

block builder webpage: <http://blocks.org/huangy22/raw/1a6a422a492dc57832e5c8e531079292/>  
data, js code: <http://blockbuilder.org/huangy22/1a6a422a492dc57832e5c8e531079292>  
python code: [https://github.com/huangy22/DataVisualization/tree/master/hw1/python\\_code](https://github.com/huangy22/DataVisualization/tree/master/hw1/python_code)

## Average/Median Monthly Income and Working Time in the Preceding Four Months by Education, 2009 (Table 2A, 2J)

This dataset includes the information of the average and median income level and working time in people with different education level at age 18 and over.

Data source: US Census Bureau, Survey of Income and Program Participation, 2008 Panel. For more information on sampling and non-sampling error see <http://www.sipp.census.gov/sipp/source.html>

### Statistics and Correlations

Before the visualization, I first use python to clean the data, transfer the numbers with “,” and “\$” to pure numbers, calculate the statistics, and correlations using pandas library. I also calculate the line of regression for some pairs of variables which I will plot later.

This table shows the statistics(mean, standard deviation, min, max, 25%, 50%, 75% quantile) of the numeric variables.

Statistics	Number	Income_mean	EarnFT_mean	EarnAll_mean	Months Worked	Income_median	EarnFT_median	EarnAll_median
mean	22624.9	3872.6	5158.8	4604.4	2.66	2869.1	3962.7	3478.1
std	17182.75	2816.89	3084.36	3045.61	0.57	1954.28	1989.96	2018.14
min	2486.0	1166.0	2136.0	1813.0	1.7	880.0	1732.0	1472.0
25%	12784.5	2021.0	3268.75	2670.5	2.38	1482.5	2641.75	2091.0
50%	18344.5	2700.5	3882.0	3253.5	2.8	2153.0	3203.0	2661.5
75%	32680.0	5237.75	6409.5	5940.75	3.08	4178.0	5151.5	4718.5
max	57880.0	9501.0	11927.0	11071.0	3.3	6125.0	7417.0	7128.0

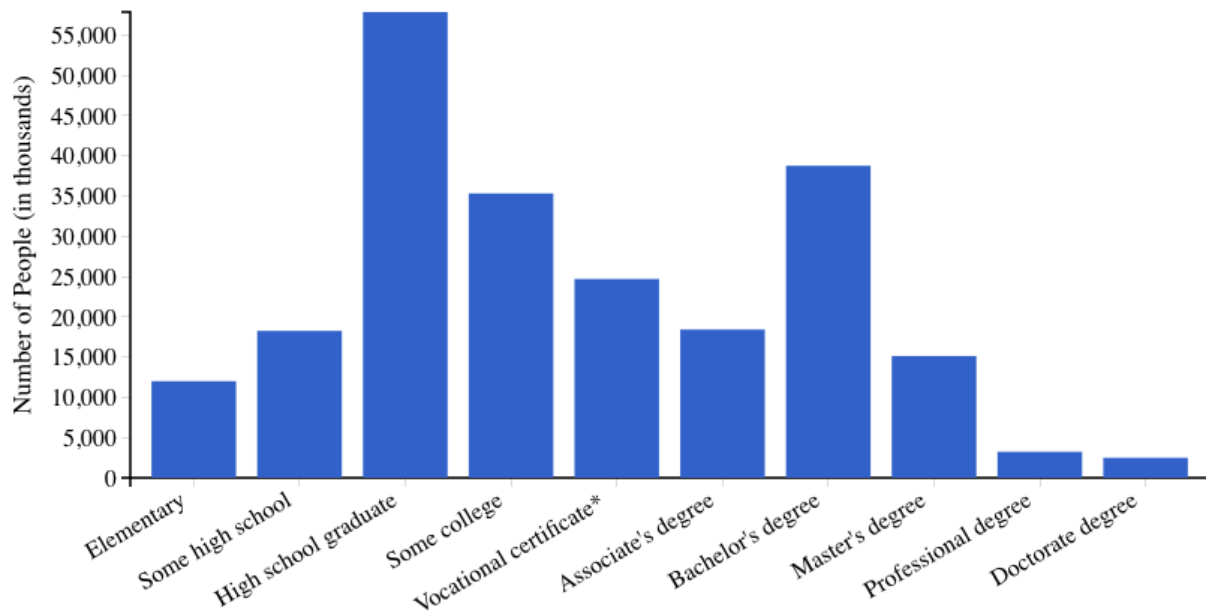
The following table shows the correlation between the numeric variables. The total income, full-time earning, all-job earning, and working time all have strong positive correlation with each other. While the number of people has a relatively weak negative correlation with the other variables.

