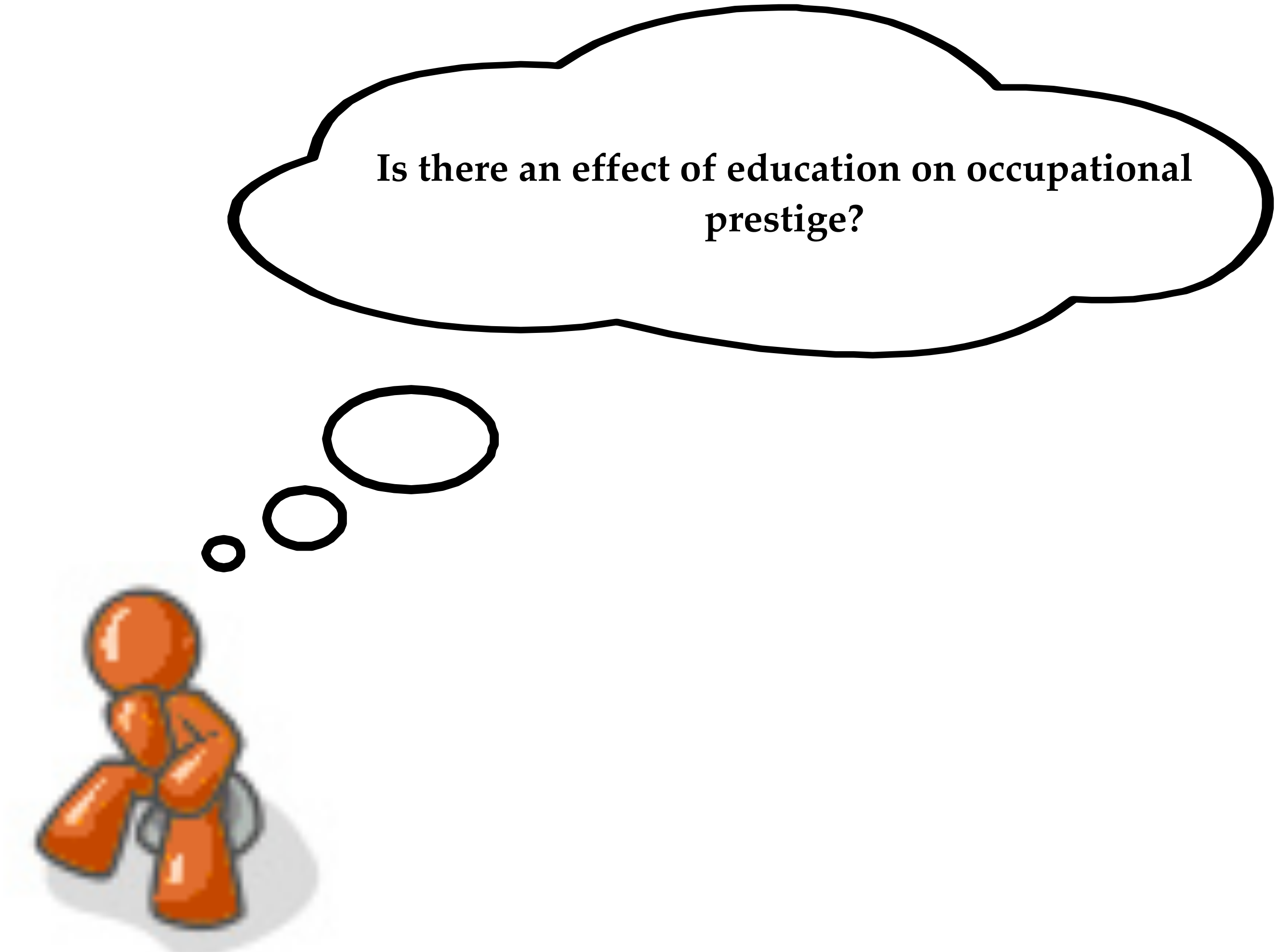
Interaction Models


Is there an effect of education on occupational prestige?

