

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

Amy Glasmeier proposed a model for the living wage [1], which she defines 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 produced data for 2014. The purpose of this project is to take this model and extend its 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. The original model used 12 different family configurations. 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

1	INTRODUCTION	1
1.1	Introduction to the Living Wage and the Living Wage Model	1
2	DATA SOURCES AND COLLECTION	3
2.1	Consumer Expenditure Report	3
2.2	USDA Food Plans	3
2.3	Free Market Rent Data From HUD	3
2.4	Most Populated Counties	4
2.5	Medical Expenditure Panel Survey from the AHRQ	4
2.6	Tax Data	4
2.6.1	Payroll Taxes	4
2.6.2	State Tax Data	4
2.6.3	Federal Tax Data	4
3	MODEL VARIABLES	5
3.1	Housing Costs	5
3.2	Food Costs	5
3.3	Child Care Cost	5
3.4	Health Insurance Costs	6
3.5	Transportation, Health Care and Other Costs	6
3.6	Taxes	6
3.6.1	Payroll taxes	7
3.6.2	State Taxes	7
3.6.3	Federal Income Tax Rate	7
3.7	Final DataFrame	8
4	INTRODUCTORY ANALYSIS	9
4.1	Living Wage in Individual Counties	9
4.2	State Averages of the Living Wage	9
4.3	Choropleth Of Counties: 2004, 2006, 2014	10
4.4	Regional Averages of the Living Wage	10
4.5	National Average Breakdown Across 2004 - 2014	10
4.6	Living Wage Distribution in Most Populous Counties	10
4.7	Living Wage Distribution by Race	10
4.7.1	Loading Race Data from 2010 Census	10
4.7.2	Living Wage County (Non-Weighted) Breakdown by Race	11
4.7.3	Living Wage County (Weighted) Breakdown by Race	11
4.7.4	Population Weighted Averages Broken Down By Race	11
5	LIVING WAGE GAP	15
5.1	Data Collection	15
5.1.1	SAIPE - Small Area Income and Poverty Estimates	15

5.1.2	Minimum Wage Data	15	
5.2	Distribution of The Median-Gap	15	
5.3	Distribution of The Minimum-Wage-Gap	15	
5.4	Percentage of Single Households At or Below the Living Wage	15	
6	RESULTS AND FUTURE WORK	17	
6.1	Overall Results	17	
6.2	Future Work	17	
I	Appendix		19
A	CODE LISTINGS	21	
A.1	Data Sources	21	
	BIBLIOGRAPHY	22	

LIST OF FIGURES

Figure 1	Living Wage Trend for Select Counties	9
Figure 2	Living Wage Trend for Select States	10
Figure 3	Choropleth of Counties based on Living Wage, 2004	11
Figure 4	Choropleth of Counties based on Living Wage, 2006	12
Figure 5	Choropleth of Counties based on Living Wage, 2014	12
Figure 6	Regional Trend of the Living Wage	13

LIST OF TABLES

LISTINGS

INTRODUCTION

1.1 INTRODUCTION TO THE LIVING WAGE AND THE LIVING WAGE MODEL

The living wage is defined by the original paper[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 9 variables:

$$\text{basic_needs_budget} = \text{food_cost} + \text{child_care_cost} + (\text{insurance_premiums} + \text{health_care_costs}) + \text{housing_cost} + \text{transportation_cost} + \text{other_necessities_cost}$$

$$\text{living_wage} = \text{basic_needs_budget} + (\text{basic_needs_budget} * \text{tax_rate})$$

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.

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.

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. [13]

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 + sub-county 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 impute 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.

