University of New Hampshire

# 2012 AND 2016 US PRESIDENTIAL ELECTIONS

## Did demographic change affect outcomes?

By: **Thomas Cook, Suzannah Hicks, Julia Vaillancourt, Fisayo Dada**

## INTRODUCTION

The 2016 US Presidential General Election results came as a surprise to most. The polls predicted a resounding Clinton victory, but as we know, they were wrong.

In addition to the polls being wrong, experts incorrectly assessed the political landscape of the United States. While there are numerous reasons why a poll could be incorrect, experts were misled by the narrative that the current demographic changes would benefit the Democratic Party. For example, the idea suggests increases in education, income and diversity would coincide with consistent Democratic dominance. On the surface, this theory appeared to hold true with back to back wins by Barack Obama, but proved false in 2016 with the election of Donald Trump and the Republican Party. Our analysis studied how demographic changes played a role in changing voting behavior between the 2012 and 2016 elections.

## MAIN CONCLUSION:

Changes in demographics did not meaningfully contribute to the differences in voting outcomes between the 2012 ad 2016 US Presidential General Elections.

## METHODS:

The study analyzed county level demographic data from the American Community Survey Fact Finder from 2012 and 2015 as well as the 2012 and 2016 election data. This analysis uses 2015 demographic data because 2016 data was not yet available. Five generally accepted demographic categories were determined to classify groups of counties. They include age, income, education, diversity, and population. The variables in each category are shown in Figure 1.

This study used principal components analysis (PCA) combined with k-means clustering to identify six distinct groups of counties based on their demographic similarity. Not all the 3,196 counties in the United States were included since Alaska does not report county level election results, Hawaii redrew its county lines between the 2012 and 2016 elections and two counties in Virginia were combined between elections and therefore left out. Using the Mahalanobis Distance and the Upper Confidence Limit Line, 302 outlier counties were removed before performing PCA and clustering. A later section analyzes the outlier counties. For the purposes of this analysis, PCA reduced the overlap in explanatory power of the variables so additional weightings were not needed, dimensions reduced and most of the information was retained. The 14 variables in this study condensed into four principal components and those four accounted for 77.8% of the information in the original 14 demographic variables in 2012 and 80.7% in 2015.

After outlier analysis and PCA, the 2012 and 2015 demographic datasets each contained 2,805 counties, four principal components, and 2012 and 2016 democratic voting data. K-means clustering identified unique groups of counties within each dataset. The process was repeated a total of three times. First, the algorithm used the 2012 principal components and 2012 voting data. Next, it used 2015 principal components and 2012 voting data. Lastly, it used 2015 principal components and 2016 voting data. Six distinct groups of counties were optimal because they were the most stable, robust, and had the highest cubic clustering criterion.

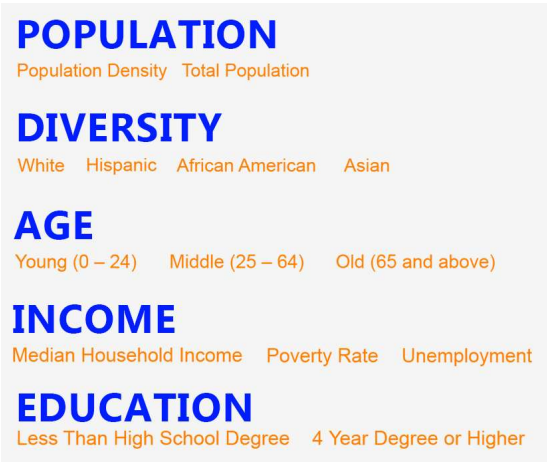


*Figure 1:Break down of Variables*

Our cluster analysis identified six distinct county groups, each with unique characteristics. These groups are Educated Professionals, Black Working Poor, Young Hispanic, Blue Collar, White Working Poor and Average Joes. The green counties were excluded from our cluster analysis because they were sufficiently dissimilar from any county group. They are outliers and were analyzed separately.

# CLUSTER CHARACTERISTICS

### Educated professionals

Counties in this cluster are defined as having the highest population and population density, suburban environments, and moderate diversity. They have the highest median household income which is 30% more than the second wealthiest group and 68% more than the poorest. 32% of the population has a BA or higher whereas all other clusters have less than 20%. They have the second highest percentage of Democratic votes and the least amount of change in voting behavior between the two elections.

### Black Working Poor

These counties have the highest rate of unemployment and the largest percentage of African Americans. A low percentage of the population have high education and median income is also very low. They have the highest percentage of Democratic votes and second lowest change in voting behavior.