Read in the *Prestige.csv* data

Pineo-Porter occupational prestige score

Is the occupation a blue-collar profession?

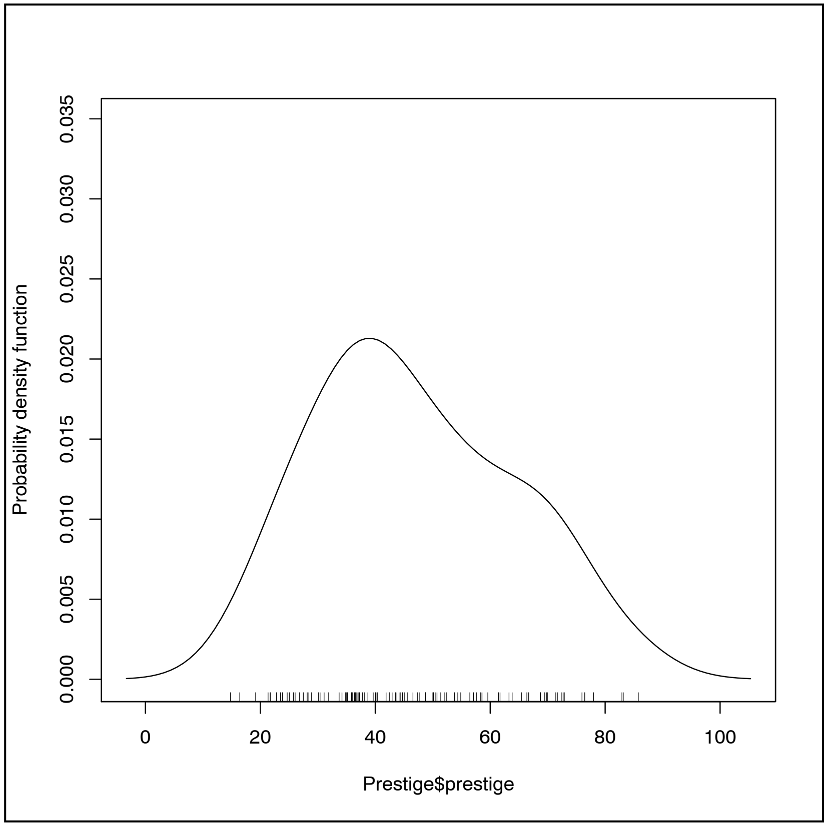
```
               occupation prestige education blue_collar income
1 government administrators     68.8     13.11           0  12351
2          general managers     69.1     12.26           0  25879
3               accountants     63.4     12.77           0   9271
4       purchasing officers     56.8     11.42           0   8865
5                  chemists     73.5     14.62           0   8403
6                physicists     77.6     15.64           0  11030
```