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

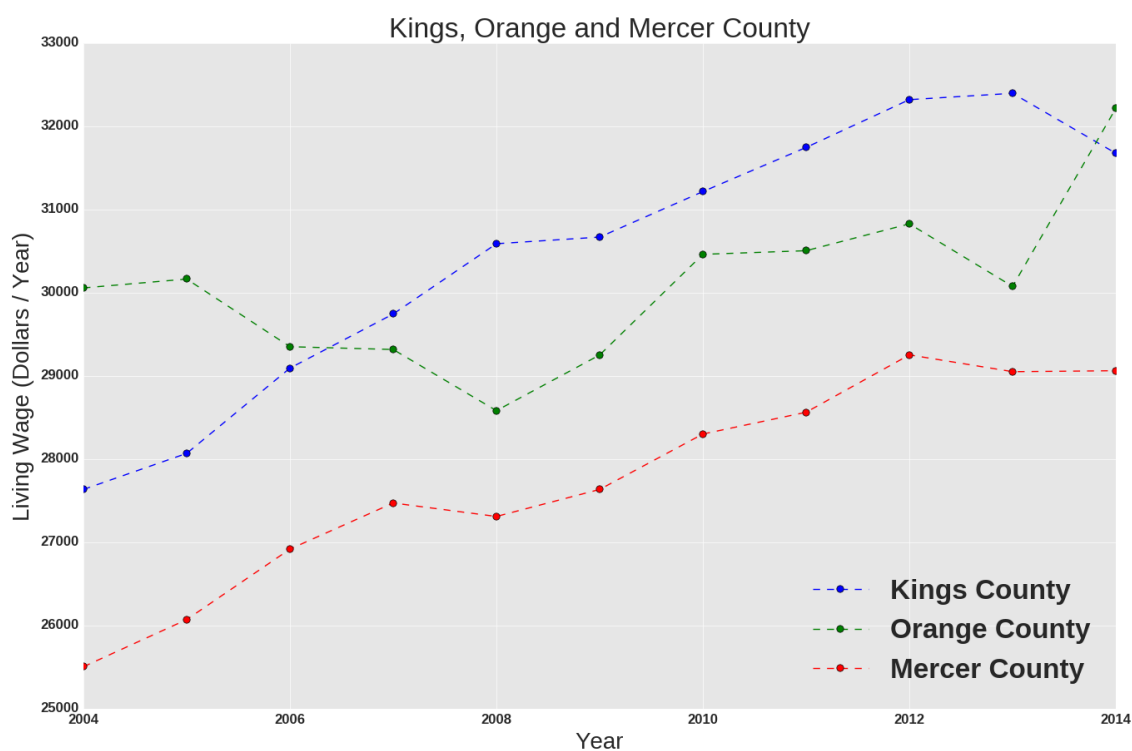


Figure 1: Living Wage Trend for Select Counties

Figure 1 shows

4.2 STATE AVERAGES OF THE LIVING WAGE

Figure 2 shows

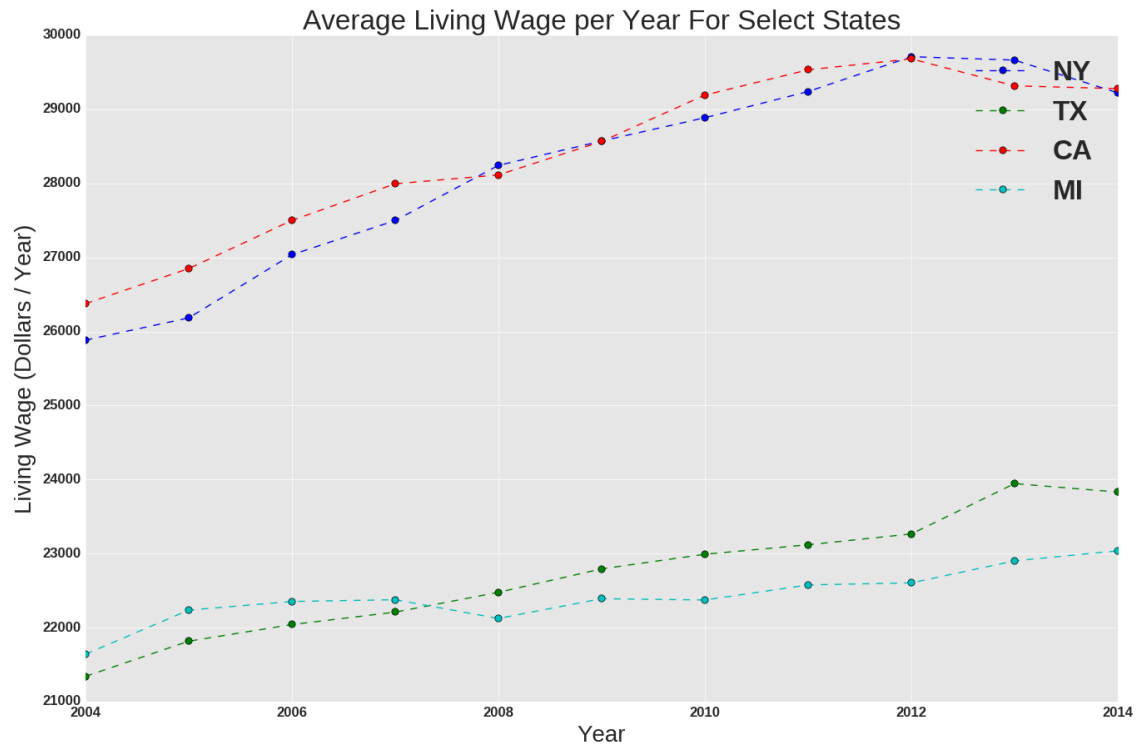


Figure 2: Living Wage Trend for Select States

4.3 CHOROPLETH OF COUNTIES: 2004, 2006, 2014

Figure 3 shows

Figure 4 shows

Figure 5 shows

4.4 REGIONAL AVERAGES OF THE LIVING WAGE

Figure 6 shows

4.5 NATIONAL AVERAGE BREAKDOWN ACROSS 2004 - 2014

4.6 LIVING WAGE DISTRIBUTION IN MOST POPULOUS COUNTIES

4.7 LIVING WAGE DISTRIBUTION BY RACE

4.7.1 Loading Race Data from 2010 Census

