

Interactive statistical graphing and analysis of poverty, population, unemployment and education data of US states

Dinesh Poddaturi and Yue Zhang

1. Introduction

In 2010, the US Census Bureau's reported 46.2 Million people being identified as "in poverty" in the nation. Family in poverty is defined as an annual income below 24,000 dollars for a four people household. It is quite interesting that only few of the people in poverty will be considered poor by most of the Americans. For example, a typical poor family lives in an air-conditioned house or apartment with cable-TV, has a car and a multiple color TV. 96% of poor parents reported that their children were never hungry at any time in the prior year. While some of the poor face significant material hardship, the long-term anti-poverty policy should address the causes of poverty carefully, rather than solely base on symptoms of poverty.

In this project, we employed the data released by the US census Bureau in 2015 and studied several key indicators that may be closely related with poverty, including the education, unemployment rate and population distribution across all US states. Several data cleaning techniques in R were used and interactive graphing was adopted to visualize the analysis results.

2. Data and Methods

The data^[1] needed for the analysis are dispersed in 4 four Excel files, which contain information on poverty, population, unemployment rate and education, respectively. The excel files are available at <https://www.ers.usda.gov/data-products/county-level-data-sets/county-level-data-sets-download-data/>. Education data are available for the years 1970-2015, unemployment data for 2007-2015 years, population estimates for 2010-2016 and poverty estimates for 2015. Each file has about 3300 rows and 50 columns with a file size of 1.5 Mb.

"tidyverse" package is used to clean the data and extract the relevant data for analysis. Following are the steps taken to clean the data.

- First we select the variables of interest from each file.
- Use 'gsub' function to replace any special characters in the names of the variables of interest.
- Use 'gather' function to combine all the variables into a single variable. (In this project the single variable name is *Type-Year*)
- Use 'separate' function to split *Type-Year* variable into two different variables, *Type* and *Year*.
- Above steps makes our data into nice long format. This data is used to do analysis in the project.

To obtain data for each individual state, the column *Area_Name* in our data is matched with available states list in *state.name*. Once we get data for each state in a data frame, we combine the data by grouping *Type* to obtain the country level data. For trends the data is combined by grouping *Type* and *Year*. In short we used split-apply-combine strategy to manipulate tidy data for this project. All the above steps are written in different functions, making the code readable and debugging easier.

The results are present using interactive statistical graphing (ggplotly). The following indicators were selected: population, education level, unemployment and poverty (all at the state level). Except the population, all other indicators were transformed into ratios for visualization in order to deliver a meaningful result.

3. Results

Fig. 1 shows the layout of the interactive graph after running the app. On the left panel, there are three drop-down lists allowing the users to select. The first one is “indicator”, which includes the education, population, unemployment rate and poverty rate. Once the indicator is selected, the corresponding levels of the indicator will be automatically updated for selection. The third drop-down list is independent of the second one, which allows the user to check all the levels of indicator in a specific state. On the right panel, there are four tabs to visualize the results. The first tab shows the trends from 1970 to 2015, in this case it is the evolution of the education levels in Fig. 1, which is also an interactive graph. It can be seen that more and more of Americans received associate or college degree in the past 40 years. The second tab shows the nationwide statistics on maps using different color scales. The third tab demonstrates the top 10 states of the indicators the user selected

on a bar-plot, while the last tab allows the users to check the corresponding indicators in a certain state.

