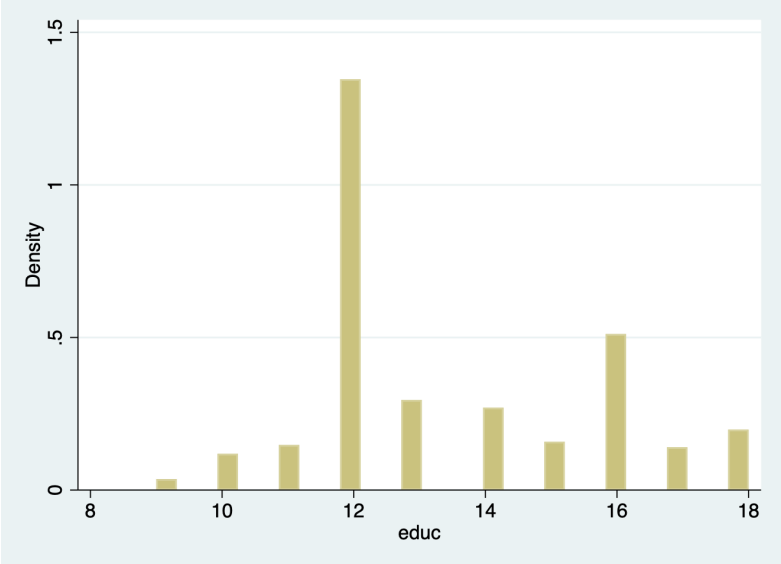
1.



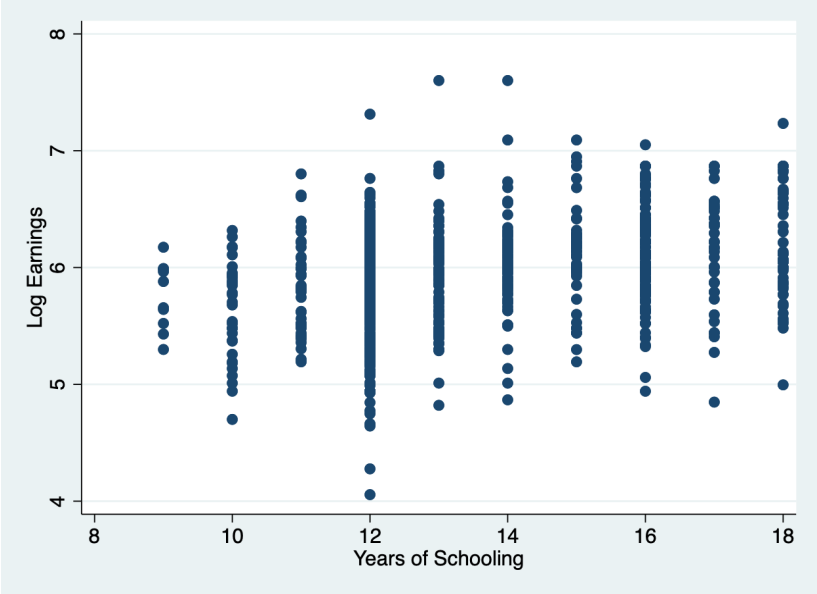
This histogram of the education variable shows that the distribution of the variable is not normal. The most likely values for education are 12 and 16. This can be explained by the natural end points for an educational career, with 12 years being the conclusion of high school and 16 years being the completion of undergraduate college. Educational careers shorter than 12 years are most uncommon, and there are conceivable explanations for the frequencies at 13,14,15,17,18, as students might complete degrees that are less than four years, drop out, or complete master's degrees.