Move to data collection section

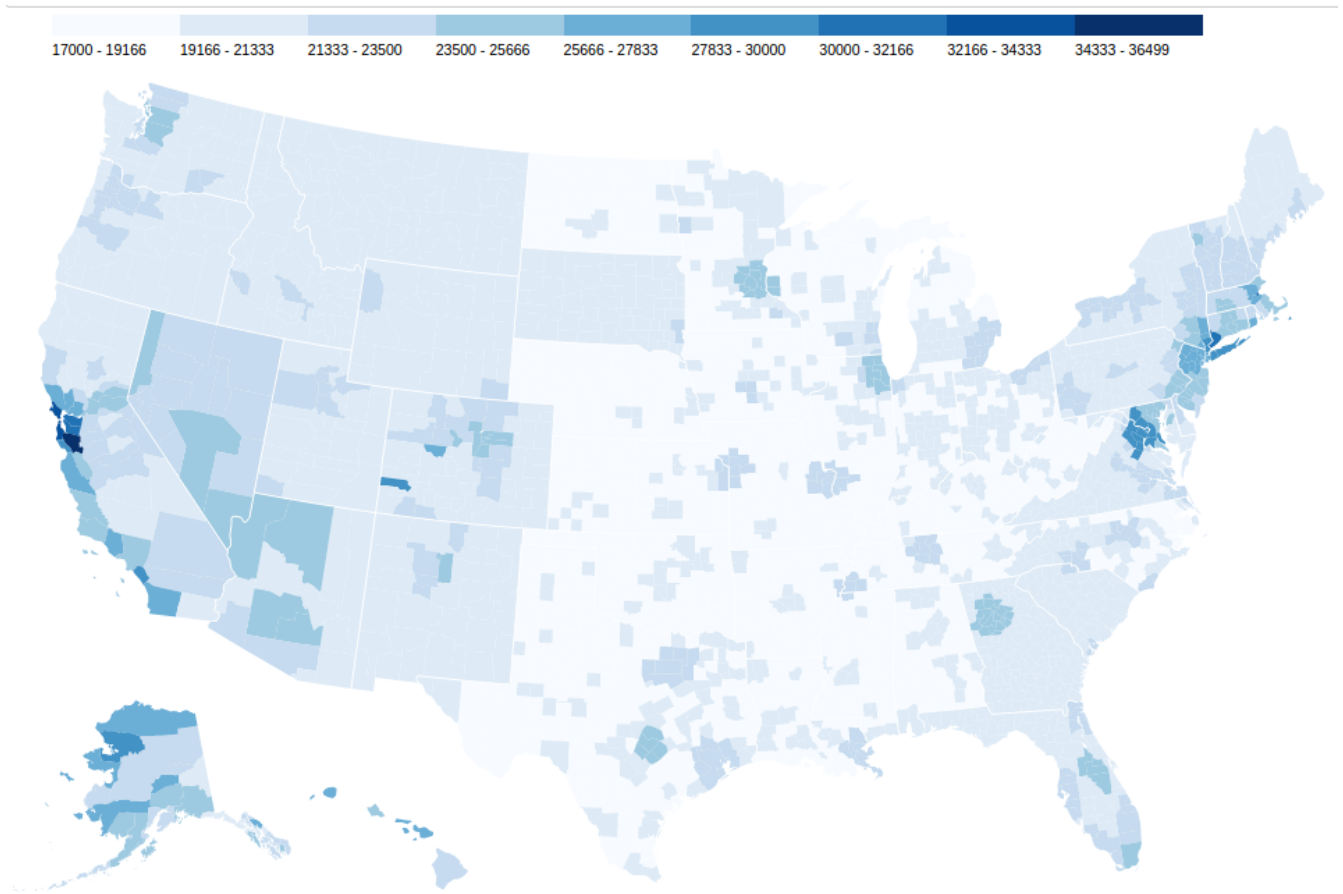


Figure 3: Choropleth of Counties based on Living Wage, 2004

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

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

4.7.4 *Population Weighted Averages Broken Down By Race*

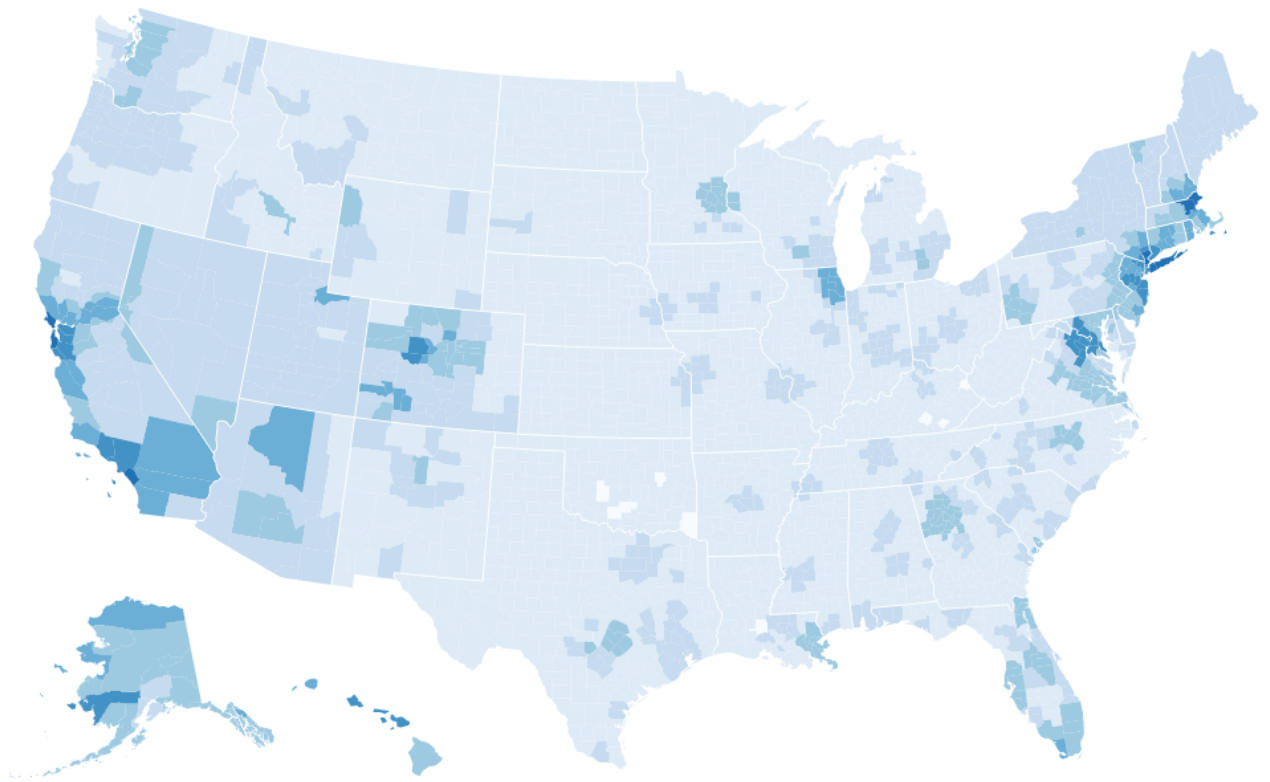


Figure 4: Choropleth of Counties based on Living Wage, 2006

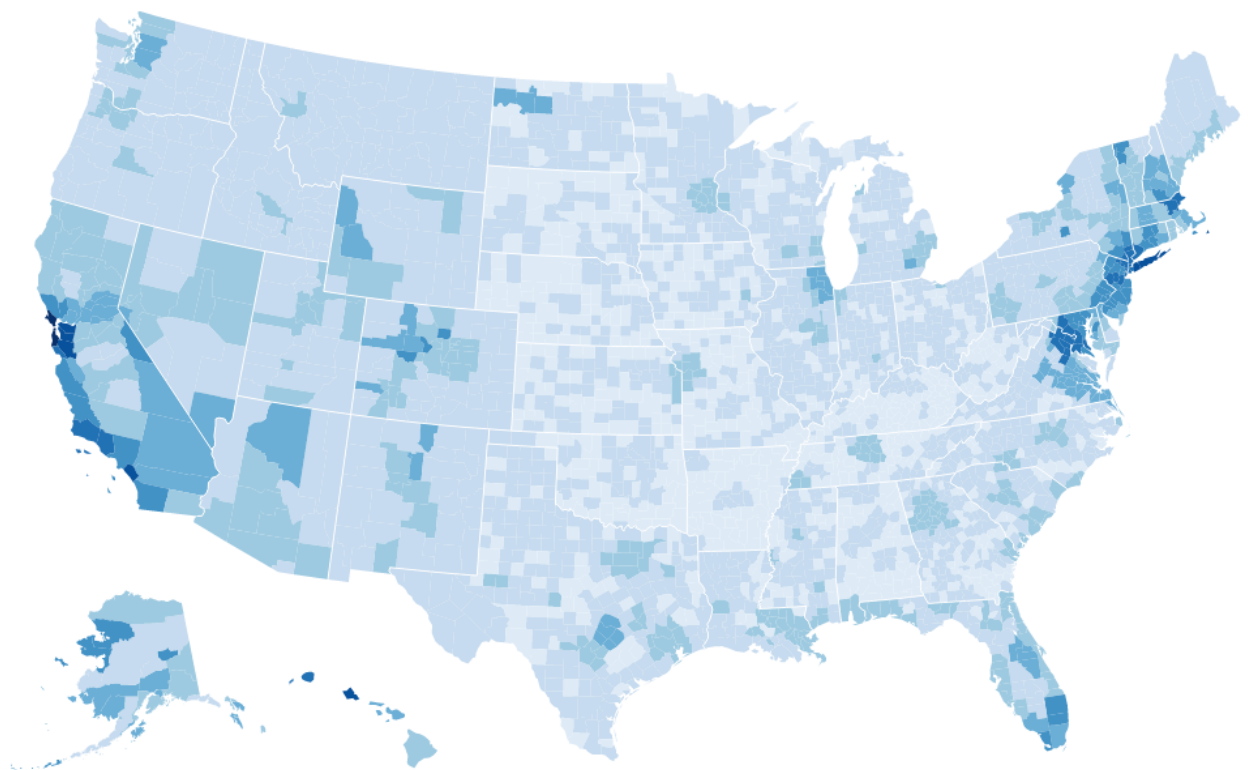


Figure 5: Choropleth of Counties based on Living Wage, 2014

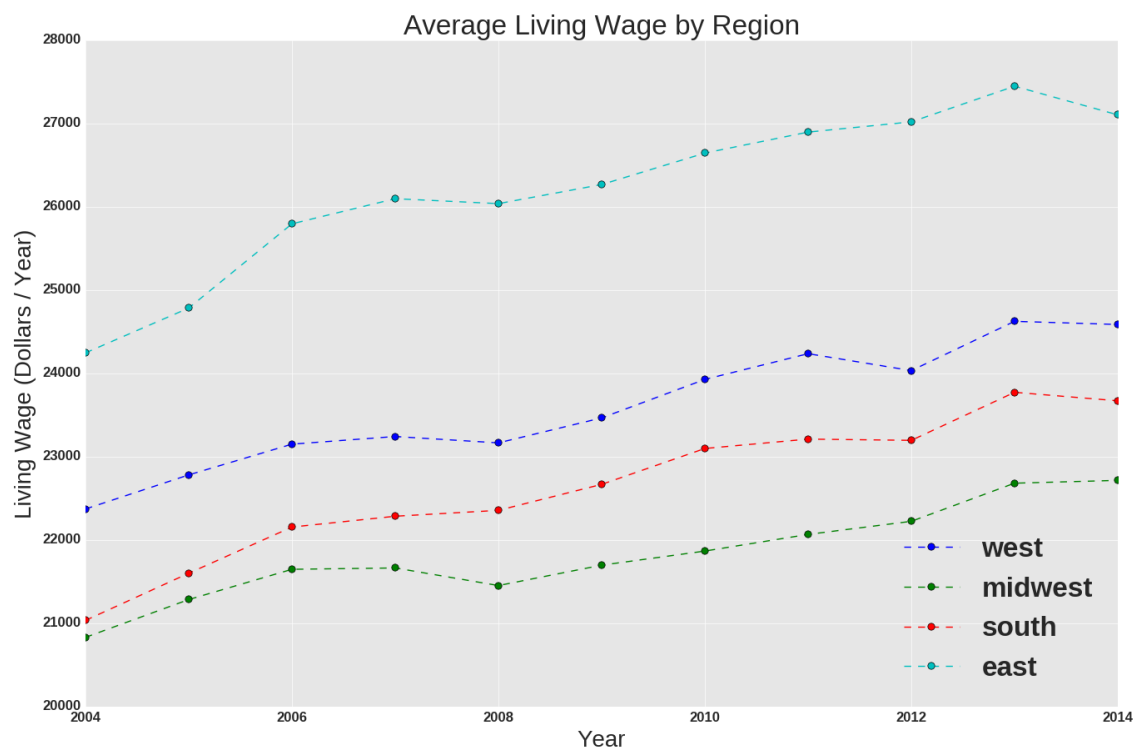


Figure 6: Regional Trend of the Living Wage

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 DATA COLLECTION