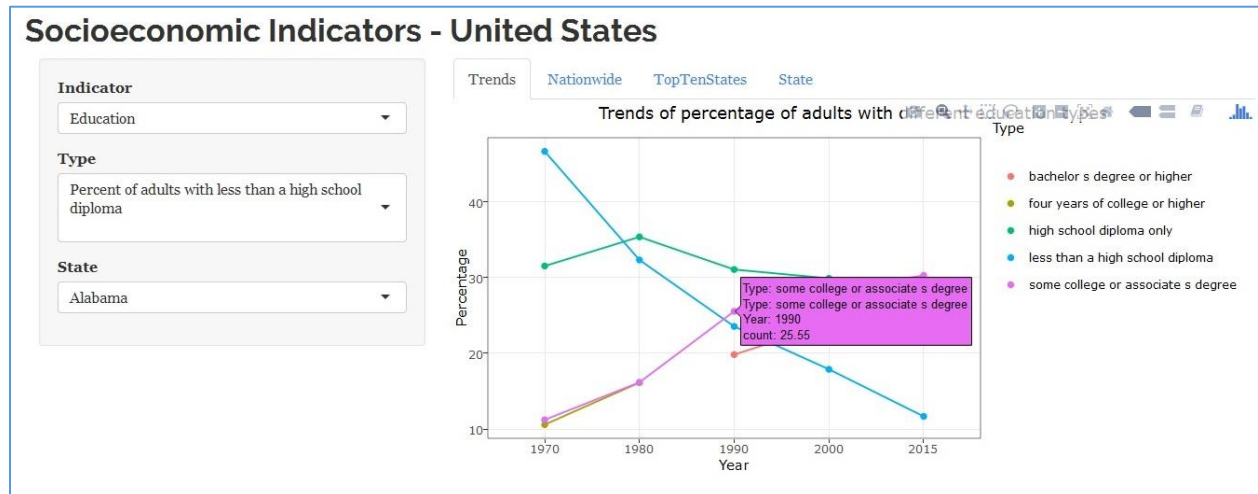


Fig. 1 The layout of the interactive graph after running the app. In this case it shows the evolution of the education levels during 1970-2015.

The birth and death rates from 2011 to 2015 in US are illustrated in Fig. 2. It can be seen that the birth rate is between 12-13% and slowly declines over the years, while the death rate is between 8% and 9%. Overall the country's population is growing steadily.

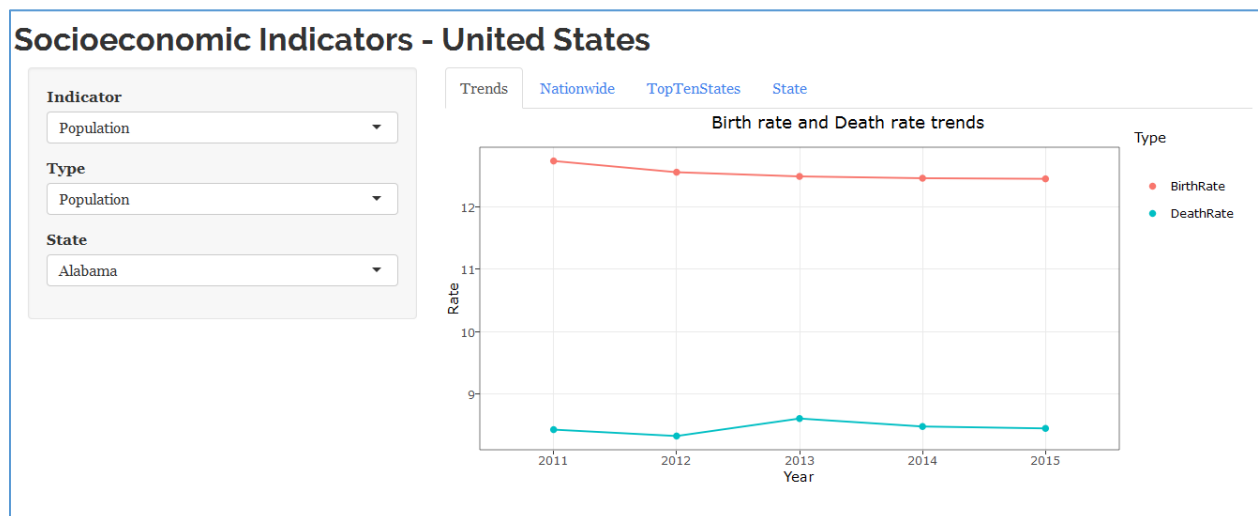


Fig.2 The birth and death rates from 2011 to 2015 in US.