Average education of occupational incumbents

Average income, in dollars,

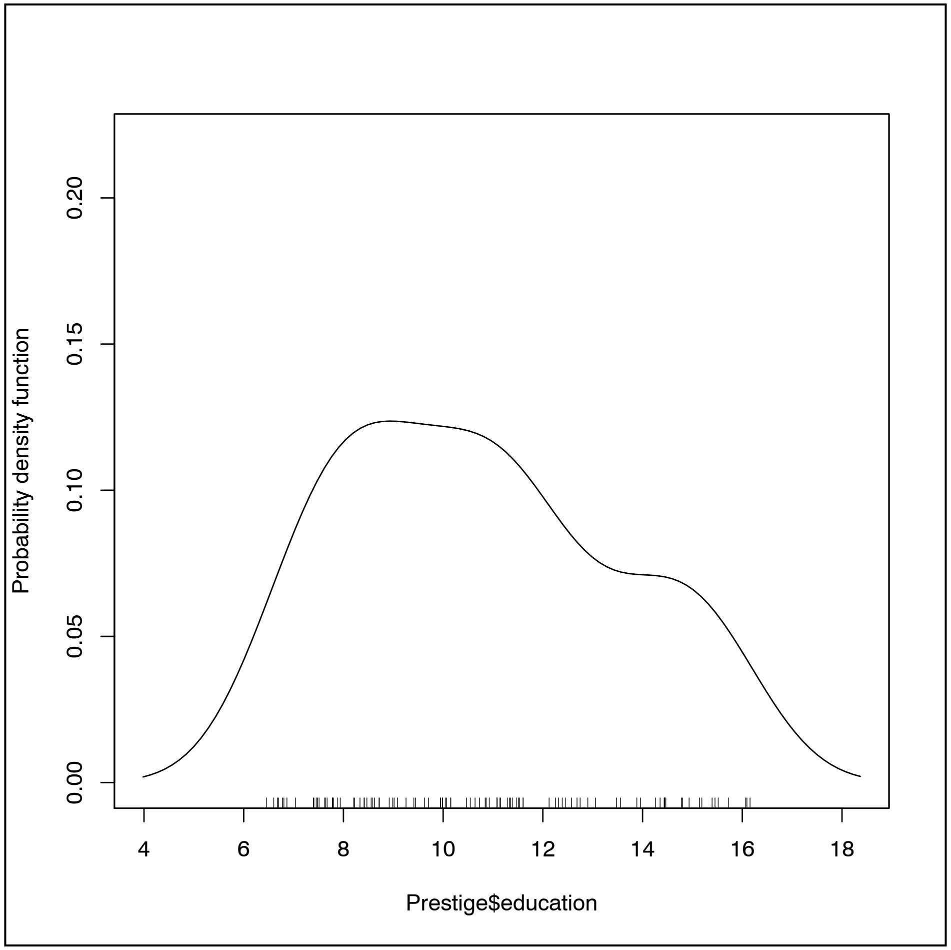
These data are Canadian Census data from 1971, and are available as part of the **car** package.