Move this to data collection section

5.1.1 SAIPE - *Small Area Income and Poverty Estimates*

5.1.2 *Minimum Wage Data*

5.2 DISTRIBUTION OF THE MEDIAN-GAP

5.3 DISTRIBUTION OF THE MINIMUM-WAGE-GAP

5.4 PERCENTAGE OF SINGLE HOUSEHOLDS AT OR BELOW THE LIVING WAGE

Note that data is broken down into buckets, so numbers are an approx. Note the household definition again

RESULTS AND FUTURE WORK

6.1 OVERALL RESULTS

Will expand on each point:

- The living wage seems to have levelled off in last few years. Not sure why, but future analysis would be interesting
- Gap between minimum wage and the living wage reached a peak, and then some gains were made due to increases in the minimum wage. However, due to rent and inflation, current levels are close to the peak again.
- The top 150 most populous counties have a much higher living wage than the rest of the country. This is mostly due to rent
- The top 150 most populous counties can be split into two groups.
- White and blacks seem to have similar living wage distributions, as they are relatively well mixed across the county. Other races, especially asians and PI, are concentrated into areas with high living wages.
- Races seem to experience the same increases over time, with their distributions being controlled by population dynamics (since the living wage changes most with location)
- When looking at **only** single households, we see that on average, 50
- The region that has the most counties that have 50% or more of their single households making the living wage or less, is the South.
- Median wage?

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.

Part I

Appendix

CODE LISTINGS

A.1 DATA SOURCES

Code for loading all the data sources can be found in the IPython notebook under "Data Collection" [14]

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] James Quacinella. Code for loading model variables. 2015.
- [14] James Quacinella. Code for data collection. 2015.