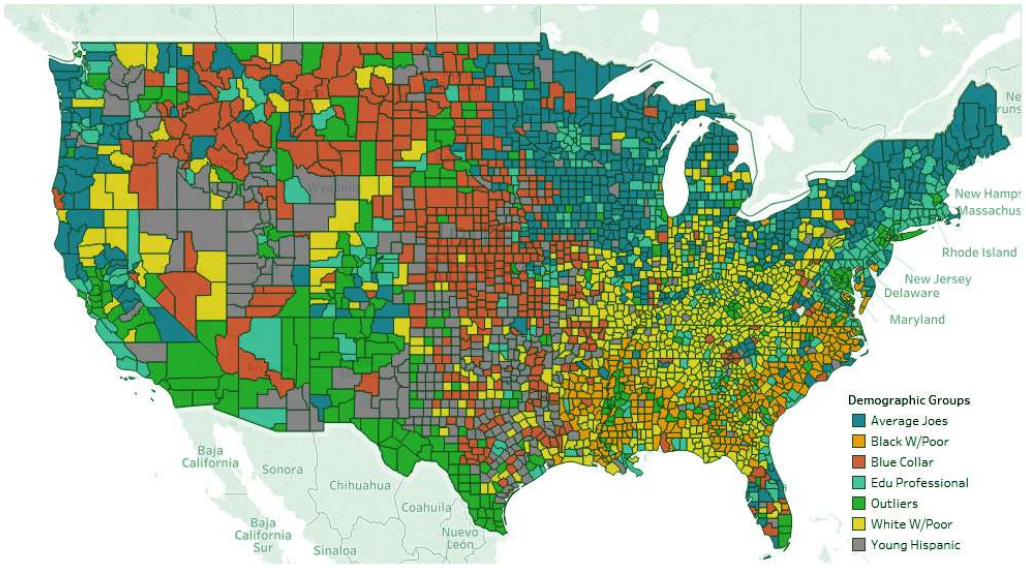


*Figure 1: Map of United States with Counties color coded by their groupings.*

### Young Hispanic

These counties are characterized by having average median income, rural environments, average education and unemployment, though they have significantly more Hispanics and tend to be younger counties. They are typically Republican counties and further increased their vote by 5% in 2016.

### Blue Collar

The Blue Collar are the most rural counties. They have an average population density of 32 people per square mile. They are 93% white and have the lowest unemployment. While they always tend to vote Republican, they decreased their democratic vote by 7% from 2012 to 2016.

### White Working poor

The white working poor is 90% white and have low income and education. Notably, these counties had the biggest percentage increase in income between the elections (10%), yet decreased their Democratic vote by 8%. Unemployment also decreased from 5.5 to 4.5%. In 2016, they voted 79% Republican.

### Average Joes

The Average Joes are 93% white and have the second highest median household income. 20% of the population has higher education. Between elections, unemployment decreased by 2.5%, yet on average these counties increased their Republican vote by 11%. These counties voted 65% for Trump.

### Outliers

The outliers are a combination of the counties with the most extreme values. For example, many of the urban centers such as Los Angeles County (Population: 9,840,024) are included in this group. Additionally, some of the most rural counties are included such as Loving County, Texas (Population: 85).

## ESSENTIAL FINDINGS

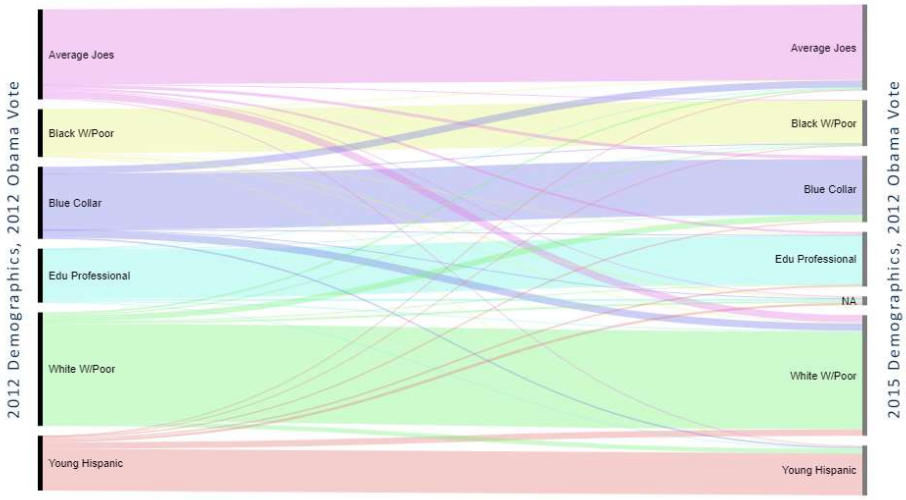


*Figure 2: Demographic changes were minor yet consistent within a county profile between elections*

To explore if changing demographics affect voting behavior, the counties in each cluster were mapped across the three K-means clustering analyses.

