DRAFT WORKING PAPER

Explaining nationalist political views: The case of Donald Trump

Jonathan Rothwell, Gallup

Pablo Diego-Rosell, Gallup

Last revised November 2, 2016[1]

Abstract

The 2016 US presidential nominee Donald Trump has broken with the policies of previous Republican Party presidents on trade, immigration, and war, in favor of a more nationalist and populist platform. Using detailed Gallup survey data for 125,000 American adults, we analyze the individual and geographic factors that predict a higher probability of viewing Trump favorably. The results show mixed evidence that economic distress has motivated Trump support. His supporters are less educated and more likely to work in blue collar occupations, but they earn relatively high household incomes and are no less likely to be unemployed or exposed to competition through trade or immigration. On the other hand, living in racially isolated communities with worse health outcomes, lower social mobility, less social capital, greater reliance on social security income and less reliance on capital income, predicts higher levels of Trump support. We confirm the theoretical results of our regression analysis using machine learning algorithms and an extensive set of additional variables.

Introduction

The 2016 Republican Party presidential nominee in the United States is Donald Trump, a man who has based his campaign largely on restricting immigration, in part by building a large wall along the border with Mexico and barring Muslims from entering the country, and restricting trade, by re-negotiating trade agreements and imposing tariffs on China and possibly other countries.[2] Already a famous businessman and television personality, Trump created headlines by accusing President Barack Obama of having conspired to forge his US-based birth certificate, despite evidence to the contrary.[3] With these positions and others, including his criticism of former president George W. Bush and the Iraq War, Trump's candidacy has attracted the support of right-wing nationalists, and provoked criticism from the Republican Party's political leaders and leading media figures.[4]

---

[1] ACKNOWLEDGMENTS: The author would like to thank a number of colleagues, scholars, interested citizens, and writers for comments and suggestions, including Frank Newport, Rajesh Srinivasan, Gary Gates, John Fleming, Bill Petti, Lydia Said, Ben Ryan, Andrew Dugan, Carol Graham, Manas Chattopadhyay, William Minter, Marcus K Felson, Jeff Guo, Robert Putnam, and James Kwak.

[2] The Associated Press, "How Donald Trump's Plan to Ban Muslims Has Evolved", available at http://fortune.com/2016/06/28/donald-trump-muslim-ban/

[3] Christina Wilkie, "Trump Admits Obama Was Born in The United States After 5 Years Of Racist Denials" Huffington Post September 19, 2016, available at http://www.huffingtonpost.com/entry/donald-trump-barack-obama-birth_us_57dbed6ee4b0071a6e068090.

[4] David A Graham, "Which Republicans Oppose Donald Trump? A Cheat Sheet" The Atlantic October 31, 2016, http://www.theatlantic.com/politics/archive/2016/10/where-republicans-stand-on-donald-trump-a-cheat-sheet/481449/

This article examines the characteristics of Trump's supporters with a view to establishing broader insight into what factors motivate nationalist political identification. Trump's nationalist appeals were evident in his acceptance speech of the Republican Party's nomination:

"The most important difference between our plan and that of our opponents, is that our plan will put America First. Americanism, not globalism, will be our credo. As long as we are led by politicians who will not put America First, then we can be assured that other nations will not treat America with respect. This will all change in 2017. The American People will come first once again."[5]

There is a large literature on historic and contemporary nationalist or nativist parities in Europe (Muddle 2007) and to some extent the United States (eg Lipset and Raab 1970). In a study of perhaps the most infamous party, the geography of voting patterns reveal that the political supporters of Hitler's National Socialist party were disproportionately Protestants, if living in a rural area, and those in lower-middle class administrative occupations and owners of small businesses, if living in an urban area (Hamilton 1982). Thus, neither the rich nor poor were especially inclined to support the Nazi Party, and even among Christians, religious identity mattered greatly. This general result is confirmed in more recent empirical work on the rise of the Nazi party (King, Rosen, Tanner, and Alexander 2008).

Among social scientists, there is a strong consensus that class or occupational-based indicators are highly relevant to broad political party preferences and this class context has mattered for decades (Brooks and Manza 1997). Yet, at the same time, racial identity and cultural attitudes clearly shape political preferences even within social classes, and the empirical evidence for this is quite strong (Ibid). For example, political support for George Wallace, a nationalism candidate in many respects, was only weakly correlated with blue collar occupational status in Northern states, but highly correlated with that status in Southern states (Hamilton 1972). Across countries, preferences for income redistribution are strongly correlated with income and education, but race, gender, and cultural traditions are also strongly predictive (Alesina et al 2011).

In work closely related to this project, Mansfield and Mutz (2009) find that ethno-centrist and isolationist world-views predict opposition to free trade, and after accounting for these factors, individual economic characteristics such as education are not significant. In a similar analysis, but of support for outsourcing, Mansfield and Mutz (2013) find that nationalism, ethno-centrism, and isolationism predict opposition to outsourcing, but objective economic threat—in terms of occupational or industrial employment—does not. Likewise, in an extensive recent review of the literature on hostile attitudes toward immigrants, Hainmueller and Hopkins (2014) conclude that labor market competition—as a cause of restrictive attitudes across a broad swatch of the population—receives very little empirical support. Instead, they argue that attitudes toward cultural assimilation are more important, whereby dislike for immigrants increases as native residents regard them as less culturally similar.

There is less empirical literature explaining why people support extreme political views generally and right-wing nationalism in particular. Recently, however, a small body of research likens the rise of more extreme politics in the United States ("political polarization") to shocks from import competition (Autor,

[5] Politico, "Full text: Donald Trump 2016 RNC draft speech transcript," July 21, 2016, available here http://www.politico.com/story/2016/07/full-transcript-donald-trump-nomination-acceptance-speech-at-rnc-225974#ixzz4LSicfyx9

Dorn, Hanson, and Majlesi 2016) or income inequality (McCarty, Rosenthal, and Poole 2006). On the other hand, Che et al (2016), however, find that, from 1992 to 2010, counties more exposed to Chinese imports became more supportive of the Democratic Party generally and candidates that oppose free trade and support economic assistance more specifically. None of these studies considers nationalist political views akin to Trump's "America first" rhetoric and proposals.

The results from the 2016 "Brexit" referendum, in which a majority of UK voters decided to leave the European Union, have been interpreted to suggest that exposure to competition from globalization was a factor motivating support, but those who have looked at the results find that it is unlikely that most Brexit voters are personally threatened by globalization. Some of the strongest support was among people with pensions and people receiving disability payments.[6]

In any case, a reasonable hypothesis for explaining the attraction of the stridently anti-establishment sentiment behind Trump's campaign is that he is disproportionately drawing support from people directly harmed by globalization or otherwise on the losing side of growing income inequality.

It may also be the case that income or other standard economic measures are not especially helpful in explaining nationalism or political preferences generally. Though there is no clear line separating cultural from economic issues, Ill-health and cultural behaviors and attitudes may be more revealing of political preferences. Beyond standard economic measures, there is evidence that whites are unusually pessimistic about their well-being, after adjusting for other factors (Graham forthcoming). Specifically, lower-income whites and older whites exhibit this pessimism compared to other groups (ibid). Along these lines, middle-aged whites have experienced a rise in mortality rates in the last decade and a half (Case and Deaton 2015).[7] The real stress, economic or otherwise, stemming from poor health or subjective anxiety, whatever its origin or relationship to objective circumstances, could very well motivate extreme political views.

Additionally, there is a large body of theoretical and empirical literature explaining the conditions of inter-group conflict. In the early 20th century, research on the military, policing, and public housing found that inter-group conflict reduced prejudice toward African-Americans (Pettigrew and Tropp 2006; 2011). The American psychologist Gordon Allport (1954) is credited with establishing this theory in the social science literature, and he further stipulated that contact reduces prejudice when certain conditions are met: equal status between groups; cooperatively working toward common goals, and under the support of an external authority (Pettigrew and Tropp 2006; 2011).

Since, then a large literature has confirmed Allport's theory, and even found that the conditions do not necessarily need to be present, at least in modern settings, in which formal civil right laws have already been established (Pettigrew and Tropp 2006). At the personal level, one recent study finds that friendly contact with other groups reduces anxiety around the threat of rejection and eases comfort with

[6] Anatole Kaletsky, "Trump's rise and Brexit vote are more an outcome of culture than economics" The Guardian October 28, 2016.

[7] This finding was called into question by Andrew Gelman, who worried that middle-aged people were older in the later period than in the former, biasing the results, but his subsequent analysis and dialog with Deaton confirmed that there was a notable increase in white mortality from 1999 to 2013. See blog from Andrew Gelman, "Age adjustment mortality update" November 6, 2015, Statistical Modeling, Causal Inference, and Social Science, available at http://andrewgelman.com/2015/11/06/age-adjustment-mortality-update/

physical and conversational engagement (Barlow et al 2009). At the scale of metropolitan areas, Rothwell (2012) finds that racial segregation but not diversity predicts lower levels of social capital, measured by trust and volunteering, in the United States.

In so far as nationalist political attitudes are characterized by suspicion of ethnic outsiders, contact theory would predict less support for nationalist political parties. In a direct test of that hypothesis, Biggs and Knauss (2012) find that neighborhood level exposure to minorities predicts lower membership rates in British nationalist parties. These neighborhood level correlations would be biased if people sort into neighborhoods based on political preferences, but Kaufmann and Harris (2015) finds no evidence of geographic sorting based on nationalist political views. In related work, Sorenson (2014) finds that an initial wave of immigration at the local level increased support for a right-wing party in Norway, but the effect quickly faded out, which the author suggests is the result of direct contact with immigrants.

Social engagement may also affect one's political identity and provide opportunities for cooperative engagement with diverse groups of people. Scholarship on this topic has focused on the other direction of this causal chain: how identification with ethnic or civic membership affects social capital. In Europe, empirical research on that topic measures nationalism as the level of importance ascribed to having the nation's ancestry (ie "French ancestry" for people living in France). Reeskens and Wright (2012) find that more nationalist individuals, in this sense, are less likely to trust others, less likely to participate in membership associations, and less likely to volunteer. By contrast, those with greater civic pride (ie rating respect for the country's institutions and laws as highly important) are generally more likely to trust others and participate in organizations (Reeskens and Wright 2012). It is unclear from this work if identity drives social engagement or the other way around. Likely, they are mutually reinforcing.

This analysis attempts to explain the characteristics of Trump supporters and test four hypotheses:

1. Social or economic hardship increases the likelihood of Trump support;
2. Exposure to globalization via foreign imports or immigrant labor increases the likelihood of Trump support;
3. Contact with immigrants or racial minorities reduces the likelihood of Trump support;
4. Living in a rich social-capital community—with respect to civic but not religious engagement--increases the likelihood of Trump support.

For the first, the analysis distinguishes traditional measures of economic hardship like income and employment status from alternative measures, such as health and intergenerational mobility. Health is a core component of well-being, and the latter may be related to one's hope for the future well-being of offspring, which may, in turn, directly affect personal well-being and level of satisfaction with the political status quo. To capture wealth effects, we examine housing price changes at the quasi-neighborhood level and separately look at credit leverage, as well as the sources of personal income see how household financial health and stability may affect political preferences.

We also analyze economic threats through completion with foreign companies (through trade) or foreign-born workers (through immigration). This is done at local level by examining the level and trends in manufacturing employment and distance to the Mexican border, which is highly predictive of Latin American immigrant population shares. We also calculate individual exposure measures using variation across states and occupations in terms of the share of workers in the manufacturing sector and the share who are foreign-born.