Features	Number	Income_mean	EarnFT_mean	EarnAll_mean	Months Worked	Income_median	EarnFT_median	EarnAll_median
Number	1.0	-0.54085	-0.52249	-0.53394	-0.25122	-0.54757	-0.52048	-0.54031
Income_mean	-0.54085	1.0	0.99401	0.99774	0.82291	0.98727	0.9915	0.99538
EarnFT_mean	-0.52249	0.99401	1.0	0.99833	0.80269	0.96526	0.97472	0.97984
EarnAll_mean	-0.53394	0.99774	0.99833	1.0	0.80157	0.97505	0.98112	0.98766
Months Worked	-0.25122	0.82291	0.80269	0.80157	1.0	0.85265	0.86431	0.84152
Income_median	-0.54757	0.98727	0.96526	0.97505	0.85265	1.0	0.99686	0.99759
EarnFT_median	-0.52048	0.9915	0.97472	0.98112	0.86431	0.99686	1.0	0.99702
EarnAll_median	-0.54031	0.99538	0.97984	0.98766	0.84152	0.99759	0.99702	1.0

## Visualizations

### The number of people with different Education Attainment

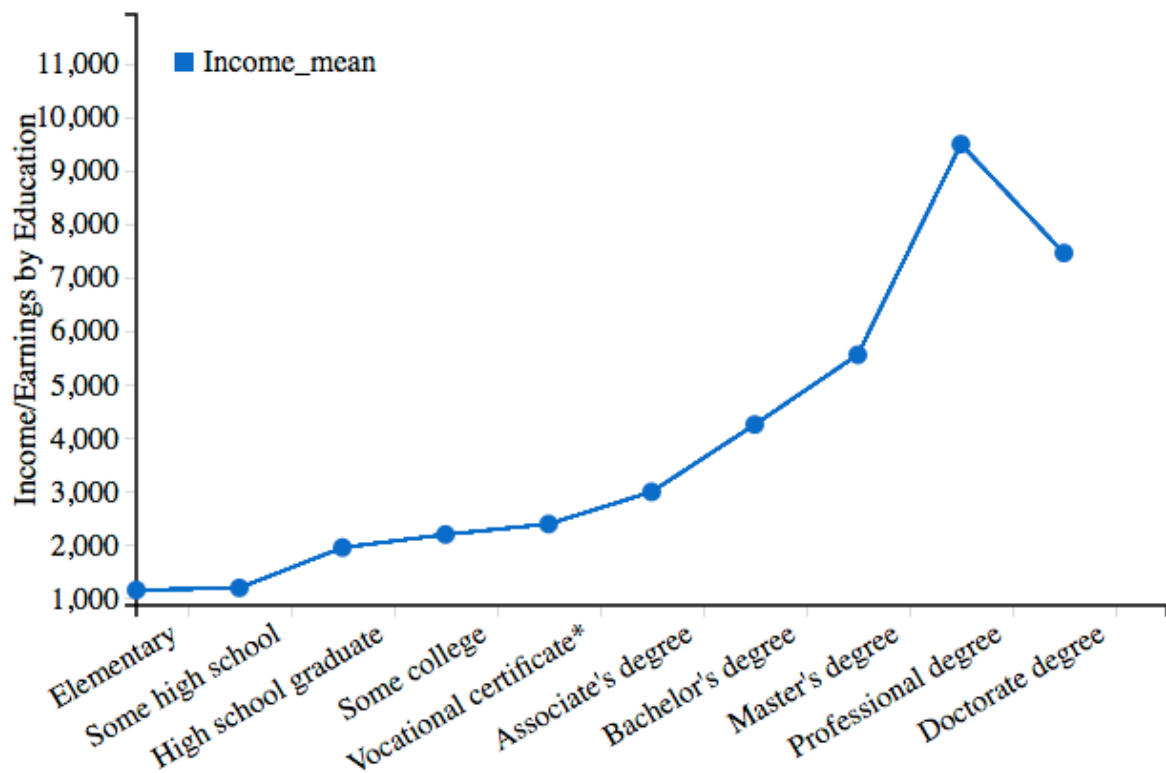
This plot shows the number of people (in thousands) with different education attainment in the study. The number of people is in thousands. From the figure, we can see that high school graduate has the highest population, while Bachelor degree is the second highest.



