2016 Election Results and County Attribute Analysis

STP 420

Instructor: Neil Hatfield

Belmina Mehmedagic, Aidan Morita, Vishal Etikala

April 2017

**Executive Summary**

Using data collected by Robert Wood Johnson Foundation and the University of Wisconsin Population Health Institute, David Leip’s Atlas of U.S Presidential Elections, and the Bureau of Economic Analysis, we investigate the relationships between county attributes and voting patterns in the 2016 U.S Presidential Election. We analyze counties along the metrics of Population, Unemployment, Per-Capita Income, Air Pollution, High School Graduation Rates, Obesity Rates, Violent Crime Rates, Child Poverty, Long Commutes, Change in Per-Capita Income since 2008, and Percent of the Population with Access to Exercise, with an eye towards their geographic distribution, and the county’s winner in the General Election.

We find statistically significant differences between Trump and Clinton counties in eight of the attributes, with Trump counties demonstrating lower Population, Unemployment, Per-Capita Income, Violent Crime Rates, and Access to Exercise, and higher air-pollution, graduation rates, and Obesity. This suggests that Trump’s campaign rhetoric on jobs, and crime found receptive audiences. The largest differences across the counties are along the measures of Population, with Clinton dominating Trump in the largest cities, and Access to Exercise, where Clinton similarly dominates. We attribute most of the variation in vote share attributed to Access to Exercise to county Population, as it seems the two are closely related. Unfortunately, we cannot identify whether county conditions are leading voters to for candidates to change the conditions, or whether conditions reflect the consequences of the policy preferences of their residents.

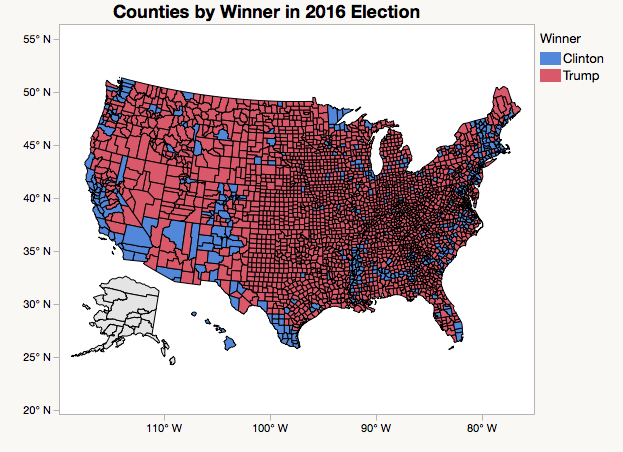
The 2016 Presidential Election was a dramatic upset unpredicted by pollsters and statisticians. No one expected a Republican win. What enabled Donald Trump’s win is a topic of discussion pored over by both political scientists and statisticians. Experts for both parties are conducting Post-Hoc analysis of the election, to identify the forces that elected Donald Trump to the Presidency. In this vein, we’re investigating national conditions and their relationship with political behavior.

When voting, citizens must order their policy preferences and compromise on some issues. They do this while evaluating the policies and outcomes of the current administration. While the voting behavior of certain demographics is well documented, we’re interested in the environmental conditions of the counties on dimensions where policy matters. We expect education, economic conditions, environmental conditions, health, safety, and social conditions to inform, and potentially be influenced by, people’s policy preferences. We will investigate these conditions, and evaluate them as indicators for Trump support. We also know from much public commentary that there is an urban-rural cultural divide, and we expect to see this reflected in voting behavior. This will act as a controlling factor for much of our analysis.

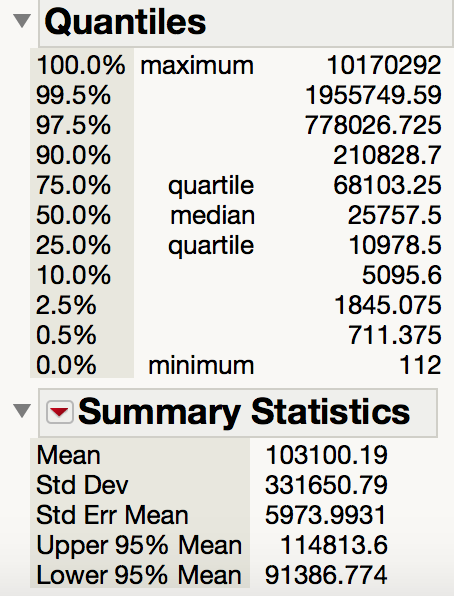
The goal of this analysis is not to identify why Trump won. We don’t have the time-series data necessary to identify trends in voting patterns. Instead, our goal is to see what conditions are indicators for Trump-like candidates to find electoral support. Our data, a combination of electoral results, socio-economic data, and health and quality of life indicators, should open a window into which conditions Trump supporting counties display. We present a basic description of each variable, as well as its geographic distribution at the county level, followed by an analysis of the question: Are counties that voted differently different across this attribute?

**County Election Results**

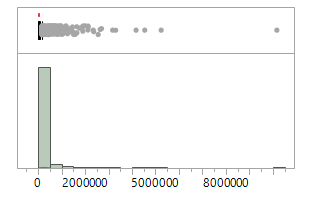
**Figure 1:**



First, we present the electoral map colored by the winner in the county following convention: Blue for Democrats, Red for Republicans. By this graph, we are able to see that Trump won far more counties than Clinton. She won most of California, half of New Mexico, most of New York, Connecticut, and Massachusetts. In addition, Clinton won multiple isolated scattered throughout the county. Clinton did better along coastal and counties along the southern border, the Gulf coast being an exception, while Trump excelled in the center of the country, and the northern border.