While these are all objective measures of social or economic hardship, previous work using the same database finds that subjective financial insecurity is particularly high among Trump's supporters and not explained by objective measures.[8] We exclude this and other subjective measures here because we consider them endogenous to political preferences. People's opinions about their finances and a variety of related issues are fundamentally tied to their opinions about politics and are informed by the cultural and media narratives that shape those political views.

To test the third hypothesis, we focus on the degree of neighborhood racial and ethnic segregation using zip code data. Aside from whatever competitive effect there is from immigration, which is debatable and varied by education and occupation, exposure to immigrants could reduce support for Trump through positive theory.

Finally, we explore community social capital measures in terms of voter participation rates and the number of non-profit associations. Communities in which these values are higher may be more engaged in civil society and more trusting of government institutions and elite organizations. We also examine religious adherence, which we posit may increase support for Trump among mainstream Christian groups, since religious communities tend to be ethnically homogenous.

The large body of theoretical and empirical literature underpinning these hypotheses represents an advantage from a scientific perspective, as the constructs and relationships in our models have been tested and validated before. It may also represent a weakness: as the complexity of the models increases, so does the possibility of model misspecification. In order to minimize this risk we compare the performance of a traditional hypothesis-driven approach with a data-driven machine learning approach. Machine learning algorithms can be a powerful tool to identify complex interactions and nonlinear effects in models with a large number of independent variables (Athey and Imbens 2016). Carefully chosen and trained algorithms will also maximize predictive performance, providing a robustness check for the hypothesis-driven model, and providing a comprehensive assessment of the relative contribution of each independent variable to the overall model.

The paper proceeds with a description of the data and methods, a summary of the ideological differences between Trump supporters and other groups, particularly other Republicans, and a comparison of Trump's supporters to others on basic economic indicators. Next, the analysis proceeds into the heart of the empirical analysis by first establishing a baseline set of models that provide tests of the relevant hypotheses and show variable importance scores. We then show a select number of robustness checks before moving into our main robustness analysis, which uses machine learning to attempt to improve upon the baseline model. The discussion concludes by summarizing what we learned using machine learning and situates the findings in a larger theoretical context of nationalism.

Data and analysis

Data

The main data are Gallup Daily Tracking survey microdata, collected from July 8, 2015 through October 11, 2016. Reached through random digit dialing of both cellular and landline phones, 132,815 American adults were asked if they hold a favorable view of Donald Trump over this period. Of these, 125,430

[8] Jonathan Rothwell, "Financial Insecurity Higher for Those Who Favor Trump" Gallup, October 10 2016.

responded either yes or no (with 63% reporting an unfavorable view). Of the entire sample, 4.2% reported no opinion, 1.4% percent said they never heard of him, and less than one percent (0.9%) refused to answer. The analysis discards observations from those who did not answer yes or no. Survey weights developed by Gallup methodologists were applied to make the sample nationally representative. This is a very large sample relative to the two thousand or less used in comparable studies (eg Mansfield and Matz 2009, 2013).

Complementing Gallup microdata, external data sources were merged into the database using respondent level information on their zip code, county, or state of residence.

Zip code level data from the IRS could be directly matched, but the US Census reports data in zip code tabulation areas (ZCTAs), which had to be mapped to traditional zip codes using a crosswalk funded by the US Health Resources and Service Administration and made available from UDS Mapper.[9] Zip code values spread across multiple ZCTAs were averaged and collapsed to single zip codes using the ZCTA share of zip code population as the weight.

County level data was used for some indicators, where finer levels of geographic detail were unavailable, but to study the economic context of respondents, counties were aggregated to commuting zones (CZs) using a county-CZ crosswalk developed by David Dorn and made available on his personal website. Commuting zones were created by researchers at the US Department of Agriculture as a more encompassing alternative to metropolitan areas, which are not available for many rural communities (Tolbert and Sizer 1996). By contrast, every US resident is assigned a CZ, though they vary greatly in size from single lightly populated counties to many densely populated counties like those surrounding New York City and Los Angeles. The method is similar to those of metropolitan areas in that counties are the fundamental unit of analysis and county to county commuting ties form the basis for whether or not they are grouped together.

Methods and sources of external data

The baseline method used is multi-variable probit regression, designed to estimate how various factors are associated with the binary probability of holding a favorable view of Trump in Gallup's Daily Tracking surveys. This provides straightforward results estimating the marginal effect of a variable on the probability of favoring Trump, which cannot be regarded as causal, given that this is an observation study.

As described below, this method is supplemented with machine learning algorithms, which will help alleviate concerns with model uncertainty, complex interaction effects, multi-collinearity, and non-linear relationships between variables. Machine learning, to be clear, does not help address concerns about variable endogeneity or causality. As a first step in that direction, the focus of this paper is to describe the objective conditions of those who favor Trump, while inferring some insights into why these factors might matter or not from a causal perspective.

Equation one shows the basic model.

$$1.\, P\left[T_{i,c}\right] = I_{i,c} + C_c + \varepsilon_{i,c}$$

[9] Accessed July 2016, available at http://udsmapper.org/zcta-crosswalk.cfm

The probability of individual i residing in commuting zone c of holding a favorable view of Trump is modelled as a function of a vector of individual characteristics (I) and geographic characteristics, including measures at the zip code, county, and commuting zone levels (C). The residual is assumed to have both an individual and commuting zone component, and errors are clustered at the commuting zone level to account for within CZ correlations. Sample weights are also applied to make the analysis representative at the national level. We include binary variables for the month-year in which the respondent was asked about Trump to account for variation in his popularity over the course of the campaign.

This basic set up will also be repeated in three different versions of the database: all respondents with complete data, white non-Hispanic respondents, and white non-Hispanics who say they are Republic Party members or lean in that direction. We also consider how these models work in explaining favorability toward other candidates beside Trump.

Individual covariates include household income—measured in 12 brackets, employment status, and 11 occupational categories, which line up with the major categories in the Standard Occupational Classification system used by the US Census Bureau, except that professions (except managers and business owners) are grouped together and most low-paying service occupations are similarly collapsed into one category in the Gallup data.

Our baseline measure of household income is the mid-point of the bracket. We explore alternative income measures, including a within-bracket imputation of median household income by state and age-cohort using data from the Current Population Survey's 2015 Annual Social and Economic Supplement, accessed via IPUMS-CPS (Flood et al 2015). income IRS data at the zip code level. These figures divide total zip code income reported on tax records by the number of

For the C set of variables, there is considerable uncertainty about which measures are the most suited to test the hypotheses described above. Economic stress, for example, could be indicated through many different channels. Modelling these channels well requires getting the right concepts, from the best available sources, and using the appropriate specifications. Machine learning aids in this empirical challenge because it does not require strong assumptions about which variables to include or not and whether they are best measured in linear or non-linear ways.

Here we briefly describe the variables used in our baseline model, as well as our more notable robustness checks, which we include as ways to potentially improve upon the baseline model.

Commuting zone level covariates

- inter-generational mobility
- distance to the Mexican border
- the manufacturing share of employment
- educational attainment

To measure how inter-generational mobility may relate to political preferences, the analysis uses data from Chetty et al (2014), which they constructed using an Internal Revenue Service database of all federal tax records for individuals born between 1980 and 1982 and the records of their parents. Intergenerational mobility is calculated as the average CZ national income rank at age 30, for individuals raised at the 25th percentile of the national income distribution, using family income between ages 15

and 20. Chetty and Hendren (2016) subsequently developed a method to estimate the causal effect of a commuting zone on intergenerational mobility, to eliminate selection effects from migration. This indicator of the "causal effect" of the CZ on intergenerational mobility is the preferred variable for this analysis, because it filters out local migratory effects. It can be interpreted as the net causal effect of growing up in a CZ on income around age 30, conditional on parental income. Chetty and Hendren (2016) describe that this variable is correlated with things like school quality, lack of concentrated poverty, and two-parent households.

Distance to Mexico was calculated by first allocating the centroid longitude and latitude coordinates for the largest county by 2010 population to each commuting zone, using data from American Fact Finder and the 2010 Decennial Census. Second, distance between these CZs and the Mexican border was approximated by grouping CZs into longitudinal regions and calculating their distance to one of five border MSAs using the "vincenty" command in STATA, based on their longitude: San Diego, for the westernmost CZs with longitudinal coordinates less than -115.345; Yuma (-115.3 to -112.8); Tucson (-112.8 to -109.0); El Paso (-109.0 to -102.2); McAllen (>-102.2).

The manufacturing share of employment is calculated using data from the Quarterly Census of Employment and Wages (QCEW). Since data suppressions are present in even the high level data file, the analysis is supplemented using manufacturing and other employment level estimates from Acemoglu et al (2016). They developed a method to impute over County Business Patterns data suppressions and have made their data available. The analysis also uses an index of Chinese import exposure from Autor, Dorn, and Hanson (2013).

CZ level educational attainment data are calculated from county data available from the 2010-2014 American Community Survey.

County level

- population density
- mortality rates, CZ level
- Non-profit associations
- Voter participation
- Debt to income ratio (robustness check)

Population density is included in the dataset by Gallup's data processing team, using the latest Census data. How density relates to politics in light of the hypotheses listed above is somewhat ambiguous, but there is strong evidence that density fosters more diverse and dynamic economic activity, and in that sense provides a relevant measure of economic opportunity. On the other hand, density is closely tied to cultural patterns and exposure to diversity, so observed effects may not be strictly economic.

Health status for likely Trump supporters is measured using the 2014 mortality rate for white non-Hispanics aged 45 to 54. These data are from the US Centers for Disease Control (CDC Wonder). County level total deaths and population were aggregated to the CZ level. All causes of death were included. To supplement the analysis, other measures of mortality were examined, as well as data from Chetty et al (2016) that measures life expectancy at birth at various income levels for the entire CZ population.

To measure social capital, we use 2009 county-level data from Rupasingha, Goetz, and Freshwater (2006). They report a social capital index, which includes, among other things, voter participation rates

and the number of non-profit associations. I focus on the latter two, which I find to be the most robust and exclude the other measures included in their broader index, except in robustness checks.

To measure consumer debt, we use county-level data measuring debt to income ratios, with data from Mian, Rao, and Sufi (2013).

Neighborhood level (zip codes)

- racial segregation
- sources of income and mortgage interest to income ratio (robustness check)
- age-adjusted disability rates (robustness check)
- housing price changes, 3-digit zip code level (robustness check)

Racial segregation in this analysis is measured by matching survey respondent zip codes to census population data (originally from ZCTAs) from the 2010-2014 5-year American Community Survey. Segregation is measured in two ways. The first uses the difference in the white share of population at the zip code and CZ levels. The second uses the ratio of CZ diversity to zip code diversity, using standard diversity index measures.

1. Racial isoaltion of whites = zip code share white – CZ share white

2. $Racial\ isolation = \frac{\text{Diversity index CZ}}{\text{Diversity index zip}}$

$$whereby\ Diversity\ index = 1 - \sum({p_g}^2)$$

For both, higher values indicate greater segregation. The overall white share of the CZ population is included as a control variable in the first and overall CZ diversity is included as a control in the second.

To better approximate the individual health of the respondent, we also measure health status by analyzing age-adjusted disability rates for the working age population at the zip code level, using the 2010-2014 American Community Survey files.[10]

To measure wealth and other household financial issues missed by income, we include data from a number of sources. Housing price appreciation from 2007 to 2015 is calculated to measure potential losses or gains in housing prices since the pre-recession peak. These data are from the Federal Housing Finance Agency and described by Bogin, Doerner, and Larson (2016). We use their three-digit zip code

[10] Because risk of disability increases sharply with age, I calculate age-adjusted disability rates for each zip code, using population data from the 2010 Decennial Census. The method predicts disability rates for the working-age population (defined here as 18 to 64) each zip code as a function of the percentage of residents aged 55 to 64, 45 to 55, 35 to 44, and 26 to 35, with the omitted group those between 18 and 25. The predicted disability rate from this population-weighted regression was subtracted from the actual rate to calculate an age-adjusted disability rate.

data, which is narrow enough geographically to approximate the local housing market but also broad enough to yield a high match rate with the Gallup database.

