THE LIVING WAGE: TRENDS AND OBSERVATIONS

JAMES QUACINELLA

Capstone Project in Partial Fulfillment of the Requirements for the Masters of Science Degree in Data Analytics (MS)

Submitted to the School of Professional Studies - City University of New York
June 2016

FOREWORD

This dissertation was done under the supervision of Professor Joshua Laurito at the School of Professional Studies in the faculty of Information Systems.

ABSTRACT

There has been much discussion in recent years regarding the minimum wage, and one recurring argument in favor of raising the minimum wage has been that it has not kept up with the overall cost of living. To bring some data to bear on these questions, Amy Glasmeier proposed a model for the "living wage" [1], to describe how much one needs to earn in wages in order to meet basic needs. The model consists of 8 variables, and uses data from various sources to estimate these variables. These data sources take into account geographical differences in these costs, and derives an estimate of the living wage on a per-county basis. The original model produced data for 2014. The purpose of this project is to take this model and extend it's use to investigate trends in the living wage for the years 2004 - 2014. This project will look into what variables are most dominant, how the living wage distribution looks across the country, and how race and population affects the living wage distribution. Since the living wage is a measure of how much one needs to earn to meet basic needs, this project will also look at how many people are earning the living wage or less, as a proxy measure of economic precarity. The original model used 12 different family configurations; however, to keep the analysis simple, only households made up of a single person are modelled. Future work would expand this analysis to family configurations including children.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 INTRODUCTION TO THE LIVING WAGE AND THE LIVING WAGE MODEL

The living wage is defined by Schultheis, Glasmeier & Nadeu [2014][1] as:

> The living wage model is an alternative measure of basic needs. It is a market-based approach that draws upon geographically specific expenditure data related to a family's likely minimum food, child care, health insurance, housing, transportation, and other basic necessities (e.g. clothing, personal care items, etc.) costs. The living wage draws on these cost elements and the rough effects of income and payroll taxes to determine the minimum employment earnings necessary to meet a family's basic needs while also maintaining self-sufficiency.

The original model proposed estimated the living wage in terms of 8 variables:

$$basic\_needs\_budget = food\_cost + child\_care\_cost + (insurance\_premiums + health\_care\_costs) + housing\_cost + transportation\_cost + other\_necessities\_cost$$

$$living\_wage = basic\_needs\_budget + (basic\_needs\_budget * tax\_rate)$$

The model, in summary, calculates the summation of common costs associated with basic living, and defines how much one needs to earn in wages to cover these costs (accounting for taxes). These variables have vary levels of coarseness: most variables are modelled at a regional level, with housing costs being the only variable modelled at the county level. This is a weakness of the model, and future work should focus on better per-county estimates of these variables.

We can also define a notion of the living wage gap, which is the difference between a household's income and the living wage. This gap is in some sense a measurement of how well a household can live above just above, or how much more income a household needs to meet, mere subsistence. This project will look at the difference between median wages earned and the living wage estimate at the county-level, and as well as the minimum wage.

The purpose of this project is to take this model and extend it's use to investigate trends in the living wage for the years 2004 - 2014. The project is structured as follows. The data sources that each of the model variables use are described in the Section 2, and the individual model variables are described in Section 3. Section 4 begins the analysis of the living wage distribution across the county. Section 5 looks at how we can compare median and minimum wage levels with the living wage to look at how well single households are doing with regards to meeting basic needs. Commentary will be made throughout the sections and summarized in the Results section, with extra commentary on where future work could go.

All code for this project can be found in the associated Github repository[2]. An alternative to this paper is the associated IPython notebook[3], which is also hosted in the Github repository.

2 DATA SOURCES AND COLLECTION

This section will outline how data was gathered for the various model parameters, as well as other data we need to calculate their values. The original model was made for 2014 data and extending this data to the past means we need to be careful that any changes in the underlying data methodology of these parameters needs to be noted. All data files mentioned here are available in the github repository, under the data/ directory. Each data source is typically loaded into a Pandas DataFrame, which can be seen in the code sections linked to in the Appendix, or via the associated IPython notebook.

2.1 CONSUMER EXPENDITURE REPORT

The Consumer Expenditure Report[4] is used by the living wage model to determine 3 variables in the model, *transportation_cost*, *health_care_costs*, and *other_necessities_cost*. From their website:

> The Consumer Expenditure Survey (CE) program consists of two surveys, the Quarterly Interview Survey and the Diary Survey, that provide information on the buying habits of American consumers, including data on their expenditures, income, and consumer unit (families and single consumers) characteristics ... The CE is important because it is the only Federal survey to provide information on the complete range of consumers' expenditures and incomes, as well as the characteristics of those consumers.

After downloading, the specific data needed for the model variables were extracted by hand. All data files are stored under the cex_survey subdirectory.

2.2 USDA FOOD PLANS

The Cost of Food project from the USDA[5] produces different food plans (The Thrifty, Low-Cost, Moderate-Cost, and Liberal Food Plans), which represent a nutritious diet at different costs. This dataset determines one variable in the model, *food_cost*. Also, the original model uses regional weighting factors to better model varying food prices across the county. [6] After downloading, the specific data needed for the model variables were extracted by hand. All data files are stored under the food subdirectory.

2.3 FREE MARKET RENT DATA FROM HUD

The U.S. Department of Housing and Urban Development produces the Fair Market Rent dataset, which the model uses as 'gross rent estimates' for the *housing_cost* variable. [7] After downloading, the specific data needed for the model variable were extracted by hand, using the FMR1 column as the best estimate for the housing costs associated with a 1-bedroom apartment. All data files are stored under the housing_cost subdirectory.

2.4 MOST POPULATED COUNTIES

An article from Business Insider lists the top 150 counties by population. [8] This project uses this to determine if there are systemic differences between the living wage with respect to county population.

2.5 MEDICAL EXPENDITURE PANEL SURVEY FROM THE AHRQ

The original model uses data from the Medical Expenditure Panel Survey (MEPS), which is done by the The Agency for Healthcare Research and Quality. From their website [9]:

> The Medical Expenditure Panel Survey, which began in 1996, is a set of large-scale surveys of families and individuals, their medical providers (doctors, hospitals, pharmacies, etc.), and employers across the United States.

This data is used for *insurance_premiums* model variable. All data files are stored under the insurance subdirectory.

2.6 TAX DATA

The following data sources are used in calculating the *tax_rate* model variable.

2.6.1 *Payroll Taxes*

Payroll tax data was manually downloaded from the Social Security and Medicare Tax Rates web page from the Social Security Administration website. [10]

2.6.2 *State Tax Data*

The Tax Foundation produces a spreadsheet of official State income tax rates. This spreadsheet is not in a useful format for analysis, so data was manually copied to formatted_state_taxes.csv file under the taxes subdirectory. [11]

2.6.3 *Federal Tax Data*

The Tax Policy Center produces a dataset called the "Historical Federal Income Tax Rates for a Family of Four". While this dataset is not quite what this model needs, since the model developed for this project only modeled single adult households, due to a lack of data for single households (which only goes back to 2011), this dataset is used. [12] Since all counties experience the same federal tax rate, inaccuracies here would not affect overall trends, but produce a worse approximation to the living wage consistently across counties.

2.7 RACE DATA FROM 2010 CENSUS

Using the Census' "American Fact Finder" web page, data on racial breakdowns per county were downloaded from the "DP-1: Profile of General Population and Housing Characteristics: 2010" dataset. [13] Code for loading this data can be found in the IPython notebook. [14]

2.8 MEDIAN WAGE DATA

For the median wage, there are two data sets from the Census.gov that are applicable. The American Community Survey provides this data, however this project will use the SAIPE data for estimates of county median income:

> Small Area Income and Poverty Estimates: The SAIPE program produces single-year estimates of median household income and poverty for states and all counties, as well as population and poverty estimates for school districts. Since SAIPE estimates combine ACS data with administrative and other data, SAIPE estimates generally have lower variance than ACS estimates but are released later because they incorporate ACS data in the models. For counties and school districts, particularly those with populations below 65,000, the SAIPE program provides the most accurate sub-national estimates of poverty. For counties, SAIPE generally provides the best single year estimates of median household income.

2.9 MINIMUM WAGE

For the minimum wage, The United States Department of Labor keeps track of state level minimum wages. [15] The data is loaded via the IPython notebook. [16] This data was created by hand, and is located under the census/MinimumWage subdirectory in the github repository.

2.10 WAGE DISTRIBUTION

It will also be instructive to see any data regarding the wage distribution per county, to get an estimate of how many people are earning a wage at or below the living wage. Data was loaded from American Fact Finder on the Census website [17] and is stored in the wage_distribution subdirectory on github.

3 MODEL VARIABLES

This section will describe each model variable, and any interesting notes about the data. All code for loading these variables can be found in the associated notebook. [18]

3.1 HOUSING COSTS

Definition from the original model:

> We assumed that a one adult family would rent a single occupancy unit (zero bedrooms) for an individual adult household, that a two adult family would rent a one bedroom apartment.

Each county is identified by the FIPS code, which is just state code + county code + subcounty code (where subcounty code is only post 2005).

3.2 FOOD COSTS

Data for the food calculations are available PDF form (see section above about data collection). From the original model documentation:

> Adult food consumption costs are estimated by averaging the low - cost plan food costs for males and females between 19 and 50

After copying the data by hand, food costs need a correction. We add 20% to the values from the data sheets, since the notes on all published PDFs state:

> The costs given are for individuals in 4-person families. For individuals in other size families, the following adjustments are suggested: 1-person - add 20 percent; ...

The notes for the model also state that regional weights are applied to give a better estimate for food costs across the nation. [6] The result of this section are values for 2014 that match exactly to the data given on the original model website, which lends confidence to the implementation of the methodology.

It should be noted that there was a change in methodology the USDA used. Starting in 2006, the USDA changed the age ranges for their healthy meal cost calculations. The differences in range are minimal and should not effect overall estimations or trend analysis.

3.3 CHILD CARE COST

Currently, we are only looking at households that contain a single adult. Therefore, we do not model the costs of raising a child. One reason why this was done is that the data source for

Child Care only goes back to 2006. Expanding on this work would find data for 2004 and 2005, and model the living wage for different family configurations.

3.4 HEALTH INSURANCE COSTS

The model uses data from the Medical Expenditure Panel Survey from the Agency for Healthcare Research and Quality (searchable here). Specifically, the model assumes a single adult's insurance costs are best estimated from Table X.C.1 Employee contribution distributions (in dollars) for private-sector employees enrolled in single coverage. This survey gives the mean cost for a single adult per state. Table X.C.1 was only added to the survey starting in 2006. There is an alternative table that appears in all years (Table II.C.2: Average total employee contribution (in dollars) per enrolled employee for single coverage at private-sector establishments), which is what is downloaded from the previous section.

One problem is that in 2007 this survey was not done. This was solved by linearly imputing data from 2006 and 2008, which seems reasonable if we can assume that costs tend to go up every year and not go down. This is true for the data that was looked at for this project.

Another problem is that some states do not appear in the earlier data due to funding issues (and not being able to get a statistically significant sample). I fix this by using the value in the data for 'states not specified' and fill in the missing states.

3.5 TRANSPORTATION, HEALTH CARE AND OTHER COSTS

The model variables for transportation, health care and other costs are all based on the Customer Expenditure Survey data. The original model defines transportation costs as sum of the costs of sub-variables (1) Cars and trucks (used), (2) gasoline and motor oil, (3) other vehicle expenses, and (4) public transportation fields under "Transportation" section in the report. The original model defines health care costs as sum of the costs of sub-variables (1) health insurance costs for employer sponsored plans, (3) medical services, (3) drugs, and (4) medical supplies under the "Health Care" section. Expenditures for other necessities are based on the sub-variables (1) Apparel and services, (2) Housekeeping supplies, (3) Personal care products and services, (4) Reading, and (5) Miscellaneous under the "Other" section.

For each sub-variable, we get the amount of money (in millions) and the percentage of that that single adults spend. After multiple those numbers (accounting for units) and dividing by the total number of single adults in the survey gives us a mean total cost per adult.

The original model takes into account regional drift by scaling based on each regions. Currently, this model does not take this to effect, as the original model is ambiguous on how to calculate it. This is a weakness in the current model, as regional weighting would help vary these variables across counties. Without it, these variables will not create any differences between counties in any given year.

3.6 TAXES

To more accurately reflect how much one needs to earn pre-taxes to earn the living wage post-taxes. From the documentation:

> Estimates for payroll taxes, state income tax, and federal income tax rates are included in the calculation of a living wage. Property taxes and sales taxes are already represented in the budget estimates through the cost of rent and other necessities.

