

How Geography Affects Voting Behavior in the 2016 Presidential Election

Background

Geography is often assumed to play a large role in how Americans vote. In the 2016 election, rural Americans voted overwhelmingly for Donald Trump, while urban Americans sided with Hillary Clinton. I hypothesize that in addition to this urban/rural dichotomy, other continuous geographic variables may affect voting behavior, such as:

- distance from the nearest metropolitan area (more distant places likely tend to be more conservative);
- distance from the nearest university (college towns are usually more liberal than their surroundings);
- distance from the nearest international border (where residents may place special importance on certain political issues); and
- distance from the coastline (which may correlate with more liberal standpoints).

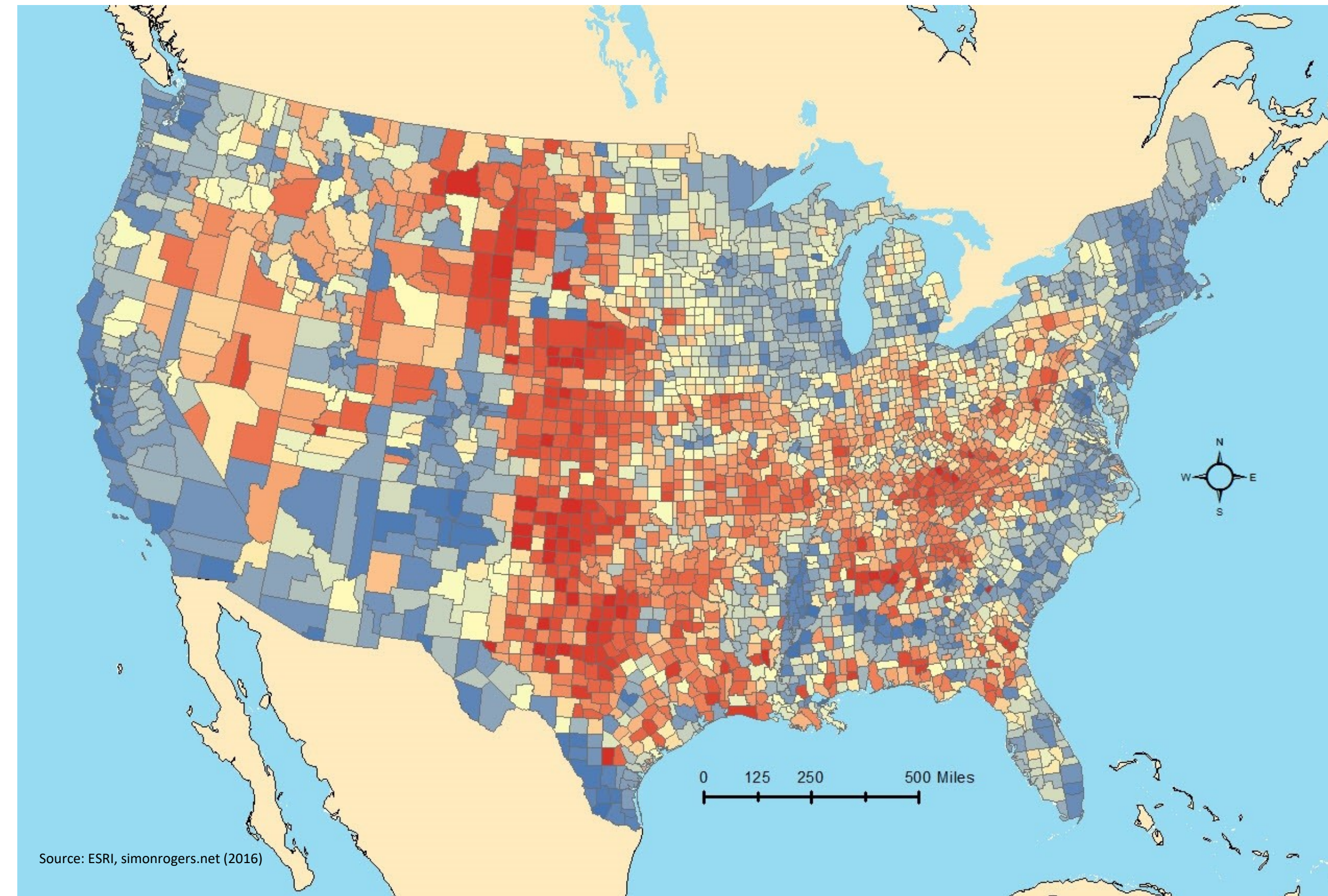
Controlling for demographic factors such as income, race, and education level, were geographic variables significant predictors of how Americans voted, at the county level, in the 2016 presidential election?