We also use zip code level data from the Internal Revenue Service (IRS) to measure a number of neighborhood level indicators of financial health: mean income, the share of income from capital—to measure exposure to financial markets or monetary policy, the share of returns with unemployment insurance income, the share of income from business partnerships, mortgage interest deduction to income ratio (as an alternative debt to income ratio), the value of the mortgage interest deduction (to proxy for wealth), and the share of income from social security.

The machine learning approach

As a complementary robustness check we introduce a supervised machine learning approach, whereby we attempt to find the combination of predictors that would most accurately classify a randomly selected training sample in the right category of Trump supporters, and we validate the performance of the resulting models against a randomly held-out validation sample. We select three algorithms that often perform well in binary classification problems such as the one at hand: C5.0 (Kuhn et al., 2015), Gradient Boosting Machines (GBM, Ridgway, 2015) and Random Forests (Wright, 2016).

- **C5.0** is an evolution of the rule-based C4.5 decision tree algorithm by Quinlan (1993). The algorithm works by iteratively splitting the sample into child nodes, based on the predictor variable that maximizes the information gain ratio. Variable importance is calculated by determining the "percentage of training set samples that fall into all the terminal nodes after the split" (Kuhn et al., 2015, p.6).
- **GBM** is an ensemble learning algorithm first described by Friedman (2001) that generates multiple decision trees, each grown on the cases that were misclassified on the previous tree. The final prediction is obtained from a weighted ensemble of all the regression trees. Variable importance is calculated by the number of times a variable is selected for splitting, weighted by the squared improvement to the model as a result of each split, and averaged over all trees (Elith et al., 2008).
- **Random Forests** is also an ensemble learning technique first described by Breiman (2001). To avoid overfitting common in single decision trees, the random forests algorithm builds a user-defined number of decision trees, taking a random subset of the training data for each tree, and a random subset of predictor variables to determine each split in the tree. Variable importance is determined by the Mean Decrease Accuracy (MDA) of the forest when the values of a predictor are randomly permuted.

The parameters of each algorithm were estimated (trained) on a randomly selected training dataset containing 80% of the sample, stratified on Trump support to ensure that the proportion of supporters is equivalent to the full dataset. Two sets of variables were used, a 'main' set containing the same 68 independent variables used in the final Probit model, and an 'extended' set, containing a total of 134 independent variables available in the full dataset. The machine learning models built on the main set of variables provide a robustness check on the Probit model specification. The models built on the extended set of variables provide a check for omitted variable bias. After eliminating cases missing data on any of the variables in the model, the training dataset using the main set of predictors contained 72,499 cases, of which 26,530 (36.6%) were Trump supporters, while the training set using the extended set of predictors contained 70,002 cases, of which 25,523 (36.5%) were Trump supporters.

To guard against overfitting, all machine learning models were trained using k-fold cross-validation with five folds, using R 'caret' package version 6.0-72 (Kuhn, 2016). K-fold cross-validation partitions the training dataset into *k* equal sized subsamples, of which *k*-1 subsamples are used to tune the model parameters, and the single remaining sample is used to validate the tuning parameters. Performance of the trained models was tested on held-out validation datasets containing 20% of the sample, stratified on Trump support to ensure that the proportion of supporters is equivalent to the full sample. The error rate found on the validation dataset, called the out of sample error or generalization error, are used to compare the predictive performance of the models. After eliminating cases missing data on any of the variables in the base model, the validation dataset using the base set contained 19,630 cases, of which 7,154 (38.3%) were Trump supporters, while the training set using the extended set contained 18,875 cases, of which 6,846 (36.3%) were Trump supporters.

Summary data

Ideology of Trump supporters

Before describing the main data, it is worth noting where Trump's supporters identify themselves on the ideological spectrum. The evidence suggest Trump's supporters are significantly further to the right than even other Republicans. The Gallup Daily Tracker asks respondents to "describe your political views" and provides options of very conservative, conservative, moderate, liberal, or very liberal. 59.1% of those who favor Trump describe their views as very conservative (15.6%) or conservative (43.5%) compared to 23.8% of those who do not favor Trump (5.0% and 18.8%, respectively). They are also less likely to identify a moderate than those who do not like Trump (26.9% vs 36.6%). Those who identify as Republicans but do not like Trump are also much more likely to view themselves as moderate ideologically (35.9%) than Republicans who favor Trump (24.3%). Finally, those who favor Trump rarely describe their party affiliation as Democratic or lean Democratic. Just 12.2% report this, and only 5.7% of those who favor Trump and no not favor his Democratic Party challenger for President, Hillary Clinton.

In addition to broad ideological orientation, Republicans who favor Trump are also significantly more likely than other Republicans to oppose trade and immigration, consistent with the candidate's public statements. The Gallup Daily Tracker reserves space for topical questions asked for brief periods of time. In May 2016, one question asks whether or not the respondent agrees or disagrees with the statement "End U.S. participation in free trade deals, such as NAFTA" and another states "Reform immigration laws to provide automatic green cards for high skilled workers" Among Republicans who favor Trump, 58 percent oppose trade deals and 57 percent oppose reforming immigration. By contrast, among Republicans who do not support Trump, 42 percent oppose trade deals and 28 percent oppose reforming immigration laws.

Main database

The primary dataset is summarized in Table 1. Each of these variables is used in the baseline model. To save space, we refer readers to the appendix for any additional information about data used in our extensive robustness checks and the machine learning models.

Prior to any formal analysis, the summary data are themselves revealing on a few of the hypotheses described above.

First, income levels are not consistent with the argument that Trump's supporters are especially stressed economically. Mean household income—using the mid-point of Gallup brackets—is $81,898 for those who view Trump favorably and $77,046 for those who do not. This gap is roughly the same if one imputes median incomes by state and age cohort within each bracket using micro-data from the 2015 CPS. The income gap in favor of those who like Trump widens even further if one attempts to adjust income for regional purchasing power, using median county rent, because Trump's supporters tend to live in areas with a lower cost of living.

Second, employment status also does not show a disadvantage to Trump's supporters. His supporters are less likely to be unemployed and less likely to be employed part-time. Labor force participation is lower among Trump supporters, but this is only an age effect. Adjusting for age, there is no difference.

Third, education and occupational differences do suggest that Trump's supporters may be disadvantaged economically, despite their relatively strong showing on income and employment. 26% of those favoring Trump hold a bachelor's degree or post-graduate degree compared to 35% of those who do not favor Trump. Just 17% of those reporting a favorable view of Trump are professionals, but 23% of those who do not report a favorable view of Trump.

Still, some caveats apply here. Among non-professionals, Trump's support is higher among those in skilled blue collar occupations, such as production, construction, installation, maintenance, and repair, and transportation, which tend to pay better than service occupations, such as food preparation, personal care, security, and cleaning. Likewise, both business owners and managers are somewhat more likely to favor Trump.

Fourth, the summary data provide mixed evidence as to whether global competition is bearing down more forcefully on Trump's supporters. They are slightly more likely to work in production occupations and construction occupations, which face stiff competition in the form of trade and immigration, but they are also more likely to be business owners and managers, groups who would be more likely to benefit from global competition.

A more systematic set of calculations reveals similar mixed evidence. To calculate exposure to immigrant labor competition, we match respondent occupational codes to the share of workers in that occupation that are foreign-born by state, using data from IPUMS-USA and the 2014 American Community Survey (Ruggles et al 2015), since state's vary greatly in the extent to which they face immigrant competition. We use the same procedure to calculate exposure to import competition, except we replace the share who are foreign-born with the share of workers in each occupation who are employed in the manufacturing sector.

Those who favor Trump work in occupations with somewhat lower shares of immigrant workers (17.8% vs 16.1%), but Trump's supporters are somewhat more likely to be in the manufacturing sector (9.6% vs 8.2%), both shares are down by roughly 1.4 percentage points since 2000. Still, given these figures, it is unlikely that trade has directly affected more than a small fraction of those who favor Trump. Most of those currently working in manufacturing or production are, if anything, beneficiaries of global trade

liberalization. The same could be said of many workers in goods transportation and repair, as well as retail and wholesale industries.[11]

Fifth, those who favor Trump are dramatically more likely to be Christian (though not Morman), heterosexual, aged 40 or older, male, and non-Hispanic white. People with these characteristics comprise 35% of those who favor Trump, though just 14% of those who do not. This speaks to the importance of cultural and philosophical issues, as opposed to economic incentives.

Sixth, the geography of those who like Trump differs in important ways even at the summary level where ecological fallacy is a concern, given the different individual compositions of these areas. Those who view him favorably live in counties with lower population density, CZs with lower college attainment rates and higher mortality rates for middle-aged whites. The zip codes of those who favor Trump are more likely to be disproportionately white relative to white population share in the CZ. In data now shown, we also calculate that neighborhood (census tract) level racial and ethnic entropy and diversity is lower among those who favor Trump, using either a Theil index or diversity index, across eight racial or ethnic groups (Minnesota Population Center, 2016).[12]

Findings

Baseline results of probit model

To make sense of which of the above factors are most important, we turn to our baseline multivariate probit analysis, which is designed to reveal the marginal predictive power of each individual variable, holding the others constant.

In Table 2, we report the specific coefficients and significance of each explanatory variable in the full sample, a subset of non-Hispanic whites, and a third subset of non-Hispanic white Republicans. This approach accounts for potentially complex interactions between race, on the one hand, and race and party affiliation, on the other. As reviewers have noted, this table is hard to interpret if one's aim is to determine the importance of specific factors, so we focus our discussion on variable importance measures using the t-statistics from our baseline regression on the full sample.

As shown in Table 3, 44 out of 52 variables reach statistical significance and 22 variables have t-statistics above 5.0. The most important explanatory factor is whether someone is male, which is highly predictive of greater support for Trump (t-stat of 34.8). Whether someone has a post-bachelor's degree is also hugely predictive (negative), along with Hispanic ethnic status (negative), whether someone is an atheist (negative), and black (negative).

As for how these findings relate to the main hypotheses, start with economic stress.

The most basic measures of economic status—income and employment—go the opposite way as the hypothesis would predict. Income is somewhat important (t-stat of 5.9) but predicts greater Trump support. Self-employment predicts greater Trump support as well and is highly important (t-stat of 7.7). Unemployment status does predict Trump support but is relatively unimportant (t-stat of 3.5). Being out

[11] Further, there is little difference in the average job growth of occupations since 2000. Both production occupations and administrative and clerical occupations have lost the most jobs since 2000. Growth, meanwhile, has been strong for managerial positions, as well as professional jobs and low-skilled service jobs.

[12] This uses data from the 2010-2014 American Community Survey.

of the labor force is the least important measure of all 52 variables and is not even statistically significant. Part-time employment status is also unimportant.

On the other hand, education status is very important and fits the hypothesis. Those with bachelor's degrees or graduate degrees are much less likely to support Trump. More subtle measures of economic status also suggest that those who favor Trump are distressed in important ways. The CZ level mortality rate is rather important (t-stat of 4.7) and consistent with the view that those who favor Trump are confronting social or material hardship. The intergenerational mobility measure is among the least important but achieves statistical significance (t-stat of -2.4) and suggests that people living in areas with diminished economic opportunity are somewhat more likely to support Trump. Education and both of the community level effects remain significant among white non-Hispanics.