3.6.1 *Payroll taxes*

The payroll tax data is simply the federal tax rate for a given year. From the model documentation:

> A flat payroll tax and state income tax rate is applied to the basic needs budget. Payroll tax is a nationally representative rate as specified in the Federal Insurance Contributions Act.

The original model used a value of 6.2%. However, the data from the SSA website states that 6.2% is the rate for just the Social Security part of the FICA tax. This might be a mistake in the original model. This project will use the combined rate (which includes Medicare's Hospital Insurance rate) when calculating final numbers for my model.

Another thing to note is that in 2011 and 2012, the rate for the Social Security part of the FICA tax was 2% lower for individuals.

3.6.2 *State Taxes*

The model also uses state tax rates in estimating the total tax rate. From the model documentation:

> The state tax rate is taken from the second lowest income tax rate for 2011 for the state as reported by the CCH State Tax Handbook (the lowest bracket was used if the second lowest bracket was for incomes of over 30,000) (we assume no deductions).

The second lowest tax bracket's rate is chosen as the rate for the state (except when the bracket is for incomes > 30k, as the original model suggests). This only came into play in the later years for Vermont, North Dakota, and RI. To be consistent, the lowest tax bracket is used for all years for these states.

Note that this project used the rate under "Single" since the model is only for adults. This is done by hand by importing correct numbers from the spreadsheet.

3.6.3 *Federal Income Tax Rate*

The model also uses state tax rates in estimating the total tax rate. From the model documentation:

> The federal income tax rate is calculated as a percentage of total income based on the average tax paid by median-income four-person families as reported by the Tax Policy Center of the Urban Institute and Brookings Institution for 2013.

It should be noted that the model authors used "Historical Federal Income Tax Rates for a Family of Four". Since I am focusing on single adults, this model should use "Historical Average Federal Tax Rates for Nonelderly Childless Households". However, that data stops at 2011 for some reason, so for consistency, this model will stick with the model definition and use the Family of Four rate.

3.7 FINAL DATAFRAME

The final data frame, that includes each individual model variable as well as the total living wage $'total\_cost'$, with a row for each county per year, is created at the end of the code section cited in this section. This DataFrame is used by the following sections of analysis.

4

INTRODUCTORY ANALYSIS

This section will start to look at the living wage data to look for trends. First, we'll look into a few individual counties, and then look at state and regional averages. Secondly, we'll develop a set of maps of counties and their associated living wage. Finally, we'll look at the living wage distribution by population and by race.

4.1 LIVING WAGE IN INDIVIDUAL COUNTIES

To start off, lets take a look at the data generated for the living wage estimates for counties. This will serve as a spot check to make sure that the data generated at least looks sensible. Figure 1 shows the living wage trend for 3 counties (Kings, Mercer and Orange County). This is not terribly interesting, but does show that the living wage estimates are within a reasonable range, and a manual inspection of the data does not show any NAs or suspicious values. It is interesting to note that both Kings and Mercer County seem to have a linearly increasing living wage, with a levelling off / decrease towards 2012 - 2014. Orange County, however, seems flatter, with a sharp increase towards the end.

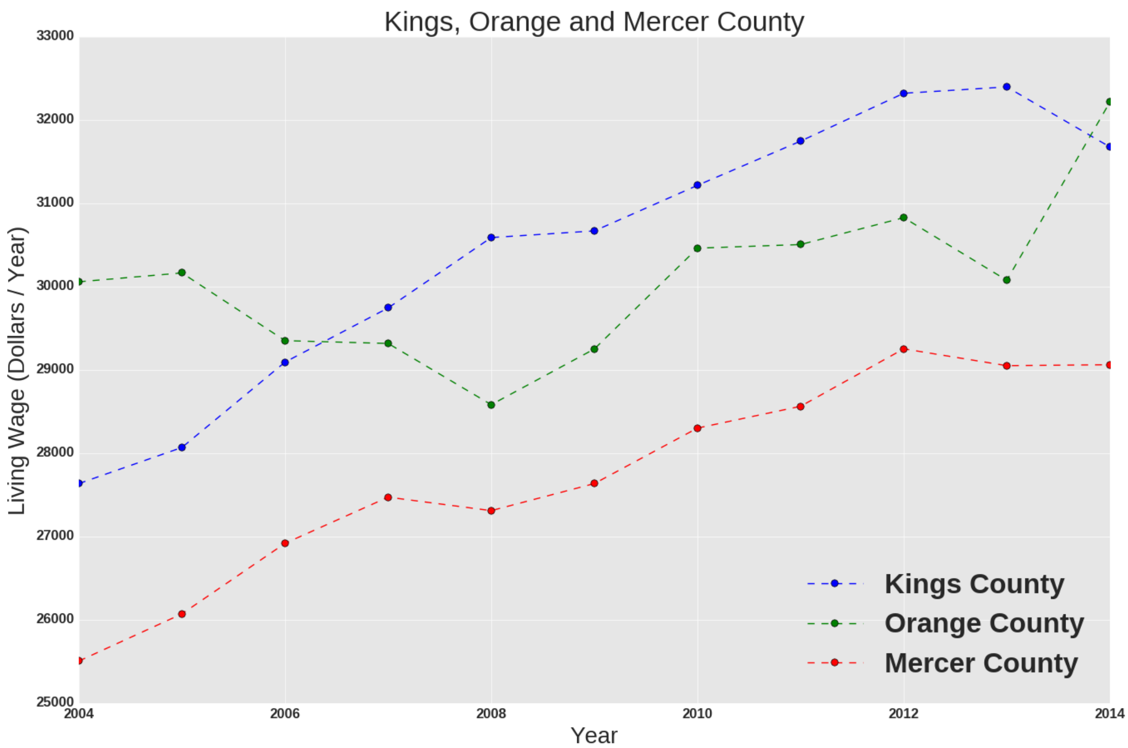


Figure 1: Living Wage Trend for Select Counties

4.2 STATE AVERAGES OF THE LIVING WAGE

Now, we can aggregate these living wages up to the state level and see if any trends emerge. We calculate the living wage averages for states, weighed by the county population. Figure 2 shows the state averages for 8 states, two from each region (indicated by color). One can see that there are stark differences between state averages.

TODO: Expand

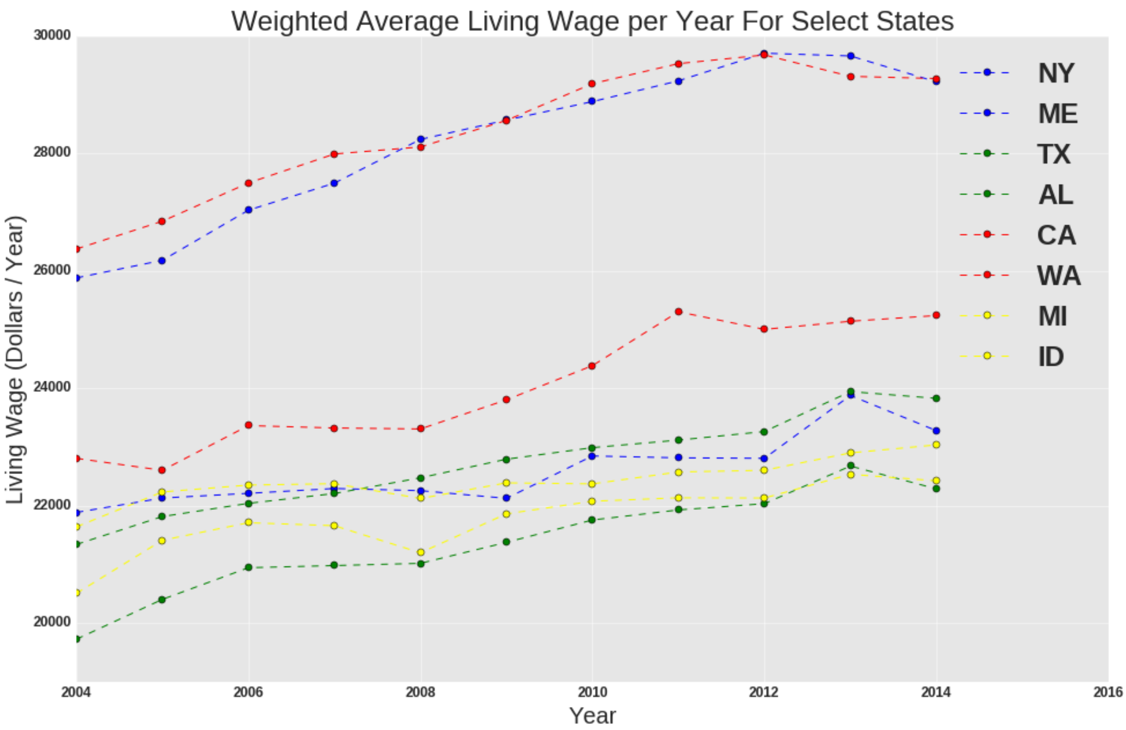


Figure 2: Living Wage Trend for Select States

4.3 CHOROPLETH OF COUNTIES: 2004, 2006, 2014

A Choropleth map of the United States is a good way to get an overall view of the living wage distribution for any given year. Using the calculated living wage estimates, choropleth maps are generated for sample years. Figure 3 shows the living wage county distribution for 2004; Figure 4 shows the same distribution in 2006; Figure 5 shows it for 2014. Once notices that large portions of the 'middle county' (Midwest and South regions) generally have lower living wages, while the East and West regions have higher living wages. Also, as would be expected, the maps get darker with time, indicating a general increase of the living wage over time.

