Looking at 2016 Presidential Election from Geography Perspective

Ziqi Li
School of Geographical Sciences and Urban Planning, Arizona State University

SPARC Spatial Analysis Research Center
Arizona State University

Introduction

This research examines the 2016 presidential election using a newly developed spatial statistical technique, Multi-scale Geographically Weighted Regression (***MGWR***), to examine spatial variations in the determinants of voting behavior (Fotheringham et al., 2017). The associations between 12 county-level ***socio-economic variables*** from the ACS (American Community Survey) and ***the percentage of the vote going to the Democratic party*** in each county were uncovered through regression with the research questions being "Do these associations vary across the country?" and "To what extent does geographical context affect the way people vote?".

Data and Model Specification

The dependent variable is county-level ***proportion of Democrat votes*** in straight fight between Democrats and Republicans after removing third parties. Eleven county-level socio-economic covariates were selected from American Community Survey 2015 5-year estimate datasets including ***sex ratio, % of population age 18 to 29, % of population age 65 and over, % of Black, % of Hispanic, household median income, % of college degrees, Gini coefficient, % of employment in manufacturing, log scale of population density and log scale of third party vote share***. Both dependent and independent variables were normalized with mean = 0 and standard deviation = 1.

Regional Regression Results

	Pacific Northwest	Pacific Southwest	South Central	North Central	Southeast	Northeast	National
Model Summary							
R-Square	0.662	0.805	0.747	0.544	0.850	0.875	0.639
Adj R-square	0.637	0.783	0.742	0.537	0.848	0.869	0.638
Parameter Estimates							
(Intercept)	1.962***	0.864**	-0.396***	0.121*	-0.138***	0.527***	0.000
sex_ratio	-0.133	-0.093	-0.045*	-0.049	-0.004	-0.063*	-0.006
age_18_29	-0.302***	-0.354**	-0.284***	-0.135***	0.001	-0.301***	-0.095***
age_65_over	-0.261**	-0.427***	-0.205***	-0.113*	0.063**	-0.329***	0.025
pct_black	2.115*	0.158	0.634***	0.481***	0.768***	0.573***	0.677***
pct_hispanic	0.160*	0.224**	0.377***	0.181**	0.388***	0.500***	0.298***
median_income	-0.445***	-0.311**	-0.530***	-0.433***	-0.179***	-0.561***	-0.192***
pct_bachelor_higher	0.626***	0.719***	0.564***	0.396***	0.388***	0.916***	0.506***
gini	0.143	0.160	-0.011	0.042	-0.021	-0.055	0.051**
pct_manuf	-0.030	-0.030	0.019	-0.024	-0.005	-0.089	0.063***
log(pop_density)	0.382***	0.120	0.147***	0.202***	0.051	0.005	0.162***
log(3rd_party)	-0.344**	-0.453***	0.348***	0.588***	0.318***	0.386***	0.372***

	0.001	0.01	0.05	Not Significant
Positive Significance	***	**	*	
Negative Significance	***	**	*	

Methods

Standard models assume the processes generating the data we observe are the same across space. **Geographically Weighted Regression (GWR)** removes this assumption and allows processes to vary over space – this is referred to as ***Spatial Heterogeneity.*** In this study, we used the recently derived Multi-scale GWR (***MGWR***) model, which allows the rate at which data are borrowed from nearby locations (***bandwidth***) to vary across the covariates. The formulation of MGWR is as follows (Fotheringham et al., 2017)

$$y_i = \beta_{bw0}(u_i, v_i) + \sum_{j=1}^{m} \beta_{bwj}(u_i, v_i) x_{ij} + \varepsilon_i$$