Canada (1971) *Census of Canada*. Vol. 3, Part 6. Statistics Canada [pp. 19-1–19-21].

Occupational Prestige



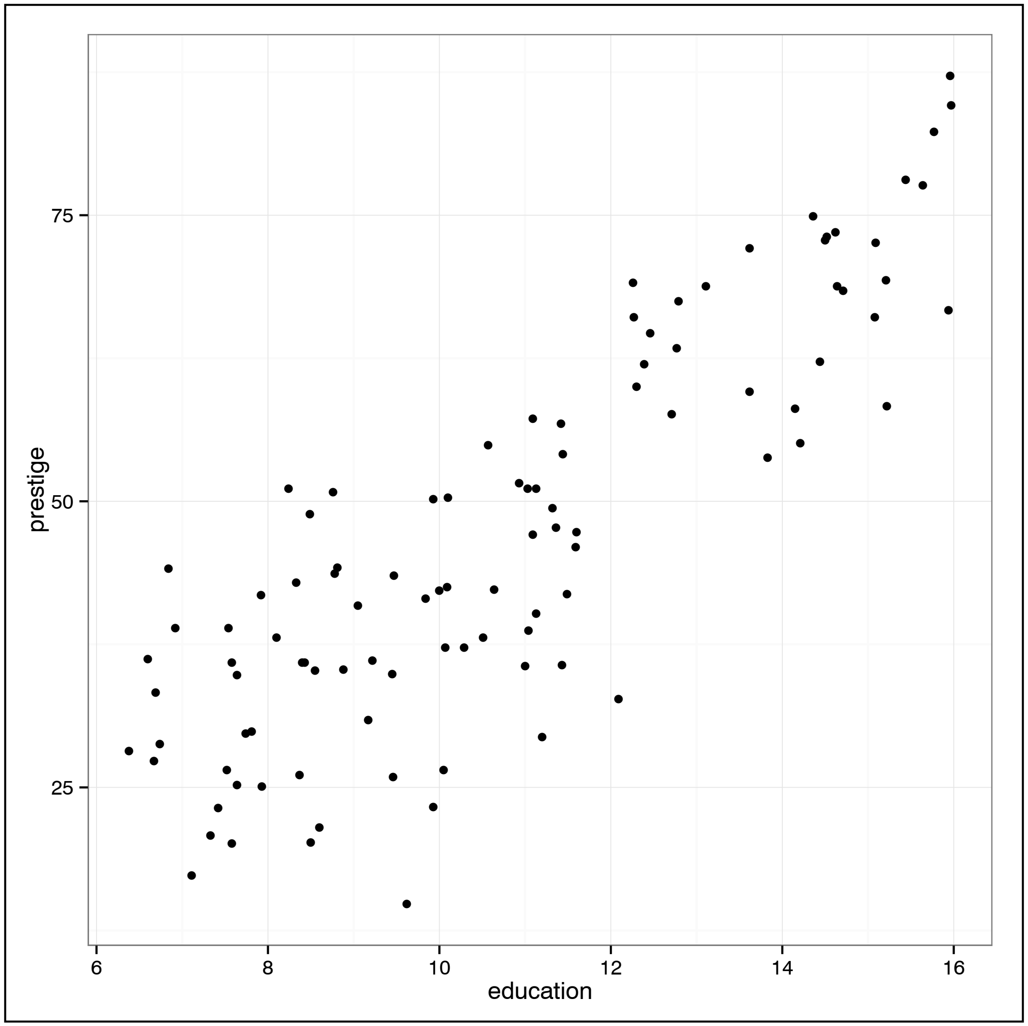
	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
1	1	102	46.83	17.2	43.6	46.2	19.2	14.8	87.2	72.4	0.33	-0.79	1.7

