.5	408,615	20.5	42.9	52.4	1	9.5	1	1	1	1	1	0
6	40,746	55.2	25,122	19.8	53.1	85.1	1	12.0	1	1	1	1	1	0
11	38,416	46	15,243	18.6	50.9	4.1	1	55.0	4.1	1	1	1	1	0
8	54,874	33.2	92,084	8.5	54.1	53.3	1	(0.8)	1	1	1	1	1	0
5	42.8	52.2	41,202	16.5	53.3	6.8	1	60.1	6.8	1	1	1	1	0
5	46,786	53.4	45,774	16.6	55.6	70.9	1	15.3	1	1	1	1	1	0
8	41,775	41.7	36,555	17.4	49.0	54.2	1	5.2	1	1	1	1	1	0
11	45,917	56.4	31,481	11.7	62.2	67.1	1	4.9	1	1	1	1	1	0
2	47.47	29.8	268,051	21	36.6	60.9	1	22.3	1	1	1	1	1	0
5	51,013	49.3	64,232	14.2	56.7	62.1	1	5.4	1	1	1	1	1	0
11	38,874	51.3	25,327	16.1	55.5	82.5	1	7.0	1	1	1	1	1	0
9	36.57	53.8	11,155	14.6	54.5	62.2	1	7.7	1	1	1	1	1	0
5	41,957	40.7	71,282	26.1	42.9	48.7	1	5.8	1	1	1	1	1	1
2	50,683	45	158,46	16.2	50.8	57.2	1	6.4	1	1	1	1	1	0
2	52,252	29.8	261,854	16	43.2	53.3	1	10.1	1	1	1	1	1	0
8	42,829	55.8	17,263	16.5	57.0	69.7	1	12.7	1	1	1	1	1	0



[illegible]