2.

```
. ttest luwe,by(eduband) unequal

Two-sample t test with unequal variances

   Group |     Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]
---------+--------------------------------------------------------------------
       0 |     374    5.851872    .0216882    .4194297    5.809226    5.894519
       1 |     142    6.119436    .0325798    .3882332    6.055028    6.183844
---------+--------------------------------------------------------------------
Combined |     516    5.925504    .0188316    .4277716    5.888508      5.9625
---------+--------------------------------------------------------------------
    diff |           -.2675635    .0391385               -.3446146   -.1905123
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =  -6.8363
H0: diff = 0                     Satterthwaite's degrees of freedom =  273.364

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

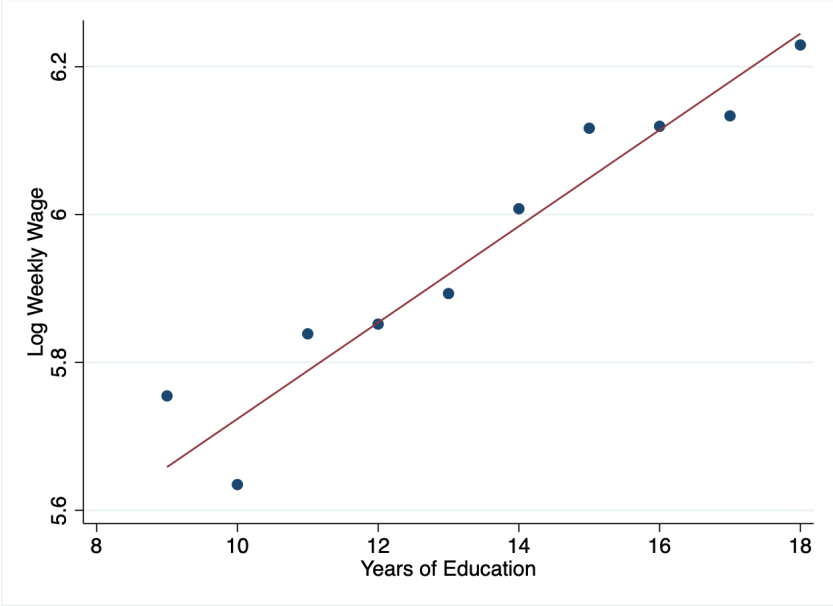
Eduband is a generated binary variable with a "1" value for 16 years of education and "0" for 12 years of education. The T-test above, in particular the 95% confidence interval and the p-values, shows that the difference between between exactly 12 and exactly 16 years of education is statistically significant.

3.



The scatter plot format and the density of the data make it hard to draw precise conclusions from this visualization. The widest range of earnings appears to be at year = 12 and looks like high school graduates tend to make more than those who did not graduate high school. There is not a clear and consistent pattern of wage increase as years of school increases.

4.



The binned scatter plot makes it much easier to draw preliminary conclusions from the data. By binning the values, we address the lack of clarity of the traditional scatter plot. With the traditional plot, a high density of values made it nearly impossible to tell how many points were at a given value. There could be five or ten or one thousand points at a given value, and there would be no way to tell from looking at the scatter plot. The binned scatter plot shows positive

correlation between years of education and wage increases. While the increase does not appear to be completely linear, the upward trend is clear.

5.

```
. regress luwe educ
```

Source	SS	df	MS	Number of obs	=	895
				F(1, 893)	=	103.69
Model	18.3402744	1	18.3402744	Prob > F	=	0.0000
Residual	157.9476	893	.176873012	R-squared	=	0.1040
				Adj R-squared	=	0.1030
Total	176.287874	894	.197190016	Root MSE	=	.42056

luwe	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
educ	.0651481	.0063978	10.18	0.000	.0525916	.0777045
_cons	5.072126	.0873479	58.07	0.000	4.900695	5.243557

Based on this regression, the log weekly wage for someone with no education would be approximately 5.07, with a standard error of .087 and this value is significant with a p-value of 0.000. Each yearly increase in education would lead to a .065 increase in log weekly wages, with a standard error of .006. With a p-value of 0.000, this is significant. This confirms the trend that we observed in the binned scatter plot.

6.

With the binary variable that I generated(eduband), I get the mean for 12 years and for 16 years in the output of the T-test. The difference between them is (.2675635). The results of the regression show that each year of education leads to an increase in salary of (.0651). If I multiply the coefficient by 4, I get .2604, which is relatively close to the value from the T-test, when considering the Standard Error of the difference.

7.

While education has a number of positive effects, a commonly-cited reason to become educated is to increase earning potential. This NLS data allows us to draw some preliminary conclusions about the relationship between earnings and educational attainment. Preliminary visualizations of the data show that the majority of subjects either completed high school or completed college, with the rest of subjects distributed between 9 and 18 years of education. This distribution is not normal, which makes sense considering the natural end points of one's education experience, after high school or after undergraduate college. A binned scatter plot shows a positive correlation between education and log earnings, but it does not appear to be linear. Because the data are not linear, using linear regression may be problematic. Additionally, heteroskedasticity undermines the accuracy of estimates drawn from linear regression, and the variance of log wages across years of education shown in the initial scatter plot appears to be heteroskedastic.

Thinking about these results from a policy perspective, I would be cautious about making too many decisions based on the analysis done above. One possible option would be to focus on those subjects who did not graduate high school. Because these students are likely part of

public school systems, it might be easier to apply a multi-year intervention. Essentially, it would be an intervention designed to increase high school graduation rates.

The results of this analysis also present questions worth exploring. The plateaus of earnings around years 12 and 16 may be worth considering. A possible explanation for these plateaus might be that earnings are based upon skill sets rather than credentials. Consider a subject with 11 years of education considering a manual labor job. In most cases a 17 year-old high school junior is as capable on a farm, in a factory, or in a trade as an 18 year old. Whether or not they've taken an additional set of high school courses isn't particularly relevant. The plateau between 15 and 17 years of education is harder to explain, but it is possible that associate's degrees or other training programs allow subjects to earn at the rate of new college graduates.

With this analysis, I would be comfortable piloting interventions to increase high school graduation rates, and also be interested in designing studies to get a better sense of what causes students to terminate their education at points other than 12 and 16 years.