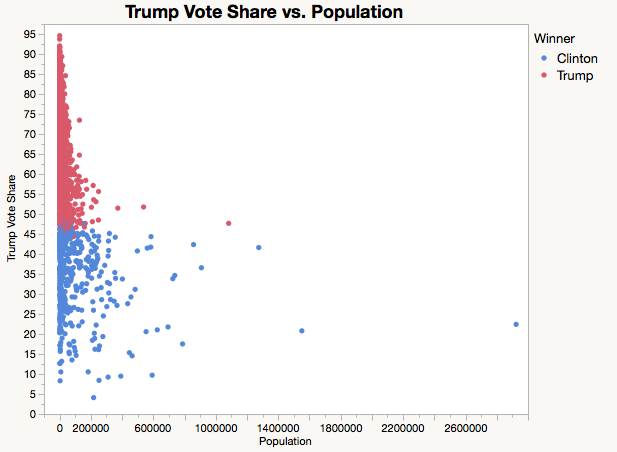
**Population:** 

**Figure 2: ‘15 Population**



Next, we investigate population. The politics of the urban-rural divide are well-documented (Gami 1). We see the vast majority of counties have populations under 2,000,000 people, and only a few counties have more, but the few that do have many, many more people living in them.

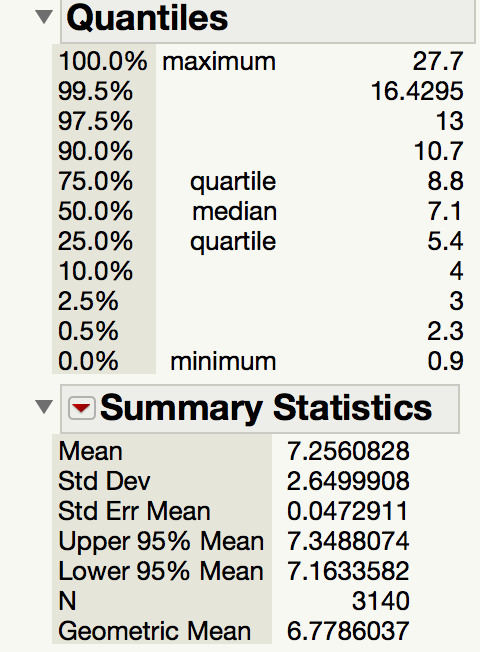
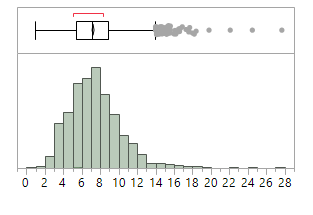
**Figure 3:**



This figure shows the share of the vote Donald Trump Received in counties, and the population of those counties. We see that in all but three of the nation’s most populous, Clinton won. Trump won mostly low density, rural counties, while in the denser, urban counties, Clinton carried. Further, for our threshold of unusualness of α=0.05, there is a statistically significant difference in the sample arithmetic mean of counties based on who they voted for (tWelch’s(493.7)=-8.9, p<.0001). This is a large effect (dcohen=-.96), and Clinton’s counties were more populous 75% of the time. The strength of this effect means that Population, and the urban-rural divide Population represents, means that for other strong effects, we should consider the impact the cultural divide may be having, and use Population as a proxy for that effect.

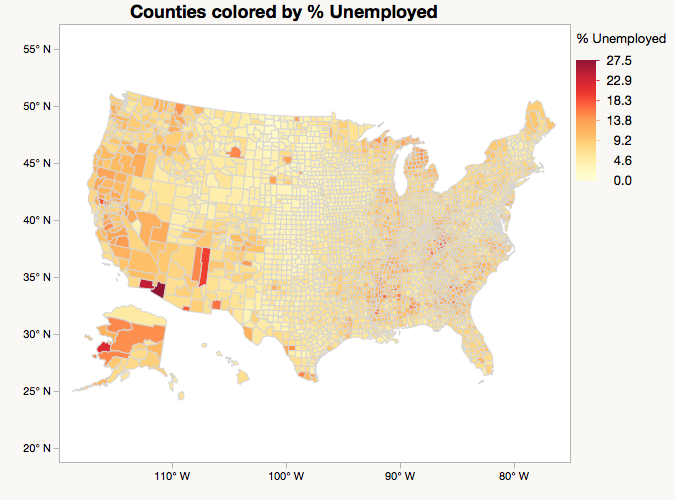
**Unemployment:**

**Figure 4: Unemployment Rate**



The values of frequencies of unemployments in each county in US are shown in the figure above. The median unemployment rate of 7.1% appears high compared to a desired unemployment rate of 3-5%.

**Figure 5**:

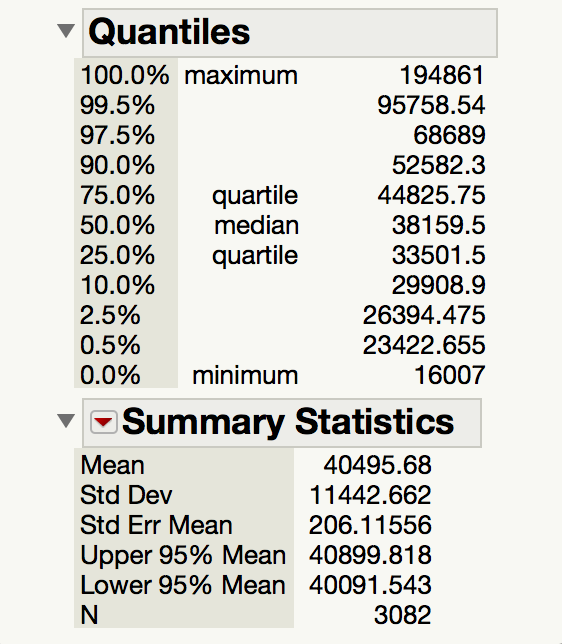
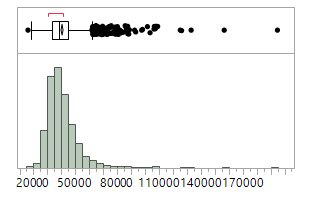


In this graph, we can see the unemployment rate in the USA, per the definition of unemployment supplied by the Bureau of Labor Statistics, which is the number of people working for pay or profit, divided by the number of people employed or actively seeking employment. First, the most unemployed county is Yuma County in AZ and trailing right behind is Imperial County, Wade Hampton, and Apache County. Interestingly, unemployment also seems to be regionally related. The small bands of high unemployment counties in appalachia and the deep south may indicate that being near a center of unemployment increases a county’s unemployment rate. Alternatively, the bands could imply that economic conditions don’t strictly follow county lines, and that regional economic conditions based on a local industry are more important. The Midwest and Northeast have a low rate of unemployment. While the West has slightly higher rates.