## Monthly Income, Full time job Earnings, Earnings for all jobs by Education (Interactive)

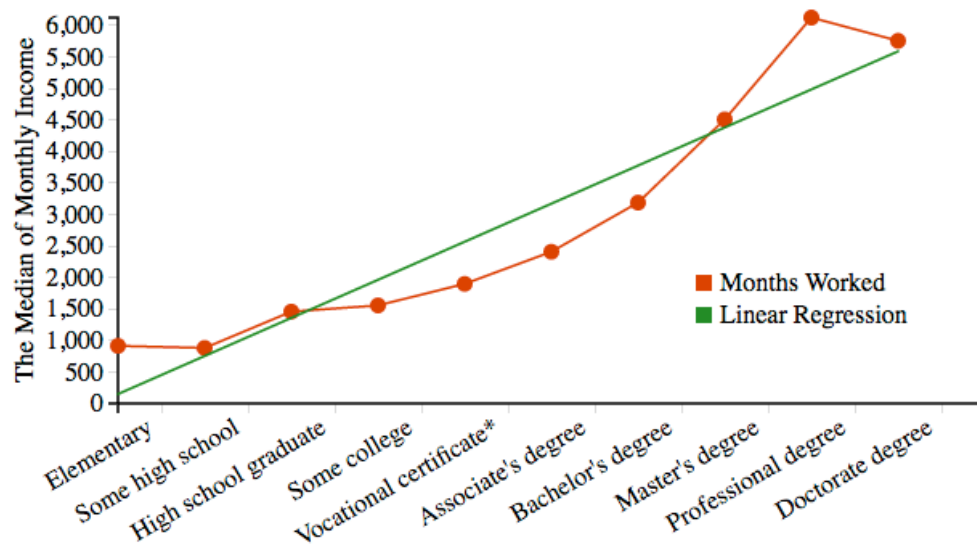
In the following plot, it shows the monthly income, full-time earnings, and all-job earnings with both average(mean) and median value. The monthly incomes and earnings are in US dollars. Earnings apply to employed population only, while Monthly income includes all kinds of incomes. Both the average value and median are shown in the following figures. The median value is generally lower than the average, indicating that there are some outliers with much higher incomes/earnings in the sample. Thus the median value is a more representative statistic than the average value.

Press the button to see the plot for the corresponding variable:



## The line of regression for Median Monthly Income by Education

The total income, full-time earning, and all-job earning have the same trend among different education levels. Since the three variables strongly correlate with each other, we can choose one variable, for example the total income and study its relation with education level. The following plot shows the linear regression of income as a function of education level, indicating a positive correlation between income and education.