Methodology

For each county in the U.S., I calculated the distance between its centroid and the following geographic features: nearest metropolitan area (population 100,000+), nearest major college/university (enrollment 5,000+), nearest international border, and nearest coastline. I then found the natural log of these distances (see maps below).

Using GeoDa, I ran an ordinary least squares regression of the percentage of the vote that Donald Trump received by county, using as independent variables three demographic variables (percentage of the county's population that is non-Hispanic white, percentage with bachelor's degrees or higher, and median household income) and the four logged centroid-to-feature distances mentioned above. Then, I ran a spatial error regression in GeoDa to see whether there were statistically significant residuals that were not explained by the independent variables in the regression.

Potential limitations include the inability to conduct this analysis at the more granular precinct level due to lack of data, and using county centroids for distance calculations, which may not be sufficiently granular.



2016 U.S. presidential results by county. Redder counties voted more strongly for Trump, while bluer counties voted at a higher rate for Clinton.

Results

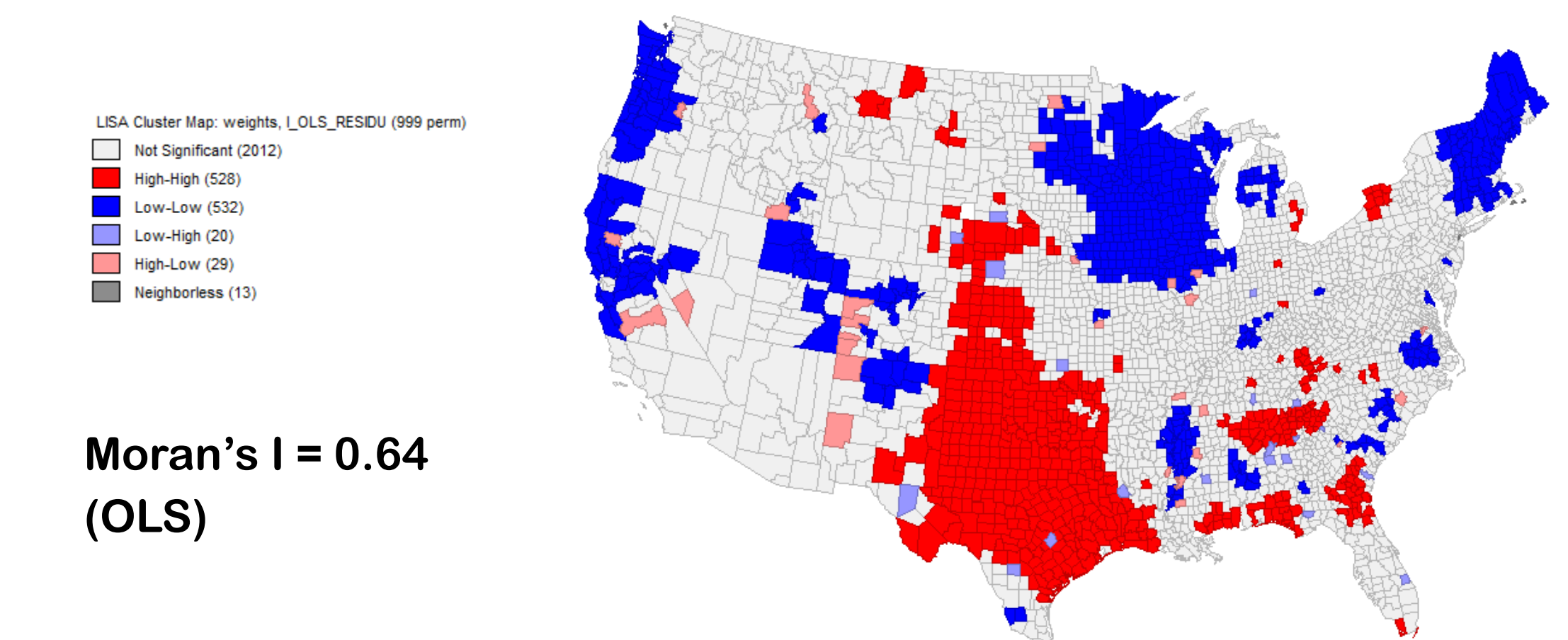
Running an ordinary least squares regression shows that the three demographic variables have strong and statistically significant effects on the dependent variable, especially the non-white Hispanic share of the population and the percentage with bachelor's degrees or higher. The logged distances are also statistically significant, but they explain much less of the variance in the dependent variable.

