U.S. State Income Growth Analysis and Visualization

Did wage growth over 2001 – 2008 impact the 2016 presidential election?

Yingying Tang
Department of Computer Science
University of Illinois at Urbana-Champaign
Richland, Washington USA
yt5@illinois.edu

Richard J. Gjeldum
Department of Computer Science
University of Illinois at Urbana-Champaign
Wheaton, Illinois USA
gjeldum2@illinois.edu

***Abstract*— This paper builts a web-based interactive data visualization tool representing, on a U.S. map, the average and median earnings for each state, by month, for the period of 2001 - 2008. Also displayed is a U.S. map of the 2016 presidential election results, by state. The goal of this study is to determine whether a meaningful relationship exists between U.S. state wage growth over the 2001 – 2008 period and the 2016 presidential election results by state. In this work, the Hadoop MapReduce framework is used for data prep and analysis. With the extracted data from MapReduce jobs, the D3 tool is used for the data visualization. The comparison of the visualization results reveals that there wasn't any clear Red / Blue state division politically based on median income growth.**

Keywords—state average wages; state median wages; 2016 presidential election; wage growth; U.S. state income.

I. INTRODUCTION

The average wage growth is a means to examine the recovery from the great recession [1]. When laid-off workers continue to re-enter the labor force after the recession, their wages tend to be well below that of peers who remained employed. As a result, these new hires bring down the average hourly wage rate. With the average growth rate over decades, whether economics is recovered and whether wages among steadily employed workers are growing at a healthy clip can be analyzed. Based on the given wages in different region in each state from the dataset "Labor Statistics Databases", the average wage is to be calculated and analyzed.

Average wages though are subject to being skewered by individuals on the high end of the pay scale. Disparities in wage growth become much clearer in measuring median wages by state [2]. Different states feature different industries, from Montana's agriculture and tourism, to Tennessee's burgeoning healthcare and transportation sectors. With the calculated and ranked median wage growth for each state, data can be visualized with different colors on a US map, which is a convenient means to find the top-ranked states with higher median wages.

Another interesting study can be conducted from visualization of the median wages on the US map is to see how the income growth is related to the presidential race. It is told by the major political narratives that the rust-belt states are more willing to support Donald Trump, because they are struggling economically. By plotting two US maps, one showing the wage growth rate in each state with different colors, and one showing the 2016 presidential election results with blue and red colors, it can reveal whether the narratives are describing reality.

Wage data was obtained from the United States Department of Labor - Bureau of Labor Statistics for each state over the years 2001 – 2008 [3]. This period of time is of particular interest as it led up to the credit crisis and Great Recession. Data for the 2016 presidential election results were obtained from the Federal Election Commission [9].

The technology stack used for the analysis was Amazon Web Services (AWS), Amazon Elastic Block Storage (EBS) for the database, Hadoop MapReduce framework for data prep and analysis, and D3 (Data Driven Documentation) for the data visualization functionality.

People who may benefit from this analysis include those 1) searching for jobs and want to understand the income trends in each state, 2) government officials who want to analyze their state's wage base for proposing new taxes for funding social programs, 3) candidates seeking political office who need to understand the financial make-up of their constituents, 4) corporations that are considering building new facilities, and 5) venture capitalists that want to start businesses in states with higher disposable incomes.

II. TECHNOLOGY STACK

We utilized Amazon Elastic Compute Cloud (EC2) servers to provide scalable computing capacity in the Amazon Web Services (AWS) Cloud. We read and preprocessed the data with the Hadoop MapReduce framework. Furthermore, the Amazon Elastic Block Store (EBS) as used as storage technology for making block level storage volumes of our public data set available to our EC2 instances. Finally, we implemented data visualization by using D3.js toolkit (https://www.d3js.org/). D3 (Data-Driven Documents or D3.js) is a JavaScript library for

producing dynamic, interactive data visualizations in web browsers. It makes use of the widely implemented SVG, HTML5, and CSS standards. Also used was *DataMaps* (http://datamaps.github.io), a customizable SVG map visualizations tool for the web in a single Javascript file, using D3.js.