The globalization theory looks weaker. The occupational categories most exposed to trade competition (production, t-stat of 3.6) and competition with foreign labor (farming and construction) are not among the most important variables (t-stats of 4.6 and 2.2). Even worse for the theory, two community level variables display the opposite relationship. The probability of favoring Trump falls as the CZ share of jobs in the manufacturing sector increases (-3.1), and the probability of favoring Trump increases as distance from Mexico increases (2.5). Neither of these variable are among the most important, but they suggest that exposure to globalization is actually lower among Trump supporters at the community level.

The results for contact theory are strong and as predicted. The racial and ethnic isolation of whites at the zip code level is one of the strongest predictors of Trump support (12.0). This result is robust to using a more general measure of diversity, which is lower in the zip codes of Trump supporters overall and relative to the CZ. The isolation effect remains significant among non-Hispanic whites and non-Hispanic white Republicans. After controlling for this relative isolation at the quasi-neighborhood level, the CZ share of whites is not significant. The distance to Mexico variable is also consistent with contact theory, and one could interpret the county population density variable as providing further evidence, since population density brings with it a more diverse mix of occupations and more opportunities for formal and informal exchange (Rothwell 2012). The share of workers with a bachelor's degree offers a complementary interpretation, though also indicates greater community affluence, consistent with the economic deprivation hypothesis.

Social capital theory also receives nuanced support in this analysis. People are far more likely to favor Trump if they affirm that religion is an important part of their daily life (17.7). These social ties are likely intra-ethnic, but secular examples of social capital at the county level indicate less support for Trump. The number of non-profit associations and voter participation rates predict lower favorability (-4.4 and -2.6, respectively).

Probit robustness checks

We ran a large number of robustness checks to determine if alternative measures of the above concepts affected the results. We briefly summarize these now, before turning to a more systematic accounting in the machine learning section.

To further explore measures of economic stress, we included IRS zip code data on the level and sources of income. People living in zip codes with greater exposure to capital income sources—as a share of

income or in terms of the percentage of tax returns—are less likely to support Trump.[13] Trump support rises as mortgage interest payment deductions (an indicator of mortgage debt) increase, which could indicate more leveraged finances. Likewise, Trump support increases in zip codes with a higher percentage of returns claiming the earned income tax credit or a higher share of income from Social Security, a program for the elderly and disabled. On the other hand, I find no evidence that Trump supporters have seen greater or lower housing price appreciation in their areas or that they are more highly leveraged through consumer credit. Taken together with the income findings above, these additional results suggest that Trump supporters may be more dependent on the government, less wealthy, and thus less secure financially, even if trends in their home values are typical.

On health, life expectancy at age 40—age and race adjusted—was not predictive, but a variety of other measures of health were. These include age-adjusted mortality rates in 2014 for all races (at the CZ level), the change in CZ level life expectancy (age-adjusted), the share of births at low-weight, and the prevalence of diabetes. On the other hand, Trump support was less common in places with excessive alcohol consumption, and no more likely in places with high rates of death from drug poisoning or high obesity rates. Most tellingly, there is a very strong and robust relationship between zip code level disability rates and support for Trump among the working-age population. This holds using the gross disability rate, as well as age-adjusted measures. Since zip-codes are more likely to measure individual circumstances than commuting zones, this provides some insight into a potential channel linking health status to nationalist political preferences.

To study community level exposure to trade, the main model used 2015 employment shares in manufacturing, but 2000 manufacturing employment shares also predict significantly lower support for Trump, whether measured by the QCEW or Acemoglu et al (2015)'s imputation of suppressions in County Business Patterns data. Likewise, an increase in the share of manufacturing employment from 2000 to 2007 (using data from Autor, Dorn and Hanson 2013) predicts higher levels of Trump support, which is the opposite of the hypothesized relationship. Overall manufacturing employment growth from 2000 to 2015 has no effect, controlling for 2015 shares, nor does the change from 1990 to 2007. Total employment growth predicts greater Trump support, suggesting that people in more economically prosperous metropolitan areas are marginally more likely to view him favorably. Finally, a measure of Chinese import exposure is unrelated to Trump support.

Likewise, we use variation across occupations and states to measure the probability that the respondent is in the manufacturing sector or works in an occupational category with a high share of immigrants. Neither of these globalization-exposure variables predict greater support for Trump.

A binary variable for residence in the southern United States enters positively and significantly in the model, but has little effect on the coefficients of the other geographic variables, with the exception of inter-generational mobility, which losses statistical significance. Southern residence predicts more conservative politics generally and is no longer significant in the sample that restricts to non-Hispanic white Republicans. Still, given the South's distinctive institutional legacy with respect to the importance

[13] This could be interpreted positively or negatively with respect to financial stress. On the one hand, wage and salary income may be more stable and predictable than capital income from stocks and business ownership, but people more reliant on capital income likely have more wealth.

of slavery and the Jim Crow system of oppressions, we include this regional dummy in our baseline probit model, which we use to compare against the machine learning algorithms.

Machine Learning Results

Overall Model Performance

Performance of the probit model was compared to the machine learning models in terms of Area Under the Curve (AUC). AUC is calculated from the Receiver Operating Characteristic curve, which plots the model's true positive rate (Sensitivity) and false positive rate (100-Specificity) for different cut-off points of the model's predicted probabilities. Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold. A perfect prediction would have an AUC of 100%, whereas a random prediction would have an AUC of 50%.

Table 4 shows the model performance for all the 'main' and 'extended' models, with an additional 'ensemble' model that is built by averaging of predictions from the C5.0, GBM and Random Forests models, with GBM as the stacking method used to weight predictions. For the main model, the gain in performance obtained from the machine learning models over the theory-driven probit model is small. Of all three algorithms, the relatively simple rule-based C5.0 algorithm provides the best performance, with an AUC = 0.7551. Ensembling the models provides the largest boost to model performance, with an AUC of 0.7619, but the difference with the probit model is small. For the extended model, the gains in performance are even more modest. In the case of the random forests model, the machine learning approach is in fact inferior to the theory-driven model. The performance of the ensemble model is minimally better than the ensemble model using the main set of variables.

Variable Importance

Table 5 shows the variable importance scores for the main models, re-scaled to the 0-100 range for comparability, as well as an average of the three importance scores. Since the method used to calculate variable importance varies substantially between the three main models, an importance rank is also calculated for robustness. Variables are sorted in Table 5 by their average rank. It is immediately obvious that individual demographic factors are key. Gender, race and age are consistently ranked among the most important variables across all models. Religion also plays an important role in all models. The importance of religion is a very important predictor in the GBM and RF models, and in the C5.0 model, religious affiliation (Muslim, Jew, other non-Christian) ranks among the top predictors by variable importance. Educational attainment, sexual orientation and veteran status are also important individual-level indicators. Among the indicators at the aggregate level, the number of non-profits in the area is the most important predictor, followed by the CZ share with BA or higher in 2014, and the Zip code isolation of whites. Out of 68 factors (including the year-month time effects), these are in the top 12.

[TABLE 5 GOES HERE]

We next look at variable the variable importance scores for the extended models (Table 6). The top variables in terms of importance remain similar, with gender, race, age and religion still at the top. The main differences between the main and the extended model is in terms of the specific race and religion attributes that reach high importance, with Hispanics and atheists having greater importance as predictors in the extended model. Among the variables that are included in the extended model and not in the main model, the most important are the log of population density, the cost of living, income

inequality (5-year Gini index), reliance on social security income, and the county diabetes rate. State dummy variables are, as a group, the least useful set of predictors.

[TABLE 6 GOES HERE]

Decision Rules

In order to better interpret the implication of the variable importance scores, we grew a C5.0 tree with 5-fold cross-validation for robustness, and a minimum child node size of 500 individuals for ease of interpretation. The importance of race becomes immediately obvious as black is the first split in the tree. The black group has a very low prevalence of Trump support (12.1%). It is also homogeneously so, with the algorithm not identifying any other variable to further split the 'y_racebk' group into more homogeneous subgroups. The role of religion also becomes immediately apparent in the next split of the tree. For the non-black, being an atheist is the next strongest split in terms of Trump support. Among non-black atheists, only 20.6% support Trump, compared to 43.1% among non-black non-atheists. Going further down the tree, Hispanics show the lowest support for Trump among non-black non-atheists, with 19.2% showing support, compared to 45.7% among non-black non-atheists non-Hispanics. The node showing the strongest rates of support for Trump includes heterosexual white males above 37 years of age, with less than a post-graduate education and for whom religion is important. Among this group, 65.8% support Trump.

Machine Learning vs Probit Discussion

The machine learning approach used in this paper performed similarly well to the more traditional probit model. This represents both a strength and a weakness. The strength is that the "agnostic" non-parametric models using largely automated algorithms can match the performance of an approach supported by a vast literature and a long tradition in econometric research. The algorithmic approach can automatically identify important interactions and non-linear effects that might have otherwise gone undetected, and the lack of a major difference in performance provides strong evidence of correct model specification.

The weakness is that the machine learning approaches are still not well integrated into social science. The standard results from these algorithms are difficult to interpret from the perspective of hypothesis testing and are not designed to communicate which direction the predictive effective goes (in this case toward or away from Trump).

The machine learning approach does, in any case, offer some interesting possibilities from an analytical perspective. The boost in predictive performance gains could prove relevant in situations where predictive accuracy is of interest, even if they do not have relevance for the advancement of theory. Predicting candidate support, turnout, and vote are all high-stakes activities where a small gain in performance could have major impact. Variable importance analysis offers insights into the relative importance of different variables in terms of their predictive accuracy, inviting further investigation using traditional methods. The automatic identification of homogeneous sub-groups represents a useful segmentation tool that can be used to conduct further sub-group analysis for research purposes, or for applied purposes, such as designing micro-targeted campaigns and voter outreach operations.

Summarizing the theoretical results across models

Given the large amount of data analyzed and presented here, we focus this discussion section on the results that complement or add nuance to those presented above.

As with the probit model, the economic hardship explanation for Trump's favorability received mixed support in the machine learning results, using either the parsimonious or extended variable list.

Income is among the 20 most important variables in both machine learning approaches, but predicts greater Trump support, as shown in the probit models, contradicting the predictions of the economic hardship explanation for nationalism, but consistent with literature on European nationalism that it often attracts people in the middle class. Meanwhile, unemployment and not working or looking for work are relatively weak predictors across the probit and machine learning models.

Greater support for the notion that economic hardship matters comes from the education results, which are very important across all three approaches. Post-bachelor's degree status is in the top 10 across the probit model and the two machine learning results. Mortality is moderately important in the parsimonious machine learning results, but other health measures rise up in the extended machine learning models. In particular, the diabetes rate at the county level (17th most important of 180 factors), the share of births at low-weight at the county level (20th), and overall age-adjusted mortality at the county level are all highly important predictors of greater Trump support.

Likewise, more refined measures of income at the zip code level predict Trump support and are highly important. Of note, reliance on social security income at the zip code level is very important (19th out of 180) and could indicate age clustering, in which case it is of not necessarily of economic importance, or it could indicate reliance on disability payments, in which case it would be of great importance. The disability rate at the zip code level is moderately important (49th).

There is also some evidence that Trump supporters are more likely to come from communities with lower economic mobility across generations, using data from Raj Chetty and his colleagues. The extended machine learnings models confirm the importance of various mobility measures. College attendance mobility, for example, defined as the probability that someone at the 25th percentile of the family income distribution will enter college by age 23, predicts lower Trump support and is the 24th most important variable across the machine learning models. Right behind it at 25th is a measure of the connection between parental incomes and adults incomes, which is higher in areas with lower mobility.