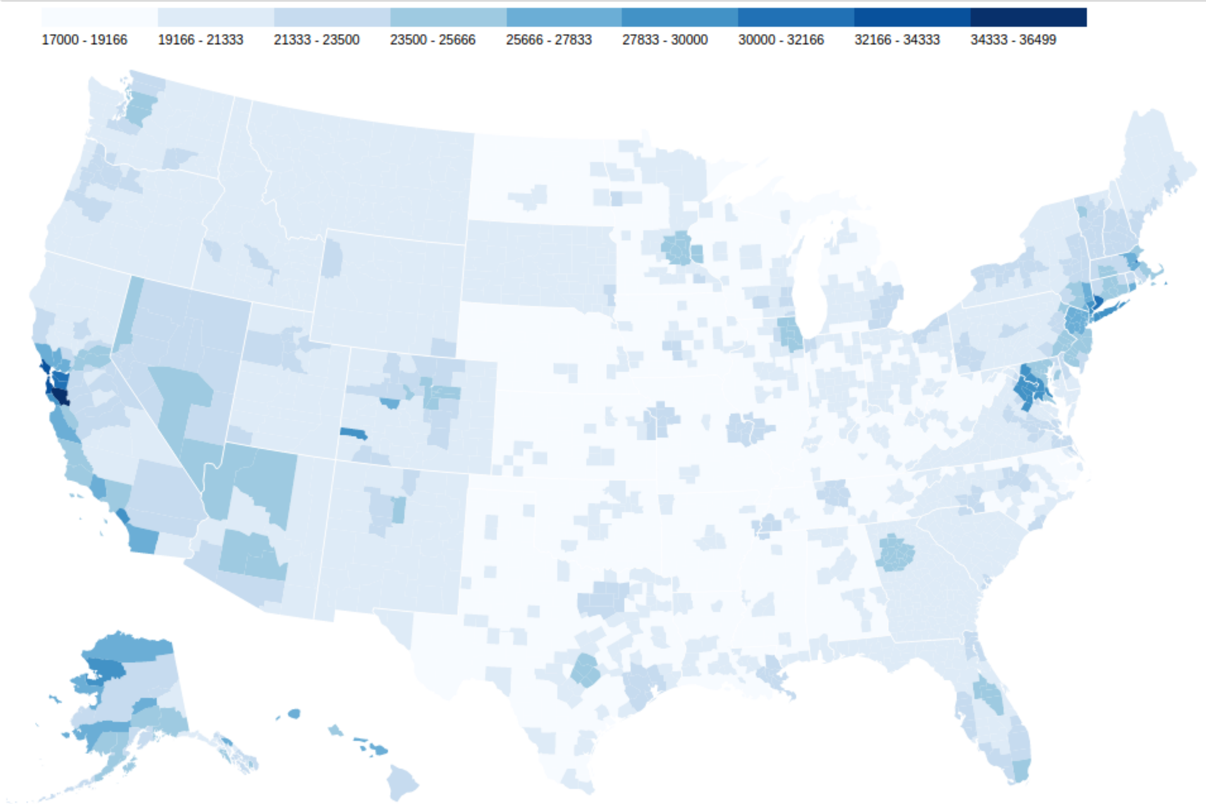


Figure 3: Choropleth of Counties based on Living Wage, 2004

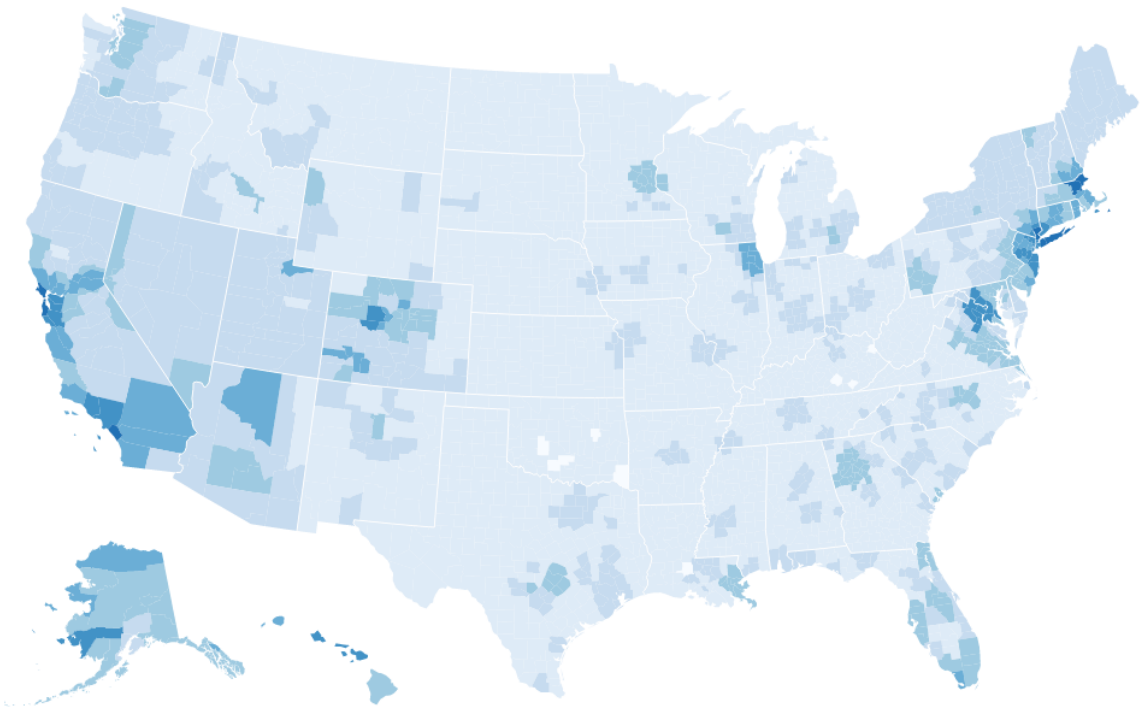


Figure 4: Choropleth of Counties based on Living Wage, 2006

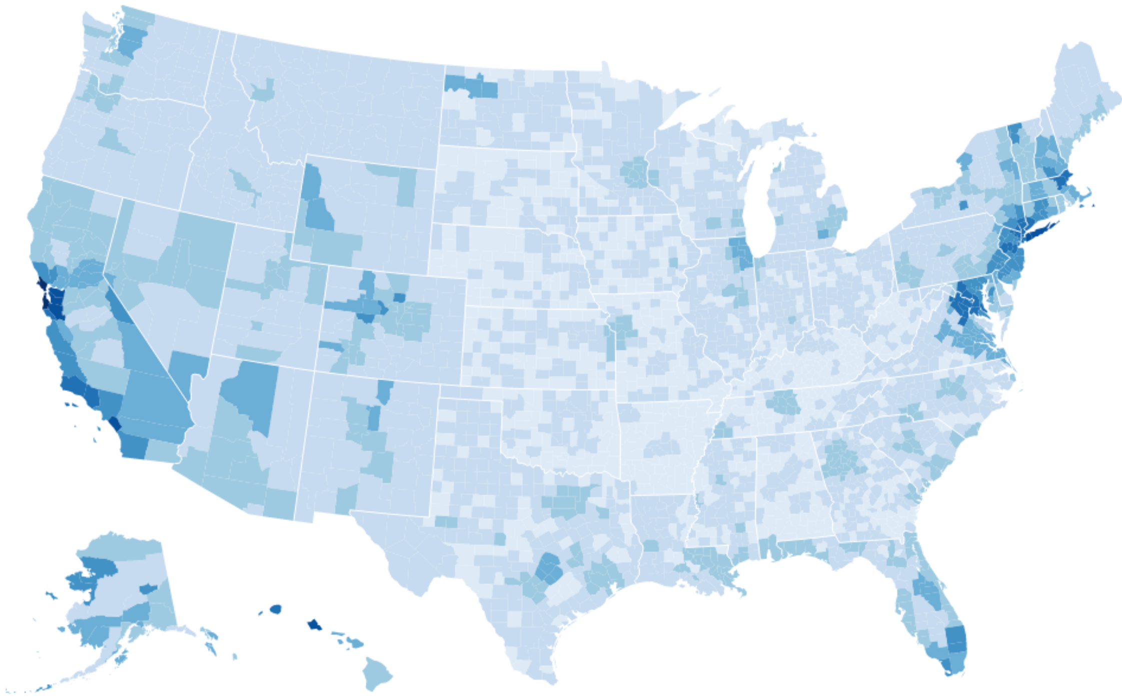


Figure 5: Choropleth of Counties based on Living Wage, 2014

4.4 REGIONAL AVERAGES OF THE LIVING WAGE

Going up to the next level of aggregation, we can also look at the trends of regional averages, using Census definitions of regions. Figure 6 shows these trends. Clearly the Eastern region is the worst off, consistently over the model time period. The South and Midwest start at similar values but a gap emerges that ends up staying consistent. As seen before, we see a general increase of the living wage over time for all regions.

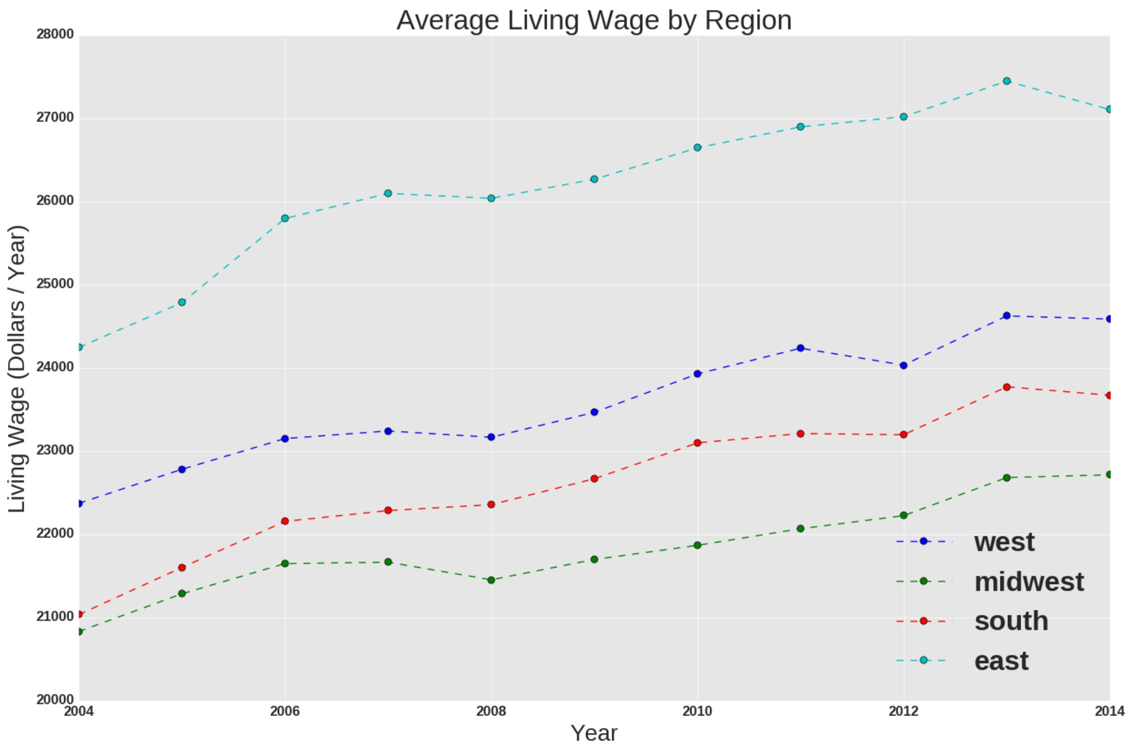


Figure 6: Regional Trend of the Living Wage

4.5 National Average Breakdown Across 2004 - 2014

Since the final value from the model is made up of 6 model variables, we can break down these averages to see which variable might be varying the most, or dominates the other variables due to its value. Figure 7 shows the trend in the national average living wage over the model years, broken down by model variable. This clearly shows that *housing_cost* is by far the most influential variable: not only does it dominate in value, but has the highest rate of change over time. The *housing_cost* value dominates due to the nature of the living wage (as rent is typically the largest amount one spends per year). The *housing_cost* variable also has the highest variance since its the only variable granular at the county level. Future versions of this model should try to find county-level estimates of other model variables to increase its accuracy.

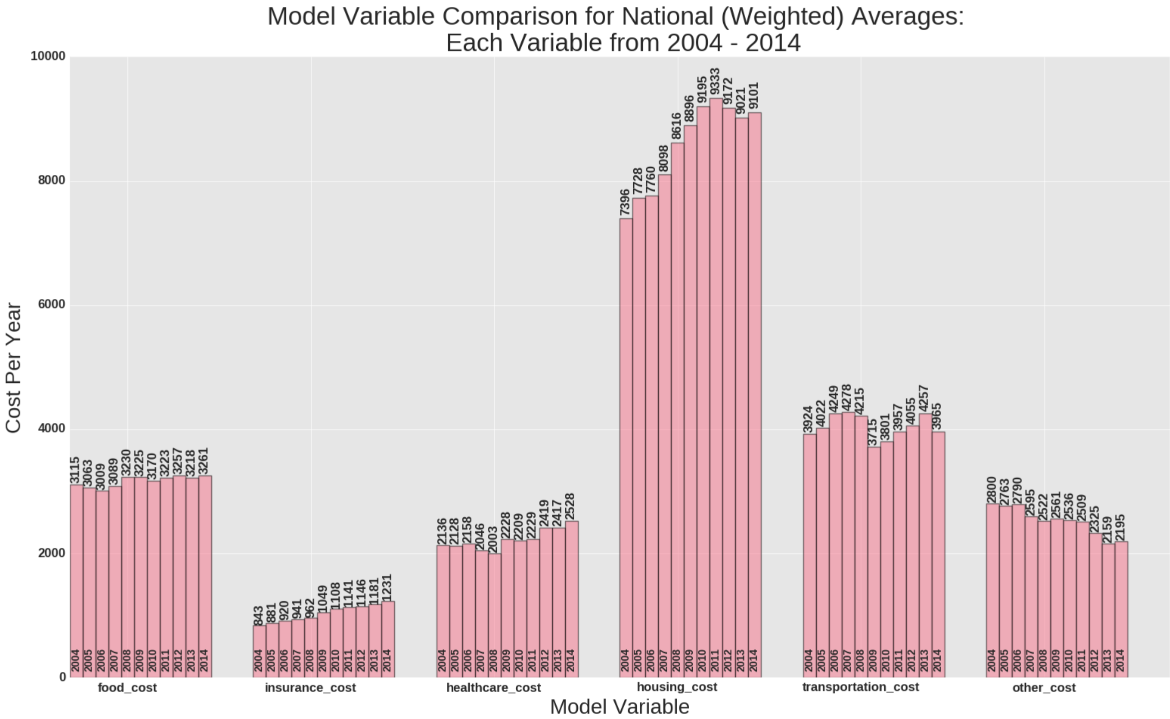


Figure 7: Breakdown of National Average of the Living Wage

4.6 living wage distribution in most populous counties

What is the relationship between a county's total population and the county's living wage estimate? An article from Business Insider [8] listed the top 150 counties by population. After importing this list of counties by hand, we can investigate the distribution of living wage estimates for counties in this group (hereafter called 'most populous') versus all the other counties (hereafter called 'non-populous'). Figure 10 shows the distribution of living wage values over time separated by this grouping. The 'most populous' county distributions (red in the figure) are spread out more (std: 3123.05) than the 'non-populous' ones (blue in the figure; std: 1791.76). The 'most populous' county distribution is also shifted to the right, indicating higher living wages for counties that are highly populated.

