Exploring the Small Area Income and Poverty Estimates (SAIPE) of States in the United States of America

Introduction

The Small Area Income and Poverty Estimates (SAIPE) is a program by the United States Census Bureau designed to derive estimates of poverty and income values over states, counties, and school districts of the United States. We looked to explore this data over the time period of 2003 to 2011. The scope of this project was limited to the state level to match other data sets. To investigate other data that may be related to the SAIPE data, we incorporated employment data and education data from other sources.

Data Sets

We made use of the SAIPE data set provided by the US Census Bureau (http://www.census.gov/did/www/saipe/data/statecounty/index.html). This data includes statistics from 2003 to 2011 for all 50 states as well as the District of Columbia and for the United States as a whole. We utilized the median household income as well as the poverty percentage values. In addition, we looked at the poverty threshold value for each of these years.

In addition, we utilized employment data from the US Department of Labor (http://www.bls.gov/lau/data.htm) and high school graduation statistics from the US Department of Education. For the high school graduation values, only the 2011 values were considered because the officially reported rate was only standardized as of 2011 (http://www.ed.gov/news/press-releases/states-report-new-high-school-graduation-rates-using-more-accurate-common-measur).
Headline 1: States show similar trends in poverty, unemployment, and median household income over the focus time period

This visualization shows a trellised view of five separate states as well as the United States as a whole. The views show the median household income, the unemployment rate, and the poverty as a percentage of the population over the nine year period. The five states were selected based on the highest and lowest representative state for each category in 2011. California had the highest unemployment in 2011, while North Dakota the lowest. New Hampshire had the lowest poverty percentage all ages in 2011 and Mississippi the highest. Maryland had the highest median income in 2011 and Mississippi the lowest. Despite the extremes values of each of these states, we can identify some common trends. The median household income for each of the states shows an increase between 2003-2008 with a dip in later years. North Dakota is an exception, showing an increase over the entire nine year period. There is an overall small increase in poverty percentage over the nine years. The exception is California, which had a small decrease from 2003 to 2007, but an increase from 2007 to 2011. The most interesting revelation from this visualization is the similarity in the unemployment trends among these five states and the United States as a whole. Despite the extreme values of unemployment and their geographic and demographic distances, they all show similar trends in the minimum dip during the 2007 and 2008 years followed by a spike in 2009 and 2010, followed by a decrease in 2011.
When constructing visualizations, we hypothesized that the percentage of population under poverty would have a negative correlation with the median household income. In other words, a low median household income would indicate a high poverty level. The above visualization would seem to support this hypothesis. However, an interesting trend is revealed when we label the data points according to state as in the visualization below. Despite the overall data as a whole indicating a negative correlation, many states show a positive correlation between median household income and poverty level. This seems counterintuitive, so we created another visualization to explore the phenomenon.
We took a closer look at the data and noticed that in 2011, states tended to have higher median household incomes. We also observed that states seemed to have lower median household incomes in 2003. By filtering out the other years, we generated the visualization below. This clearly shows the discrepancy between 2003 and 2011. Not only are median household incomes higher in 2011, but so are poverty levels. This drives the individual states to show a positive correlation, even though the states as a whole show a negative correlation in a given year. It is possible that being limited to only nine data points for each state is causing the counterintuitive results. By aggregating all states, the data points show a clear negative correlation. We propose a possible explanation of the trend as the distribution of wealth widening. Those above the poverty threshold became richer, increasing the median household income at a rate that outpaced the increase in those losing money and falling below the poverty threshold.
As shown in the above figure, the poverty percent of all ages is negatively correlated with the graduation rate. This is an expected correlation, as states with a better educated population will
presumably have an easier time finding employment. From another perspective, we can hypothesize that low poverty rates are an indicator of a more affluent community that will be able to provide better educational facilities. This would help students in performing well and graduating high school.

The unemployment rate is also negatively correlated with the graduation rate. It is intuitive that the high graduation rate will result in more capable labor force. This labor force would be able to more easily obtain and hold employment. An outlier state is California, which has high graduation rate but still has high unemployment rate. This implies that people with a high school education may be unable to obtain employment. We can hypothesize that because California has a lot of high-technology companies, employment is very fluid. Small technology based companies can have their prospects vary widely, causing their workforce to expand or shrink rapidly. In addition, California has a very high population and several densely populated cities, which may contribute to high competition for employment.

The median household income is positively correlated with the graduation rate, although at a weak level. It can be interpreted that a better educated workforce is better able to obtain employment, thus producing a low unemployment rate. We can also speculate that households with higher income are better able to provide resources that students need to perform well in an academic environment.

**Critique of Tableau**

Tableau is a powerful and versatile tool. Although it took some time to become familiar with the interface, Tableau proved to be very flexible. We found it easy to create a wide range of visualizations and explore the data sets. In particular, it was easy to apply filters and use the “quick filter” option to vary the visualization and observe the data. In addition, it was very convenient to export visuals as image files as the application provided all necessary titles and labels.

We found it somewhat frustrating to link data into Tableau. We thought that the program should be able to join tables from different Excel files, given that the files have identical headings and format. Instead, we had to manipulate the data by hand to combine data sets into a single Excel file in multiple worksheets.