There is a statistically significant difference between the unemployment rates of counties who voted for Clinton and Trump, using the Kruskal-Wallace two sample test to compare the median rates (S = 893,404, p<0.0001). Clinton’s counties had higher unemployment rates (H-L(Trump-Clinton) = -0.9), and Clinton’s counties had higher unemployment rates 60% of the time. A median difference of .9 percentage points is a substantial difference, and represents a substantial difference in the strength of the labor market and people’s employment prospects in different counties. Clinton succeeded where more people were out of work.

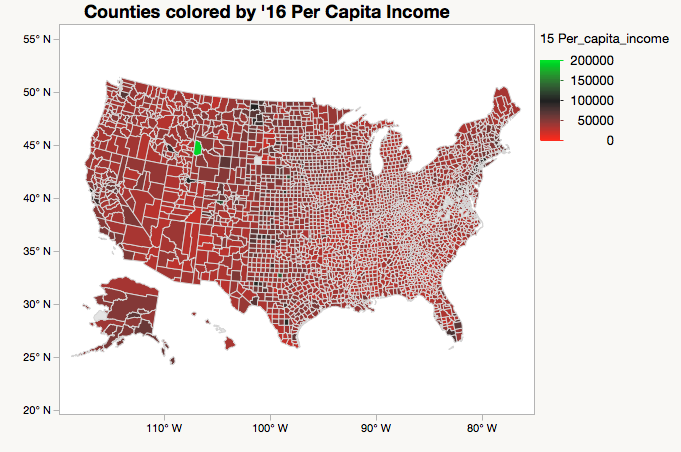
**2015 Per Capita Income:**

**Figure 6: Per Capita Income**



The economic debate surrounding tax cuts and inequality were major planks of both candidates’ platforms. So we investigate per-capita income, a measure of individual wealth to investigate whether the income level of counties differed by electoral choice. In 2015, the threshold for medicaid eligibility, 138% of the federal level poverty level, was $16,243 for a household with one person. The minimum of $16,007 is lower than this threshold, meaning a high proportion of people in that county would receive medicaid.

**Figure 7:**

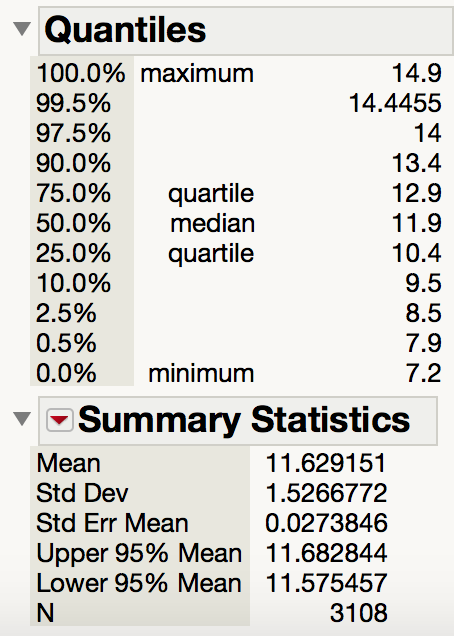
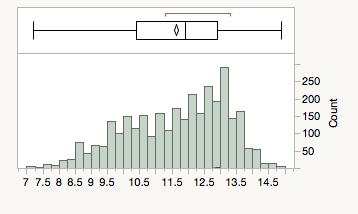


In this graph, we can see the per capita income per each county, as defined as the net income of the county divided by the population. Most values of income per capita are between $20,00 and $70,000, and very few are above $100k. The bright green county, Teton County, WY, is a bit of a mystery. Teton’s population of approximately 21,000, implies a net income of 4.2 billion dollars, so a very rich person must indicate Teton, WY as his or her home county. Income does not seem to follow an obvious geographic pattern.

For an unusualness threshold of α = 0.05, we reject the null hypothesis that there is no difference in the per-capita income of counties based on their electoral choice. There is a statistically significant difference in the 2015 per-capita income of Trump counties and Clinton counties. Trump counties have a lower per-capita income (tWelch’s(518.5275)=-6.3, p<.0001). This is a small effect (dCohen=-0.463), and Clinton’s counties have higher per-capita incomes 63% of the time. This is surprising given the unemployment data. Clinton’s won both in counties that were richer per-capita, as well as in counties with worse employment.

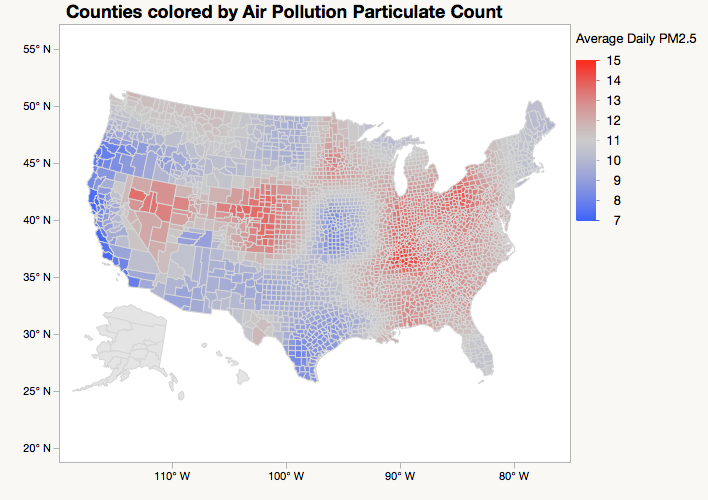
**Air Pollution:**

**Figure 8: Air Pollution in Micrograms per Cubic Meter**



Both candidates ran on issues regarding environmental policy: Clinton ran on a platform of limiting climate change and environmental protection, while Trump ran on expanding mining permits and loosening regulation to mobilize the economy. We use air pollution as a measure of interest in environmental protection, to investigate whether counties actually differ politically on the issue. Air pollution is measured as the density of air particulate matter in micrograms per cubic meter of air. There’s a lot of variation in air quality, without much clustering of the values.