Also, it seems that the 'most populous' county distributions are almost bimodal, with peaks labelled in the figure. In light of this, the 'most populous' counties are separated into two subgroups in the subsequent visualizations and analysis, where 'moderately populous' are the counties in the 'most populous' group but with a total cost in the bottom portion of the group; 'highly populous', the top portion of the group. The split is done based on a given year's minimum value of the kernel density estimate between the peaks, highlighted in the figure

Figure 8 shows the trend in the distribution peaks, labelled by population group. What is interesting here is that the different between the peak in the least-populous county group and the two peaks in the most-populous group is very steady in value, despite a general increase over time. This seems to insinuate that living in a more populated area comes at the cost of an increase in the living wage needed, and that this cost is maintained over time. Also, from 2009 to 2012, the 'highly populous' group of counties had a 'jolt' in the equilibrium of the living wage estimate; there is no associated jolt in the other county groupings.

Table 1 shows variables in the model to account for the differences in the living wage across the least and most populated counties. Figure 9 shows this same breakdown, but as relative percentages. As you can see, rent is the dominate variable in explaining the differences between population group.

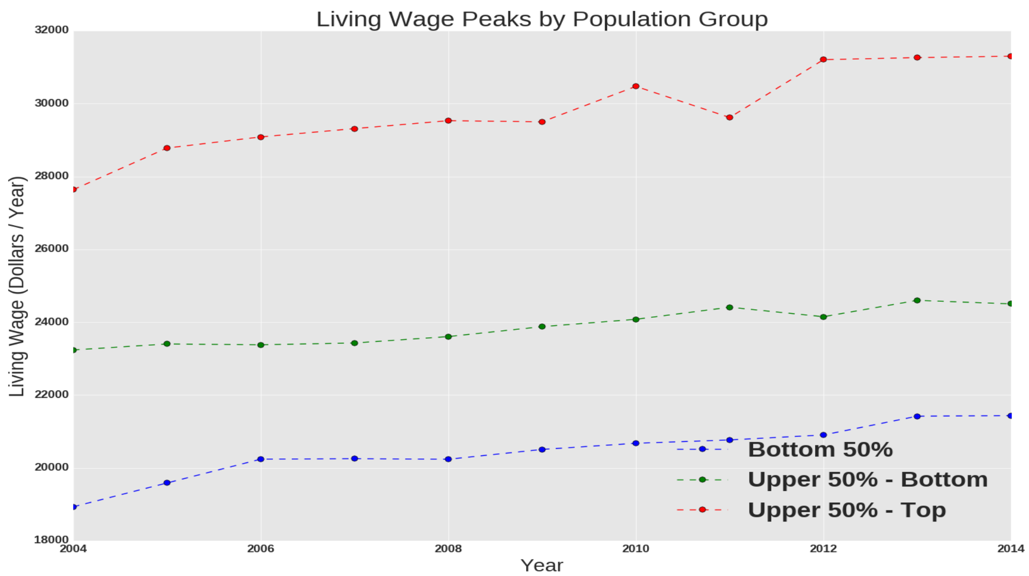


Figure 8: Living Wage Peaks by Grouped By Population

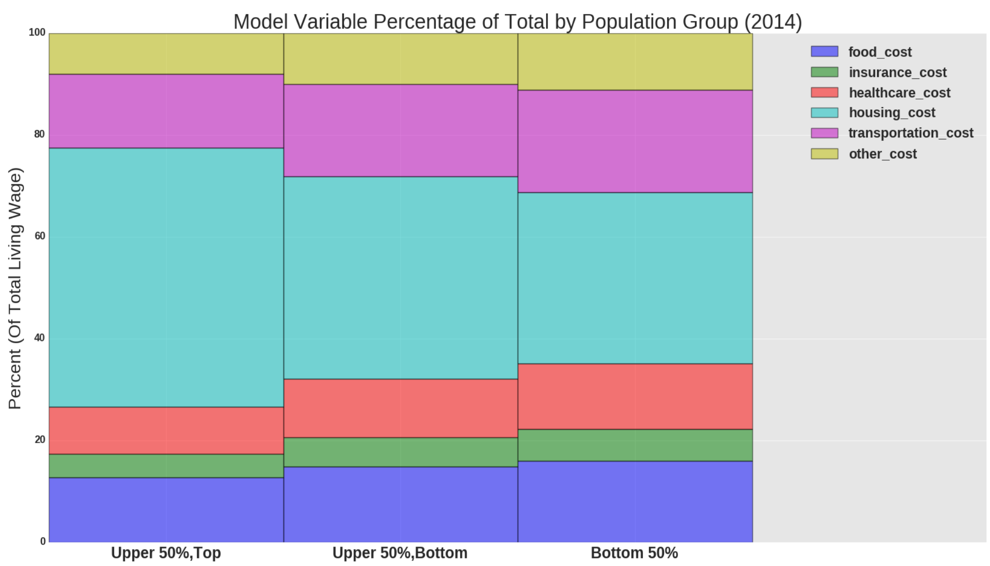


Figure 9: Model Variable Percentage of Total by Population Group (2014)

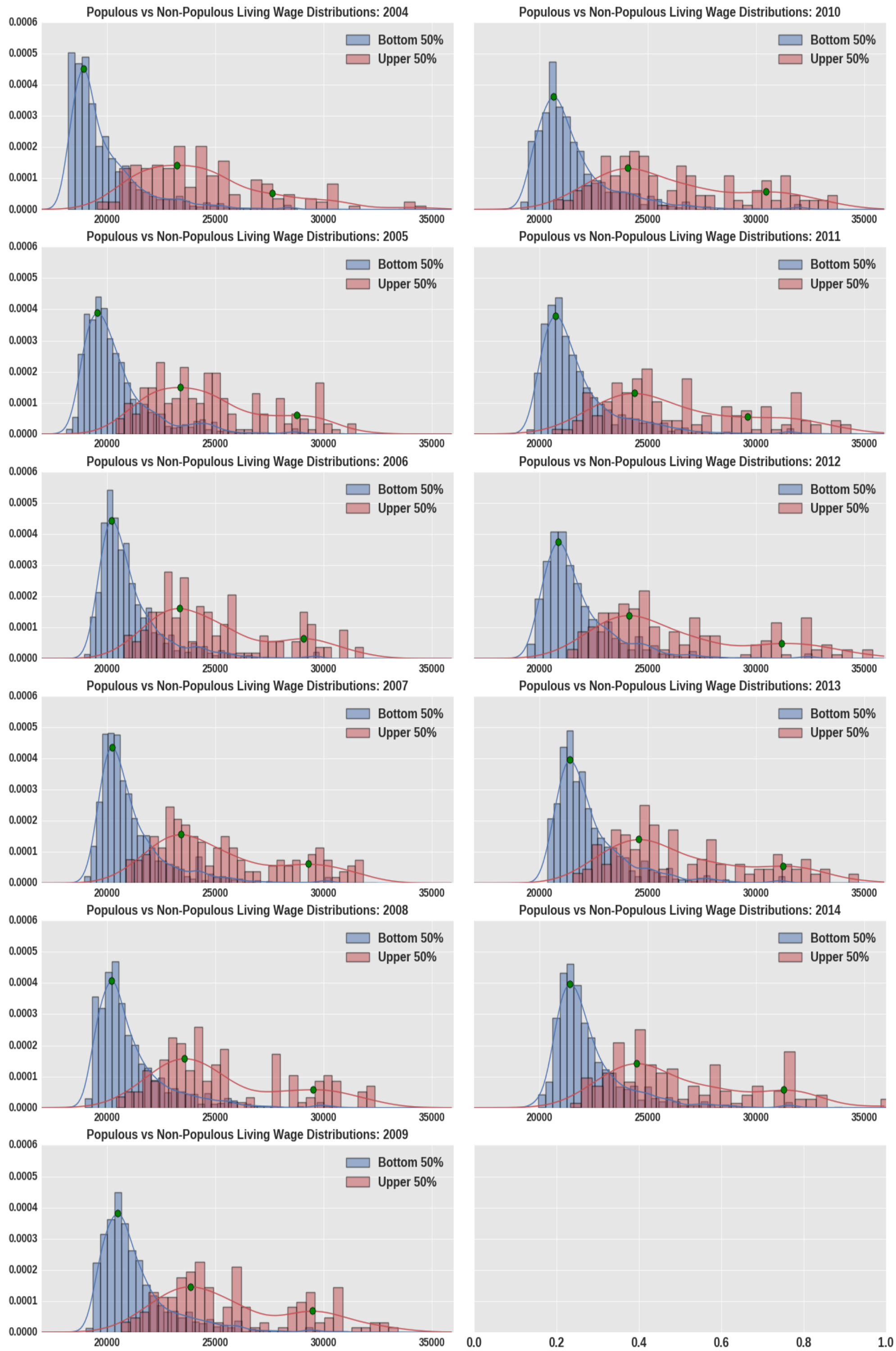


Figure 10: Living Wage Distribution of Top and Bottom

Variable	"Upper 50%,Top" - "Upper 50%,Bottom"	"Upper 50%,Top" - Least Pop	"Upper 50%,Bottom" - Least
food_cost	$197.52	$306.93	$109.41
insurance_cost	$48.47	$67.46	$18.99
healthcare_cost	$0.00	$0.00	$0.00
housing_cost	$5,190.87	$7,668.49	$2,477.62
transportation_cost	$0.00	$0.00	$0.00
other_cost	$0.00	$0.00	$0.00
total_cost	$6,119.75	$9,053.42	$2,933.67

Table 1: Average values across counties in different population groups for all model variables

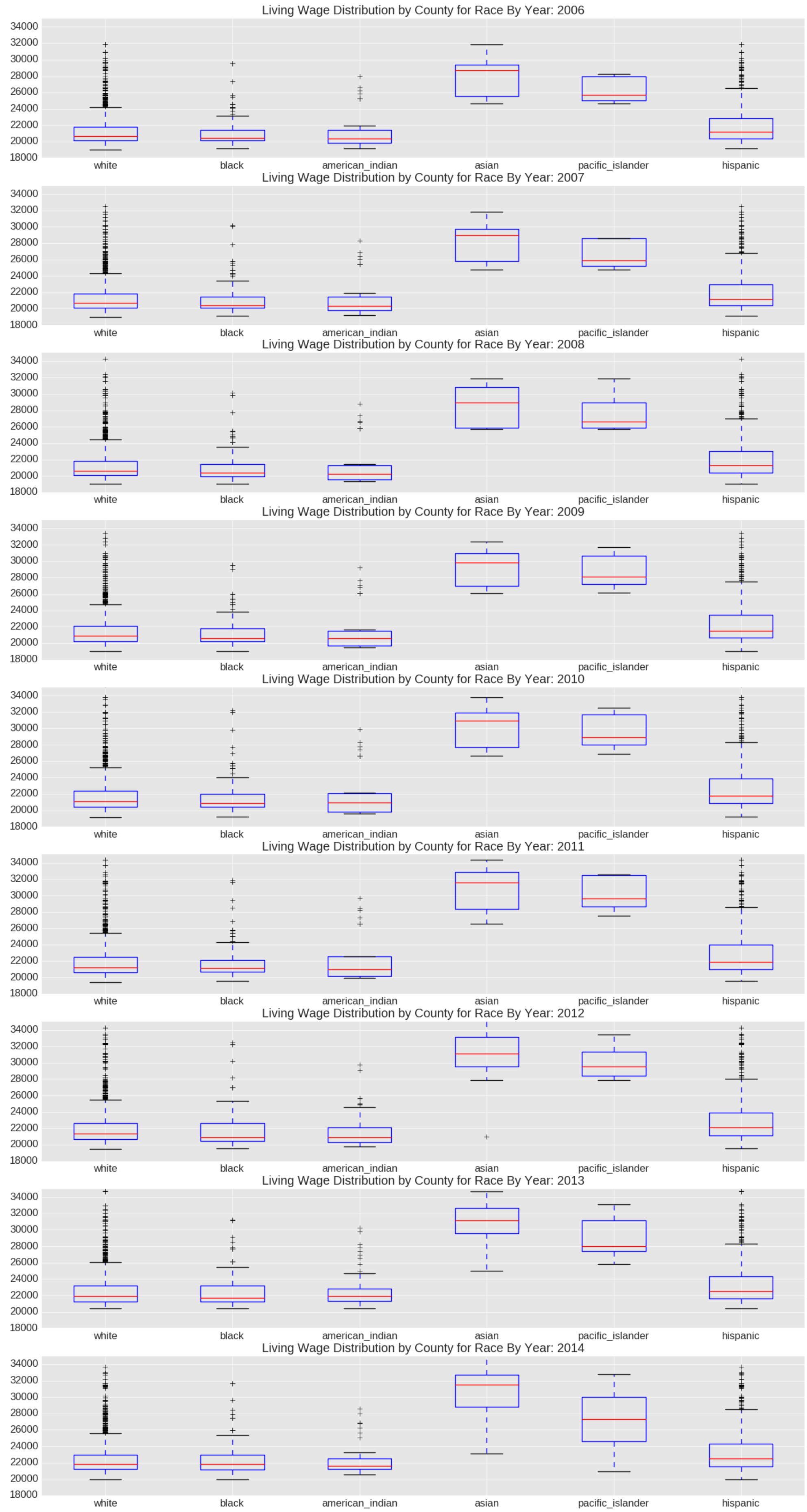
4.7 LIVING WAGE DISTRIBUTION BY RACE

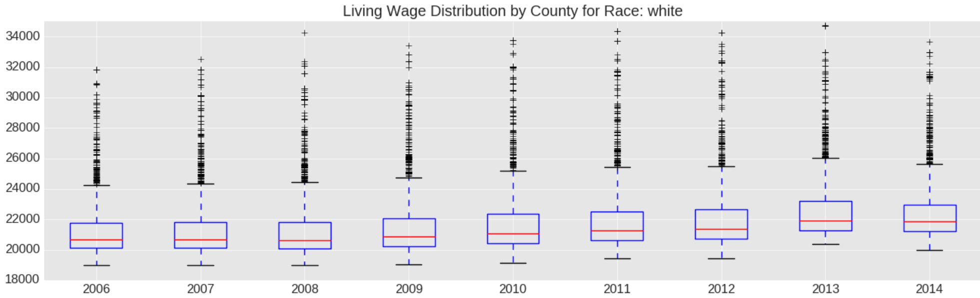
This section will investigate how the living wage breaks down across the country by race. A note about the methodology here: racial breakdowns in each county come from the 2010 Census data, which means this model is a bit inaccurate since these proportions change over time. For a more accurate measurements, more data would need to be gathered about how these racial proportions change over time.