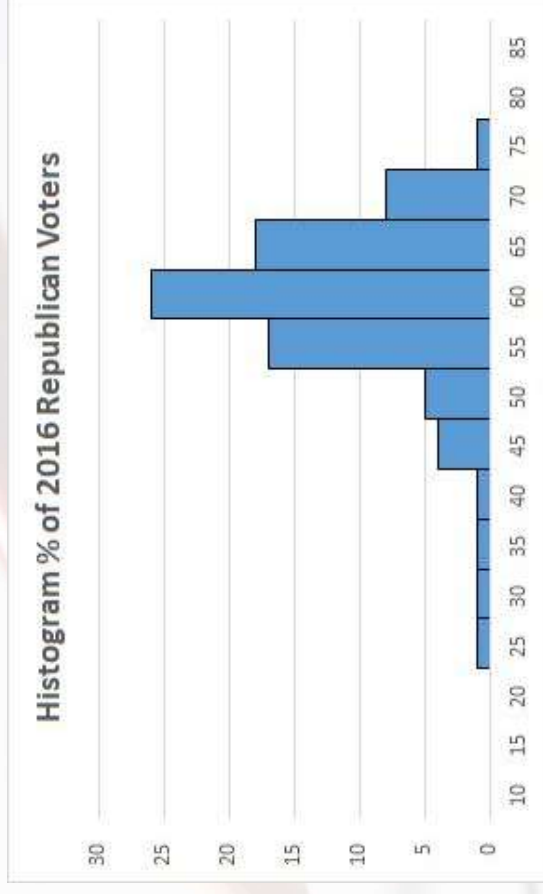
2012 & 2016 Data: Histogram



Mean **52.61**

Median **53.00**

Stdev **6.88**



Mean **58.64**

Median **61.20**

Stdev **8.9**

2016 Data Analysis: Best Model

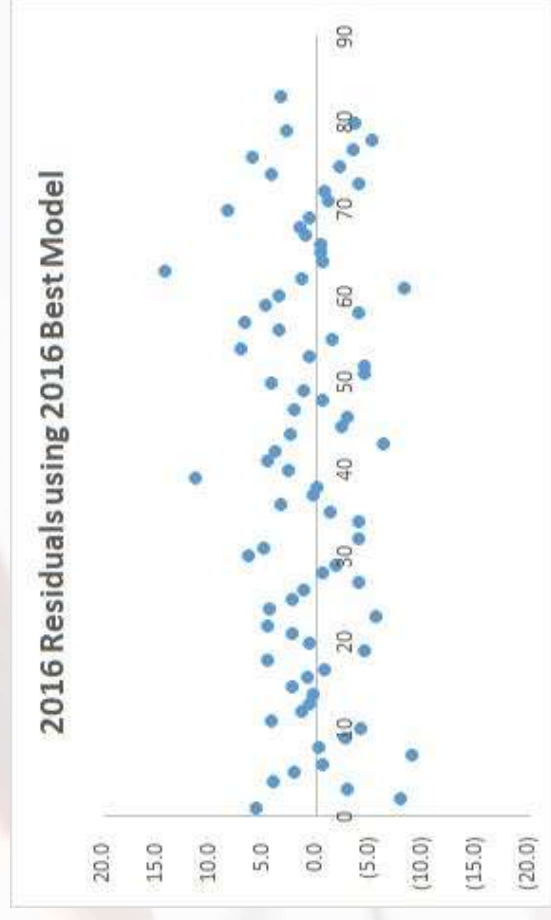
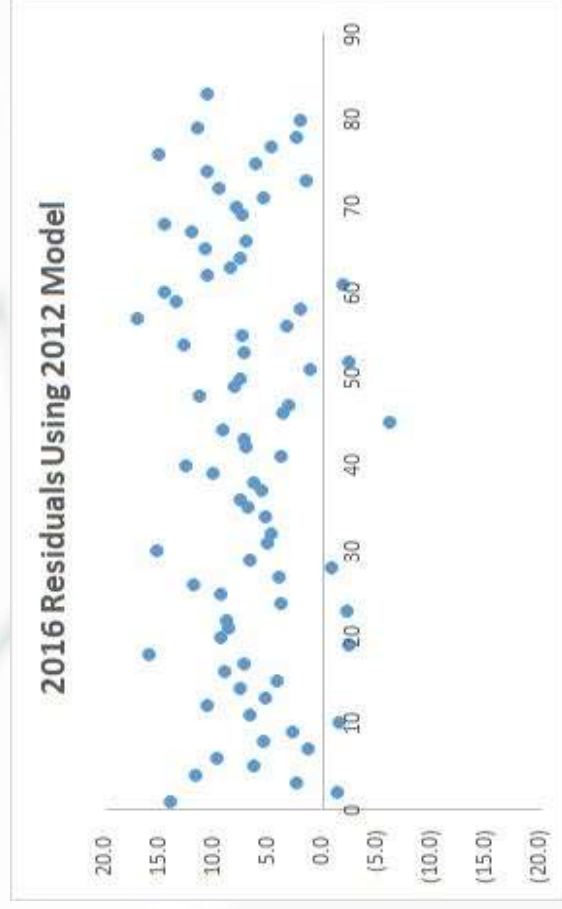
- The best 2016 regression model accurately predicts 78 of 83 countries (93%)