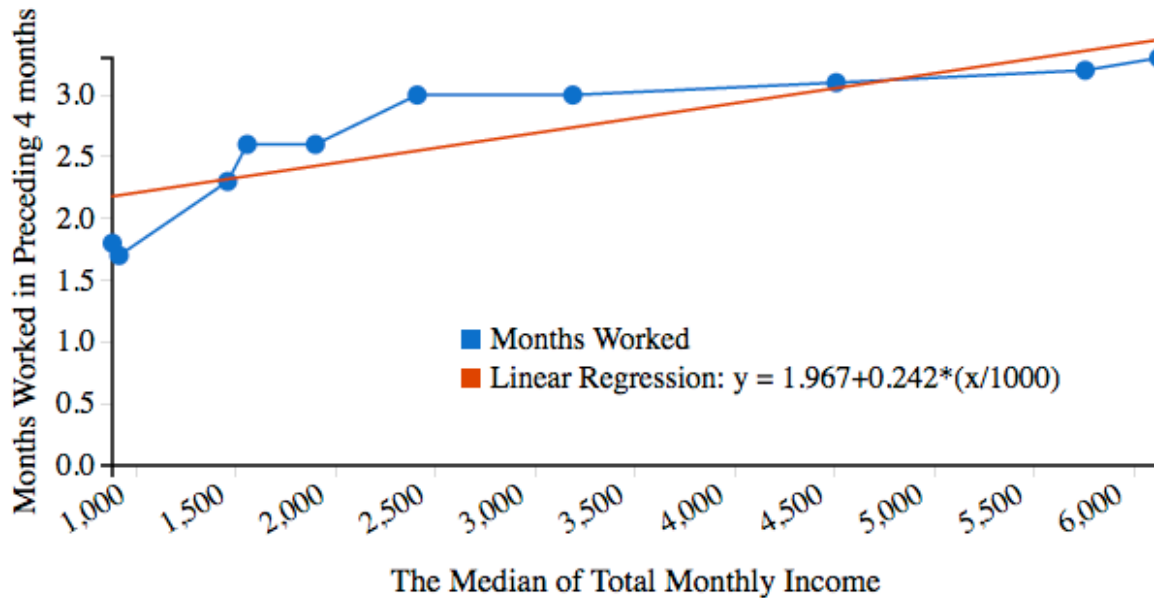
## Working time by Education

The following plot shows the average time of having a job in the last four months as a function of the education level. The plot also shows a linear regression of the curve, featuring a positive correlation as well.



## The line of regression for Months Worked by Income level

From the correlation table we already see that the months of having a job and the income level are positively correlated. Here we plot the months worked as a function of the median of monthly income and its linear regression. The result confirms the positive correlation observed from the correlation table.



## Educational Attainment in the United States, 2015 (Table 1)

This data set is from a study in 2015 of people at age 18 and over, including the data of their educational attainment, sex, age, race and hispanic origin. From this data set, we could learn about how people's educational status depend on their age, sex and race.