Contradicting this evidence somewhat is the fact that higher test scores, adjusted for parental income, predict higher not lower Trump support. This may be the result of racial segregation, which scholars have found continues to lead to lower educational opportunities for black and Hispanic children, but it contradicts the narrative that Trump's supporters are socially marginal.

The globalization hypothesis performs no better in the machine learning models. The first community level indicator of manufacturing to show up is ranked 70th (the share of jobs in manufacturing in 2000, as measured by Autor, Dorn, and Hanson) and other measures perform much worse, and as noted, predict the wrong preferences for the hypothesis to hold. Whether someone works in a production occupation rates 94th. The distance to Mexico rates 140th. Taken together with the result above, there is no evidence to suggest that direct exposure to globalization is an important factor in explaining people's view of Trump.

The probit results for contact theory and the modified interpretation of social capital theory hold up to the rigors of the machine learning algorithms. Indicators of racial isolation and non-profit exposure rank 9th and 11th in the extended models. The importance of religion to Trump's base, however, is a reminder that not all social capital has equal effects.

Conclusions

These results do not present a clear picture between social and economic hardship and support for Trump. The standard economic measures of income and employment status show that, if anything, more affluent Americans favor Trump, even among white non-Hispanics. Employment status and occupational categories are among the least important explanatory variables for Trump favorability.

Surprisingly, there appears to be no link whatsoever between greater exposure to trade competition or competition from immigrant workers and support for nationalist policies in America, as embodied by the Trump campaign. These results make it very unlikely that direct exposure to harm from globalization could be a causal factor in motivating large numbers of Trump's supporters.

Yet, various measures of health, longevity, and intergenerational mobility at the community or zip code level do relate to the likelihood of viewing Trump favorably, and these data indicate that low levels of social or economic well-being are a factor in his support. Moreover, higher disability rates and greater reliance on Social Security income and the Earned Income Tax Credit are predictive of Trump support. Education is also hugely important predictor of one's views on Trump and suggests that his supporters are worse off than many Americans on that dimension.

The causal mechanisms linking health and intergenerational well-being to political views are relatively unexplored in the social science literature. The most straightforward interpretation is that ill-health leads to disaffection with the political status quo, in the same way as low-income might. With intergenerational mobility, it may be that parents see their children or those in their community failing to reach milestones predictive of success and find fault with the political status quo. In any case, these results warrant further study.

The analysis provides clear evidence that those who view Trump favorably are disproportionately living in racially and culturally isolated zip codes and commuting zones. Holding other factors, constant support for Trump is highly elevated in areas with few college graduates and in neighborhoods that standout within the larger commuting zone for being white, segregated enclaves, with little exposure to blacks, Asians, and Hispanics.

This is consistent with contact theory, which has already received considerable empirical support in the literature in a variety of analogous contexts. Limited interactions with racial and ethnic minorities, immigrants, and college graduates may contribute to prejudicial stereotypes, political and cultural misunderstandings, and a general fear of not-belonging.

Finally, the results concerning race, age, gender, ethnicity, and religion imply that political behavior is not related directly and neatly to economic self-interest or class position. One could argue that blue collar men would disproportionately benefit from Trump's policies on trade and immigration, but as discussed above, these factors are relatively unimportant. As others have found, cultural views and social identity are likely quite important in affecting political preferences.

Given the complexity of political behavior and the large number of unanswered questions and unknown interactions, we also believe that machine learnings offers an important complement to traditional social science methods in cases such as these. We have used it here to shed light on the importance of alternative measures of various concepts and hope others will pursue some of these findings in greater detail.

References

Acemoglu, Daron, David Autor, David Dorn, Gordon H. Hanson, and Brendan Price, "Import Competition and the Great US Employment Sag of the 2000s," Journal of Labor Economics 34:S1 (2016): S141-S198

Alesina, Alberto, Paola Giuliano, A Bisin, and J Benhabib. 2011. "Preferences for Redistribution." Handbook of Social Economics, 93-132. North Holland, 93-132.

Allport, Gordon. 1954. The Nature of Prejudice (Addison-Wesley).

Athey, Susan and Imbens, Guido. 2016. "The State of Applied Econometrics - Causality and Policy Evaluation" Working Paper, https://arxiv.org/pdf/1607.00699v1.pdf

Autor, D., Dorn, D., Hanson, G., & Majlesi, K. (2016). Importing Political Polarization?, available at http://www.kavehmajlesi.com/uploads/7/2/8/9/7289299/adhm-politicalpolarization.pdf ;

Autor, David David Dorn, Gordon Hanson, "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," American Economic Review, 103(6) (2013), 2121-2168.

Barlow F. K., Louis W. R., Hewstone M. (2009). Rejected! Cognitions of rejection and intergroup anxiety as mediators of the impact of cross-group friendships on prejudice, British Journal of Social Psychology, 48, 389-405.

Biggs, Michael and Knauss, Steven. 2012. Explaining Membership in the British National Party: A Multilevel Analysis of Contact and Threat, European Sociological Review 28 (5): 633-646

Bogin, A., Doerner, W. and Larson, W. (2016). Local House Price Dynamics: New Indices and Stylized Facts. Federal Housing Finance Agency, Working Paper 16-01. The working paper is accessible at http://www.fhfa.gov/papers/wp1601.aspx.

Breiman, L. 2001, Random Forests, Machine Learning 45(1), 5-32.

Brooks, Clem and Jeff Manza, "Class Politics and Political Change in the United States, 1952-1992" Social Forces 76 2 (1997): 379-408

Burke, B.L; Kosloff, S; Landau, M. 2013. Death Goes to the Polls: A Meta-Analysis of Mortality Salience Effects on Political Attitudes 34 (2): 183-200.

Case, Ann and Deaton, Angus. 2015. Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century, PNAS 112 (49) 15078-15083.

Che, Yi, Yi Lu, Justin R. Pierce, Peter K. Schott, Zhigang Tao. 2016. "Does Trade Liberalization with China Influence U.S. Elections?" NBER Working Paper No. 22178

Chetty, R. and Hendren, N. 2016. "The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates." Harvard University and NBER.

Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., & Cutler, D. 2016. The association between income and life expectancy in the United States, 2001-2014. *JAMA*, *315*(16), 1750-1766.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez, 2014. Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States, Quarterly Journal of Economics 129(4): 1553-1623

Elith, J., Leathwick, J. R., & Hastie, T. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*(4), 802-813.

Flood, Sarah, Miriam King, Steven Ruggles, and J. Robert Warren. 2015. Integrated Public Use Microdata Series, Current Population Survey: Version 4.0. [Machine-readable database]. Minneapolis: University of Minnesota.

Friedman, J. H. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.

Graham, C. forthcoming. The Unequal Pursuit of Happiness? Inequality in Agency, Optimism, and Access to the American Dream.

Hamilton, Richard F. 1982. Who Voted for Hitler? (Princeton NJ: Princeton University Press).

Hamilton, Richard F. 1972. Class and Politics in the United States (New York: John Wiley and Sons).

Kaufmann, Eric and Harris Gareth. 2015, "White Flight" or Positive Contact? Local Diversity and Attitudes to Immigration in Britain, Comparative Political Studies October 2015 48: 1563-1590

King, Gary, Rosen, Ori, Tanner, Martin, and Wagner, Alexander. 2008. "Ordinary Economic Voting Behavior in the Extraordinary Election of Adolf Hitler" Journal of Economic History 68 (4): 951-995.

Kuhn, M., Weston, S., Coulter, N., & Quinlan, R. (2015). C50: C5.0 decision trees and rule-based models. *R package version 0.1. 0-24. URL https://cran.r-project.org/web/packages/C50/C50.pdf*.

Kuhn, M. (2016). Package 'caret'. *Classification and Regression Training 6.0.72 1-204. URL ftp://cran.r-project.org/pub/R/web/packages/caret/caret.pdf*.

Lipset, S. M., and Raab, E. (1970). The politics of unreason: right wing extremism in America, 1790-1970 (Vol. 5). New York: Harper & Row.

Mansfield, Edward D. and Mutz, Diana C. Mutz 2013. "US versus Them: Mass Attitudes toward Offshore Outsourcing," *World Politics* 65 (571-608).

Mansfield, Edward D. and Mutz, Diana C. Mutz 2009. "Support for Free Trade: Self-Interest, Sociotropic Politics, and Out-Group Anxiety," *International Organization* 63 (425-457).

McCarty, Nolan, Howard Rosenthal, and Keith T. Poole. 2006. Polarized America: The Dance of Ideology and Unequal Riches. MIT Press.

Mian, A. Rao, K, and Sufi, A. "Household Balance Sheets, Consumption, and the Economic Slump" The Quarterly Journal of Economics (2013) 128 (4): 1687-1726.

Miller, David. 2000. Citizenship and National Identity. London, UK: Polity Press

Minnesota Population Center. 2016. National Historical Geographic Information System: Version 11.0 [Database]. Minneapolis: University of Minnesota, http://doi.org/10.18128/D050.V11.0

Muddle, Cas. 2007. Populist Radical Right Parties in Europe (Cambridge University Press).

Pettigrew T. F., Tropp L. R. 2006. A meta-analytic test of intergroup contact theory, Journal of Personality and Social Psychology, 90, 751-783.

Pettigrew T and Tropp, L. 2011. *When Groups Meet: the Dynamics of Intergroup Contact* (Psychology Press, 2011)

Putnam, R. D. 2007, *E Pluribus Unum*: Diversity and Community in the Twenty-first Century The 2006 Johan Skytte Prize Lecture. Scandinavian Political Studies, 30: 137–174

Quinlan, R (1993). C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, http://www.rulequest.com/see5-unix.html

Reeskens, Tim and Wright, Matthew Wright. 2012. "Nationalism and the Cohesive Society: A Multilevel Analysis of the Interplay Among Diversity, National Identity, and Social Capital Across 27 European Societies" *Comparative Political Studies* 20 (10): 1-29.

Ridgeway, G. 2015. Package 'gbm'. *R package version 2.1.1 1-34. URL https://cran.r-project.org/web/packages/gbm/gbm.pdf*.

Rothwell, J. T. 2012. The effects of racial segregation on trust and volunteering in US cities. Urban Studies, 49 (10): 2109-2136.

Rupasingha, A., Goetz, S. J., and Freshwater, D. 2006. The production of social capital in US counties. Journal of Socio-Economics, 35, 83–101

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek.2015 *Integrated Public Use Microdata Series: Version 6.0* [Machine-readable database]. Minneapolis: University of Minnesota.

Sorenson, Rune Jorgen. 2014. "After the immigration shock: The causal effect of immigration on electoral preferences," Electoral Studies 44 (1-14).

Tolbert, C. and Sizer, M. 1996. *U.S. commuting zones and labor market areas: A 1990 update*. Washington, DC: US Department of Agriculture.