The unemployment rate from 2007 to 2015, which is a very important macroeconomic matric, is shown in Fig. 3. It can be seen that the employment rate was around 4.5% in 2007, which is the lowest among the years. During the great recession in 2008 and 2009, the unemployment rate drastically increased to 8-9%. It decreased almost linearly these years to 5% in 2015, which is very close to the lever before the recession.

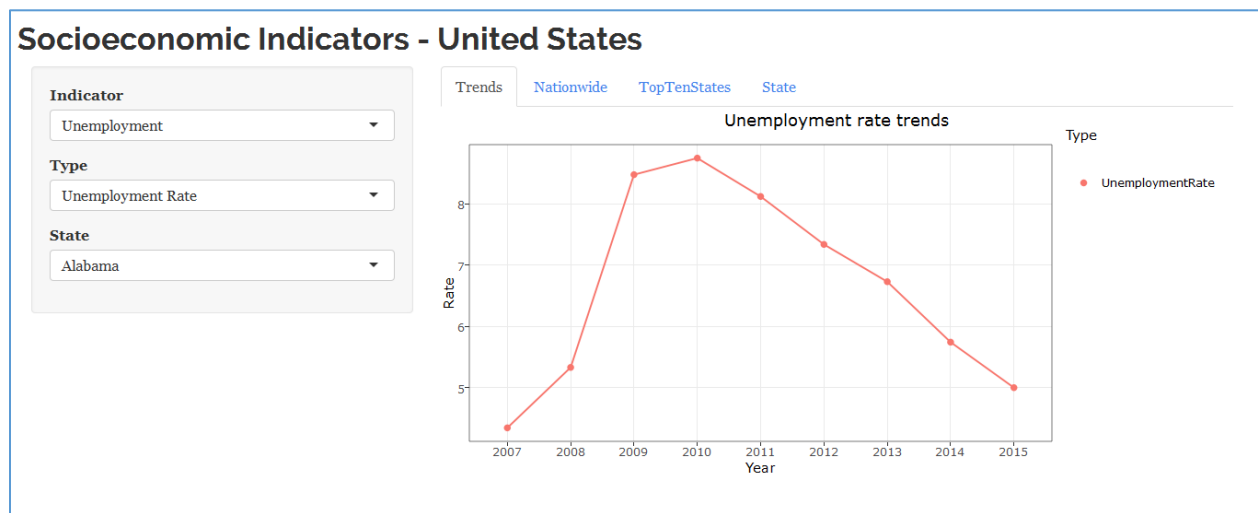


Fig. 3 The US unemployment rate from 2007 to 2015

After the nation-level analysis, we would like to check the results at the state level. First, the population of each US state in 2015 is shown on Fig. 4. The map shows that California, Texas and Florida are the states with the largest populations, while the majority of the Midwest or Mideast states have low populations. Next the unemployment in 2015 is illustrated on Fig. 5, it shows that the unemployment in New Mexico was quite high in 2015, which is about 6.6%.