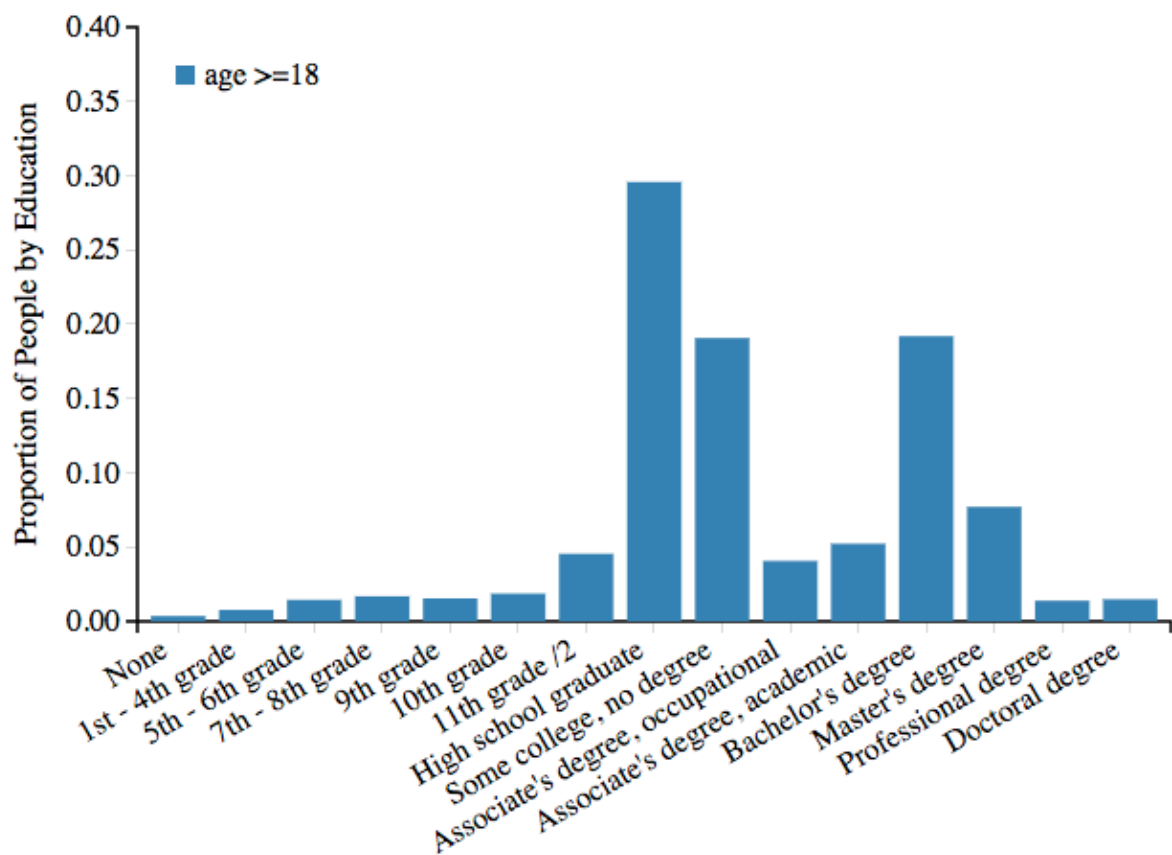
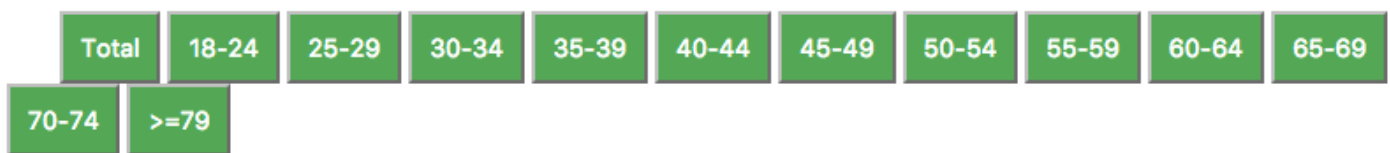
Before the visualization, I use python to clean the data, and combine the both sex, female, and male tables into one data frame (I use pandas library).

# Visualizations

## Education by Age (Interactive)

The following bar plot is the histograms of people's educational attainment at different age. The number of people has been normalized respect to the total number of people in the selected age range.

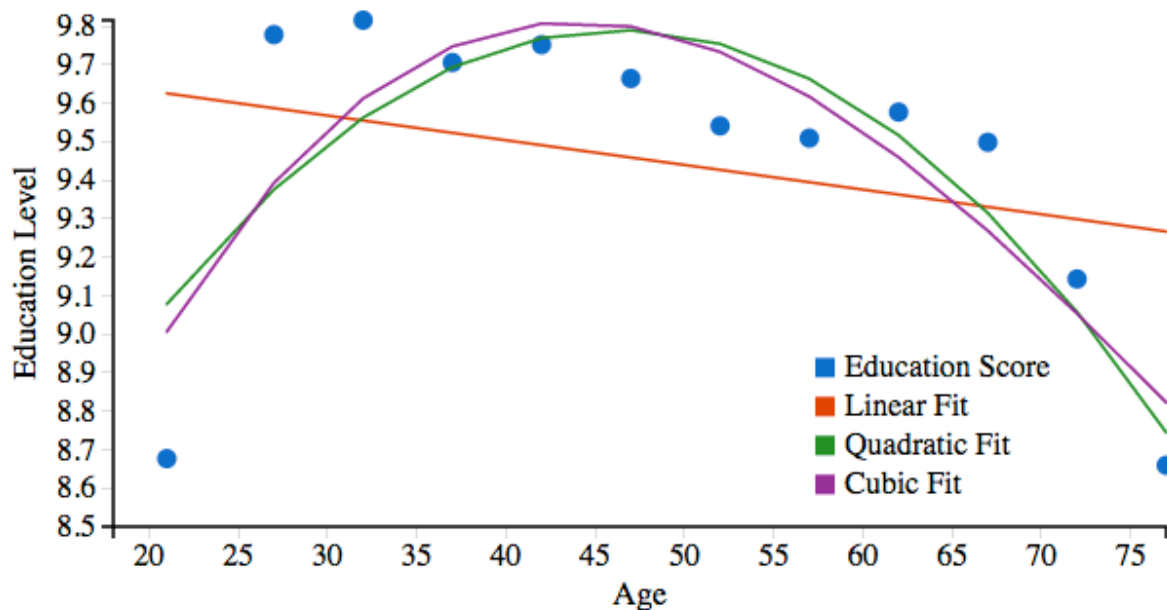
Press the button to see the plot for the corresponding age range:



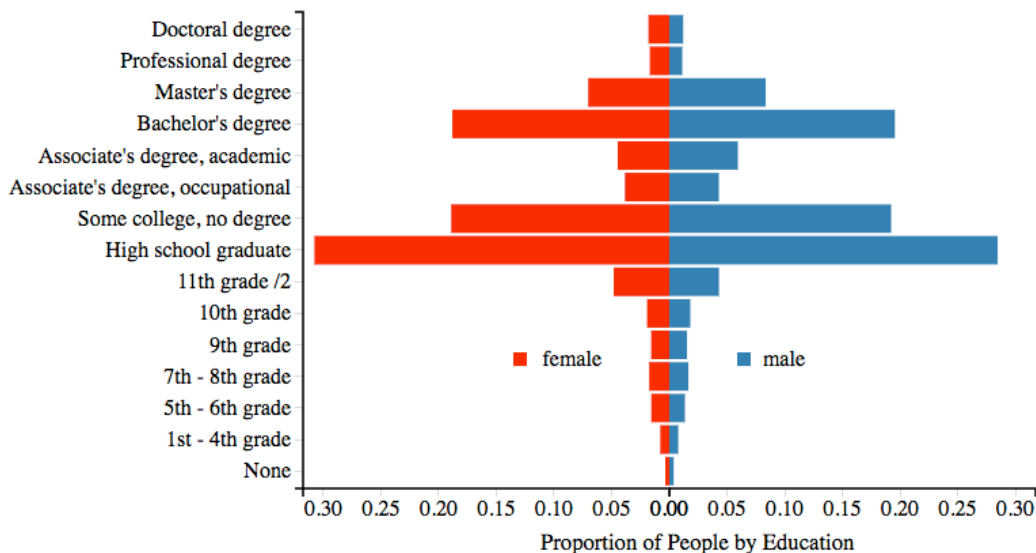
## The correlation between the Education Attainment and Age

From the above bar plots, we can clearly see that the education attainment of people at different age are changing with a pattern. In order to study the correlation in detail, I first change the education level to a numeric score from 1 to 15, and then calculate the average education score in different age range from the histogram. I also change the age range variable to a average age variable.

The following plot shows the educational score as a function of age, along with the lines of regression. From the scatterplot, we can see that the correlation is not linear, but rather a bell shape curve. Thus I tried linear, quadratic, and cubic fit for the data. The linear fit is clearly not representative for this data, while the quadratic and cubic fit behaves similarly and capture the overall shape of the data. The bell-shape curve can be understood. For people at age 18-30, some of them are still in the middle of education process, therefore the education level increases with age. While for people between age 30 and 60, most of them have already finished education. The change of education level with age is not significant. While for older people, at the time when they are at school, the educational resources are not as good as today, therefore the education level decreases with age.



The following is a bar plot showing the educational attainment of male and female at all ages over 18. The education level in female and male are nearly equal with a slightly higher rate of female with professional Degree and Doctoral degree. This is a quite interesting observation.



This is a grouped bar plot showing the educational attainment of different races at all ages over 18. It can be seen from the plot that asian and asian comb people have a much higher proportion with higher education level.