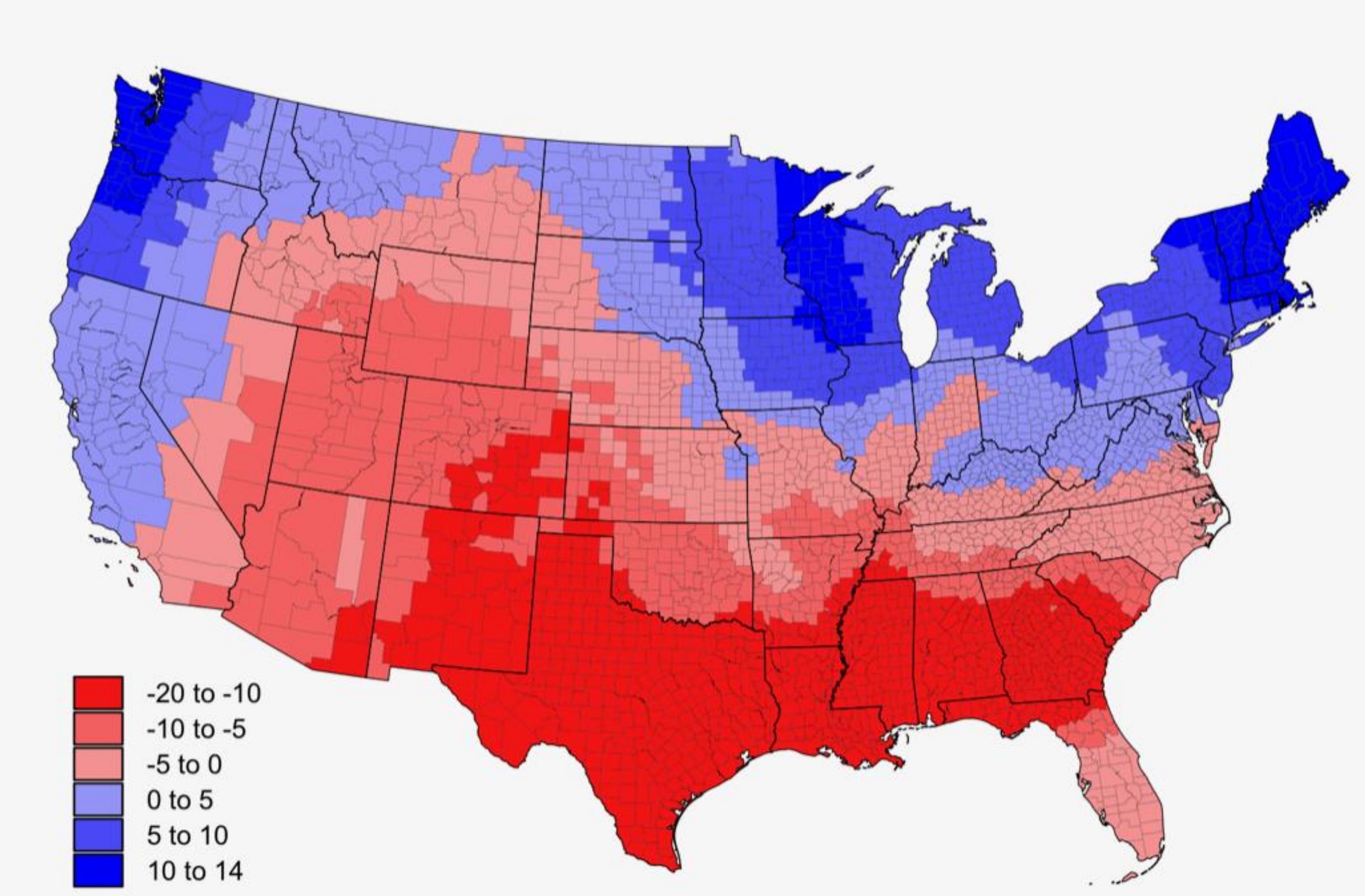
Since all the covariates are normalized, we establish two scenarios:

- **If all counties had an average socio-economic composition**, then the covariates $\boldsymbol{x_{ij}}$ will be all be zero and $y_i = \beta_{bw0}(u_i, v_i) + \varepsilon_i$ so that the predicted local intercept describes the **geographical contextual effect in voting Democrat**.
- **Conversely, if geography did NOT influence voting behavior** and we remove the intercept term, then $y_i = \sum_{j=1}^{12} \boldsymbol{\beta_{bwj}(u_i, v_i) x_{ij}} + \varepsilon_i$. This predicted value is then the democrat share of the vote which results solely from the socio-demographic composition of each county's population and has nothing to do with geography.