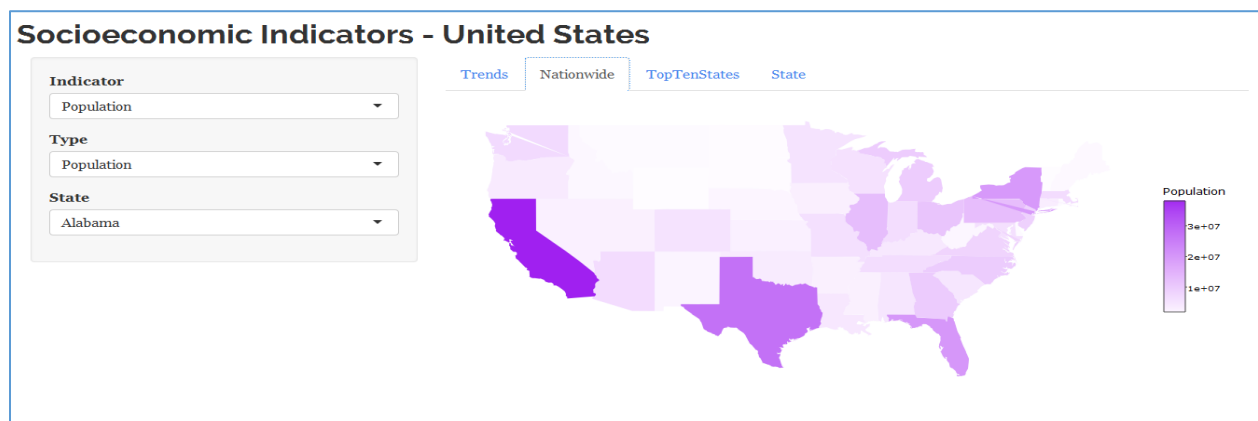


Fig. 4 The population of US states

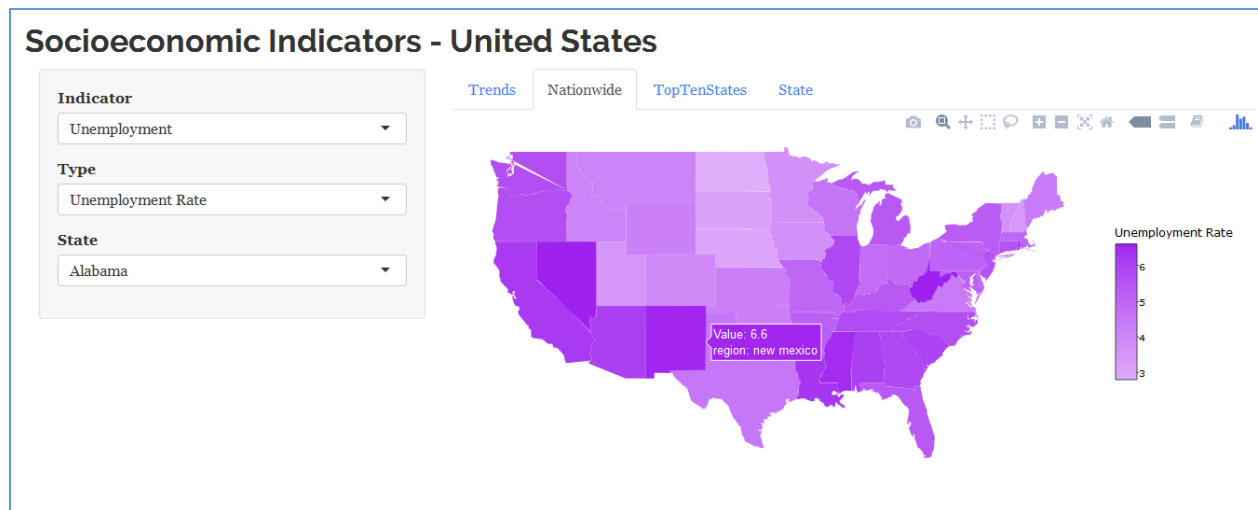


Fig. 5 The unemployment rates of US states

After examining the state-level data on the maps, we would like to find the top 10 states of each indication at all their levels. First the top 10 states with the highest percentage of adults with a bachelor's degree or higher in 2015 are shown on Fig. 6, it is found that adults in Massachusetts, Colorado and Maryland received higher education in average. However, the most populated states such as California, Texas and Florida are not on this list.

