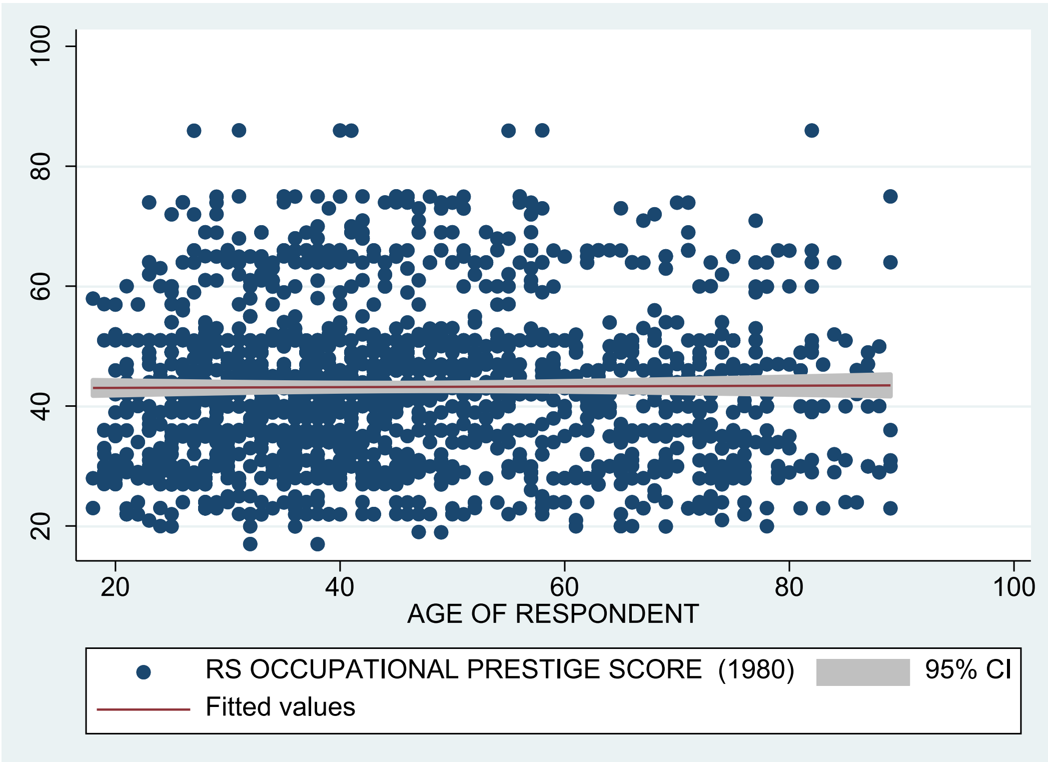
Amanda Conley
Soc 211
Problem Set 4

1.

A and B.



C. Well. I think you could argue that there is heteroskedasticity here. As age increases, we see the confidence interval widening, as the scores increase in variance, or distance, from the line of best fit. When you look at the far left of the line, you see the same thing as age decreases, though not as intense. However, this variance doesn't necessarily take on a bowtie shape, though the confidence interval seems to, you can see that as age decreases, between 18-20, occupational prestige scores are closely concentrated. The opposite occurs as age increases at the far end of the line of best fit. The variance is not randomly distributed, heteroskedasticity.

2.

A.

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

         Ho: Constant variance

         Variables: fitted values of prestg80

         chi2(1)      =     0.75

         Prob > chi2  =   0.3853
```

Amanda Conley
Soc 211
Problem Set 4

B.

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

         Ho: Constant variance

         Variables: fitted values of prestg80_log

         chi2(1)      =     0.84

         Prob > chi2  =   0.3597
```

I don't think it's appropriate to log occupational prestige, but I wanted to try it out with this test. No, logging it only increased the chi-square score, and we cannot reject the null hypothesis of homoskedasticity.

3. I am not able to get the code to work in Stata--I've played with it but Stata is consistently giving me the error: "variable random already defined." I'll follow up and see what I need to do to fix this in Stata and then come back to it and incorporate the results in questions four and five.

4.

A.

```
quietly reg prestg80 age, robust

quietly eststo
```

B. (I need to fix the bootstrap model code before I can do this completely)

Table 1: Multiple Regression Model

	OLS Model	Robust Model
AGE OF RESPONDENT	0.008	0.008
	(0.020)	(0.020)
Constant	42.850***	42.850***
	(0.975)	(0.960)
R^2	0.000	0.000
Adjusted R^2	-0.001	-0.001
BIC	1.2e+04	1.2e+04
Observations	1526	1526

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

C. Well. Using the robust command with the regression slightly reduced the standard error. I think if there was greater heteroskedasticity in the conventional OLS model, then I would have potentially seen a larger reduction.

5.

A. I don't think that given my variables, it is necessary to attempt a regression cluster analysis because I am not attempting to understanding occupational prestige at a higher, connected level. There is no cluster in this dataset, and I have no real theory or basis for doing it. But I am going to try it anyway to see what I get if I use education as the basis of the cluster, and to try to understand the code in Stata.

B.

Table 1: Multiple Regression Model

	OLS Model	Robust Model	Cluster Model
AGE OF RESPONDENT	0.008	0.008	0.008
	(0.020)	(0.020)	(0.039)
Constant	42.850***	42.850***	42.850***
	(0.975)	(0.960)	(2.997)
R^2	0.000	0.000	0.000
Adjusted R^2	-0.001	-0.001	-0.001
BIC	1.2e+04	1.2e+04	1.2e+04
Observations	1526	1526	1526

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

C. Given what I see here, the best model would be the robust model, as it has reduced the standard error. However, I don't think they differ that much from the conventional OLS model. While I think you could argue that there is a bit of heteroskedasticity in the model, I think you could use conventional OLS without restricting assumptions.