**Figure 9:**

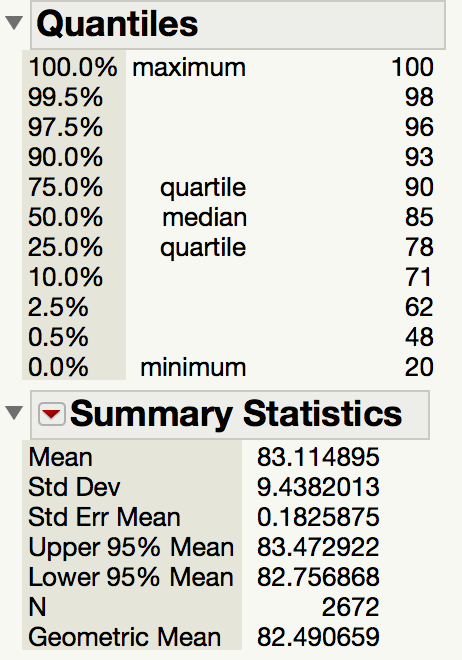
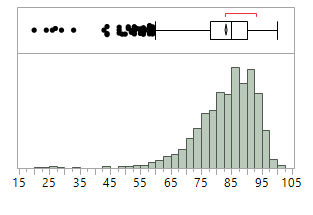


We can notice that the pollution is not as bad near surrounding waters and that the pollution is worst in a center point and just dies of slowly in the counties near by. Pollution does seem to be at least regionally connected, and follow state lines. This implies some level of political impact on air pollution, either by reporting, or actual environmental policy.

For an unusualness threshold of α=0.05, we reject the null hypothesis that there is no difference in the sample arithmetic means of air particle concentrations. There is a statistically significant difference in the sample arithmetic means of air particle count of Trump counties and Clinton counties. Air particles are additively cumulative, so Welch’s t can be used. Trump counties have a higher air particle count (tWelch’s(678.58)=4.27, p<.0001). This is a small effect (dCohen = 0.21), and Trump’s counties have air particle counts 55% of the time. This implies that at the county level, people do care about the environment as a political issue, and vote on environmental issues, but not as a high priority.

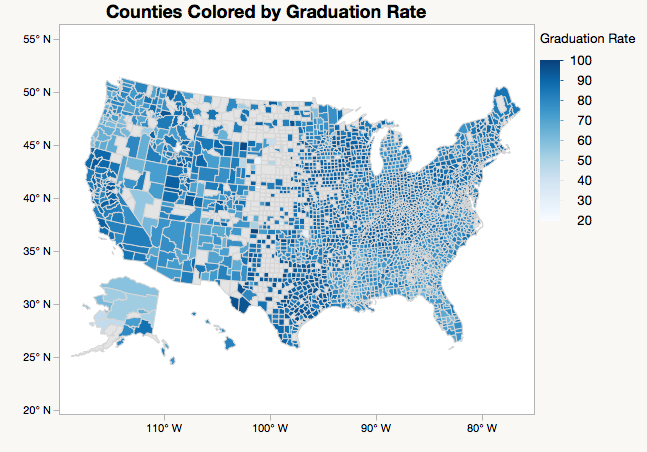
**Graduation Rate:**

**Figure 10: Graduation Rate**



We investigate the graduation rate as a measure of school quality. Graduation rate is measured as the percent of students who complete high-school. Graduation rates are complicated both as measures of education quality and economic conditions, but they do answer the question ‘Are students completing school?’ We expect counties to prefer higher graduation rates to lower, but we can’t identify whether voting behavior is caused by, or a response to, graduation rates. Most values of the graduation rate are between 80% to 95%, so in many counties, 10% to 15% of students do not graduate.

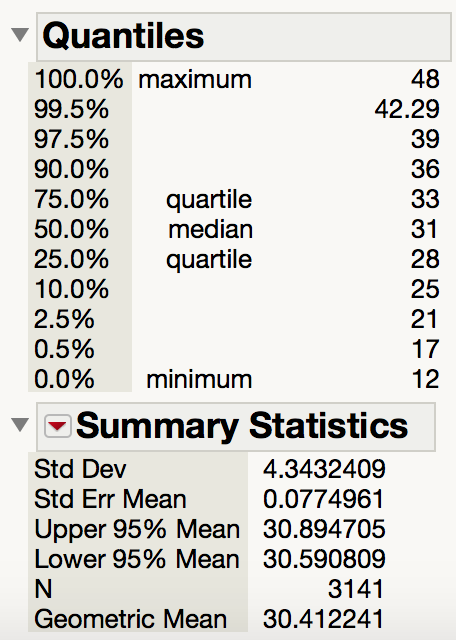
**Figure 11**:

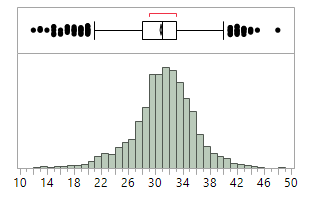


In this graph, we take a look at the Graduation rate from high school. We can notice that California has a high graduation rate where are many other states are scattered with the amount of graduates. The counties with no color are not zero graduation rates, but simply non-reporting counties. Graduation rate does not seem to follow an obvious geographic pattern.