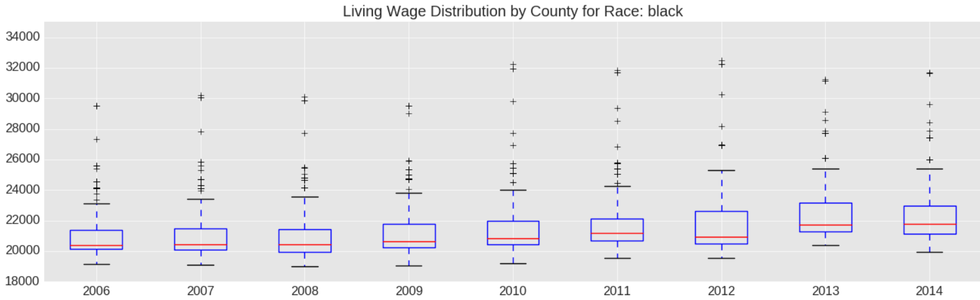
4.7.1 *Living Wage County (Non-Weighted) Breakdown by Race*

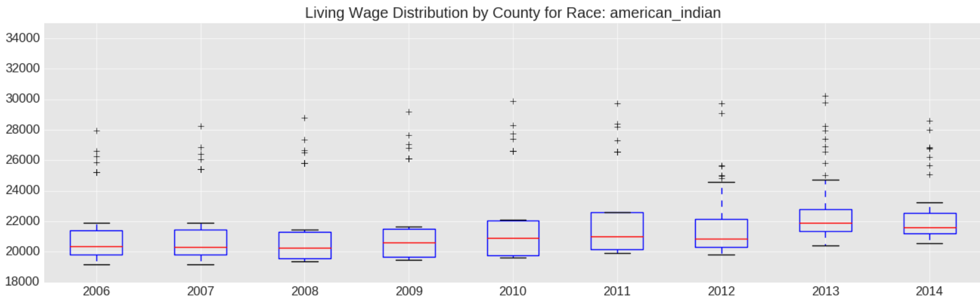
This section will look at the distribution of the living wage across time and race, with the living wage not weighed by population. A county is counted in the average for a given race if that give race had a population that meets a threshold. The living wage is then averages across the counties that 'count' for that race. This was initially done as a way of starting analysis despite the lack of support in matplotlib for generating weighted box plots or violin plots. The next subsection will explore a population weighted average

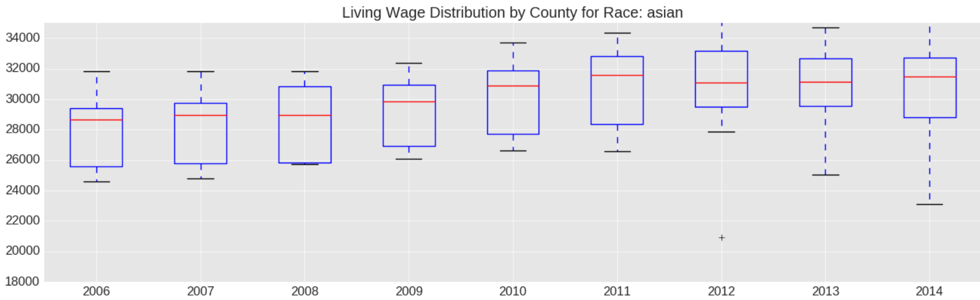
Figure 11 and Figure 12 shows boxplots of the distribution of the living wage, broken down by year and race. Both figures are produced to make comparing distributions across races or comparing across years for the same race easy. Figure 13 shows the same information as the previous figures, but as density plots overlaid on each other. It becomes a bit difficult to compare across races since they are overlaid, so the "Asian" and "Pacific Islander" aces are put into their own figure.

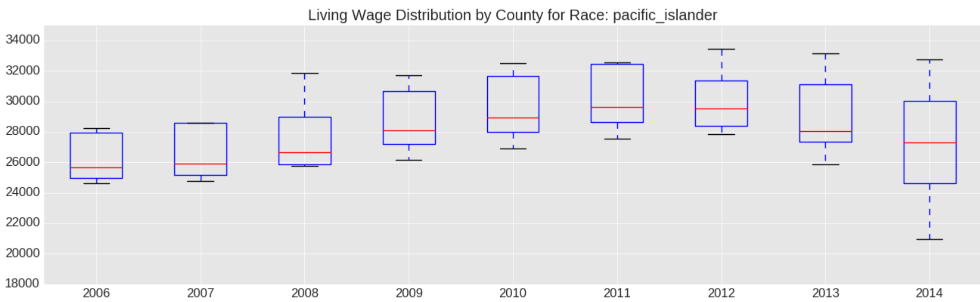

Living Wage Distribution by County for Race By Year: 2006
Living Wage Distribution by County for Race By Year: 2007
Living Wage Distribution by County for Race By Year: 2008
Living Wage Distribution by County for Race By Year: 2009
Living Wage Distribution by County for Race By Year: 2010
Living Wage Distribution by County for Race By Year: 2011
Living Wage Distribution by County for Race By Year: 2012
Living Wage Distribution by County for Race By Year: 2013
Living Wage Distribution by County for Race By Year: 2014
34000
32000
30000
28000
26000
24000
22000
20000
18000
white
black
american_indian
asian
pacific_islander
hispanic

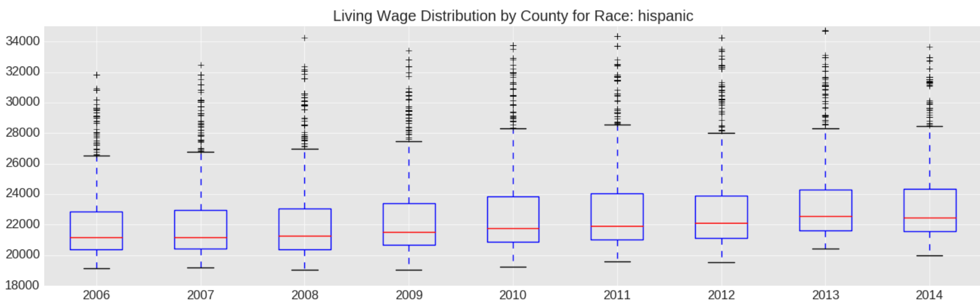

Living Wage Distribution by County for Race: white
34000
32000
30000
28000
26000
24000
22000
20000
18000
2006
2007
2008
2009
2010
2011
2012
2013
2014


Living Wage Distribution by County for Race: black
34000
32000
30000
28000
26000
24000
22000
20000
18000
2006
2007
2008
2009
2010
2011
2012
2013
2014


Living Wage Distribution by County for Race: american_indian
34000
32000
30000
28000
26000
24000
22000
20000
18000
2006
2007
2008
2009
2010
2011
2012
2013
2014


Living Wage Distribution by County for Race: asian
34000
32000
30000
28000
26000
24000
22000
20000
18000
2006
2007
2008
2009
2010
2011
2012
2013
2014


Living Wage Distribution by County for Race: pacific_islander
34000
32000
30000
28000
26000
24000
22000
20000
18000
2006
2007
2008
2009
2010
2011
2012
2013
2014


Living Wage Distribution by County for Race: hispanic
34000
32000
30000
28000
26000
24000
22000
20000
18000
2006
2007
2008
2009
2010
2011
2012
2013
2014

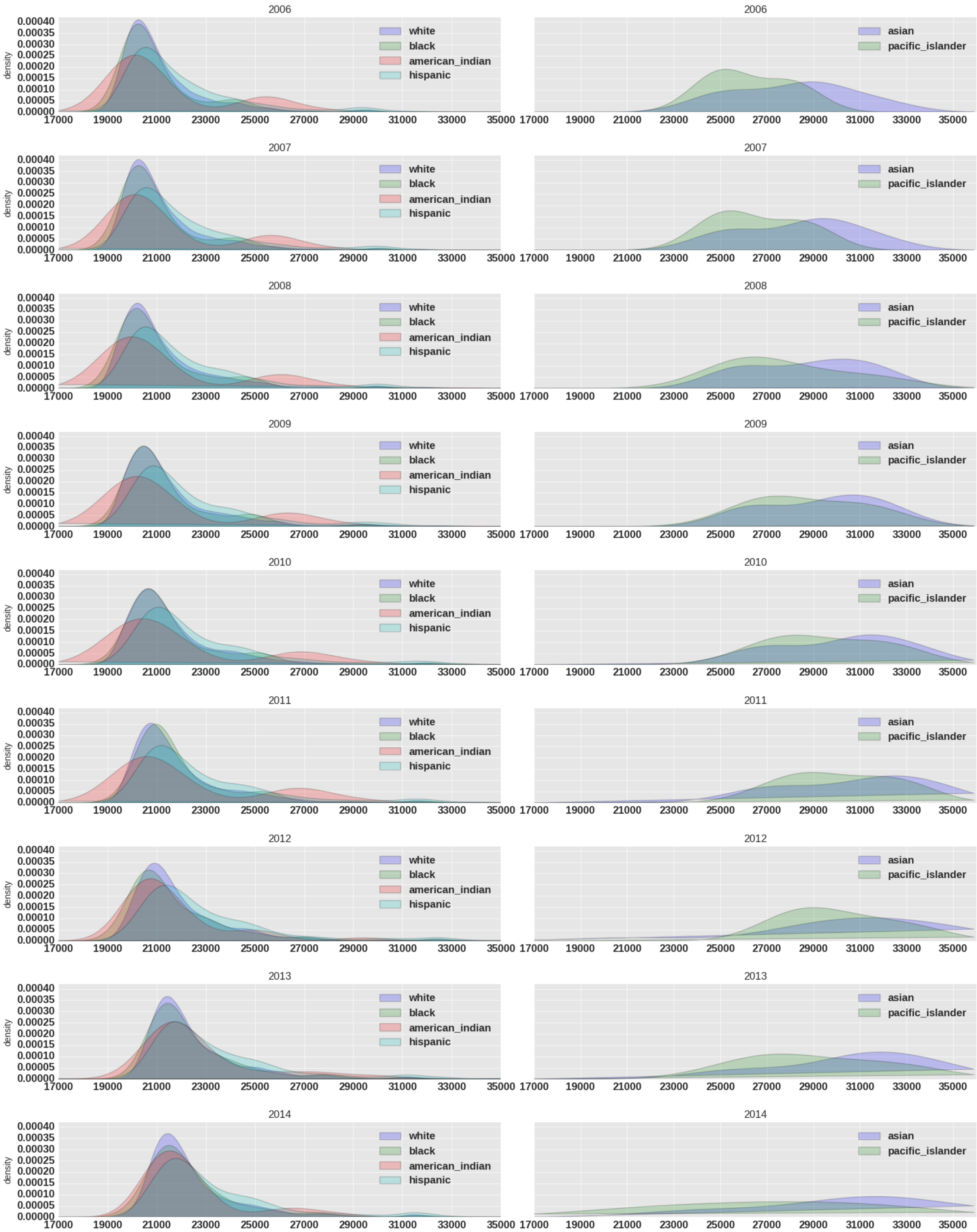


Figure 13: Living Wage Density Plots for Race By Year

4.7.2 *Living Wage County (Weighted) Breakdown by Race*

This section will look at the distribution of the living wage across time and race, with the living wage not weighed by population. Every county's population is broken down by race, and the living wage is then averages across the counties weighed by the population of that race. Matplotlib does not support generating weighted box plots or violin plots, so a custom implementation was created. This code can be found in the IPython notebook.