A. Hadoop MapReduce Components

Using the Hadoop MapReduce framework, we read through all the states' given wages data, and calculated the average and median prices for each year, each month, for each state, from 2001 to 2008, in the dataset Labor Statistics Databases. Figure 1 shows an example of what included in the dataset. In each file, a value was given for each *series_id* in each year and month. The *series_id* contains the information of state code, area code, and data type. Totally, 7 data types were defined in the dataset. Only the type 1, representing the total employees, and type 6, representing the average hourly earnings, were needed for this work. Figure 1 shows part of the given data for the state Alaska. In this figure, since the series_ids listed are all ended with type 1, the corresponding data of year, period (month) and value are recorded and used.

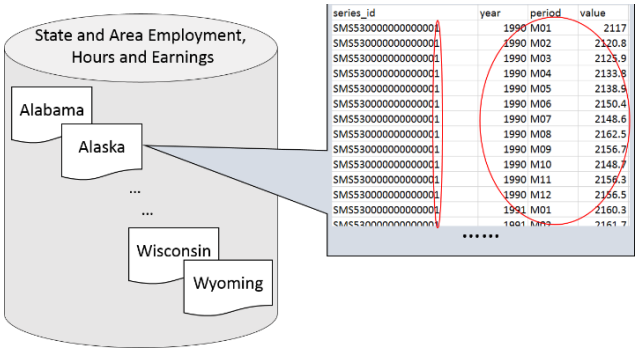

Figure 1 – Dataset Labor Statistics Databases

We divided the data reading and data processing tasks into two jobs, *LaborStatsArea* and *LaborStatsTime*. Figure 2 and 3 shows respectively the flowchart of the two jobs.

In the Mapper class of the first job (*LaborStatsAreaMap*), files containing wage information for each state were read in. In order to store the related data together, two custom Hadoop Writable data types, *YearMonthArea* and *SalaryWorkerNumWritable*, were defined and used as key and value outputs respectively. With each line read by the Mapper class, the year, month and area code (each state is composed of several areas) data were stored into each *YearMonthArea* object; the values related to data types 1 or 6, if found, were stored into each *SalaryWorkerNumWritable* object.

In the Reducer class of the first job (*LaborStatsAreaReduce*), the hourly wage and numbers of the employees given in the same time and state were combined together as Strings, and output as a Text data type. The corresponding information of year, month and state were also combined and saved as output.

The benefit of using the costumed Hadoop Writable data types in the Mapper class is to speed up the processing time; However, when the first job ends, and the outputs need to be transferred from the first job to the second job, only the data types defined in Hadoop need to be used.