For an unusualness threshold of α=0.05, we reject the null hypothesis that there is no difference between the sample median of graduation rates in counties based on their electoral choice. There is a statistically significant difference in graduation rates between Clinton counties and Trump counties for reporting counties. Using 1500 simulations on the bootstrap seed of 1, we obtain a 95% bias corrected CI of (-5, -3) for the difference in the sample median (Clinton - Trump). So of the reporting counties, Trump’s counties have higher graduation rates. While we cannot interpret any causality, it is interesting that Trump county schools were not struggling as hard as Clinton’s on this measure. A difference in graduation rate between 3% and 5% is substantial, especially because Trump campaigned on this issue of school quality.

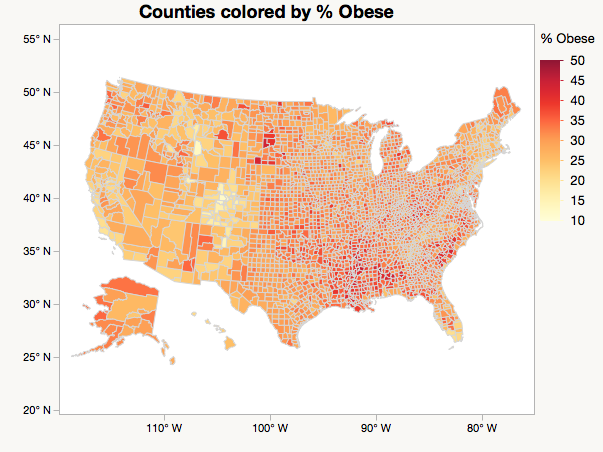
**Obesity:**

**Figure 12 : Obesity Rate**



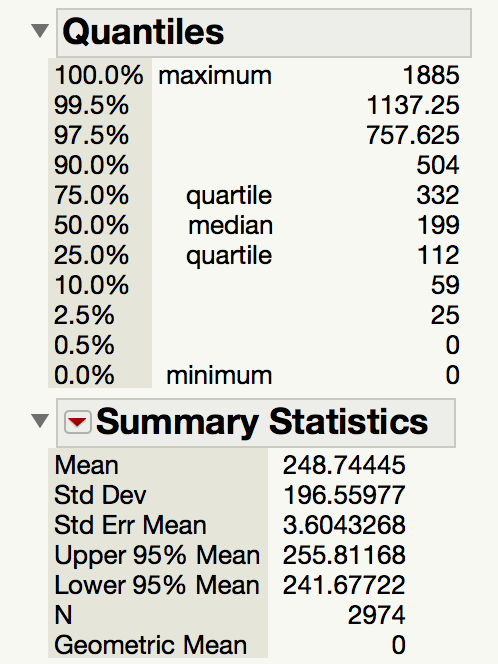
Previous results by The Economist showed health indices to be strong predictors for counties to switch party from the previous Presidential Election(“The Presidential 1”). We disaggregate their indices to investigate whether any particular health variable distinguishes counties. We investigate the obesity rate, the percent of people medically obese. The rate seems clustered around 30%, around the sample geometric mean. The maximum, a county with a 48% obesity rate is highly alarming.

**Figure 13:**

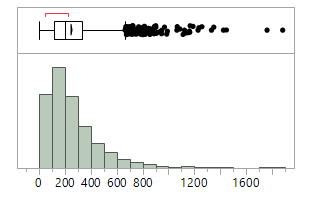


In this graph, we can see the percent of the population considered medically obese. Obesity is highest in counties in Alabama, Mississippi, Louisiana and South Dakota. Obesity seems to be centered in the Deep South, but there are also bands of high obesity counties in South Carolina, the Dakotas, and Appalachia. This geographic clustering indicates some sort of environmental factor, otherwise we would expect a more even spread of high and low obesity counties.

For an unusualness threshold of α=0.05, we reject the null hypothesis that there is no difference between the sample median of obesity rates in counties based on their electoral choice.There is a statistically significant difference in obesity rates between Clinton counties and Trump counties. Using 1500 simulations on the bootstrap seed of 1, we obtain a 95% bias corrected CI of (-3, -2) for the difference in the sample median (Clinton - Trump). So Trump’s counties have slightly higher obesity rates. A difference in the sample median of a few percent is not massive, but could represent substantial medical costs in the county.

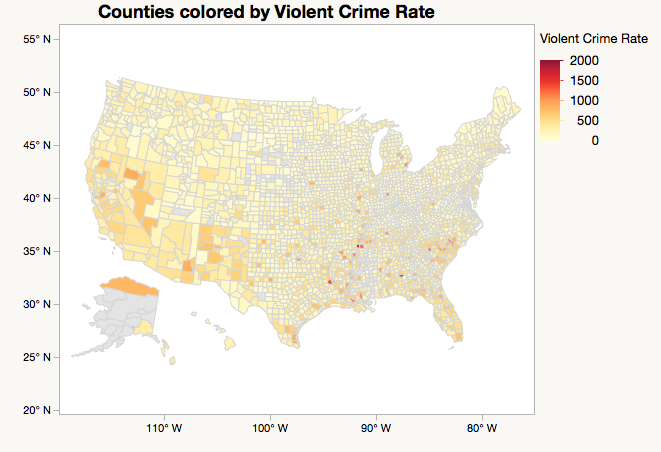
**Violent Crime:**

**Figure 14: Violent Crime per 100,000 population**



Trump ran on a ‘Law and Order’ Campaign, declaring the murder rate at an all time high. We investigate the violent crime rate per 100,000 people to see whether counties did differ in electoral behavior in response to local violent crime. Most counties have fewer than 200 violent crimes committed per year per 100,000, although the tail is long to the right.