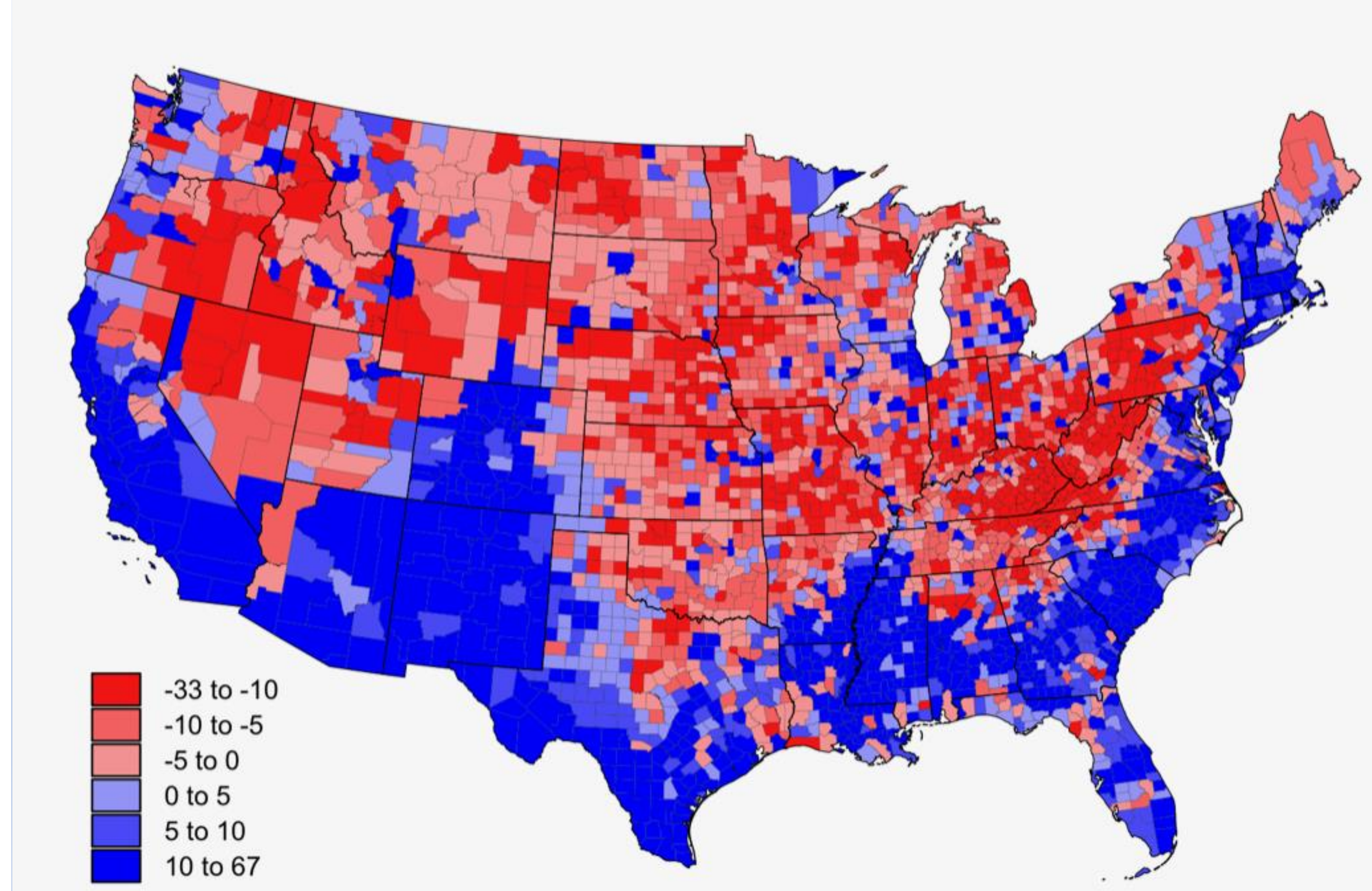
Multi-scale Geographically Weighted Regression Results