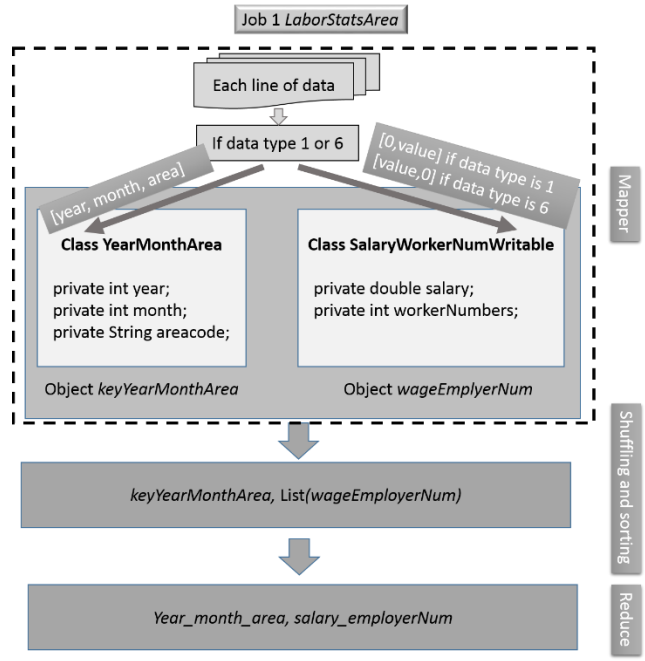

Figure 2 – Description of MapReduce job 1 *LaborStatsArea*

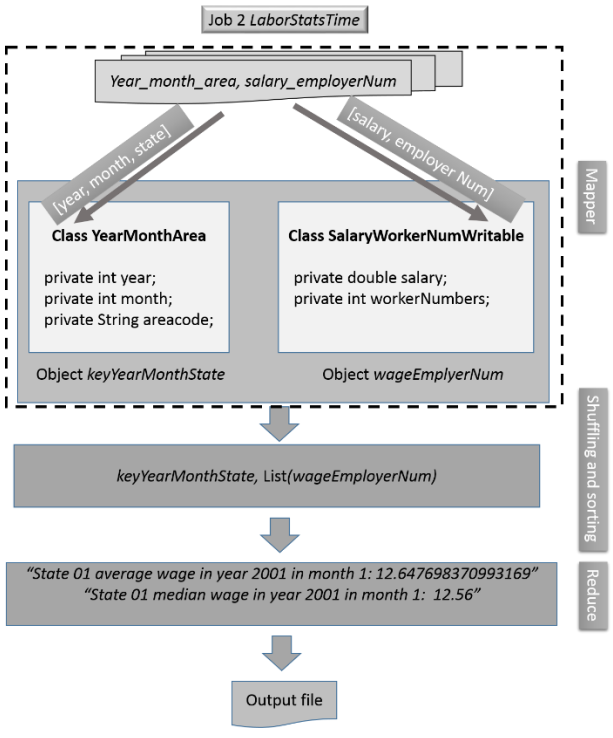

Figure 3 – Description of MapReduce job 2 *LaborStatsTime*

In the Mapper class of the second job (*LaborStatsTimeMap*), the outputs from the first job were received and converted back into the data types *YearMonthArea* and *SalaryWorkerNumWritable*. In the *YearMonthArea* object in this second job, the state code instead of the area code is stored. In this way, each reduce task in the second job will receive the the hourly wage and numbers of the employees given in the same time and state.

In the Reducer class of the second job (*LaborStatsTimeReduce*), the average and median wages in each state were calculated, based on the average hourly wages and numbers of employees in each area of the same state, and were written to an output file.

B. *Data Visualization Components*

With the output file generated by the second MapReduce job, average and median hourly wages in each state are obtained. In order to understand the relationship between the wages and the president election results, a data visualization is needed to visualize the wages of each state using different colors on the US map. We used Javascripts to plot the US map based on presidential election results. The tool *DataMaps* found in http://datamaps.github.io/ was used for generating the map.

The data visualization work is composed of three parts: *readInLaborStats.js*, *USmap.html*, and *plotMap.js*, as shown in Figure 4.

In *readInLaborStats.js*, the output file from the MapReduce jobs is read, and written into a JSON file. The format of the JSON file makes it easier for parsing wage data in specific year, month and state.

The file *USmap.html* defines the *content* of the webpage. The webpage is divided into three parts, the first part is the US map showing the president election result in 2016 (Democratic won in the blue-colored state, and Republican won in the red-colored state); the second part of the page visualizes the average wages in each state on the US map. The state with relatively lower average wage value is plotted in red colors, and the color of the state will gradually turns to blue with higher average wage observed; the third part of the page presents the median wages in each state on the US map. Similar to the second US map, a red to blue gradient is used when a relatively higher median wage is observed. Titles and subtitles of each part in this webpage are given.

The file *plotMap.js* defines the webpage's functionality. A function *slider* is used, so that user can drag and choose the year and month he/she wants to check the wages at that time. In the second and the third plot, each state is colorized based on average/median wages, and the colors is unique for every value. For this purpose, a palette is created using the minimum and maximum value in each specific time, with the red to blue gradient. Whenever user move drag and change the time (either year or month), the corresponding wage value for each state will be extracted from the JSON file, based on the updated year and month data. Then, the palette is created, and each state will be represented by a color representing its wage level compared to other states.