Wright, M. 2016. Package 'ranger': A Fast Implementation of Random Forests. *R package version 0.6.0 1-19. URL*

Table 1. Summary statistics of primary data used in analysis for full sample, for those who favor Trump, and for those who do not favor Trump

				Favorable view of Trump		Unfavorable view of Trump	
	Obs	Mean	St. deviation	Mean	St. deviation	Mean	St. deviation
Percent who view Trump favorably	125,430	0.35	0.48	1.00	0.00	0.00	0.00
Income	106,091	$78,716	$63,495	$81,898	$61,953	$77,046	$64,227
Income, purchasing power parity adjusted	106,091	$80,097	$64,520	$87,736	$66,554	$76,087	$63,057
Income, census imputation within brackets	105,438	$80,215	$72,204	$83,345	$71,345	$78,570	$72,597
Income, census imputation within brackets, ppi adjusted	105,438	$81,451	$72,530	$89,144	$75,893	$77,410	$70,364
self-employed, full-time	125,430	0.06	0.24	0.08	0.27	0.05	0.22
employed part-time	125,430	0.12	0.33	0.11	0.31	0.13	0.34
unemployed	125,430	0.04	0.19	0.03	0.17	0.04	0.19
not in labor force	125,430	0.31	0.46	0.33	0.47	0.30	0.46
Catholic	125,430	0.23	0.42	0.23	0.42	0.23	0.42
Jewish	125,430	0.02	0.14	0.01	0.12	0.02	0.15
Muslim	125,430	0.01	0.09	0.00	0.05	0.01	0.10
Mormon	125,430	0.02	0.14	0.02	0.14	0.02	0.14
Other non-Christian	125,430	0.03	0.16	0.01	0.12	0.03	0.17
Atheist	125,430	0.18	0.39	0.11	0.31	0.22	0.41
Religion is important	125,430	0.64	0.48	0.72	0.45	0.59	0.49
Veteran or family member of veteran	125,430	0.26	0.44	0.33	0.47	0.22	0.41
Works for government	125,430	0.10	0.30	0.09	0.29	0.10	0.30
Gay, lesbian, or trans-sexual	125,430	0.04	0.19	0.02	0.13	0.05	0.22
male	125,430	0.49	0.50	0.58	0.49	0.44	0.50
Has children	124,803	0.32	0.47	0.30	0.46	0.33	0.47
Married	125,430	0.52	0.50	0.61	0.49	0.47	0.50
Was married	125,430	0.15	0.36	0.16	0.37	0.15	0.36
union member, non-government	125,430	0.04	0.19	0.04	0.19	0.04	0.19
Occupation: Business owner	125,430	0.02	0.14	0.03	0.16	0.02	0.13
Occupation: Managers	125,430	0.05	0.23	0.06	0.23	0.05	0.23
Occupation: Professional	125,430	0.21	0.40	0.17	0.37	0.23	0.42
Occupation: Sales	125,430	0.05	0.22	0.05	0.22	0.05	0.22
Occupation: Clerical	125,430	0.04	0.19	0.03	0.17	0.04	0.20
Occupation: Construction	125,430	0.04	0.19	0.05	0.22	0.03	0.18
Occupation: Manufacturing/production	125,430	0.03	0.17	0.03	0.18	0.03	0.16

Occupation: Transportation	125,430	0.02	0.15	0.03	0.17	0.02	0.13
Occupation: Installation, maintenance, repair	125,430	0.02	0.14	0.03	0.16	0.01	0.12
Occupation: Farmer	125,430	0.02	0.12	0.02	0.14	0.01	0.11
Hispanic	125,430	0.14	0.35	0.06	0.24	0.18	0.39
Less than High School diploma	125,430	0.10	0.30	0.08	0.26	0.11	0.31
High school diploma	125,430	0.29	0.46	0.33	0.47	0.27	0.45
Technical degree program	125,430	0.03	0.18	0.04	0.20	0.03	0.16
Bachelor's degree	125,430	0.18	0.38	0.16	0.37	0.19	0.39
Post-bachelor's degree	125,430	0.14	0.34	0.10	0.30	0.16	0.37
Age	125,430	48.3	19.0	52.3	18.3	46.2	19.1
White	125,430	0.77	0.42	0.89	0.31	0.70	0.46
Black (or White and Black)	125,430	0.13	0.34	0.05	0.21	0.18	0.38
Asian (or White and Asian)	125,430	0.03	0.18	0.04	0.18	0.03	0.18
County: Population density	125,245	2102	6783	1274	4543	2540	7675
CZ: Bachelor's degree or higher attainment, 25 and older share of pop	125,417	0.29	0.08	0.28	0.08	0.30	0.08
CZ: Share employed in Manufacturing, 2015	125,420	0.11	0.06	0.11	0.07	0.10	0.06
CZ: Causal effect of growing up on income at age 30	124,339	0.00	0.28	0.02	0.32	0.00	0.26
County: Voter participation, 2009	124,045	0.57	0.09	0.58	0.09	0.57	0.09
County: Non-profit associations	124,045	4619	7570	3271	5932	5336	8222
(zip code share white) - (CZ share white)	121,597	0.01	0.20	0.05	0.17	-0.01	0.21
CZ: White share of population, 2014 5-yr	125,417	0.64	0.18	0.68	0.18	0.63	0.19
CZ center distance to Mexico in miles	125,430	1015	527	1025	498	1010	541
CZ 2014 white mortality rate per 100,000 pop, 45 to 54	123,940	403	117	426	126	390	109

Table 2. Baseline models--probit regression of Trump favorability on individual and geographic level characteristics

	Dependent variable=Probability of holding a favorable view of Trump		
	Full sample	White non-Hispanics	White non-Hispanic Republicans
	1	2	3
Individual level variables			
ln of household income	0.0437***	0.0350***	0.00797
	(0.00746)	(0.00792)	(0.0127)
self-employed	0.184***	0.147***	0.135***
	(0.0240)	(0.0266)	(0.0438)
employed part-time	-0.0378*	-0.0894***	-0.0561
	(0.0195)	(0.0235)	(0.0370)
unemployed	0.134***	0.0982**	0.137
	(0.0389)	(0.0470)	(0.0896)
not in labor force	-0.00890	-0.0559**	-0.00212
	(0.0201)	(0.0231)	(0.0393)
Catholic	-0.0300	0.0251	0.121***
	(0.0221)	(0.0212)	(0.0294)
Jewish	-0.303***	-0.330***	0.0917
	(0.0347)	(0.0308)	(0.0688)
Muslim	-0.567***	-0.883***	-0.325
	(0.0858)	(0.147)	(0.397)
Mormon	-0.239***	-0.309***	-0.527***
	(0.0505)	(0.0523)	(0.0625)
Other non-Christian	-0.369***	-0.517***	0.0200
	(0.0417)	(0.0430)	(0.114)
Atheist	-0.340***	-0.395***	0.0322
	(0.0180)	(0.0206)	(0.0447)
Religion is important	0.285***	0.339***	0.0727**
	(0.0161)	(0.0162)	(0.0289)
Veteran or family member of veteran	0.179***	0.160***	0.107***
	(0.0118)	(0.0133)	(0.0217)

Works for government	0.0222	0.0114	0.0147
	(0.0234)	(0.0240)	(0.0373)
Gay, lesbian, or trans-sexual	-0.331***	-0.535***	-0.289***
	(0.0352)	(0.0383)	(0.0974)
male	0.386***	0.378***	0.214***
	(0.0111)	(0.0117)	(0.0223)
has children	0.0456***	0.0909***	0.00831
	(0.0149)	(0.0174)	(0.0305)
married	0.0702***	0.0787***	0.00272
	(0.0152)	(0.0180)	(0.0330)
was married	0.0307	0.00957	-0.0452
	(0.0189)	(0.0210)	(0.0374)
union member, non-government	-0.104***	-0.143***	0.0821
	(0.0331)	(0.0373)	(0.0644)
Occupation: Business owner	0.0919**	0.0550	0.0705
	(0.0410)	(0.0435)	(0.0657)
Occupation: Managers	0.0200	-0.00294	-0.0166
	(0.0248)	(0.0278)	(0.0471)
Occupation: Professional	-0.0474***	-0.0705***	-0.0689**
	(0.0167)	(0.0194)	(0.0337)
Occupation: Sales	0.111***	0.112***	0.0764
	(0.0264)	(0.0296)	(0.0530)
Occupation: Clerical	-0.0385	-0.0422	-0.0796
	(0.0312)	(0.0370)	(0.0643)
Occupation: Construction	0.142***	0.191***	0.0537
	(0.0306)	(0.0375)	(0.0664)
Occupation: Manufacturing/production	0.119***	0.170***	0.154**
	(0.0334)	(0.0376)	(0.0673)
Occupation: Transportation	0.203***	0.306***	0.191***
	(0.0378)	(0.0447)	(0.0720)
Occupation: Installation, maintenance, repair	0.267***	0.317***	0.156**
	(0.0440)	(0.0532)	(0.0753)
Occupation: Farmer	0.102**	0.186***	0.105

	(0.0471)	(0.0515)	(0.0789)
Hispanic	-0.657***		
	(0.0279)		
Less than High School diploma	-0.0966***	0.0270	0.0895
	(0.0316)	(0.0372)	(0.0682)
High school diploma	0.0806***	0.0970***	0.0752***
	(0.0146)	(0.0155)	(0.0257)
Technical degree program	0.102***	0.143***	0.193***
	(0.0272)	(0.0307)	(0.0518)
Bachelor's degree	-0.249***	-0.288***	-0.282***
	(0.0161)	(0.0166)	(0.0269)
Post-bachelor's degree	-0.487***	-0.546***	-0.461***
	(0.0169)	(0.0198)	(0.0325)
age	0.0446***	0.0392***	0.0510***
	(0.00658)	(0.00791)	(0.0130)
age^2	-0.000645***	-0.000521***	-0.000641***
	(0.000126)	(0.000151)	(0.000244)
age^3	2.98e-06***	2.18e-06**	2.48e-06*
	(7.33e-07)	(8.79e-07)	(1.41e-06)
White	0.260***		
	(0.0236)		
Black (or White and Black)	-0.655***	-0.680***	-0.446**
	(0.0349)	(0.0866)	(0.189)
Asian (or White and Asian)	0.162***	0.0580	0.0804
	(0.0333)	(0.0424)	(0.0707)
Geographic level effects			
County: Population density	-0.0203***	-0.0391***	-0.0472***
	(0.00618)	(0.0107)	(0.00982)
CZ: Bachelor's degree or higher attainment, 25 and older share of pop	-0.0809***	-0.0958***	-0.0445**
	(0.0148)	(0.0164)	(0.0213)
CZ: Share employed in Manufacturing, 2015	-0.0319***	-0.0389***	-0.0150
	(0.0103)	(0.0113)	(0.0156)
CZ: Causal effect of growing up on income at age 30	-0.0168**	-0.0174**	-0.0168
	(0.00715)	(0.00751)	(0.0104)

County: Voter participation rate, 2008 presidential election	-0.0309***	-0.0378***	-0.0306*
	(0.0118)	(0.0122)	(0.0184)
County: Non-profit associations	-0.0422***	-0.0703***	-0.0387***
	(0.00962)	(0.0141)	(0.0137)
Racial isolation (zip code share white) - (CZ share white)	0.0817***	0.0597***	0.0352***
	(0.00682)	(0.00923)	(0.0129)
CZ: White share of population, 2014 5-yr	-0.0162	-0.0666***	-0.0747***
	(0.0163)	(0.0191)	(0.0234)
CZ center distance to Mexico in miles	0.0398**	0.0394**	0.0488**
	(0.0159)	(0.0178)	(0.0246)
CZ 2014 white mortality rate, 45 to 54	0.0547***	0.0495***	0.0631***
	(0.0118)	(0.0125)	(0.0173)
Constant	-2.145***	-1.701***	-0.680***
	(0.118)	(0.149)	(0.256)
Observations	101,547	78,832	28,610
Pseudo R-squared	0.156	0.128	0.0707

Robust standard errors in parentheses, clustered on CZ. *** p<0.01, ** p<0.05, * p<0.1. CZ level variables are standardized to have mean of zero and standard deviation of one. All models include fixed effects for the month of the survey. The omitted occupational references category is “service” (eg personal care, food preparation, security, building cleaning and maintenance). The omitted educational reference category is “some college, with no degree.”