% of Democratic Vote = Geographic Context + Socio-Economics

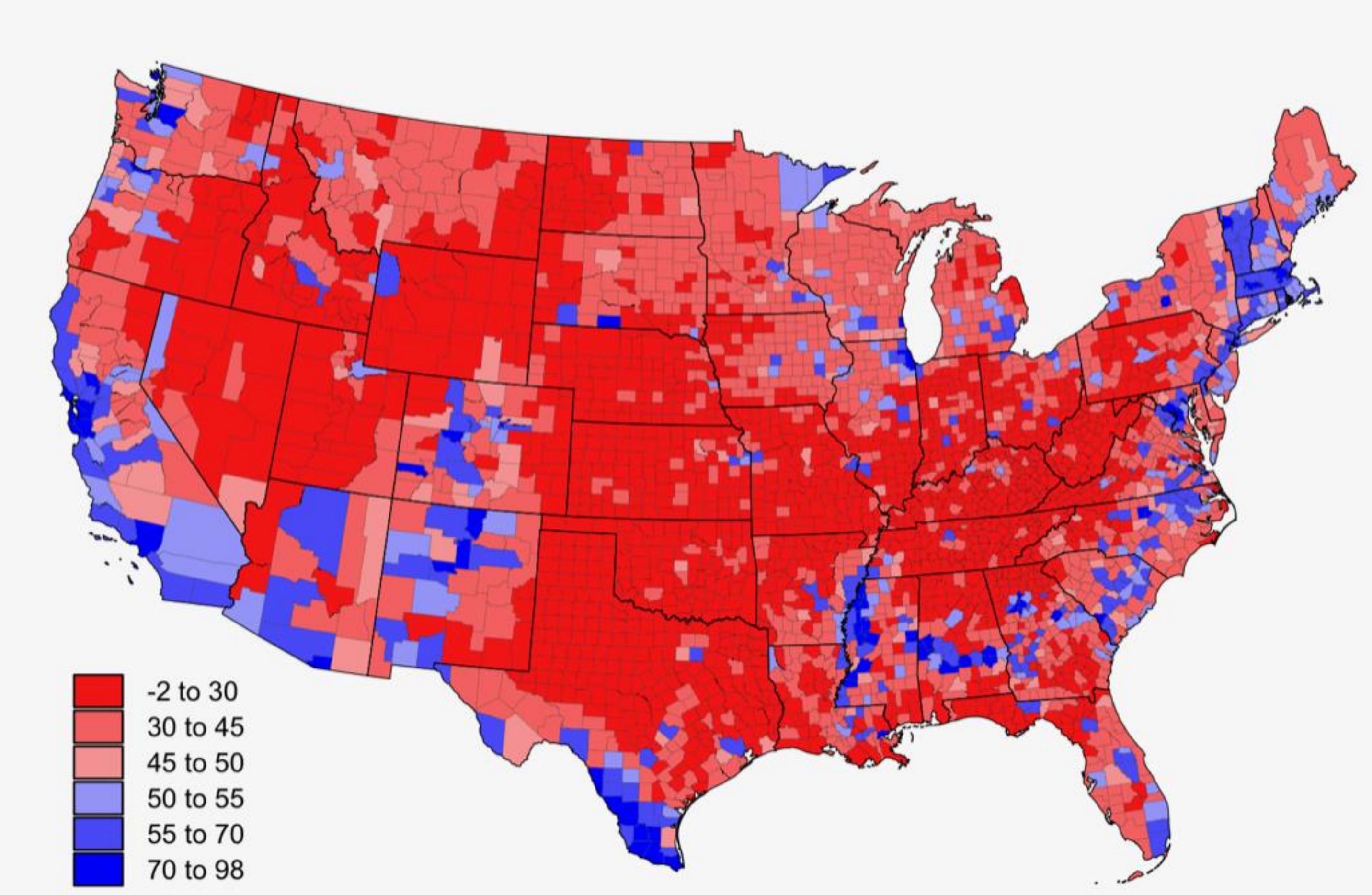
% Vote Gained/Lost due to Geography



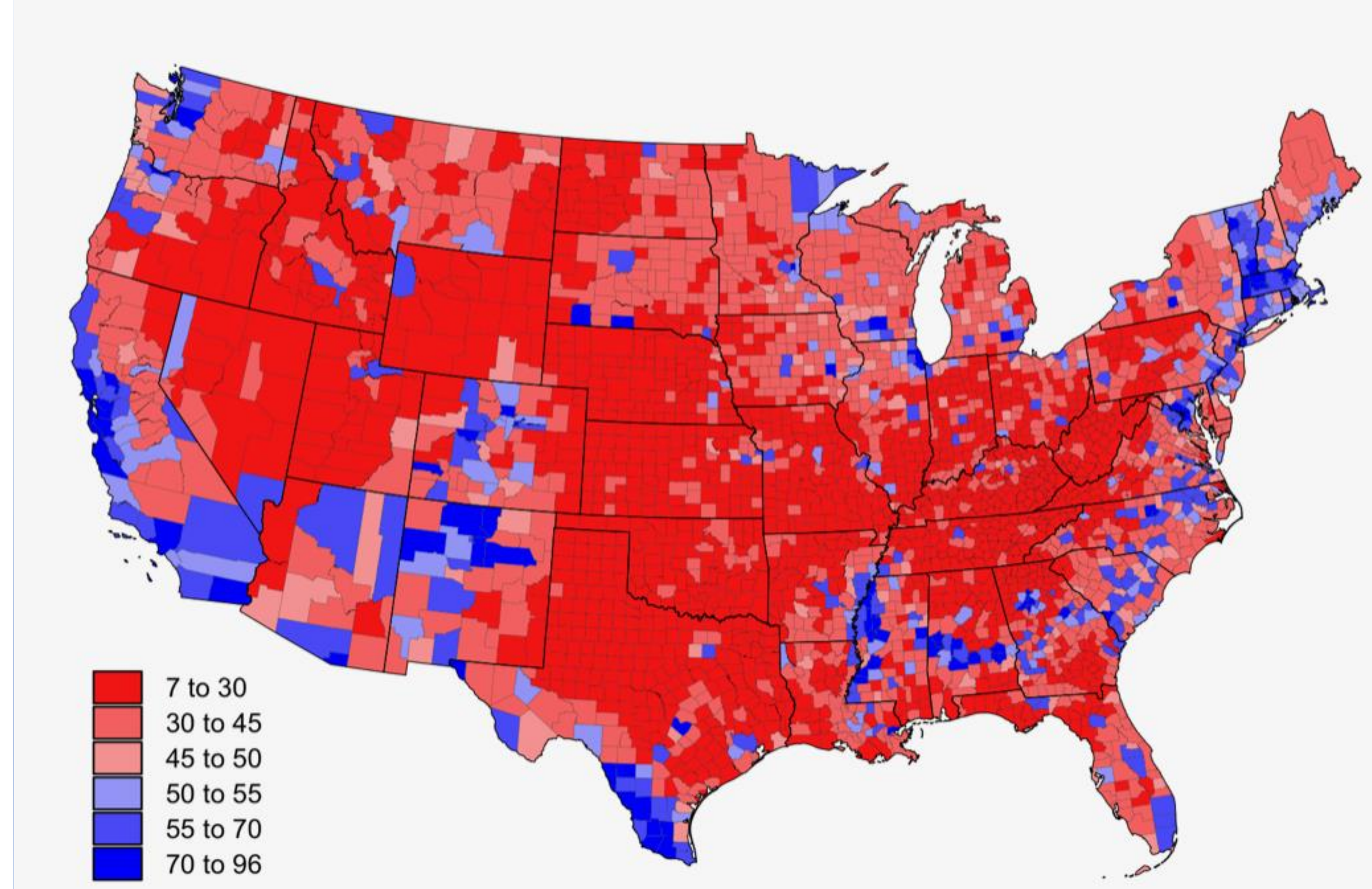
% Vote Gained/Lost due to Socioeconomics



Predicted % Vote for Democratic



Observed % Vote for Democratic



Conclusion

- Geographic context influences voting behavior significantly. Voters in southern US counties inherently favor the Republican party while voters in northern and west coast counties inherently favor the Democratic party.
- The socio-economic composition of southern US counties should favor the Democratic party, while Central and Northern US counties should favor the Republican party but these influences are outweighed by geography.
- The MGWR model replicates 94% of the variance in voting behavior compared to only 64% by the traditional OLS model.

Acknowledgement

I would like to thank Dr. Fotheringham to advice me on this study and provide critical comments and suggestions in making this poster.

Reference

Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale Geographically Weighted Regression (MGWR). *Annals of the American Association of Geographers*, *107*(6), 1247-1265.