Figure 4 shows the flow chart of the three parts for the data visualization process. In the screenshot of the finished web-based interactive data visualization tool, the data in December 2008 are chosen to be displayed on the U.S. maps with the colors changing from red to blue.

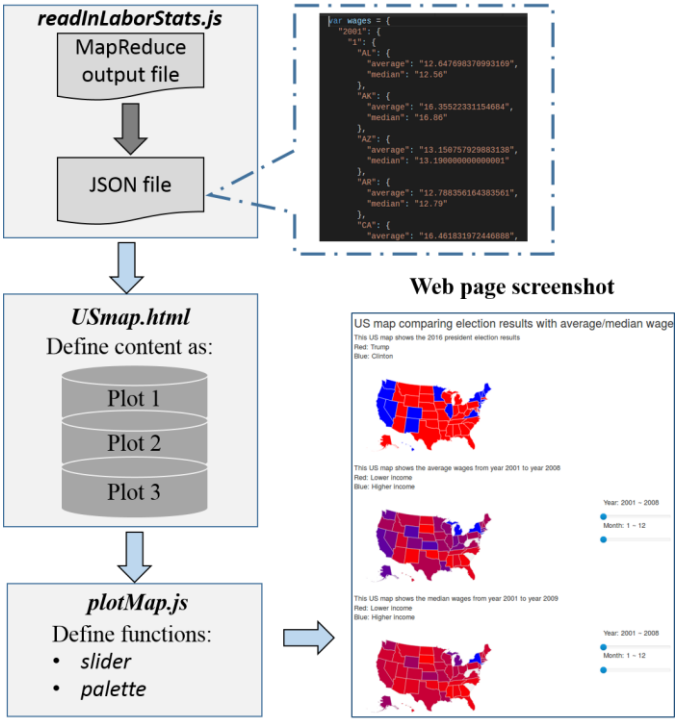

Figure 4 – Flow chart of data visualization process

III. Experimental Results and Evaluation

Figure 5 compares the U.S. map of the 2016 presidential election result, with the one showing the median wage of each state in December 2008. It clearly reveals that the election results are not closely related to the median wage salaries. For example, the states that voted for Clinton such as California, Oregon and Nevada have a relatively lower median wage; in the other hand, states such as Alaska, Wyoming and Utah are presented as relatively higher income states in the right U.S. map, but they voted for Trump as shown in the left U.S. map.

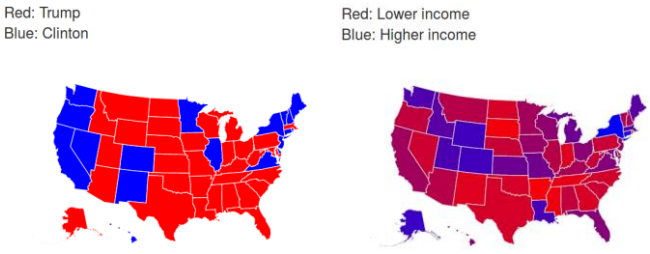

Figure 5 – Comparison of the 2016 presidential election result (left) and median wage of each state in December 2008 (right)

Figure 6 compares the average wages in January 2001 to the ones in December 2008. It can be seen that New York is always one of the highest average wage states in these years. In addition, Michigan state has an obvious decrease of average wage, of a 4.26% decrease rate. In the other hand, Idaho state has a high average wage growth rate of 46.28%.