**Figure 15:**

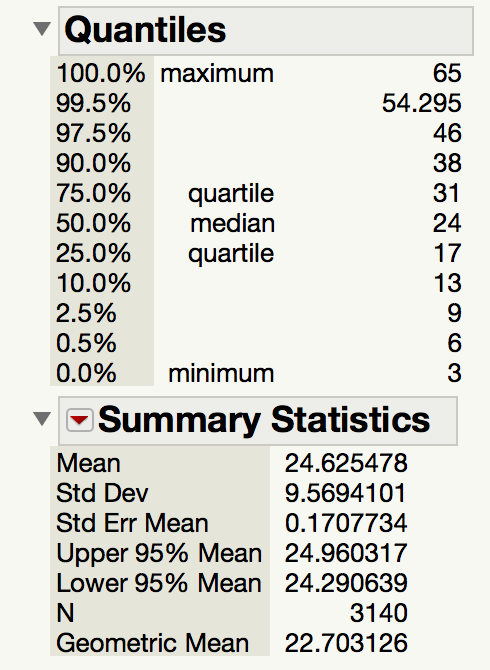
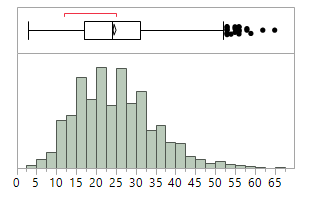


Here, we see what the Violent Crime Rate per 100,000 people is per county. The highest rate is in St. Louis, Missouri. By and large, we see that violent crime rates are low across the country, although St. Louis’s two crimes per 100 people is slightly alarming. Crime seems to be centered in specific counties without much effect on their neighboring counties.

For an unusualness threshold of α=0.05, we reject the null hypothesis that there is no difference between the sample median of violent crime rates in counties based on their electoral choice. There is a statistically significant difference in violent crime rates between Clinton counties and Trump counties. Using 1500 simulations on the bootstrap seed of 1, we obtain a 95% bias corrected CI of (118, 190) for the difference in the sample median (Clinton - Trump). So of the reporting counties, Clinton’s counties have higher violent crime rates.

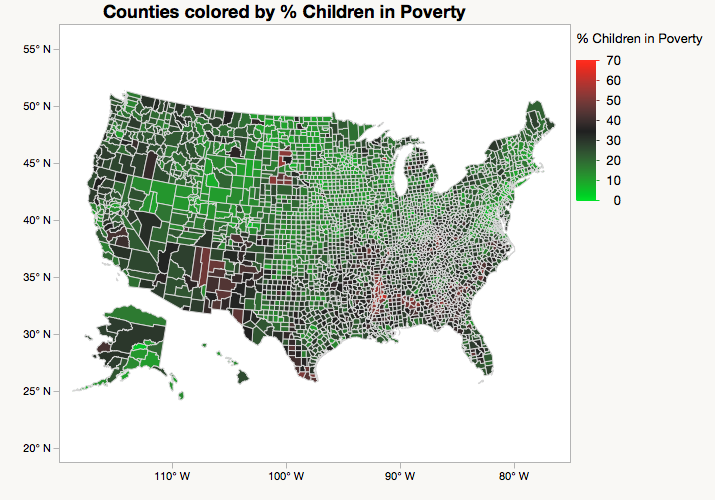
**Child Poverty:**

**Figure 16: Percent of Children living in Poverty**



Child poverty is both an economic, educational, and public health problem, with sympathetic victims. We investigate the county differences based on winner to determine whether either party more successfully campaigns on the issue. There is a lot of variation in this attribute, as seen in the extrema. Most values are clustered around the geometric mean, which at 22% per county is alarmingly high.

**Figure 17:**

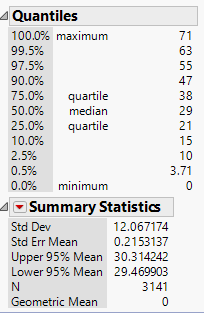
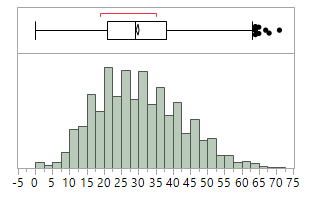


In this graph, we are looking at the percentage of children living in poverty per county, where poverty is defined as having a family income three times the price of a minimum food diet, price adjusted using the CPI from the minimum diet from 1963. The highest rate is in Humphreys County in Mississippi with 65% and the lowest rate being in False Church City County in VA with 3%. Child poverty does seem to be somewhat geographically centered. The pockets of high child poverty in Alabama, Mississippi, the Dakotas, and Arizona and New Mexico imply some kind of environmental factor that crosses county lines, or a spillover effect of poverty, causing poverty in neighboring counties.

There is no statistically significant difference in child poverty rates rates between Clinton counties and Trump counties for reporting counties. Using 1500 simulations on the bootstrap seed of 1, we obtain a 95% bias corrected CI of (-0, 3) for the difference in the sample median (Clinton - Trump). So we fail to reject the null hypothesis that there is no difference for a threshold of unusualness of α =0.10.

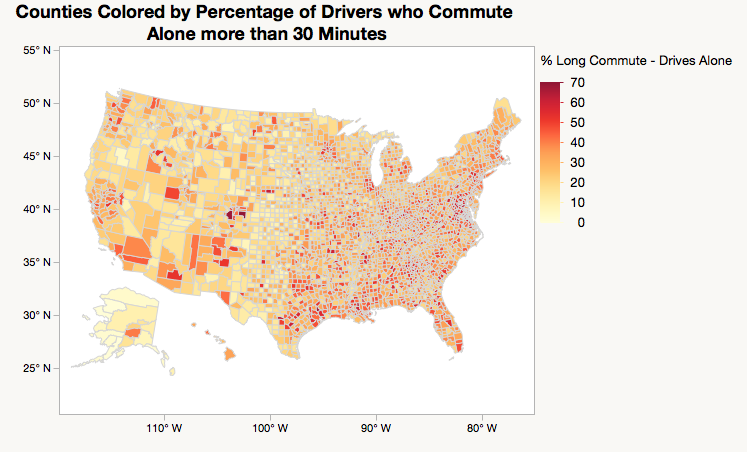
**Long Commutes:**

**Figure 18: Percent of People who commute alone more than 30 minutes**



Post-election punditry identifies talk-radio as sources of motivation for Trump voters (Dougherty 1) (Glass, Chace). We conjecture that spending long periods of time alone in the car may be a measure of these factors. We observe lots of variation, from 0% in some counties to over 50% in the top 2.5%. Most counties lie between 20% and 35%, but this demonstrates the prominence of the automobile in American life.