Figure 14 and Figure 15 shows violin plots of the distribution of the living wage, broken down by year and race. Both figures are produced to make comparing distributions across races or comparing across years for the same race easy. Figure 16 shows the same information as the previous figures, but as density plots overlaid on each other. It becomes a bit difficult to compare across races since they are overlaid, so the "Asian" and "Pacific Islander" aces are put into their own figure.

Some interesting trends and talking points emerge from these plots. As noted from earlier, we can see a general increase in the living wage across time. Another thing to notice is that all races seemed to have had a faster increase from 2004 - 2006, than across other years. This can be seen by the sharper shift to the right in Figure 15 for this years.

Whats more interesting is how different these distributions are across races. Asians, Hispanics and Pacific Islanders have distributions which a shifted much further up the living wage scale for any given year than other races. Where one lives is a huge factor in the living wage, since housing costs are the highest impact variable in the model, and housing costs vary a lot across counties. These populations, especially Asian and Pacific Islanders, are more like to be concentrated into certain areas. If these areas have high housing costs, like big cities, then as a result, their distributions would skew upwards. Whites and Blacks are more evenly distributed throughout the county, and therefore have a lower and wider distribution. Native Americans also have one of the lower living wages distributions, which result from population dynamics as well.

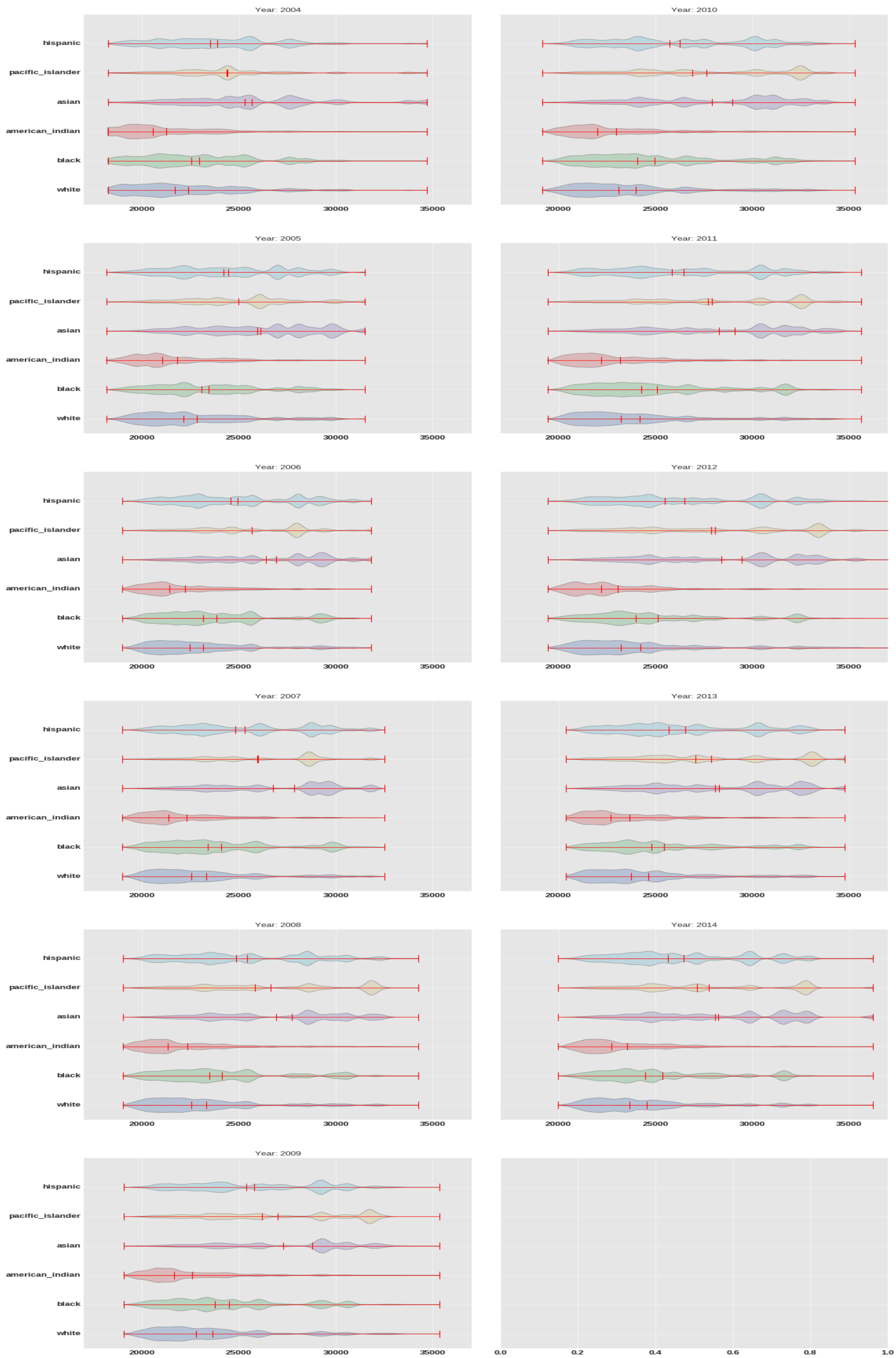


Figure 14: Violin Plots of Living Wage (Weighted by Racial Population of County)

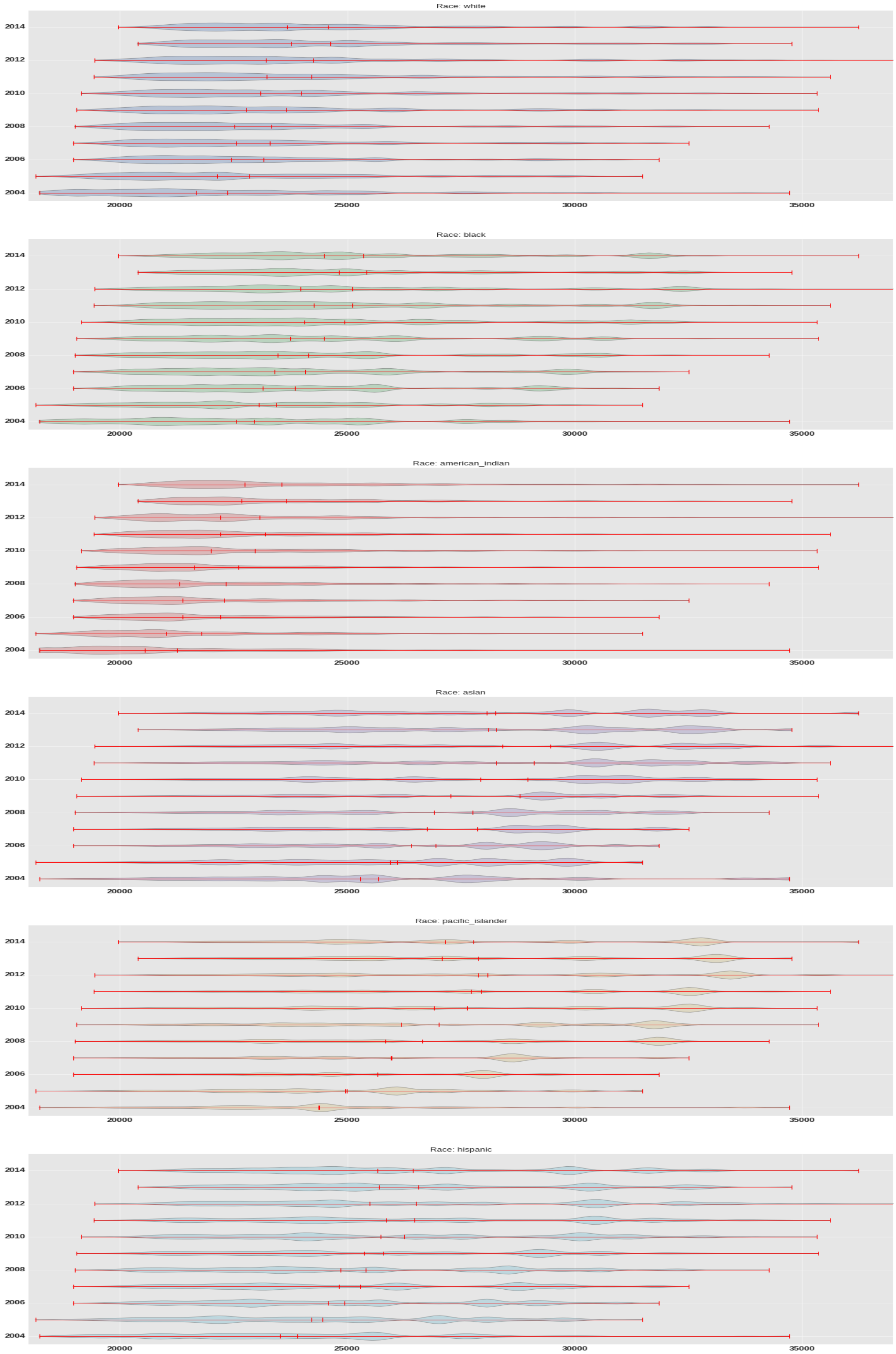


Figure 15: Violin Plots of Living Wage (Weighted by Racial Population of County)

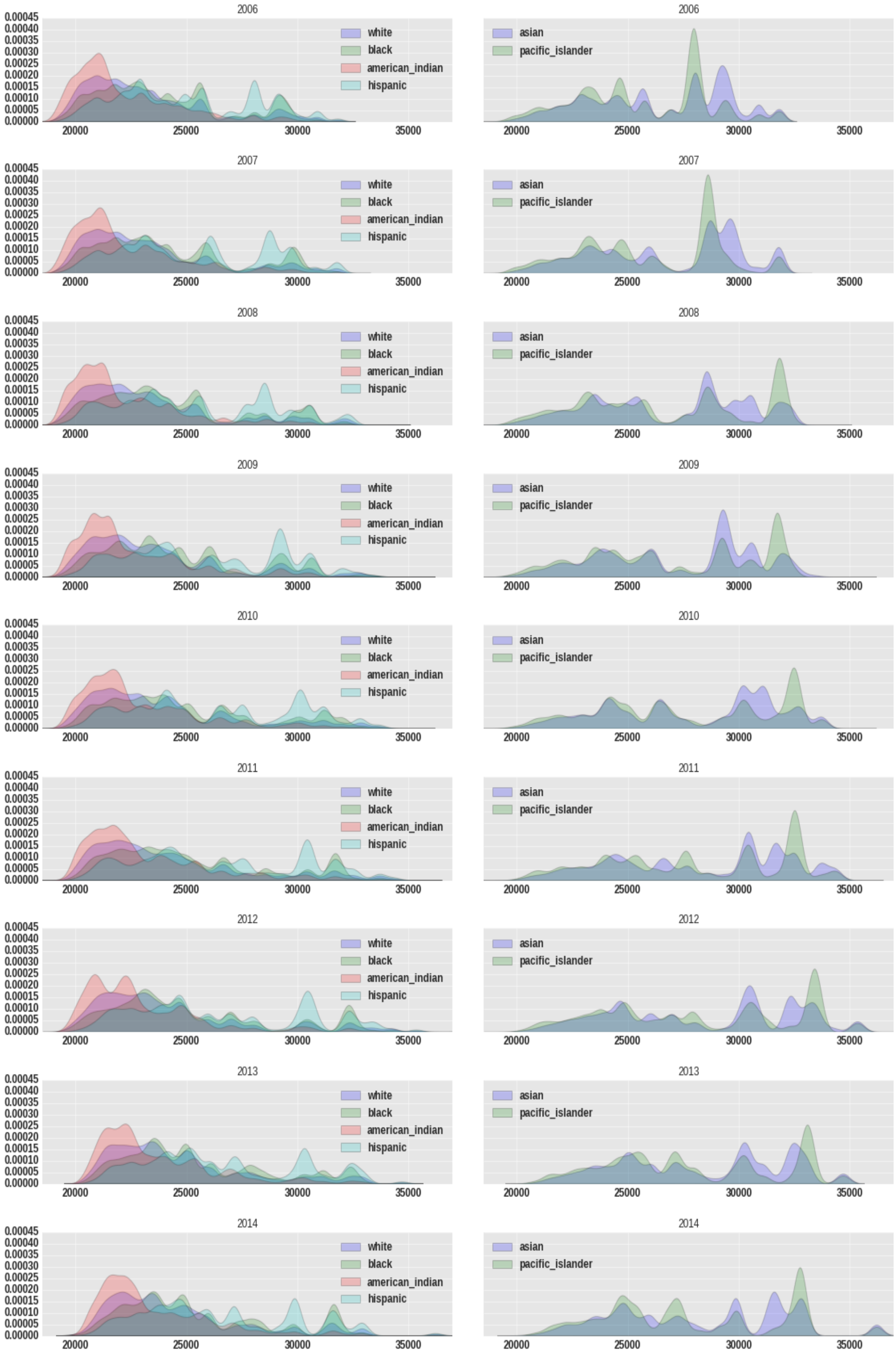


Figure 16: Density Plots of Living Wage (Weighted by Racial Population of County)

4.7.3 *Population Weighted Averages Broken Down By Race*

Figure 17 shows the living wage (weighted) average for each race across years. As can be seen, Asian, Hispanic and Pacific Islanders all have higher averages for any given year. Each race seems to be experiencing the same rate of growth in the living wage average, with Pacific Islanders having a slightly steeper slope than the other races.