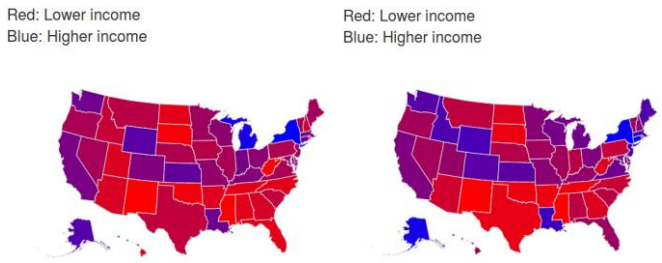

Figure 6 – Comparison of the average wage in each state in January 2001 (left) and in December 2008 (right)

Figure 7 compares the average wages to the median wages in December 2008. Although average wages are subject to being skewered by individuals on the high end of the pay scale, not much difference is seen between the average and median state wage incomes in Figure 7.

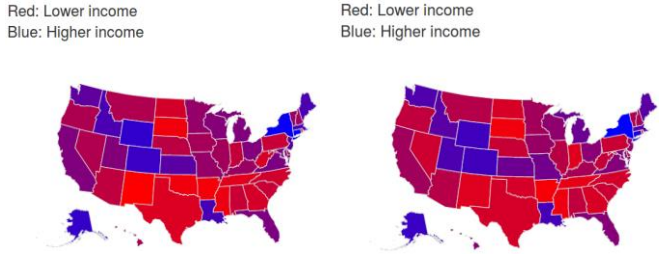

Figure 7 – Comparison of the average wage (left) to the median wage (right) in each state in December 2008

There are some other interesting observations, for example, as shown in Figure 8, the median wage in New York is much higher than the one in other state, in January 2001. However, the median wage in New York decreases from 24.97 $/hour in the year 2001 to 23.42$/hour in the end of year 2008, whereas several states have large median wage growth rate, such as Idaho, Colorado, Utah and Louisiana, as indicated from the right plot in Figure 8.

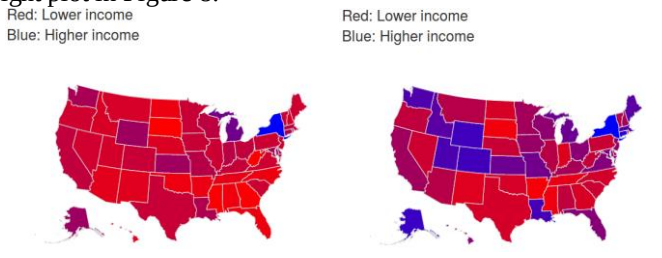

Figure 8 – Comparison of the median wages in January 2001 (left) to the ones in December 2008 (right)

Except for these analyzed comparisons, user can use this web-based interactive data visualization tool to check and compare the wage changes over the years, by dragging the slider. This tool has been provided on the GitHub website [10].

IV. Other Studies

A recent article on the International Monetary Fund Blog points out that U.S. Wage growth has been quite elusive since the Great Recession. There were 3 primary reasons cited: 1) laid-off workers continue to re-enter the labor force by accepting wages well below that of their peers who remain employed, 2) lower worker productivity – the more productive workers are the more they earn, and 3) reduced worker mobility which often results in higher wages by changing jobs (20% of workers changed jobs per year in the 2000's while only 12% have in recent years) [1].

The *Governing Data* web site graph showing changes in real hourly wages (by state) since 2003. They point out that hourly wages in most states have climbed in recent years as the economy recovered, but over the longer term, median wages generally haven't increased much [2].

The Economic Policy Institute observed that slow wage growth is a key sign of how far the U.S. economy remains from a full recovery. Nominal wage growth, as seen in their has a nominal wage tracker graph, has been far below target in the recovery as seen in a graph of year-over-year change in private-sector nominal average hourly earnings from 2007 – 2017. 3.5 percent nominal wage growth is consistent with the Federal Reserve Board's 2 percent inflation target, 1.5 percent productivity growth, and a stable labor share of income [4].