**Figure 19:**

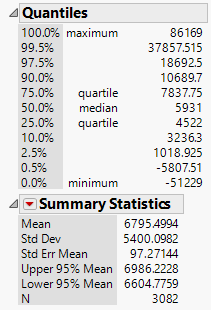
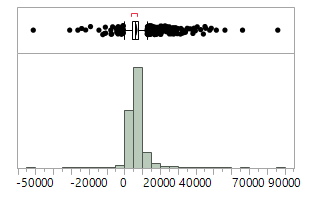


In this graph we can observe the percentage of drivers that commute alone for more than 30 minutes. The longest percentage of people that commute are in Elbert County in Colorado, and the shortest commutes are in Alaska. In addition Alaska has low commute rate for the whole state aside from only one county. We can see that the south and southeast have more counties with people who commute longer than thirty minutes, whereas the midwest has a lower percentage rate and west coast has a lower percentage rate as well.

For an unusualness threshold of α=0.05, we fail to reject the null hypothesis that there is no difference in the sample median of percent of people who are solitary long distance drivers between Trump counties and Clinton counties. There is a statistically significant difference in the percent of people who drive alone long distances between Clinton counties and Trump counties. Using 1500 simulations on the bootstrap seed of 1, we obtain a 95% bias corrected CI of (-.876, 2) for the difference in the sample median (Clinton - Trump). So of the reporting counties, there is no statistically significant difference. Commute time may be reflective of many things - urban density, traffic flow, and may simply not be an indicator for social isolation.

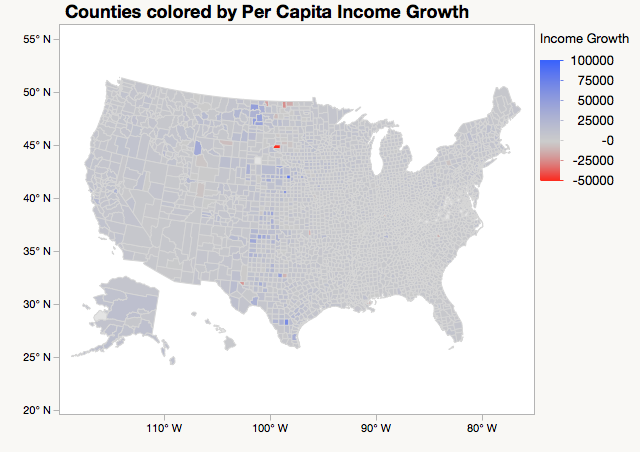
**Income Growth Since 2008:**

**Figure 20: Per-Capita Income Growth from 2008 to 2015**



Presidential elections are a referendum on the previous administration, and people measure the administration on whether or not they feel better off than they did at the start of the administration. Per-capita income growth over a Presidency is a reasonable measure of people’s impression of effectiveness. We investigate whether there is a relationship between income growth over Obama’s term and electoral behavior. We see most income growth between $0 per person and $10,000 per person, which is modest, but substantial. A few counties stand out as miracles and disasters, gaining almost $90,000 per person, or losing $50,000 per person. These enormous swings are mysterious.

**Figure 21:**

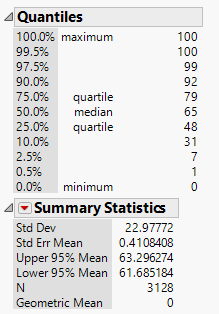
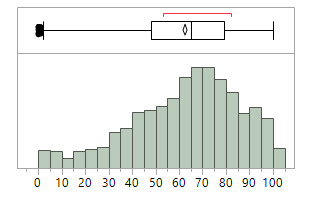


This graph shows the income growth change from Obama’s starting term in 2008 to when the Presidential Primaries began in 2015. The county that grew the most was Wheeler county in Nebraska and the county that lost most growth per capita was Sully County in South Dakota. Most counties saw moderate growth around $6000, as demonstrated by the mostly gray map. There are more slightly blue counties than slightly red, but the overall effect is of a country reasonably better off over 8 years.

There is not a statistically significant difference in per-capita income growth between Trump counties and Clinton counties (tWelch’s(723.4)=-0.19, p=.84), so we fail to reject the null hypothesis that there is no difference between the two for an unusualness threshold α = 0.05.

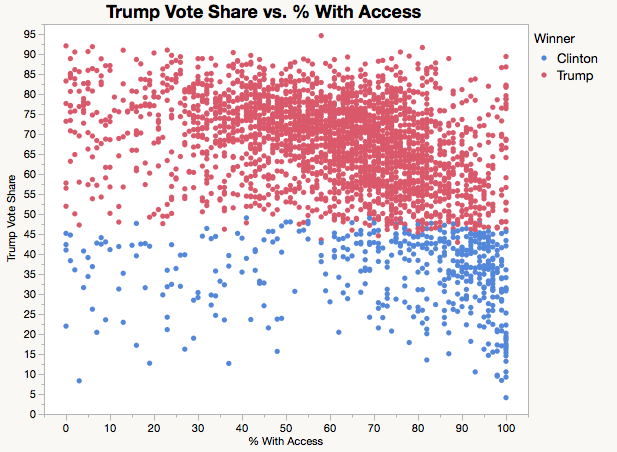
**Access to Exercise:**

**Figure 22: Percent of the Population with Access to Exercise**



The argument over health care, and the right to coverage vs. the right to access lies at the core of the battle over Obamacare. Preventative care, like exercise, provides people the ability to take control of their health, and not rely on policy decisions. So the question of access to health resources was politically salient in the election, and might reveal have an electoral difference. Access to exercise, as defined by survey takers is the attribute of interest.