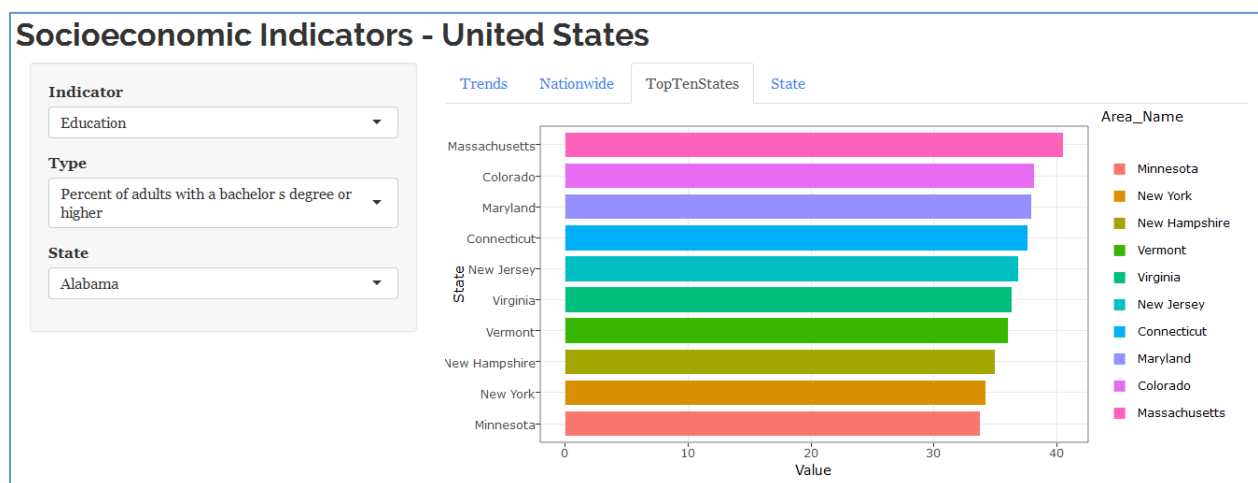


Fig. 6 The top 10 states with the most adults having a bachelor's degree or higher in 2015

The top 10 states with the highest unemployment rates in 2015 are demonstrated on Fig. 7. The top three are West Virginia, Nevada and New Mexico, followed by Mississippi and Alaska. California was ranked No. 7.

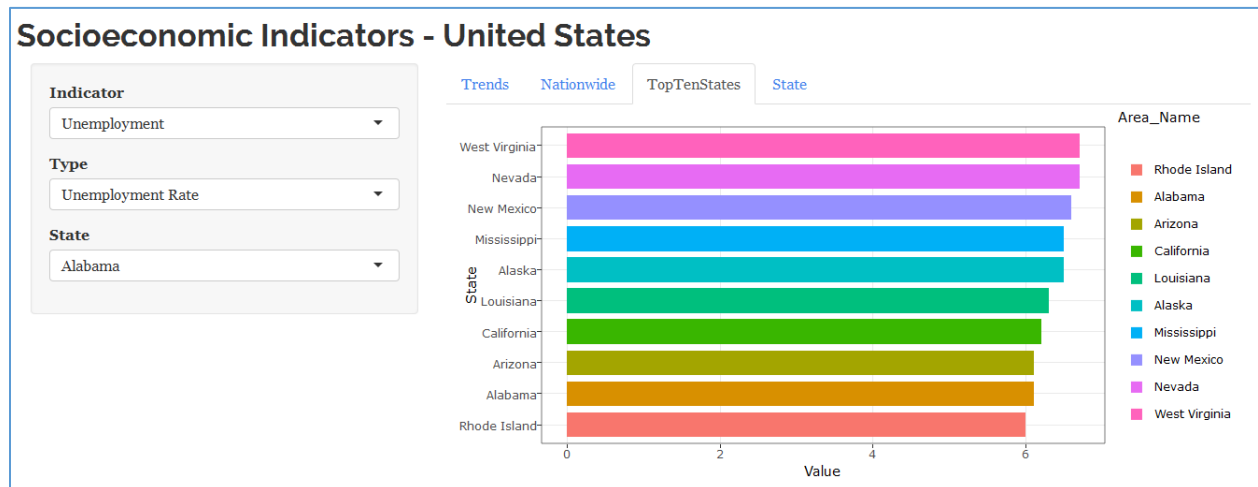


Fig. 7 The top 10 states with highest unemployment rate in 2015

The top 10 states with the highest poverty rates in 2015 are demonstrated on Fig. 8. The top three are Mississippi, New Mexico and Louisiana, followed by Arkansas and Alabama. It is interesting that although the poverty is closely related with unemployment rate, California, which has the largest population and high unemployment, wasn't on this list. It seems that the relationship is very strong in the southern states, but not for the states on the east or west coasts. However, the states with the most of adults having college degree or higher, were all not on the poverty list, indicating that education is also a very important factor to poverty.