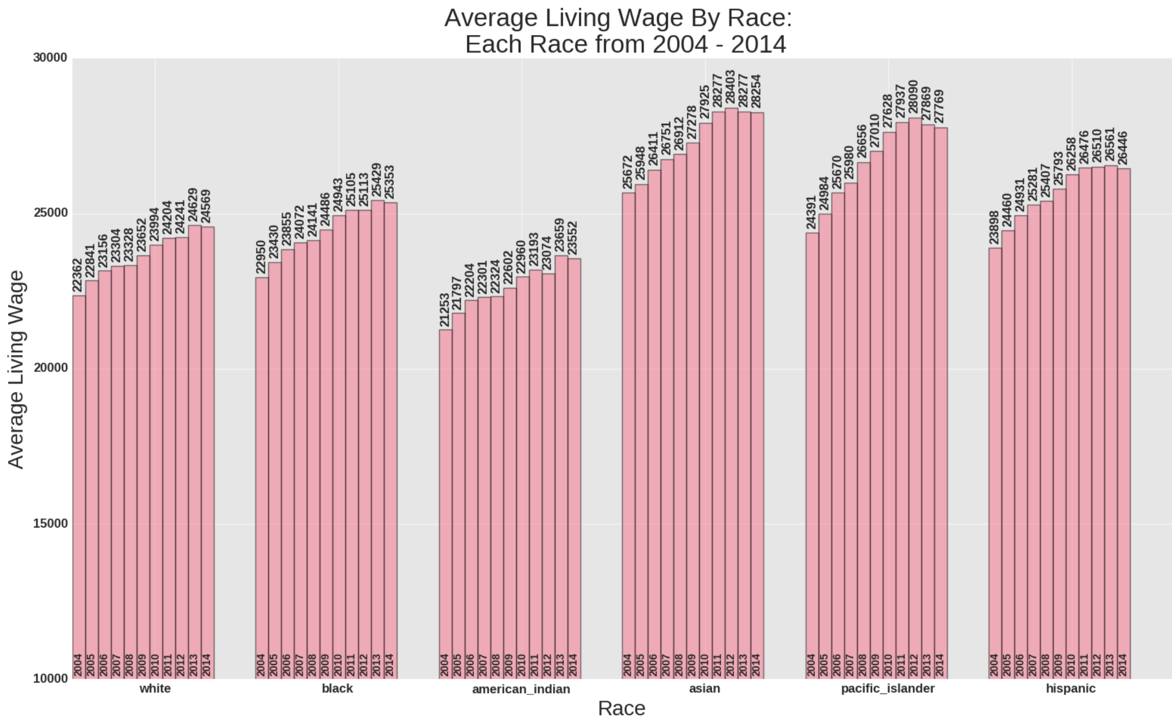


Figure 17: National Living Wage Average Breakdown by Model Variable

5 LIVING WAGE GAP

This section will start to look at the living wage gap. First, we will come up with two definitions for the living wage gap, one based on the median wage, and one defined by the minimum wage. After that, we will look at distributions of both definitions. Finally, we will look at the distribution of households who earn the living wage or below.

5.1 DISTRIBUTION OF THE MEDIAN-GAP

Figure 18 shows the living wage gap distribution across counties when the gap is between the living wage the county's median wage, as defined by the Small Area Income and Poverty Estimate data set described earlier. This gap has maintained stability over time, fluctuating around $30,000. This means, for those making the median wage in their county, on average, they have $30,000/year left over above and beyond their living costs.

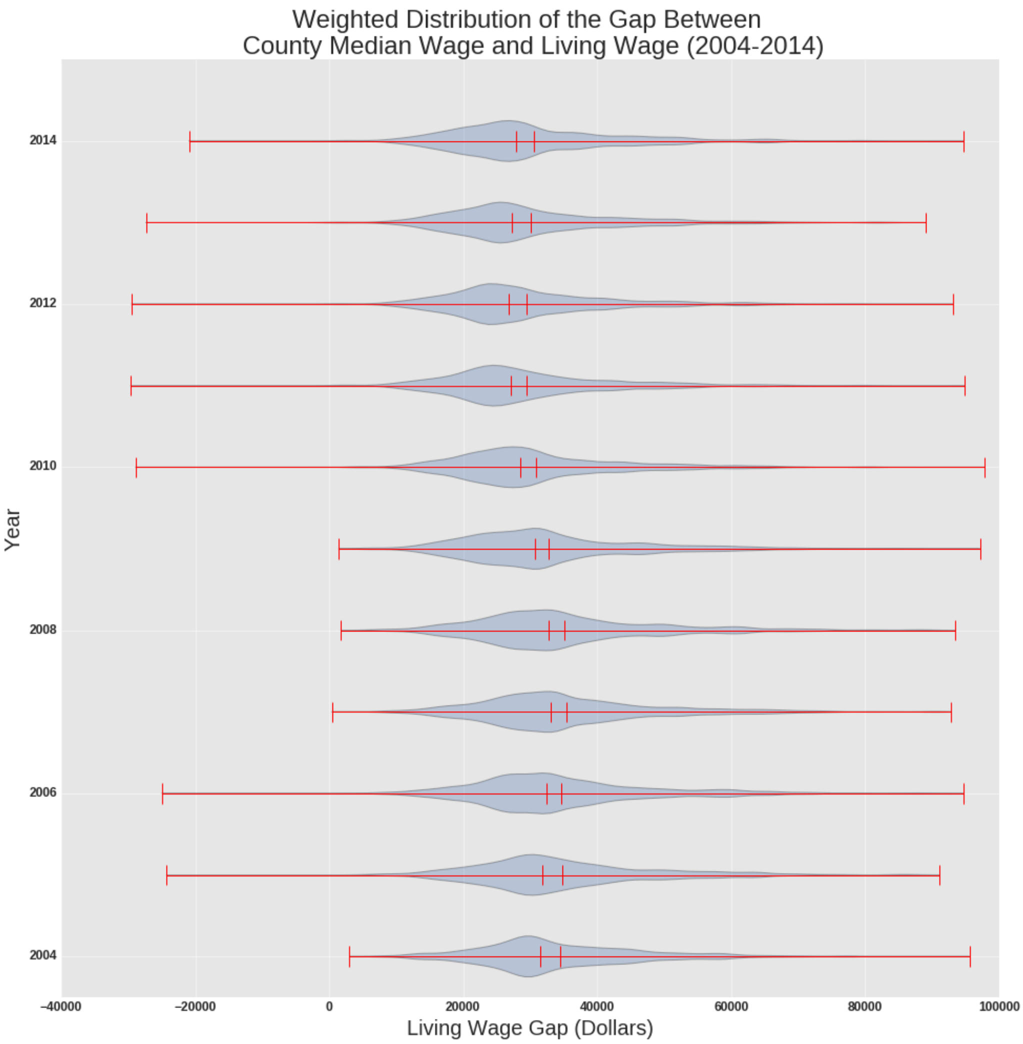


Figure 18: Weighted Distribution of the Gap Between County Median Wage and Living Wage (2004-2014)

5.2 DISTRIBUTION OF THE MINIMUM-WAGE-GAP

Figure 19 shows a similar figure to Figure 18, but for the living wage gap for the minimum wage. Using the minimum wage data discussed in the previous chapter, we find the difference between the county's applicable minimum wage and its minimum wage. A line showing the break even point is shown, with the green plots showing what the living wage gap would look like with a $15.00/hour minimum wage for that year (inflation causes that $15 to be worth less over time, which explains their trend towards the left over time). The figure shows that this gap seems to have peaked in 2007, but the gains made from 2007 - 2010 are being eaten away again. The distribution in 2014 is very similar to 2008, and if the trend continues, will match the peak in 2007. Future extensions to this project should look into the most recent data available to see how the trend continues.

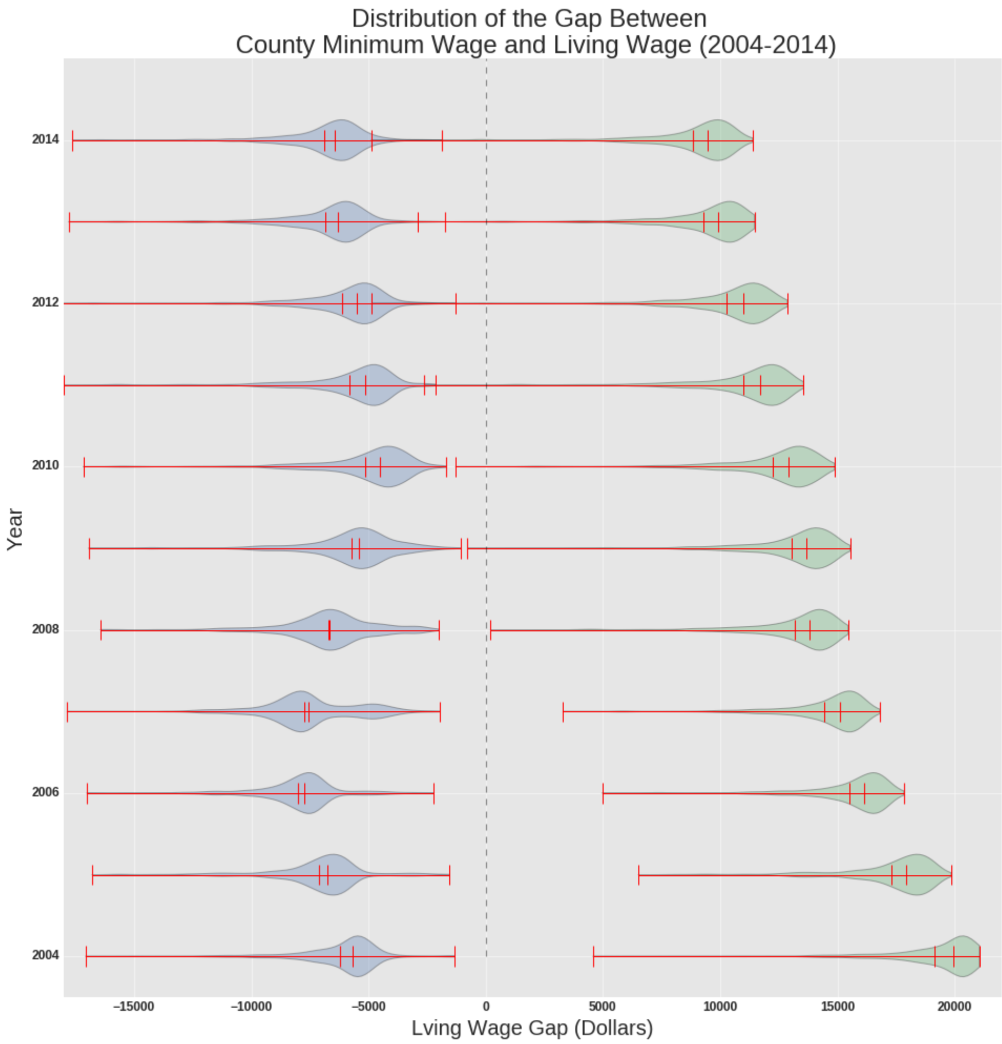


Figure 19: Distribution of the Gap Between County Minimum Wage and Living Wage (2004-2014)

5.3 Percentage of Single Households at or below the Living Wage

Using the wage distribution data from the Census, we can see what percentage of people in a county are making the living wage or less. Figure 20 shows this distribution as a histogram and a violin plot. A note on methodology: this uses the non-family household column in the data to match our model definition, since this model is for single adults with no children. Also the data from the census only breaks down the wage distribution based on their own bucketing. This percentage therefore is an approximation that can err on the higher or lower side. To compensate, the green distribution is the under-estimate, while the blue distribution is the over estimate. Neither prints a pretty picture, as there seems to be significant portions of the population that are making the living wage or below. To see details of how this was generated, please consult the associated IPython notebook.

