Table 1:

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
logwage	1,309	1.622	0.389	0.005	1.354	1.935	2.261
hgc	1,676	12.987	2.502	0.000	12.000	15.000	18.000
tenure	1,667	5.947	5.508	0.000	1.583	9.333	25.917
age	1,678	39.132	3.046	34	36	41	46

560 missing value in the logwage. the missing rate is 560/2229=0.2512. it should be MNAR.

the Bata1 for in complete case 0.735, when we use mean to impute missing value in logwage, the Bata1 we got is 0.869446. when we use the first regression to impute NA in logwage. the bata1 we got is 0.833384. the bata1 when we use mice package are 0.735965, 0.689307, 0.726557, 0.6561954, and 0.692572. to compare all the impute above, I feel that using the regression to impute na is the most appropriate way to deal with missing value. because, the regression we got is related with the data set.

the R package I am going to use is called ttbbeer. in this package I could find the Alcohol Tax and Trade Bureau (TTB) collects data and reports on monthly beer industry production and operations. Ideally, in my project I can general data for alcohol consumption in the past and the consumption after the new policy. Also, I will compare tax and consumption with other states to predict the income for government from new tax low.

Table 2:

	Dependent variable:		
	logwage		
	(1)	(2)	(3)
hgc	0.062***	0.049***	0.051***
	(0.006)	(0.004)	(0.005)
collegenot college grad	0.117***	0.160***	0.131***
	(0.039)	(0.026)	(0.031)
tenure	0.023***	0.015***	0.016***
	(0.002)	(0.001)	(0.001)
age	−0.003	−0.001	−0.002
	(0.003)	(0.002)	(0.003)
marriedsingle	−0.023	−0.029**	−0.030*
	(0.020)	(0.014)	(0.017)
Constant	0.735***	0.833***	0.869***
	(0.165)	(0.115)	(0.139)
Observations	1,296	2,229	1,662
R^2	0.204	0.132	0.150
Adjusted R^2	0.201	0.130	0.148
Residual Std. Error	0.347 (df = 1290)	0.311 (df = 2223)	0.327 (df = 1656)
F Statistic	66.264*** (df = 5; 1290)	67.496*** (df = 5; 2223)	58.546*** (df = 5; 1656)
Note:			*p<0.1; **p<0.05; ***p<0.01