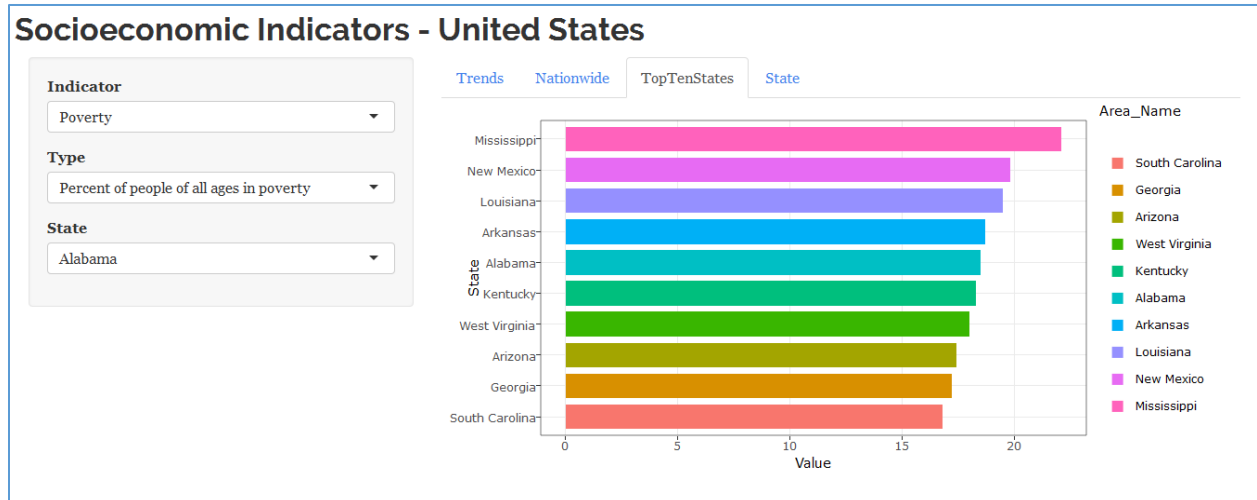


Fig. 8 The top 10 states with highest poverty rate in 2015

After the comparison of all US states, we would like to check the statistics of the indicators in an individual state. First the distribution of the poverty people in Mississippi was shown on Fig. 9. It can be seen that about 22% of people in this state was regarded as in poverty in 2015, which was quite high. Above 30% of the people below 17 years old were in poverty, making their future less promising.