Accurate	78
Missed	5
Total	83
Pred R, D	3
Pred D, R	2

2016 ANALYSIS										0.33376		Accurate	78
*16 Coefficients												Missed	5
0.83 0.70 0.44 (0.58) (0.01) 13.37												Pred R,	3
UIC												Pred D,	2
	Median Income	% <= HS	% Bachelo	Population	Predicted '16	'16 % Re Error	Pred Rep	'16 Act Re	Error				
10	37,618	51.4	14.3	10,352	62.3	68.0	1	1	1	0			
10	43,341	54.4	18.4	9,219	65.2	57.3	1	1	1	0			
5	55,611	48.4	21.5	115,548	64.2	61.3	1	1	1	0			
8	40,665	43.7	16.5	28,704	57.9	61.9	1	1	1	0			
10	49,118	43.6	25.3	23,144	60.3	62.4	1	1	1	0			
7	37,12	57.7	11.7	15,122	63.7	63.1	1	1	1	0			
10	42,389	61.7	13.1	8,503	70.9	61.9	1	1	1	0			
2	57,665	45.3	19.2	59,702	63.6	63.4	1	1	1	0			
2	48,256	46.3	18.3	104,747	56.2	53.5	1	1	1	0			
8	48,127	42.8	24.2	17,572	58.3	54.2	1	1	1	0			
2	46,588	41.9	26.1	154,01	49.5	53.8	1	1	1	0			
5	47,146	52.8	13.5	43,427	65.6	66.9	1	1	1	0			
2	43,864	45.7	20.1	134,386	53.0	53.6	1	1	1	0			
2	43,889	48.1	16.3	51,599	60.9	61.2	1	1	1	0			
11	48,267	39.6	27.9	26,174	57.2	53.5	1	1	1	0			
11	41,036	50.3	18.1	25,401	62.7	63.6	1	1	1	0			
8	42,473	48.8	19.1	37,724	59.8	59.1	1	1	1	0			
6	33,25	55.1	11.6	30,358	59.0	63.6	1	1	1	0			
2	65,051	34.7	29.7	77,888	57.8	53.3	1	1	1	0			
12	40,307	47.9	17.2	13,744	63.0	63.7	1	1	1	0			
8	43,02	43.2	16.9	36,202	57.8	60.2	1	1	1	0			
8	47,616	46.8	22.4	25,535	60.7	65.3	1	1	1	0			
2	56,76	36	24.7	103,16	55.2	43.6	1	1	1	0			
2	49,774	32.5	33.2	33,182	52.0	56.5	1	1	1	0			
2	44,181	43.5	19.3	408,615	50.1	52.4	1	1	1	0			
6	40,746	55.2	12.3	25,122	63.9	65.1	1	1	1	0			
11	38,416	46	16.3	15,243	58.9	55.0	1	1	1	0			
8	54,874	33.2	31.5	92,084	53.8	53.3	1	1	1	0			
5	42.8	52.2	14.1	41,202	62.0	60.1	1	1	1	0			
5	46,786	53.4	15.5	45,774	64.4	70.9	1	1	1	0			
8	41,775	41.7	31.1	36,555	49.2	54.2	1	1	1	0			
11	45,917	56.4	14.1	31,481	71.0	67.1	1	1	1	0			
2	47.47	23.8	37.2	288,051	37.0	60.9	1	1	1	0			
5	51,013	49.9	14.9	64,232	66.0	62.1	1	1	1	0			
11	33,874	51.9	16.1	25,327	63.7	62.5	1	1	1	0			
9	36.57	53.8	19.3	11,195	58.9	62.2	1	1	1	0			
5	41,957	40.7	27.3	71,282	48.3	48.7	1	1	1	0			
2	50,683	45	20	158,46	57.2	57.2	1	1	1	0			
2	52,252	29.8	35.1	261,654	41.8	53.3	1	1	1	0			
8	42,829	55.8	12.7	17,263	67.1	63.7	1	1	1	0			



2016 Data Analysis: Residuals



2012 model applied to 2016 has a democrat bias (positive error means actual % republican higher than predicted). This suggests we did not account for a incumbent party factor

Data Limitations

- There was no measure, explicit or proxy, to factor incumbency and any potential natural desire for change
- Also note that the population of the state voting determines electoral votes, not counties



Conclusion

- Our model had significant predictive power for 2016, in fact, as demonstrated by its lower adjusted R^2 , our model predicted 2016 better than it did 2012.
- Despite national discussion around “economic anxiety,” the 2016 model was improved by removing the % poverty as a variable and adding a variable denoting higher education (bachelors degree)



Michigan Election Results 2012 to 2016