When comparing the clusters created with 2012 demographic data and 2012 voting data to the clusters created with 2015 demographic data and 2012 voting data (Figure 2), any change in a county's cluster membership was due to a demographic change that was relatively different than the other counties in that cluster. For example, on average Blue Collar counties increased income by 6.9% and decreased unemployment by 18.18% from 2012 to 2015. However, a subset of Blue Collar counties increased their income by 7.5% and decreased unemployment by 9%. This subset was identified as Average Joes when clustered using 2015 demographic data and 2012 voting data because their demographic changes were sufficiently different than the demographic changes of other Blue Collar counties. The alluvial chart shows that the counties generally remain in their respective clusters, meaning counties within a cluster experienced similar demographic changes between 2012 and 2015. If the narrative that changing demographics affect voting behavior holds, one would expect the counties within a cluster to change their voting behavior similarly from 2012 to 2016.

When comparing the clusters created with 2015 demographic data and 2012 voting data to the clusters created with 2015 demographic data and 2016 voting data (Figure 3), any change in a county's cluster membership is due to a change in voting behavior that was relatively different than the other counties in that cluster.

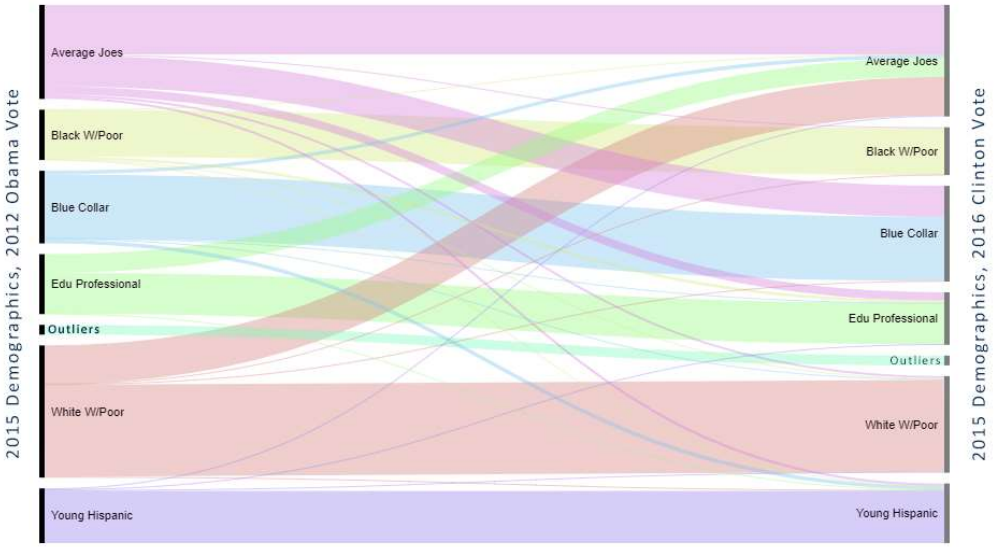


*Figure 3: Movement of Counties between clusters due to different rates of change in voting behavior*

For example, on average the Educated Professional counties decreased their democratic vote by only 3% from 2012 to 2016. However, a subset of Educated Professional counties decreased their vote by 5.5%. These counties were identified as Average Joes when clustered using 2015 demographic data and 2016 voting data because their change in voting behavior was different than the other Educated Professional counties. Since the change in voting behavior cannot be explained by differing demographic changes, something else must account for the unexpected change in voting behavior. Interestingly, this subset of counties had a lower proportion of highly educated population. Only 26.66% of the population had a Bachelor's Degree or higher, whereas on average 32% of a county's population within the Educated Professional cluster has a BA or higher.

Similarly, a subset of counties changed from the White Working Poor to the Average Joes. They decreased their Democratic vote by a smaller margin (7.9% compared to 9%) and had a higher proportion of highly educated population (18.32% compared to 14.63%).