Dependent Variable :	TrumpVote	Number of Observations :	3134
Mean dependent var :	63.5523	Number of Variables :	8
S.D. dependent var :	15.5899	Degrees of Freedom :	3126
R-squared :	0.607412	F-statistic :	690.936
Adjusted R-squared :	0.606533	Prob(F-statistic) :	0

Variable	Coefficient	Std. Error	t-Statistic	Probability
CONSTANT	7.9413	1.81137	4.38414	0.00001
WhiteNonHispanic	0.38759	0.0102968	37.6418	0.00000
Bachelors	-0.935339	0.0285886	-32.7172	0.00000
MedianIncome	0.000248147	2.04536e-005	12.1322	0.00000
UnivNear	1.01811	0.26401	3.85635	0.00012
BorderNear	2.85556	0.178649	15.9842	0.00000
CoastlineNear	0.609725	0.147257	4.14056	0.00004
MetroNear	1.74892	0.232497	7.52234	0.00000

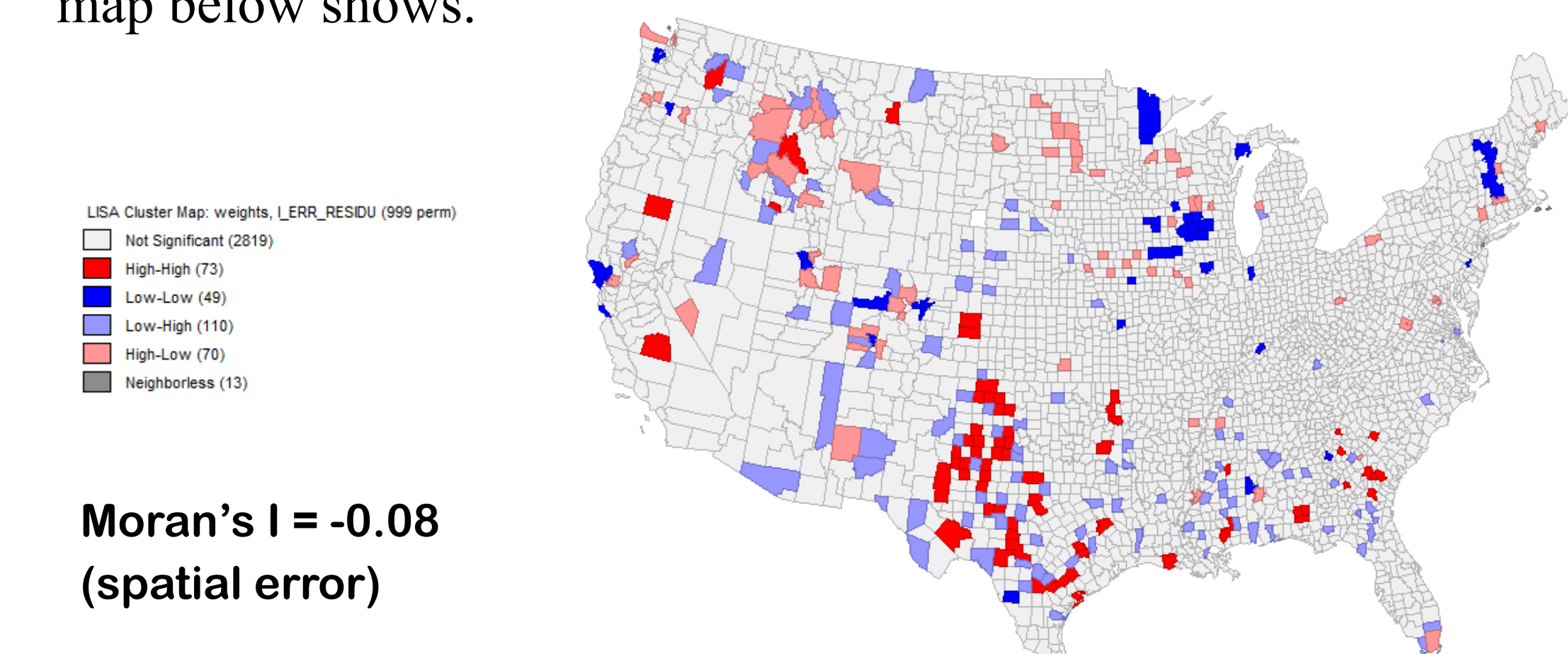
As I hypothesized, the further a county is from universities, metropolitan areas, coastlines, and international borders, the more likely it is to have voted for Trump, holding demographic variables constant. This effect is more pronounced for the metro area and international border distances, and less so with the university distance.