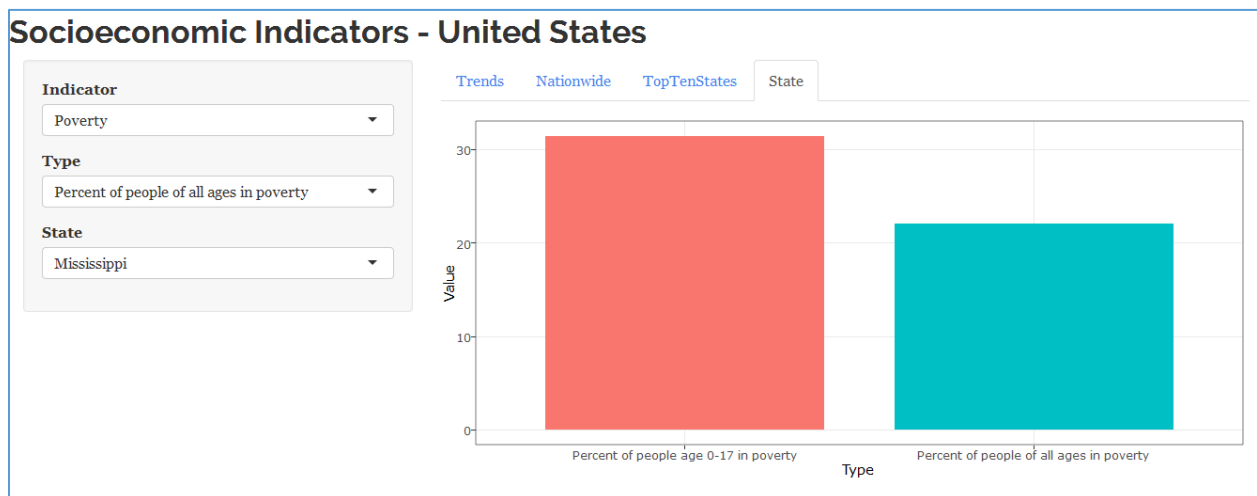


Fig. 9 The statistics of people in poverty in Mississippi in 2015

Finally, we go back to the home state, Iowa. The statistics of education in Iowa in 2015 was illustrated on Fig. 10. It shows that more than 50% of the adults in Iowa have achieved some college degree, bachelor's degree or higher in 2015.

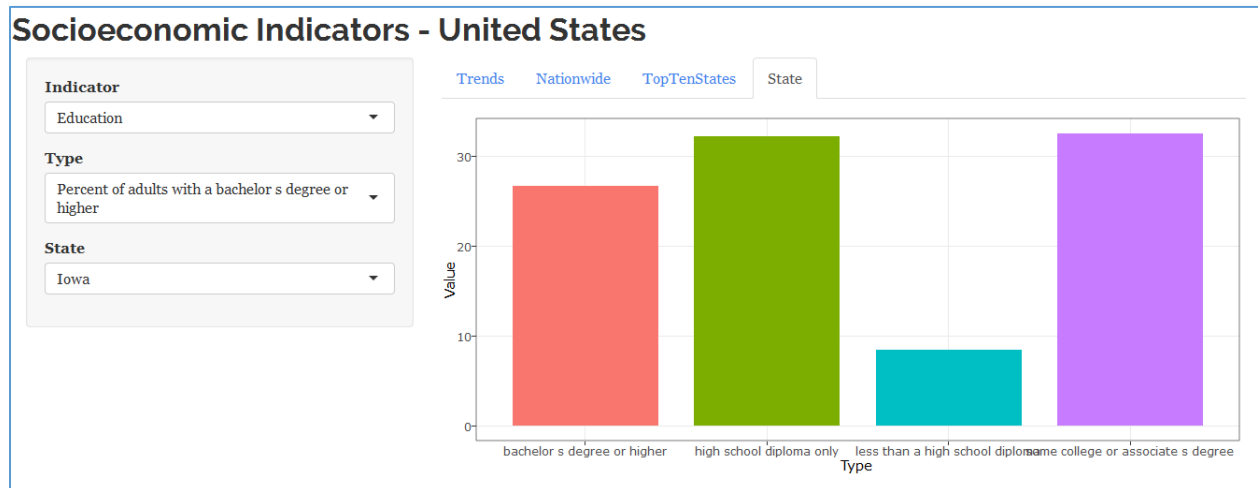


Fig. 10 The statistics of education achievements in Iowa in 2015

4. Discussion and Conclusion

From the results we can observe that the poverty estimates are highly correlated with education and unemployment rate. Population doesn't seem significant in explaining the poverty rate. However, we cannot get into the proper conclusions without any numerical analysis. One option is to fit a multiple linear regression with education, unemployment rate, and population as explanatory variables and see which one is more significant explaining the poverty rate in each state and overall nationwide. Once we know the significant variables we can predict the poverty rate of different states for subsequent years.

References

[1] <https://www.census.gov/library/publications/2016/demo/p60-256.html>