Fortune Magazine determined which states had the fastest and slowest growth in median income from 2014 - 2015. They determined that the median increase of 5.2 percent for all Americans was the first increase since 2007 and the largest ever annual increase since the Census Bureau started tracking the statistic in 1967. The states with the highest median income in 2015 were Massachusetts, Connecticut, District of Columbia, New Jersey, Maryland, Hawaii, and Alaska. These states were primarily in the northeast corrider of the U.S. or happen to benefit from oil or tourism. But even after 2015's record year, the typical American's income was still lower in 2015 than in 2007. The magazine went on to point out that there wasn't any clear Red / Blue state division politically based on median income growth from 2014-2015. In fact, rust-belt Wisconsin ranked as one of the states where income grew the fastest, while Ohio, Pennsylvania, New York, and Michigan were in the middle of the pack. Also noteworthy was wealthy, blue New Jersey ranked last in median income growth, right behind ruby-red Idaho. These observations can be also found from the web-based interactive data visualization tool developed in this work. They concluded the 2016 presidential race was much more about cultural and regional differences than economic misfortune [6].

V. Concusion

This work has presented the importance of analyzing average and median state wages, the procedure of extracting and visualizing the wage data in years, and the evaluations of the wage data by comparisons. By using the developed web-based interactive data visualization tool in this work, users can get a

better understanding of the wage growth in each state in each year. Moreover, this tool clearly reveals that there is no direct relation between the wage growth rate and the 2016 presidential election results.

Since the dataset provided [3] only contains the wage data from 2001 to 2008, more interesting results, such as how each state recovers from the great recession in 2008, can not be displayed. However, once the wage data are available, the tool developed can be used to read in and visualize the data.

In the future work, a wage growth tracker will be provided in the tool, displaying the value of average or median growth rate in the past years. Users can choose the state to be analyzed. Except for the wage changes, other properties can also be displayed for each state, either in the U.S. map, or in the growth tracker, such as employment numbers and college graduates numbers. Changes of different properties can be placed in the same chart, and relationships between each property can be clearly observed.

ACKNOWLEDGMENT

The authors would like to thank Dr. Mark Bennett, Senior Data Scientist at a major investment bank in Chicago, IL for his encouragement during this study and for the time he spent discussing possible research topics and best practices for data analysis.

REFERENCES

[1] iMF Direct - The International Monetary Fund's global economy forum No Victory Lap Yet: U.S. Wage Growth Elusive, by Stephan Danninger, September 20, 2016.
https://blog-imfdirect.imf.org/2016/09/20/no-victory-lap-yet-u-s-wage-growth-elusive/

[2] Governing The States and Localities
Average Wages By State, 2012.
http://www.governing.com/gov-data/wage-average-median-pay-data-for-states.html

[3] United States Department of Labor – Bureau of Labor Statistics
https://www.bls.gov/data/
https://download.bls.gov/pub/time.series/sm/

[4] Economic Policy Institute
NOMINAL WAGE TRACKER: Nominal wage growth has been far below target in the recovery Year-over-year change in private-sector nominal average hourly earnings, 2007–2017.

http://www.epi.org/nominal-wage-tracker/

[5] Trading Economics
United States Wages and Salary Growth, 1960-2017
http://www.tradingeconomics.com/united-states/wage-growth

[6] Fortune Magazine
The States Where Paychecks Grew the Fastest and Slowest Last Year, Chris Matthews Sep 14, 2016
http://fortune.com/2016/09/15/states-income-growth/

[7] 24/7 Wall St., LLC
States Where Incomes are Booming (or Not); Evan Comen, Thomas C. Frohlich, Samuel Stebbins and Michael B. Sauter, May 5, 2016
http://247wallst.com/special-report/2016/05/05/states-where-incomes-are-booming-or-not-2/11/

[8] NCSL – National Conference of State Legislatures
STATE MINIMUM WAGES | 2017 MINIMUM WAGE BY STATE, Jan 5, 2017
http://www.ncsl.org/research/labor-and-employment/state-minimum-wage-chart.aspx

[9] Federal Election Commission
www.fec.gov/pubrec/fe2016/2016presgeresults.pdf

[10] https://github.com/yingjf/laborStat.git