However, conducting a Moran's I test shows that even controlling for these variables, there is still a high degree of spatial autocorrelation between counties. The map below shows that counties in New England, the Upper Midwest, and the West Coast vote more heavily Democratic than the model predicted, while areas of Texas and the Deep South vote more Republican than predicted.



Moran's I = 0.64 (OLS)

After spatial autocorrelation is taken into account by using a spatial error regression, these high-high and low-low clumps virtually disappear, as the map below shows.

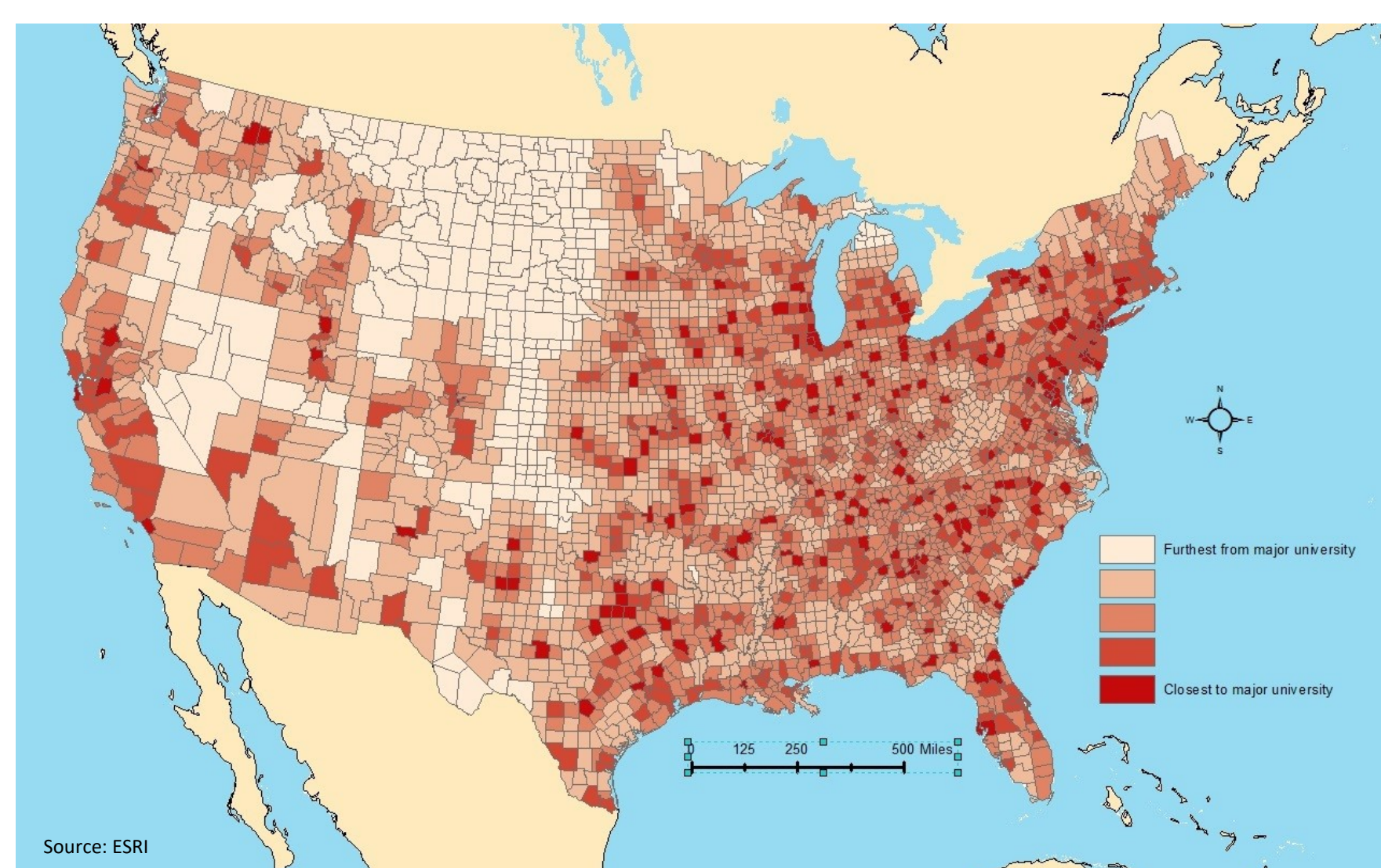


Moran's I = -0.08 (spatial error)