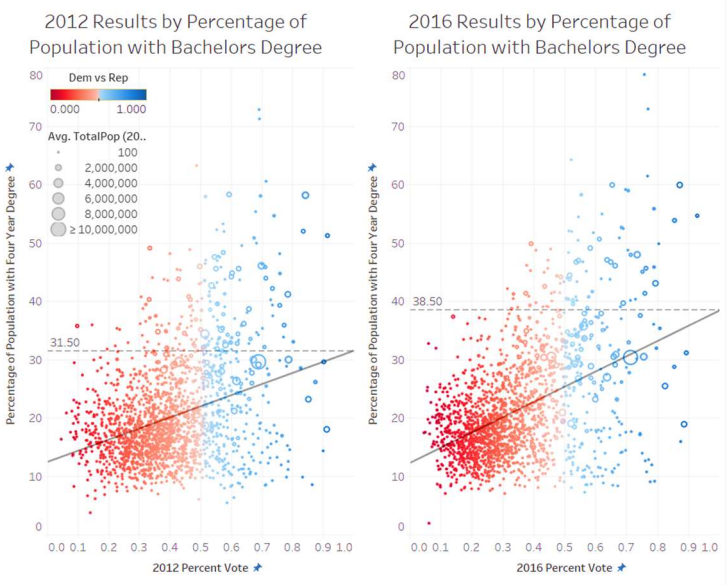


*Figure 4: The association between the percentage of counties with a four year college education and voting behavior increased in 2016.*

## OBSERVED TRENDS

The next step in the analysis looked for clues as to what changes in voting behavior occurred to account for the difference in outcomes between the 2012 and 2016 elections.

In Figure 4, the slope of the line steepened from 19.56 in 2012 to 28.47 in 2016. The association between the percentage of people with a four-year degree and voting behavior increased between the two elections.

Additional examination of the percentage of people with a four-year degree in counties with a population of at least 50,000 supports this conclusion. Of the top 25 counties by education 23 out of 25 increased their share of vote for the Democrats and the top 25 increased their share of the vote by 3.1% for the Democrats versus an average increase of a 4.2% for the Republicans election over election. 20 of the top 25 most educated 2012 Republican voting counties increased their share of vote for the Democrats as well by 2.3% versus the national increase of 4.2% for the GOP. On the other side, all 25 of the least educated counties in the country voted more strongly for the Republicans by an average 5.7% versus 4.2% nationally. Finally, all 25 of the least educated counties that voted for the Democrats in 2012 increased their share of vote for the GOP by a staggering average of 7.2%. The defection of the lower educated Democratic voters ultimately broke the so-called Blue Wall and led to the Republican victory.

## OUTLIER ANALYSIS:

Interestingly, each of the six clusters voted, on average, for the Republican candidate in both 2012 and 2016. Where in this analysis lies the Democratic counties? The outliers. Of the 302 outliers, 161 voted for the Democrats and 141 voted for the Republicans in 2012. The Democratic voting counties contained nearly 85 million people whereas the Republican outliers only contained 18.5 million. The outlier counties represent every major city in the country and, as a result, the Democrats' most important counties. The problem for the Democrats in the 2016 election was that the location of these counties led to an electoral college loss since too many of them are located within the same states. So even though the strength of the vote in the outlier counties contributed to a Democratic popular vote win, the electoral college map favored the GOP.

For the GOP, the outlier counties contained the few metropolitan cities they won and many sparsely populated counties. The outlier cities that voted for the Republicans in 2012 increased their share of vote for the Democrats in 2016. Two such examples are Maricopa County, AZ and Harris County, TX. Phoenix, AZ and Houston, TX reside in these two counties, respectively. Both cities contain larger than average Hispanic populations, have a higher share of the population with a four-year degree and both increased their share of vote for the Democrats in 2016.

A lot of the discussion on the outliers focused on the population variable because of the outsized voting impact of having so many additional people. In addition to containing more people, the outliers tend to differ in other ways as well. The high population outliers contain more diverse populations, are higher educated, have higher poverty rates, and higher median incomes than the rest of the country. On the other hand, the low population outliers tended to have lower median income, high poverty rates, lower rates of college education, and a variety of both extremely white and very diverse. In other words, the outliers contained extreme values or counties that contained an unusual combination of characteristics compared to the nonoutlier counties.

## CONCLUSION:

Demographic changes did not impact the election outcome in 2016. Certain demographic variables are correlated with voting behavior, and the strength of the correlation changed between the elections. As discussed, the correlation between education levels and voting behavior increased from 2012 to 2016.