Table 3. Rank importance of explanatory variables using t-statistics on baseline probit model

	t-stat
male	34.75
Post-bachelor's degree	-28.89
Hispanic	-23.54
Atheist	-18.91
Black (or White and Black)	-18.79
Religion is important	17.72
Bachelor's degree	-15.43
Veteran or family member of veteran	15.19
Racial isolation (zip code share white) - (CZ share white)	11.99
White	11.03
Gay, lesbian, or trans-sexual	-9.41
Other non-Christian	-8.85
Jewish	-8.72
self-employed	7.67
age	6.79
Muslim	-6.60
Occupation: Installation, maintenance, repair	6.06
ln of household income	5.85
High school diploma	5.54
CZ: Bachelor's degree or higher attainment, 25 and older share of pop	-5.47
Occupation: Transportation	5.36
age^2	-5.12
Asian (or White and Asian)	4.86
Mormon	-4.72
CZ 2014 white mortality rate, 45 to 54	4.65
Occupation: Construction	4.63
married	4.63
County: Non-profit associations	-4.39
Occupation: Sales	4.21
age^3	4.06
Technical degree program	3.74
Occupation: Manufacturing/production	3.56
unemployed	3.46
County: Population density	-3.29
union member, non government	-3.15
CZ: Share employed in Manufacturing, 2015	-3.11
has children	3.07
Less than High School diploma	-3.05
Occupation: Professional	-2.83
County: Voter participation rate, 2008 presidential election	-2.62
CZ center distance to Mexico in miles	2.49
CZ: Causal effect of growing up on income at age 30	-2.35
Occupation: Business owner	2.24
Occupation: Farmer	2.18
employed part-time	-1.94
was married	1.62
Catholic	-1.36

Occupation: Clerical	-1.23
CZ: White share of population, 2014 5-yr	-0.99
Works for government	0.95
Occupation: Managers	0.81
not in labor force	-0.44

T-statistics calculated using robust standard errors, clustered on commuting zones

Table 4. Accuracy and AUC of Main and Extended Models

	Main	Extended
Probit	0.7506	0.7506
C5.0	0.7551	0.7583
GBM	0.7539	0.756
Random Forests	0.7514	0.7335
Ensemble	0.7619	0.7632

Table 5. Variable Importance Score and Rank by Algorithm (Main model)

	Importance Score				Importance Ranking			
	C5.0	GBM	RF	Average	C5.0	GBM	RF	Average
male	100.0	80.7	78.9	86.5	1	2	3	2.0
white	100.0	73.7	88.8	87.5	1	4	2	2.3
black	98.5	80.3	68.1	82.3	7	3	4	4.7
age	100.0	43.6	62.6	68.8	1	8	8	5.7
religion is important	83.5	100.0	100.0	94.5	19	1	1	7.0
number of nonprofits	92.6	54.0	59.2	68.6	10	6	9	8.3
Post-BA	89.6	32.7	65.5	62.6	14	9	6	9.7
LGB	100.0	17.6	34.9	50.8	1	13	16	10.0
Bachelor's	94.4	22.9	56.2	57.8	8	11	12	10.3
veteran or family member	89.3	50.2	58.7	66.0	16	7	10	11.0
CZ share with BA or higher 2014	92.8	11.3	56.9	53.7	9	15	11	11.7
Zip code isolation of whites	87.9	59.5	50.3	65.9	17	5	13	11.7
Other non-Christian	91.2	27.2	23.7	47.3	11	10	21	14.0
muslim	100.0	9.9	9.0	39.6	1	16	31	16.0
jew	100.0	2.8	19.6	40.8	1	26	24	17.0
region is south	79.1	9.1	50.2	46.2	26	17	14	19.0
causal effect of growing up in CZ on income	82.1	8.7	18.9	36.6	21	18	26	21.7
ln hhincome (midpoint of Gallup brackets)	82.0	5.9	22.7	36.8	22	22	22	22.0
installation occupation	78.0	17.6	11.3	35.6	29	12	28	23.0
married	69.8	12.7	24.6	35.7	37	14	19	23.3
Share pop white CZ	83.4	4.1	13.3	33.6	20	23	27	23.3
technical degree or associates	89.5	6.0	5.1	33.6	15	21	35	23.7
transportation occupation	86.5	6.6	4.8	32.6	18	20	36	24.7
distance to Mexico	89.9	3.7	4.1	32.6	13	24	37	24.7
self-employed	77.1	3.2	23.8	34.7	31	25	20	25.3
High school diploma	71.8	2.2	25.5	33.2	35	29	18	27.3
professional occupation	60.5	6.8	19.9	29.1	42	19	23	28.0
employed part-time	90.7	2.1	1.4	31.4	12	31	47	30.0
pop density of county	51.4	2.4	40.6	31.5	56	27	15	32.7
construction occupation	78.4	1.0	5.3	28.2	27	40	34	33.7

catholic	76.8	2.4	2.7	27.3	33	28	41	34.0
production occupation	78.2	2.1	2.4	27.6	28	30	44	34.0
have children	76.8	1.6	3.6	27.3	34	34	38	35.3
Asian	77.0	1.6	2.9	27.2	32	35	39	35.3
not in labor force	79.6	1.6	2.3	27.8	25	36	45	35.3
share in mfg 2015 CZ	58.4	0.9	18.9	26.1	45	42	25	37.3
Hispanic	45.6	0.0	66.3	37.3	61	50	5	38.7
white mortality 2014, age 44-55	38.9	1.5	26.6	22.3	65	37	17	39.7
atheist	0.0	0.0	64.4	21.5	68	50	7	41.7
no HS diploma	67.8	0.6	2.7	23.7	38	47	40	41.7
time13	80.5	0.4	1.1	27.3	24	49	52	41.7
time11	54.9	1.9	2.5	19.8	52	32	43	42.3
unemployed	77.4	0.4	1.3	26.4	30	48	49	42.3
sales occupation	48.3	1.0	10.1	19.8	57	41	30	42.7
time8	59.9	1.2	1.9	21.0	43	39	46	42.7
Morman	44.4	1.5	8.8	18.2	63	38	32	44.3
time15	81.6	0.6	0.4	27.5	23	46	65	44.7
time5	59.4	0.7	1.4	20.5	44	44	48	45.3
causal effect of growing up in county on income	56.7	0.0	2.7	19.8	47	50	42	46.3
time2	71.7	0.0	0.9	24.2	36	50	54	46.7
voter participation 2008 election county	26.3	0.6	10.2	12.4	67	45	29	47.0
clerical occupation	55.4	1.8	0.8	19.3	50	33	58	47.0
time3	67.5	0.0	0.9	22.8	39	50	56	48.3
time10	35.8	0.0	6.4	14.1	66	50	33	49.7
in nonpublic sector union	55.0	0.0	1.1	18.7	51	50	51	50.7
time12	66.8	0.0	0.6	22.5	40	50	63	51.0
time9	55.9	0.8	0.6	19.1	49	43	62	51.3
time14	53.6	0.0	1.0	18.2	53	50	53	52.0
time6	58.0	0.0	0.7	19.6	46	50	60	52.0
time7	55.9	0.0	0.8	18.9	48	50	59	52.3
owner occupation	63.5	0.0	0.4	21.3	41	50	66	52.3
farmer occupation	53.3	0.0	0.8	18.0	55	50	57	54.0
works in govt	47.6	0.0	0.9	16.2	58	50	55	54.3
time4	43.0	0.0	1.2	14.7	64	50	50	54.7
age^3	53.5	0.0	0.5	18.0	54	50	64	56.0
management occupation	47.4	0.0	0.6	16.0	59	50	61	56.7

was married	45.9	0.0	0.4	15.4	60	50	67	59.0
age^2	44.8	0.0	0.0	14.9	62	50	68	60.0

Table 6. Variable Importance Score and Rank by Algorithm (Extended model)

	Importance Score				Importance Ranking			
	C5.0	GBM	RF	Average	C5.0	GBM	RF	Average
black	100.0	69.8	100.0	89.9	1	4	1	2.0
male	100.0	79.0	69.3	82.8	1	3	4	2.7
atheist	100.0	46.3	60.5	68.9	1	7	5	4.3
white	99.2	81.2	86.1	88.8	11	2	3	5.3
age	100.0	41.2	52.1	64.4	1	8	7	5.3
Post-BA	100.0	51.7	48.7	66.8	1	6	10	5.7
Hispanic	100.0	52.2	45.1	65.8	1	5	12	6.0
religion is important	94.2	100.0	99.1	97.8	16	1	2	6.3
number of nonprofits	93.5	24.1	43.3	53.6	17	10	13	13.3
log of population density	100.0	2.7	41.9	48.2	1	31	14	15.3
Zip code isolation of whites	87.7	13.3	52.8	51.2	33	13	6	17.3
veteran or family member	95.5	25.7	25.5	48.9	15	9	30	18.0
gini coefficient hh income 1999	100.0	1.6	20.2	40.6	1	42	40	27.7
cost of living CZ	99.8	1.0	25.6	42.2	10	49	29	29.3
ln hhincome (midpoint of Gallup brackets)	97.7	6.1	12.8	38.9	12	21	56	29.7
CZ share with BA or higher 2014	77.1	6.9	37.2	40.4	55	18	19	30.7
share diabetic, county	72.8	17.7	38.7	43.1	67	11	16	31.3
K-12 test scores adjusted for poor student share	93.4	2.2	20.2	38.6	18	35	41	31.3
reliance on SSI income zipcode	72.2	9.2	46.3	42.6	69	16	11	32.0
low birth rate share, county	91.6	3.2	14.6	36.5	23	27	49	33.0
high school diploma	89.1	4.3	14.6	36.0	27	24	50	33.7
age adjusted mortality rate 2014	83.5	1.3	35.7	40.2	41	45	20	35.3
LGB	100.0	9.8	5.2	38.4	1	15	91	35.7
upward mobility of college attendance, CZ	80.7	2.6	22.2	35.2	48	32	34	38.0

income at 30 regressed on parental income, county	78.3	2.4	21.0	33.9	52	33	38	41.0
region is south	82.6	5.9	11.7	33.4	44	22	61	42.3
white mortality 2014, age 44-55	67.6	2.7	29.8	33.4	81	30	25	45.3
self-employed	91.7	6.5	3.5	33.9	22	20	94	45.3
professional occupation	84.6	4.2	9.2	32.7	37	25	77	46.3
jew	96.0	3.0	1.9	33.6	13	28	100	47.0
income diversity zipcode	81.3	0.3	35.4	39.0	47	79	21	49.0
mortage interest payment to income ratio zipcode	69.2	2.0	21.4	30.9	76	37	36	49.7
Bachelor's	82.4	14.7	4.6	33.9	46	12	92	50.0
obesity rate, county	88.4	0.2	25.5	38.0	31	91	31	51.0
married	66.0	6.9	15.3	29.4	86	19	48	51.0
ln median hhincome income (ZCTA)	58.7	2.8	29.2	30.2	99	29	27	51.7
no HS diploma	93.3	1.6	2.6	32.5	19	41	97	52.3
share poor or fair subjective health	84.4	0.3	21.3	35.3	38	89	37	54.7
income to debt ratio 2006 county	70.9	1.8	13.5	28.7	73	39	52	54.7
pop density of county	46.3	8.6	41.7	32.2	133	17	15	55.0
social capital index county	49.6	11.6	22.2	27.8	119	14	33	55.3
sales occupation	92.2	1.7	1.4	31.8	21	40	105	55.3
installation occupation	86.5	2.4	1.6	30.1	35	34	104	57.7
diversity index CZ	86.7	0.0	15.3	34.0	34	94	47	58.3
reliance on capital income, zipcode	48.7	4.1	30.3	27.7	126	26	24	58.7
muslim	95.5	1.8	0.5	32.6	14	38	127	59.7
other non-Christian	76.0	5.4	2.6	28.0	59	23	98	60.0
CZ diversity divided by zipcode diversity	76.0	0.0	28.9	35.0	60	94	28	60.7
share disabled in zipcode	67.4	0.4	29.3	32.4	82	76	26	61.3
technical degree or associates	88.4	1.0	1.7	30.4	30	52	102	61.3
construction occupation	84.0	1.2	1.8	29.0	40	47	101	62.7
diversity index, zipcode	48.8	1.5	31.1	27.1	124	43	23	63.3
transporation occupation	85.3	1.2	1.2	29.2	36	46	108	63.3
social capital index county 2009	56.3	1.4	16.1	24.6	104	44	45	64.3
pct drink excess, county	54.6	0.9	20.4	25.3	106	56	39	67.0
catholic	76.8	0.8	7.7	28.4	57	59	90	68.7
age^3	55.2	0.0	51.3	35.5	105	94	8	69.0
Time10	78.0	2.1	0.7	26.9	53	36	119	69.3
share of children living with single moms 2014	75.3	0.0	13.2	29.5	62	94	53	69.7
number of clubs	59.3	0.4	17.0	25.6	98	74	42	71.3

