Assignment-3

1. Estimate an OLS model that estimates the effect of years of schooling on crime (as measured by being in prison). Include state and year fixed effects (i.e. as.factor(state), etc. . .).

Table 1: OLS Model

	Dependent variable:
	Prison Sentence
Education	-0.001^{***}
	(0.00001)
Observations	3,610,667
R^2	0.003
Adjusted R^2	0.003
Residual Std. Error	0.081 (df = 3610613)
F Statistic	219.944^{***} (df = 53; 3610613)
Note:	*p<0.1; **p<0.05; ***p<0.01

2. Explain why you cannot interpret the coefficient on educ as the causal effect of education on crime.

It may be that the effects of education on crime rates differs by state and/or year. It would be incorrect to assume that the interpretation above definitely applies equally to all states.

3. Explain why the compulsory years of schooling is plausibly a good instrument to solve this problem. Discuss this with reference to the inclusion and exclusion conditions.

"Years of compulsory schooling" could very well be part of the error term which effects this exact issue. it both causes education, and it may be the case that states which have the kind of political climate to enforce schooling, also have systematically different laws for sentencing. It is also the case that "Years of compulsory schooling" probably has no correlation with the error term (given its exclusion from said error term).

4. Estimate an instrumental variables model that estimates the effect of years of schooling on crime. Interpret the results.

See Table 2.

Interpretation: Education *does* have an effect on crime. The effect is inversely proportional.

5. Provide a plot and a regression estimation that establishes that the instruments satisfy the inclusion condition.

I generally think it's probably for the best to consider each group of compulsory education, 4 years, 6 years, etc. . . to be a categorical variable. This is because, after reviewing the data, each group seems to act differently, with no clear pattern between them, but they all have a statistically significant non-minor effect on education

See Figure 1. and Table 3. (Only the first 100,000 observations are used in generating the plot, otherwise it drastically bloats the pdf file size and load times. It's insane to me that's how ggplot works, instead of just baking a png or svg, but it is what it is. To see the "true" figure one, compile this document removing the line "`head(100000) |>`" in the Figure 1. block.)

Table 2: Instrumental Variable Model

	Dependent variable:
	Prison Sentence
Education	-0.002^{***}
	(0.0003)
Observations	3,610,667
R^2	0.001
Adjusted R^2	0.001
Residual Std. Error	0.081 (df = 3610613)
Note:	*p<0.1; **p<0.05; ***p<0.01

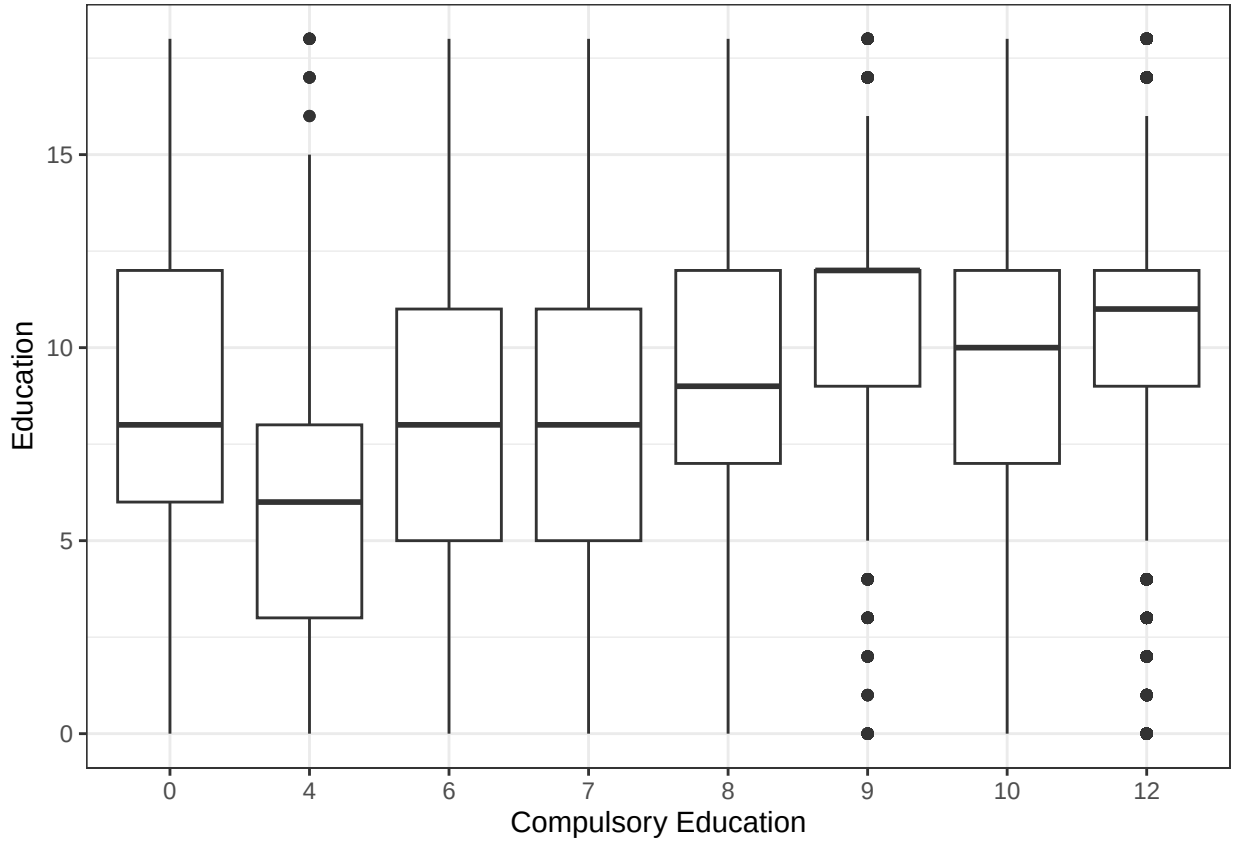

Figure 1: Years of Education by Compulsory Education

Table 3: Demonstration of Correlation

	Dependent variable:
	Education
Compulsory Education: 4 Years	-2.533^{***}
	(0.061)
Compulsory Education: 6 Years	-0.433^{***}
	(0.008)
Compulsory Education: 7 Years	-1.817^{***}
	(0.010)
Compulsory Education: 8 Years	-0.116^{***}
	(0.005)
Compulsory Education: 9 Years	-0.072^{***}
	(0.010)
Compulsory Education: 10 Years	-0.565^{***}
	(0.024)
Compulsory Education: 12 Years	0.525^{***}
	(0.005)
Observations	3,610,667
R^2	0.087
Adjusted R^2	0.087
Residual Std. Error	3.125 (df = 3610607)
F Statistic	$5{,}809.323^{***}$ (df = 59; 3610607)
Note:	*p<0.1; **p<0.05; ***p<0.01

6. Do an over-identification test and interpret the results.

Table 4: Over-Identification Test

	Dependent variable:
	Prison Sentence
Education	-0.001^{***}
	(0.00001)
Observations	3,610,667
R^2	0.002
Adjusted R^2	0.002
F Statistic	$8{,}715.844^{***}$ (df = 1; 3610613)
Note:	*p<0.1; **p<0.05; ***p<0.01