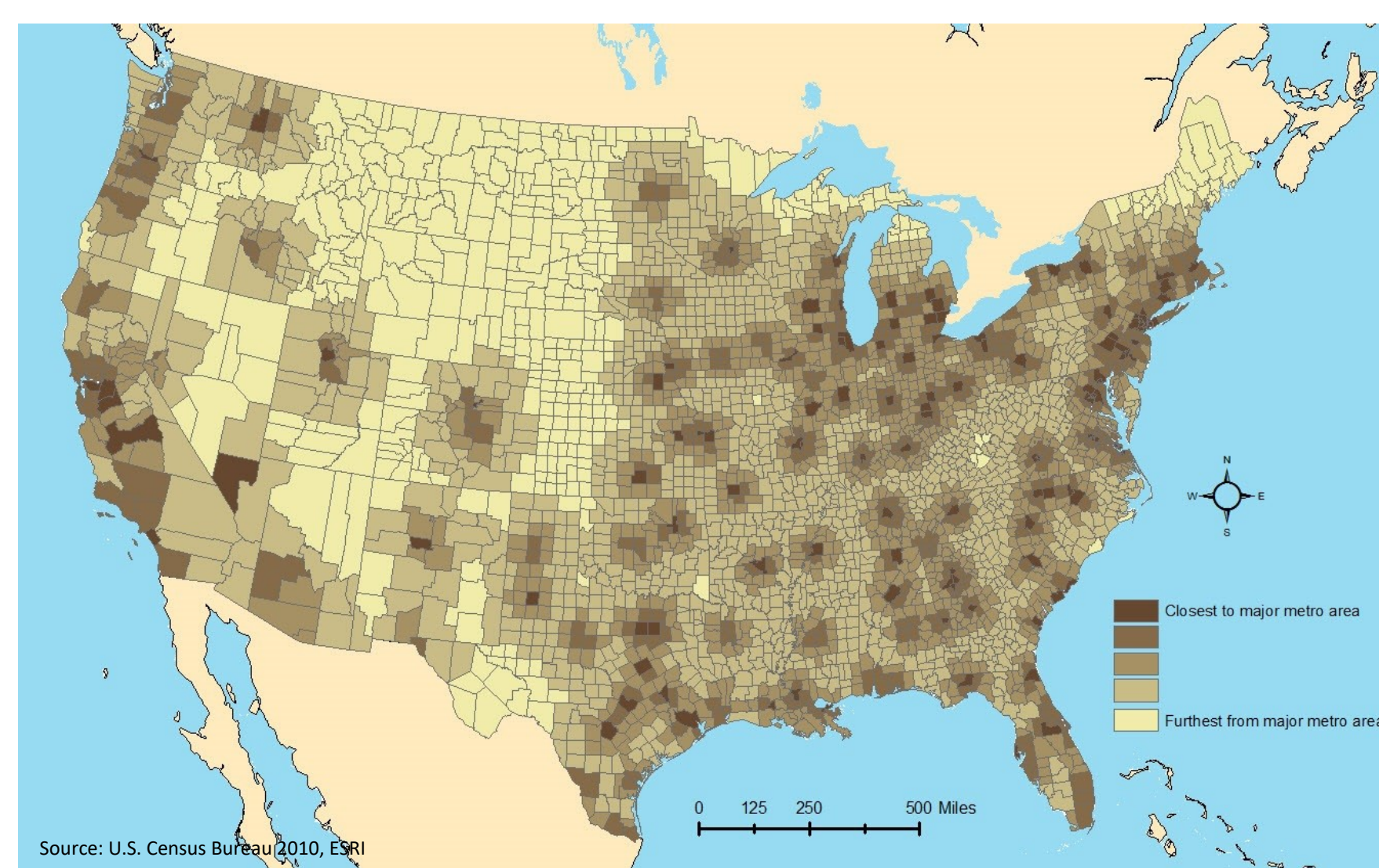
This study has shown that geographic location affects voting behavior, though not to the same extent that demographics do. Future studies could explore the reasons for the high spatial autocorrelation in certain regions, possibly due to omitted variables such as religiosity or regional identity.