Asian	84.4	0.7	0.8	28.6	39	61	114	71.3
housing price growth since 2007 zipcode3	57.5	0.4	16.0	24.6	101	71	46	72.7
Morman	77.5	0.8	1.0	26.4	54	58	109	73.7
upward mobility at age 30, county	49.1	0.9	13.1	21.0	121	53	55	76.3
have children	67.8	1.0	2.2	23.7	80	50	99	76.3
share with poor mental health	63.1	0.3	12.5	25.3	93	84	58	78.3
ln IRS mean hh income zipcode	47.1	0.0	38.6	28.6	128	94	17	79.7
Time4	90.0	0.0	0.7	30.2	26	94	120	80.0
age adjusted disabiltiy rate age 18-64 zipcode	21.3	1.1	34.7	19.0	171	48	22	80.3
share mfg 2000	61.2	0.4	9.5	23.7	96	72	75	81.0
growth total employment 2000 to 2014	63.8	0.3	9.9	24.7	92	81	73	82.0
Time5	72.3	0.6	0.9	24.6	68	66	112	82.0
share in mfg 1991 CZ	66.7	0.0	10.4	25.7	83	94	70	82.3
trade exposure to china 1999 2011	66.7	0.4	8.3	25.1	84	77	87	82.7
lnpop 2014 CZ	49.0	0.0	21.6	23.5	122	94	35	83.7
Time11	75.1	0.5	0.6	25.4	63	68	123	84.7
state = OR	90.0	0.0	0.5	30.2	25	94	135	84.7
voter participation 2008 election county	52.4	0.2	14.4	22.4	114	90	51	85.0
life expectancty CZ	54.2	0.0	13.2	22.5	107	94	54	85.0
state = FL	79.6	0.0	0.9	26.8	50	94	111	85.0
works in govt	82.5	0.0	0.8	27.8	45	94	116	85.0
state = LA	88.8	0.0	0.5	29.8	29	94	133	85.3
unemployed	87.7	0.3	0.4	29.5	32	86	139	85.7
Time9	92.9	0.0	0.4	31.1	20	94	143	85.7
ln IRS mean adjusted income	39.1	0.0	37.7	25.6	146	94	18	86.0
share poor physical health	49.0	0.0	16.7	21.9	123	94	43	86.7
causal effect of growing up in CZ on income at age 30 if born at 25th pct	53.5	0.5	8.6	20.9	109	70	82	87.0
CZ share in mfg 2000	50.3	0.4	10.5	20.4	118	75	69	87.3
share mfg 1990	65.1	0.0	8.7	24.6	89	94	79	87.3
response rate to census 2010 county	61.9	0.2	9.5	23.9	95	92	76	87.7
employed partime	71.5	0.3	0.8	24.2	72	78	113	87.7
method2 age adjusted disabiltiy rate age 18-64 zipcode	27.2	0.5	24.2	17.3	164	69	32	88.3
state = NJ	76.8	0.3	0.6	25.9	56	85	125	88.7
production occupation	76.4	0.0	0.8	25.7	58	94	115	89.0
causal effect of growing up in county on income at age 30 if born at 75th pct	32.6	1.0	11.7	15.1	156	51	62	89.7
state = MA	88.9	0.0	0.4	29.8	28	94	147	89.7

change mfg share 1991 to 2011	56.9	0.0	8.8	21.9	102	94	78	91.3
causal effect of growing up in county on college enrollment by age 23 if born 25th pct	33.2	0.9	10.8	15.0	155	55	65	91.7
age^2	15.7	0.0	51.3	22.3	174	94	9	92.3
state = IL	82.8	0.0	0.4	27.8	43	94	140	92.3
growth mfg employment 1999-2011	57.5	0.0	8.4	22.0	100	94	85	93.0
Time6	72.9	0.0	0.6	24.5	66	94	121	93.7
causal effect of growing up in county on income at age 30 if born at 25th percentile	38.2	0.7	10.0	16.3	148	62	72	94.0
share in mfg 2015 CZ	39.8	0.4	10.6	16.9	145	73	67	95.0
change trade exposure to china 1990 to 2007	37.9	0.8	8.7	15.8	149	57	80	95.3
state = UT	65.8	0.5	0.5	22.3	87	67	132	95.3
state = OK	80.0	0.0	0.4	26.8	49	94	145	96.0
state = CA	62.3	0.0	1.4	21.2	94	94	106	98.0
Time8	90.6	0.0	0.2	30.3	24	94	176	98.0
causal effect of growing up in county on college enrollment by age 23 if born 75th pct	39.9	0.0	12.6	17.5	144	94	57	98.3
CZ share in mfg 2014	42.3	0.3	10.0	17.5	141	83	71	98.3
farmer occupation	68.4	0.0	0.6	23.0	79	94	122	98.3
state = MN	83.1	0.0	0.3	27.8	42	94	159	98.3
state = MO	79.2	0.0	0.4	26.5	51	94	151	98.7
hhincome gini 2014 5-year	46.8	0.0	9.9	18.9	130	94	74	99.3
growth in jobs in goods industries 2000 to 2014 CZ	49.3	0.0	8.4	19.2	120	94	84	99.3
state = GA	68.4	0.0	0.5	23.0	78	94	126	99.3
state = AZ	68.6	0.0	0.5	23.0	77	94	128	99.7
state = NC	74.2	0.0	0.4	24.9	65	94	142	100.3
state = NY	64.4	0.0	0.7	21.7	91	94	118	101.0
growth in mfg jobs 2000 to 2015	30.9	0.7	8.6	13.4	160	63	81	101.3
upward mobility at age 30, CZ	26.9	0.3	12.5	13.3	166	80	59	101.7
growth in mfg jobs 1991 to 2011	29.4	0.7	8.6	12.9	163	60	83	102.0
not in labor force	31.9	0.9	3.4	12.1	157	54	96	102.3
region is midwest	53.3	0.0	1.6	18.3	110	94	103	102.3
growth total employment 1999-2011	31.1	0.7	8.3	13.4	159	64	86	103.0
state = CO	70.3	0.0	0.4	23.6	74	94	141	103.0
owner occupation	66.0	0.0	0.5	22.2	85	94	131	103.3
state = MS	71.8	0.0	0.4	24.0	71	94	146	103.7
state = VA	74.6	0.0	0.3	25.0	64	94	155	104.3
Share pop white CZ	3.8	0.0	16.4	6.7	177	94	44	105.0
causal effect of growing up in county on income at age 30 if born at 25th percentile	29.9	0.3	10.6	13.6	161	87	68	105.3

distance to Mexico	9.3	0.3	11.3	7.0	176	82	63	107.0
state = OH	60.8	0.0	0.5	20.4	97	94	130	107.0
share in mfg 2011 CZ	25.0	0.0	12.0	12.3	168	94	60	107.3
was Married	45.3	0.0	3.4	16.2	136	94	95	108.3
Time2	71.8	0.3	0.3	24.1	70	88	167	108.3
change share mfg 1990 2007	16.1	0.7	7.7	8.2	173	65	89	109.0
Time3	75.4	0.0	0.2	25.2	61	94	174	109.7
ln CZ distance from Mexican border	17.8	0.0	11.3	9.7	172	94	64	110.0
Time14	52.6	0.0	0.6	17.7	113	94	124	110.3
state = WV	69.8	0.0	0.3	23.3	75	94	164	111.0
share in mfg 1999 CZ	15.2	0.2	10.7	8.7	175	93	66	111.3
region is west	35.5	0.0	4.2	13.2	150	94	93	112.3
x share mfg 2000	31.5	0.0	8.0	13.2	158	94	88	113.3
state = MI	48.8	0.0	0.5	16.4	125	94	129	116.0
state = TX	38.9	0.0	0.9	13.3	147	94	110	117.0
state = PA	42.7	0.0	0.7	14.5	140	94	117	117.0
state = NM	53.6	0.0	0.4	18.0	108	94	150	117.3
state = AR	52.6	0.0	0.4	17.7	112	94	149	118.3
state = IA	52.8	0.0	0.3	17.7	111	94	153	119.3
in nonpublic sector union	65.3	0.0	0.2	21.8	88	94	177	119.7
state = TN	46.7	0.0	0.5	15.7	131	94	137	120.7
Time15	64.9	0.0	0.2	21.7	90	94	178	120.7
state = NH	56.9	0.0	0.3	19.1	103	94	168	121.7
state = IN	44.1	0.0	0.5	14.9	138	94	134	122.0
state = ID	51.5	0.0	0.3	17.3	115	94	157	122.0
state = SC	48.3	0.0	0.4	16.2	127	94	148	123.0
state = MD	46.4	0.0	0.4	15.6	132	94	144	123.3
region is north	23.7	0.0	1.4	8.3	170	94	107	123.7
Time7	39.9	0.0	0.5	13.4	143	94	136	124.3
state = KS	44.7	0.0	0.3	15.0	137	94	152	127.7
state = NV	46.0	0.0	0.3	15.4	134	94	156	128.0
Time12	34.3	0.0	0.4	11.6	153	94	138	128.3
clerical occupation	50.7	0.0	0.2	17.0	117	94	175	128.7
management occupation	51.4	0.0	0.0	17.1	116	94	180	130.0
state = ME	45.4	0.0	0.3	15.2	135	94	166	131.7
state = CT	40.4	0.0	0.3	13.6	142	94	160	132.0

state = KY	34.7	0.0	0.3	11.7	152	94	154	133.3
Time13	47.0	0.0	0.0	15.7	129	94	179	134.0
state = NE	35.0	0.0	0.3	11.8	151	94	158	134.3
state = DE	43.9	0.0	0.3	14.7	139	94	172	135.0
state = WA	33.6	0.0	0.3	11.3	154	94	173	140.3
state = MT	24.9	0.0	0.3	8.4	169	94	161	141.3
state = WI	29.7	0.0	0.3	10.0	162	94	169	141.7
state = RI	27.1	0.0	0.3	9.1	165	94	171	143.3
state = SD	25.6	0.0	0.3	8.6	167	94	170	143.7
state = DC	0.0	0.0	0.3	0.1	178	94	162	144.7
state = ND	0.0	0.0	0.3	0.1	178	94	163	145.0
state = VT	0.0	0.0	0.3	0.1	178	94	165	145.7