Introduction

Education continues to play a critical role in both individual success and the economic and political health of today’s societies. In the United States enrollment in all degrees, as Fig. 1 shows, has grown over the past two decades. However, the direct impact of such increase on the well-being indexes of individuals and the society as a whole remains in question. In this study, by looking into the statistical data of US education at various levels and by different demographic factors such as gender and geographical location, we have tried to find the potential or significant trends in educational attainment as well as its relationship with other important social and economical factors such as employment rates, income, GDP, poverty, and crime rates.

Figure 1: School Enrollment in All levels (1,000), With Projections: 1965 to 2016
Dataset

All datasets were extracted from the 2008 Statistical Abstract ¹ in US Census Bureau ², which is the authoritative and comprehensive summary of statistics on the social, political, and economic organization of the United States.

The education section in the US Census Bureau presents data primarily concerning formal education as a whole, at various levels, and for public and private schools. This data includes school enrollment, educational attainment, education personnel, and financial aspects of education.

For the scope of this project we decided to gather two types of data, as described in below, to explore with Tableau and HCE, respectively,

- Temporal datasets with concentration on degrees and income earned at different educational levels by gender.
- Multi-dimensional data including different educational, economical, and social attributes by state.

Moreover, our multi-dimensional dataset contains following attributes:

- Educational Attainment-high school grad or more: 2006
- Educational Attainment-BS grad or more: 2006
- Enrollment in degree Granting Institutions: 2005 (1000)
- No of degree Granting Institutions: 2005 (1000)
- State and Local Financial Support for Postsecondary Education (per full time): 2006($)
- Employed/pouplation ratio: 2005
- Household income: 2005-2007 median($)
- Daily and Sunday Newspapers-Number-2006
- Population: 2006 (1000)
- GDP- 2006 (million dollars)
- Patents total- 2006
- Crime Rates per 100,000 population-2004-2005
- Hate crimes incidents reported- 2005 (1000)
- Fraud Consumer Complaints- 2006

¹http://www.census.gov/compendia/statab/index.html
²http://www.census.gov/
Interesting findings

The Crossover in Female-Male Degree Earning Rates

College attendance in US has grown over the past few decades. However, increase in total demand for higher education has not been the same for males and females. We used Tableau to visualize the degree earning trends by gender and in the course of history. Figure 2 shows the percentages of men and women who earned Bachelor’s, Master’s, and Doctoral degrees since 1970. The horizontal axis shows the temporal data and the vertical axis represents the percentages of men and women who conferred their degrees with respect to the total number of degree holders in each year. As the figure shows, since 1986, the proportion of women who earned Bachelor’s and Master’s has exceeded the degree earning rates for men, and the gap has widened over time. Moreover, women are catching up with men in earning Doctoral degree. This represents a major shift in the gender balance at U.S. colleges and universities. Several reasons have been cited for this phenomenon, including gender differences in academic achievement (girls do better in high school than boys), changes in societal values, and a shift in womens expectations for future employment.

Figure 2: Percentage of earned degrees by sex, 1970 to 2005
Women Lead in Earning Degrees but not Money

Considering the increase in the rate of educational attainment at all levels among women, now the question is if the same upward trend exists or is projected on their earnings. To verify this, we used Tableau to visualize the earning of men and women in different educational levels. Moreover, we explored the trends in annual earning by gender since 1947. Figure 3.1 shows the mean earning of both genders by educational level. As can be seen, while the incomes of both men and women increased significantly with higher educational attainment (the increase becoming larger with each level), the income gap between genders remained at each educational level. Fig. 3.2 shows that this gap also becomes larger with each level. This finding indicates that womens earnings, relative to those of men, have not kept up with their gains in educational attainment.

![Figure 3: (1) Mean Earnings of Full-Time Year-Round Workers in Current Dollars, 2005 (2) Difference in income between genders by educational level](image-url)
Figure 4 visualizes the temporal data regarding the median income by sex since 1947. We can see that since 1970 the income gap between male and female workers has decreased but remains relatively large. Part of this difference reflects the higher concentration of men in higher-paying fields, including the engineering, computer and information sciences. Even at the college level, it is observed that fewer women than men take courses in science-related fields. The U.S. economy can benefit greatly from women’s educational gains, but only if women are working in occupations that can use their specialized knowledge and skills.

![Figure 4: Median income in constant (2005) dollars by gender.](image)

**Education and Socio-economic indexes: how strongly related?**

To explore how the education statistics varies across the different geographical locations and if/how they relate to some other main social and economical characteristics such as median household income, poverty, employment, crime, and population, we used HCE3.5 (Hierarchical Clustering Explorer) to visualize and cluster a multidimensional data set containing 17 attributes across all 51 states.

Different linkage method and similarity measurements were tried for clustering the data. By some minor differences, the linkage method of Shneiderman’s 1 by 1 and similarity measurement of Pearson Correlation Coefficient seemed to give the most reasonable clustering results. It was amazing to see how the color coding and reordering of the attributes by itself were helpful in tracking and digesting some interesting patterns within different clusters as follows: The states with highest poverty levels and lowest educational attainment level were clustered together. As for states in the most left side cluster of the dendrogram view [Fig. 5] such as New Hampshire and Vermont were the ones with highest educational attainment and lowest crime and poverty rates while their median household income stays low or average in the nation. On the other hand, big and populated states like California, New York, and Florida with highest amount of net worth of wealth in them are in a cluster of their own with their educational attainment level (high school and up) below the average and their employment and median household income only in the mediocre range.

Next the histogram (1D) and scatter plot(2D) orderings /rankings of all attributes were examined. Among some interesting patterns, we could point out that the educational attainment of high school degree or more strongly correlates with the ratio of employment per population with a high score.
of 0.965 [Fig. 6]. On the same token, the rate of college education attainment of BS degree or more is in strong positive correlation with median household income in each state.

In the course of examining relationships between each pair of attributes of our data (136 combinations), as a side note, it was inevitable to notice while the strongest correlation of poverty is with the crime rates in each location, the weakest link among all pairs of attributes belongs to poverty and the median household income. This implies why the indexes by which the wealthiest states are measured and published each year need to be revised!

**Tool Critique**

**Tableau**

**Pros**

- Having the capability of dragging and dropping the attributes over row and column makes Tableau simple and user-friendly.
- Tableau makes it easy to edit the aliases in the figures which is very helpful.
- Tableau enables users to export their visualizations in the image format.
Figure 6: Correlation between educational attainment and employment/population ratio.

**Cons**

- The separation between measure and dimension in Tableau causes some difficulties, especially for time series data.
- We could not find a good description on data preparation for Tableau. There are some general explanations on data formats that Tableau can visualize but it does not show the appropriate data format for creating the visualization that you have in your mind.
- Tableau includes a limited number of visualizations and some of them just have limited applications.
- Tableau doesn’t export images in JPG format.

**HCE**

**Pros**

- HCE is a very powerful tool for analyzing multidimensional data.
• HCE has the capability of visualizing data in several formats and clustering them using various criteria.

• HCE enables users to investigate relationships among different attributes of data and find possible causes and effects. By this way, it helps users to analyze their datasets.

• The color coding in HCE is really helpful for extracting the correlations and highlighting the outliers.

Cons

• The clustering parameters need to be set every time the user tries to repeat the clustering function. No previous setting is kept by HCE.

• It would be helpful to be able to capture and export a detached/single larger view of histograms or scatter plots on ordering tabs.