Sam Bollier, UEP 294 - Advanced GIS Date: December 20, 2017
Data Sources: ESRI, U.S. Census Bureau, simonrogers.net
Projection: North America Equidistant Conic

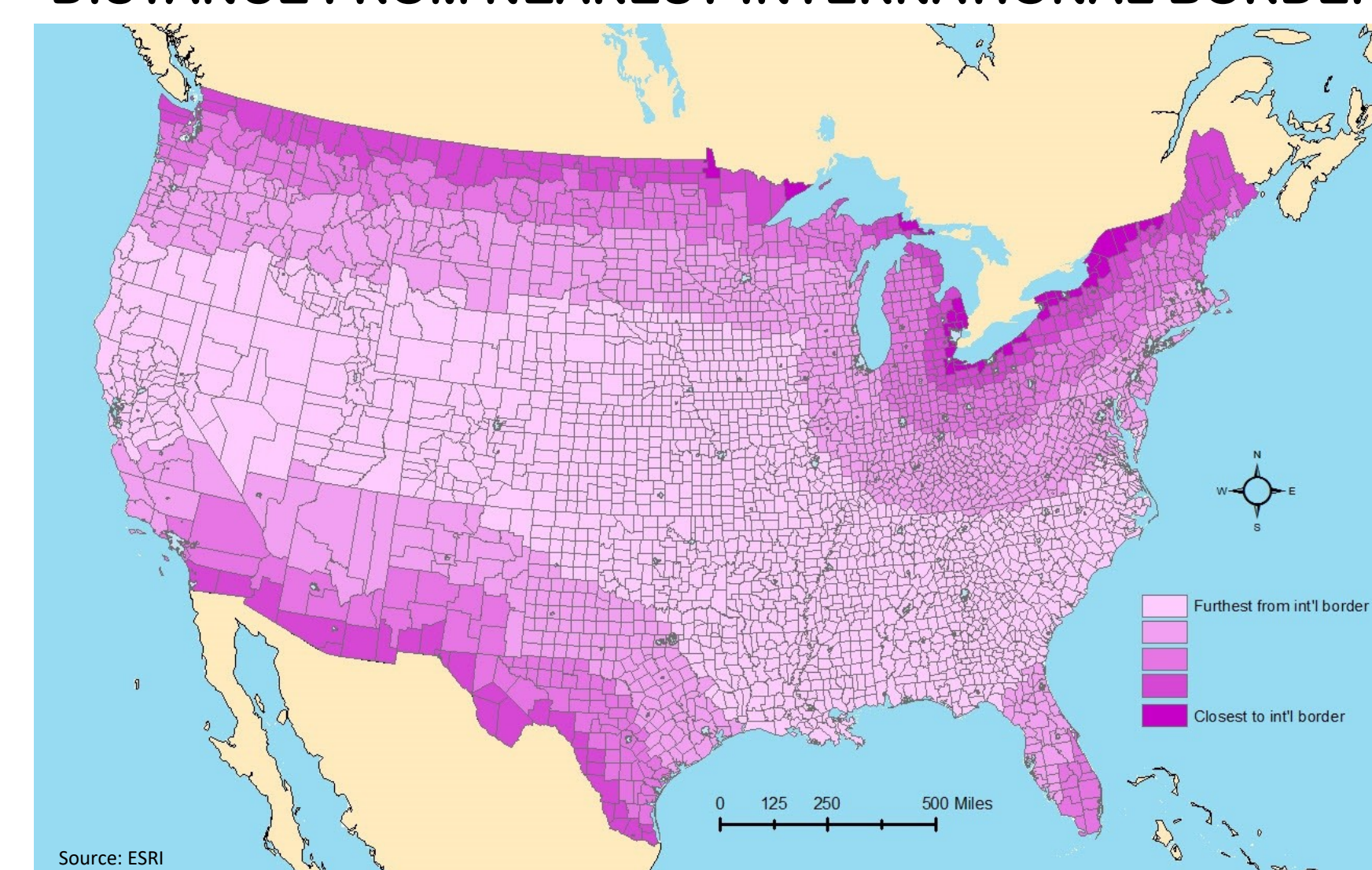
DISTANCE FROM NEAREST MAJOR UNIVERSITY



DISTANCE FROM NEAREST MAJOR METRO AREA



DISTANCE FROM NEAREST INTERNATIONAL BORDER



DISTANCE FROM COASTLINE