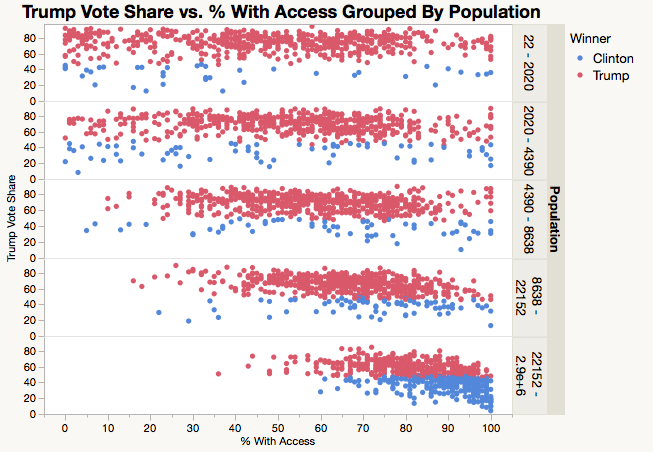
**Figure 23:**



Counties Trump won were very diverse, but here, an interesting relationship is visible. Access to exercise opportunities is negatively correlated to Trump’s vote share. This is by no means causal: people didn’t vote for Trump out of the expectation he’d increase their access to exercise opportunities, or even more egregiously, that people in poor health voted for Trump. Instead, more likely is that rural and urban counties seem to be correlated with access to exercise.

There is a statistically significant difference in the sample median of the percent of people with access to exercise across county victors. We reject the null hypothesis of no difference at our threshold of unusualness of α=0.05. For 1500 Bootstrap simulations of seed 1, we get a 95% bias corrected CI of (20, 23.38). This is an enormous difference. 20% more of population of the median Clinton county has access to exercise than those living in Trump counties. The percent of the population with access to exercise varies immensely between Trump and Clinton counties, and deserves further investigation.

**Figure 24:**



This figure confirms the indications of the previous two figures. As the population of a county increases, the amount of the population with access to exercise opportunities increases, and Clinton wins more the of the counties. This indicates pure access to exercise is related to both political preference and urbanism, with urbanism explaining a large share of the variation. Limited access exercise in rural counties is slightly confusing, because the popular conception of rural areas as having lots of access to nature introduces interesting questions about the interpretation of exercise by the participants.

**Conclusion**

To answer the question ‘Are there differences in environment across counties that voted for Clinton and for Trump, the answer is simple: there are some dimensions along which Trump voting counties and Clinton voting counties are different, and others along which the two are the same. Trump Counties are less populous. They have lower per-capita income, higher air pollution, higher obesity rates, and less access to exercise. They also have higher graduation rates, lower unemployment, and less violent crime.

This paints a rather nuanced and complicated picture of the differences of the counties according to political preference. Trump campaigned on improving the job market, but the counties he won had low unemployment. This suggests either that his jobs argument was not a motivating factor for his voters, or a measurement problem in the unemployment statistic (people who have given up looking for work are not counted). The lower per-capita income suggests this mismeasurement interpretation. His counties also have lower crime rates despite his tough on crime rhetoric, suggesting again, either an ineffective argument, or that his supporters were not reacting to crime in their neighborhoods, but to national crime reporting.

The health problems in Trump voting counties - higher obesity rates and less access to exercise are obviously related. Both lead to higher health expenditures, leading to a greater reliance on the healthcare system. It seems that voters with higher obesity rates either tolerated or supported repealing the Affordable Care Act.

The graduation rate measure raises the critical remaining question of our analysis. Are people’s voting responses responses to local conditions or an indication of their preferences? Do Democratic counties have lower graduation rates because they vote for Democrats? Or do people vote for Democrats because graduation rates are low? Or even, does policy not matter and Democrats just happen to be popular with groups who happen to live in areas with low graduation rates? Unemployment, per-capita income, access to healthcare, and violent crime all fall in this same dilemma. All we can observe with this study are the differences, not the causes or mechanisms. Air pollution, and environmental policy fortunately sidestep this issue. Democrats run on cleaner environments, and have cleaner air. Republicans run on industry, and tolerate more pollution. This difference is reasonably attributable to policy decisions, which indicate voter preferences.

The results of our analysis indicate that voters don’t just react to conditions in their county, they react to things nationally as well. There are significant differences between counties across multiple metrics, but the biggest difference is in population and access to exercise - both measures of urbanism. Our analysis indicates that campaign rhetoric works - campaigning on crime can work even where crime is low. Republicans succeeded both in ‘struggling counties’ with lower per-capita income and higher obesity rates, as well as in ‘successful counties’ with high graduation rates and low violent crime. The quality of the unemployment measure, and the causal ordering of voter preferences and county conditions still require further investigation.

Works Cited

Dougherty, Michael Brendan. "Why Talk Radio Coronated Donald Trump." The Week - All You Need to Know about Everything That Matters. The Week, 08 Mar. 2016. Web. 03 May 2017.

Gamio, Lazaro. "Urban and Rural America Are Becoming Increasingly Polarized." The Washington Post. WP Company, 17 Nov. 2016. Web. 03 May 2017.

"The Presidential Election: Illness as Indicator." The Economist. The Economist, 16 Nov. 2016. Web. 3 May 2017.

Glass, Ira, and Zoe Chace, prods. "This American Life: The Beginning of Now." *This American Life*. Chicago Public Media and Public Radio Exchange. Chicago, 28 Apr. 2017. Radio.