Figure 21 shows how this breaks down over region, by showing the percentage of counties in a region that have more than 50% of their constituents making the living wage or less. This was done using the over-estimated values; using the under-estimated values and changing the threshold to the living wage at the peak (approx $30,000) creates a similar pie chart.

The last figure shows the percentage of people in a county making the living wage or less, distributed across the map of the united states. This is an interesting plot, as it shows that areas in the deep south have a very high proportion of people in a precarious situation. In contrast to Figure 5, these areas don't have very high living wages, which shows that these areas might be plagued with such low wages that many people cannot make enough to get by.

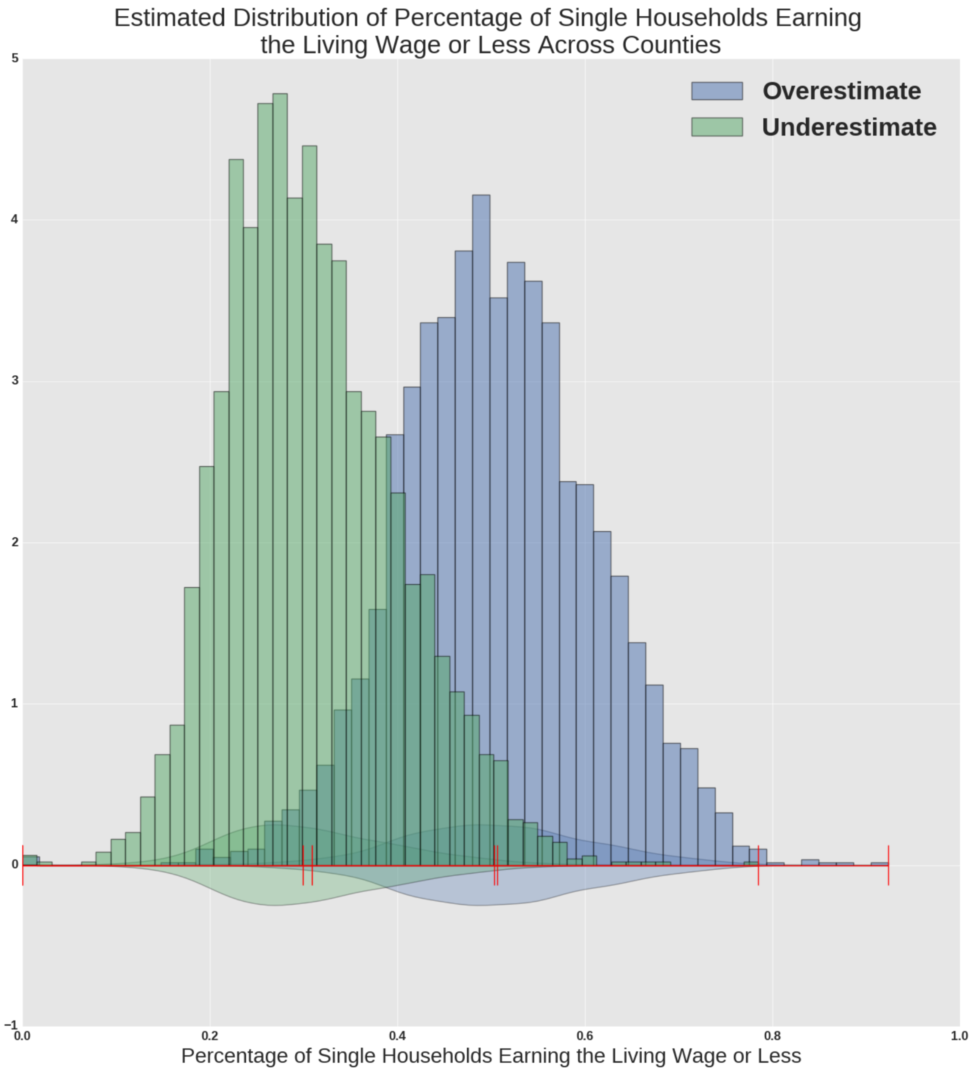


Figure 20: Estimated Distribution of Percentage of Single Households Earning the Living Wage or Less Across Counties

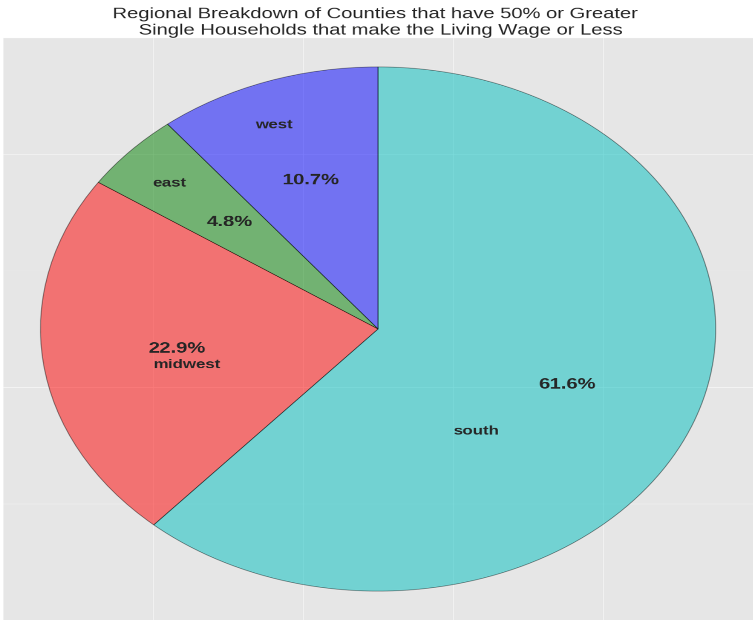


Figure 21: Regional Breakdown of Counties that have 50% or Greater Single Households that make the Living Wage or Less

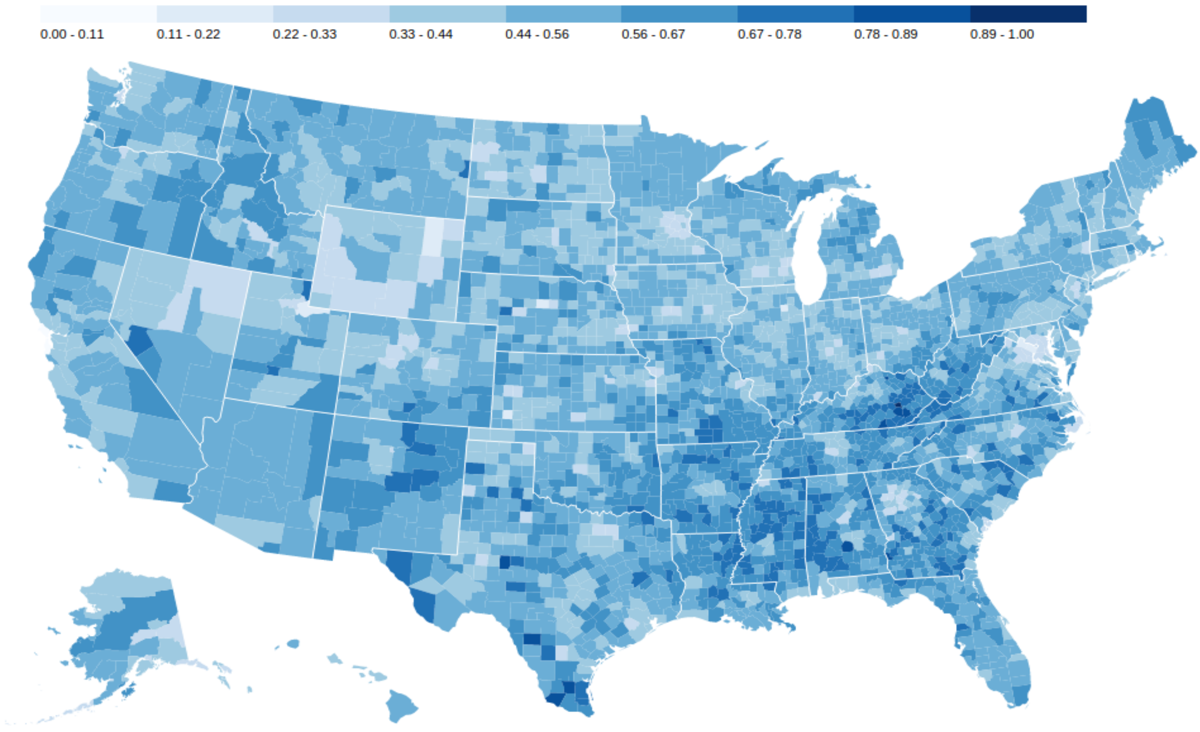


Figure 22: Map of Counties Showing Percentage of Single Households Earning The Living Wage or Less

6 RESULTS AND FUTURE WORK

6.1 OVERALL RESULTS

Here is a summary of points visited in the previous sections:

- The living wage seems to have levelled off in last few years. The dominate variable in this model (discussed below) is *housing_costs*, and a levelling off of average *housing_costs* (as shown in Figure 7 seems to explain this. Further analysis would look into why housing costs have stopped their steep decline in recent years.
- The gap between minimum wage and the living wage reached a peak in 2006, and then some gains were made due to increases in the minimum wage across the nation. However, due to higher rent and inflation, current levels are close to the 2006 peak again. See Figure **??** for reference.
- The top 150 most populous counties have a much higher living wage than the rest of the country. This is mostly due to rent being higher in densely populated areas, as rent is the model variables with the highest impact.
- The top 150 most populous counties can be split into two groups (bi-modal). This shows that even in the top 150 counties by population, there are a handful of counties with a very high living wage.
- White and Blacks seem to have similar living wage distributions, as they are relatively well mixed across the county. Other races, especially Asians, Hispanics and Pacific Islanders, are concentrated into areas with high living wages. Population dynamics plays a huge role in determining the living wage.
- Races seem to experience the same increases over time, with their distributions being controlled by population dynamics.
- The gap between a county's median wage and it's living wage seems stable over time. This might indicate that those making a median wage for their area are not so concerned about the living wage. Analysis of the gap between the living wage and the minimum wage show that the minimum wage in all areas is not enough to support a living wage for a single adult household.
- When looking at **only** single households, we see that on average, 30% - 50% of them are making the living wage or less. This range is due to inaccuracies in determining the wage distribution per county, but still shows that many areas in the country have significant portions of the population 'living on the edge'.
- The region that has the most counties that have 50% or more of their single households making he living wage or less, is the South.

6.2 FUTURE WORK

- If we could get wages broken down by race **and** by county, this would allow us to see how the living wage gap have evolved over time between races.
- Data that supports the model variables but at a more granular level would help increase the model's accuracy overall.
- Better data regarding wage distributions per county would help tighten up the analysis of those making the living wage or less.

REFERENCES

[1] Eric Schultheis Amy K. Glasmeier, Carey Anne Nadeau. Living wage calculator: User's guide / technical notes. 2014.

[2] James Quacinella. Associated github repository. 2015.

[3] James Quacinella. Associated ipython notebook. 2015.

[4] Consumer expenditure survey. 2015.

[5] Usda food plans: Cost of food. 2015.

[6] Stretching the food stamp dollar: regional price differences affect affordability of food. *Economic Information Bulletin Number 29-2*, 2007.

[7] Fair market rents from the u.s. department of housing and urban development. 2015.

[8] http://www.businessinsider.com/half-of-the-united-states-lives-in-these-counties-2013-9. 2015.

[9] Medical expenditure panel survey home. 2015.

[10] Social security and medicare tax rates. 2015.

[11] The tax foundation's center for state tax policy. 2015.

[12] Historical federal income tax rates for a family of four. 2015.

[13] Profile of general population and housing characteristics: 2010. 2010.

[14] Code for loading race data from the 2010 census. 2015.

[15] Changes in basic minimum wages in non-farm employment under state law: Selected years 1968 to 2016. 2015.

[16] Code to load minimum and median wage data. 2015.

[17] S1901: Income in the past 12 months (in 2014 inflation-adjusted dollars). 2014.

[18] James Quacinella. Code for loading model variables. 2015.