Education



	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
1	1	102	10.74	2.73	10.54	10.63	3.15	6.38	15.97	9.59	0.32	-1.03	0.27

Effect of Education on Occupational Prestige

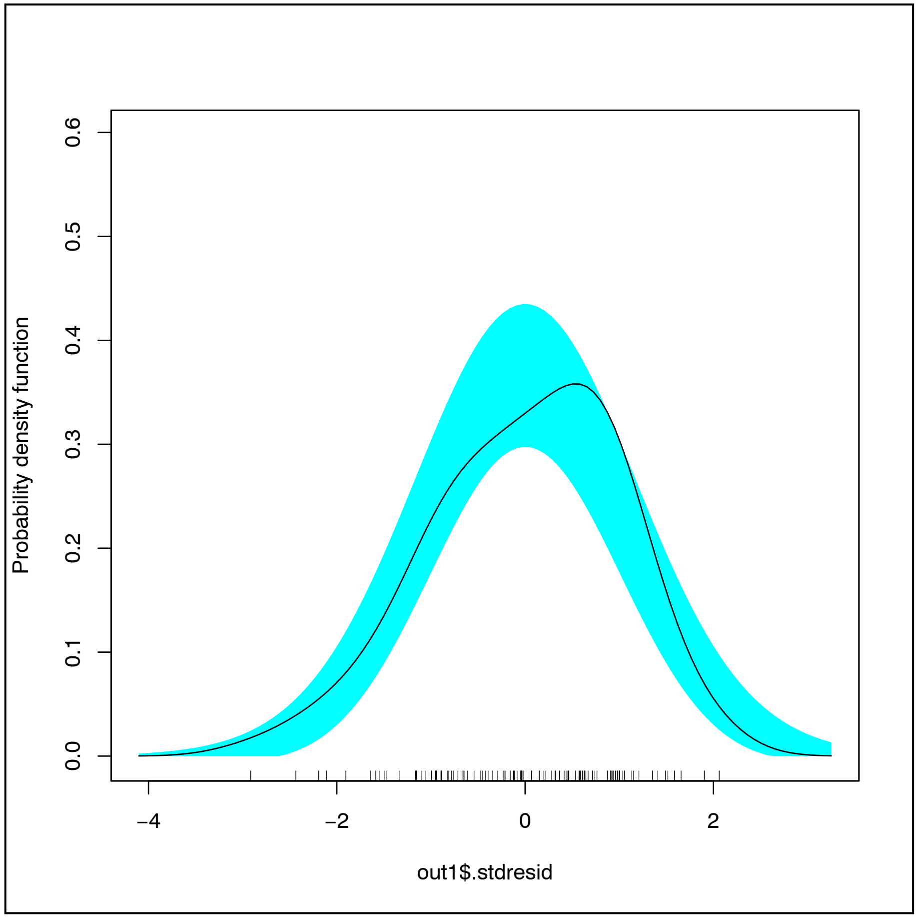


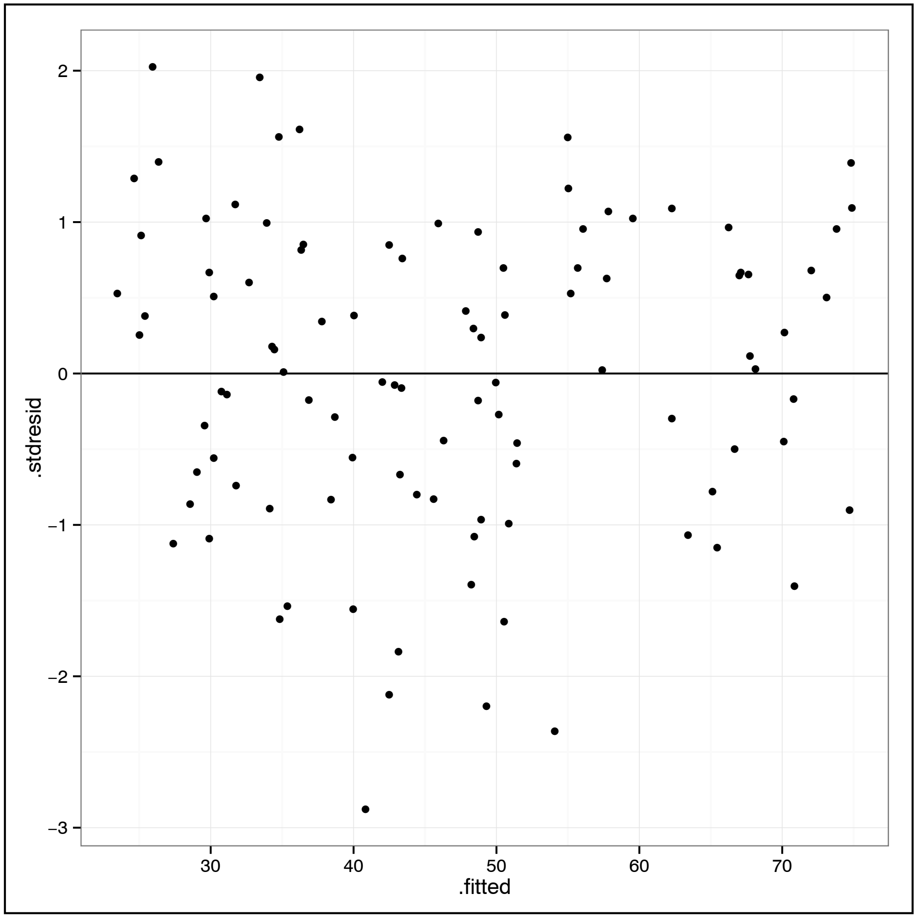
	prestige	education
prestige	1.0000000	0.8501769
education	0.8501769	1.0000000

Model

`prestige ~ 1 + education`

Residuals


Probability density function
0.0
0.1
0.2
0.3
0.4
0.5
0.6
−4
−2
0
2
out1$.stdresid


.stdresid
2
1
0
−1
−2
−3
30
40
50
60
70
.fitted

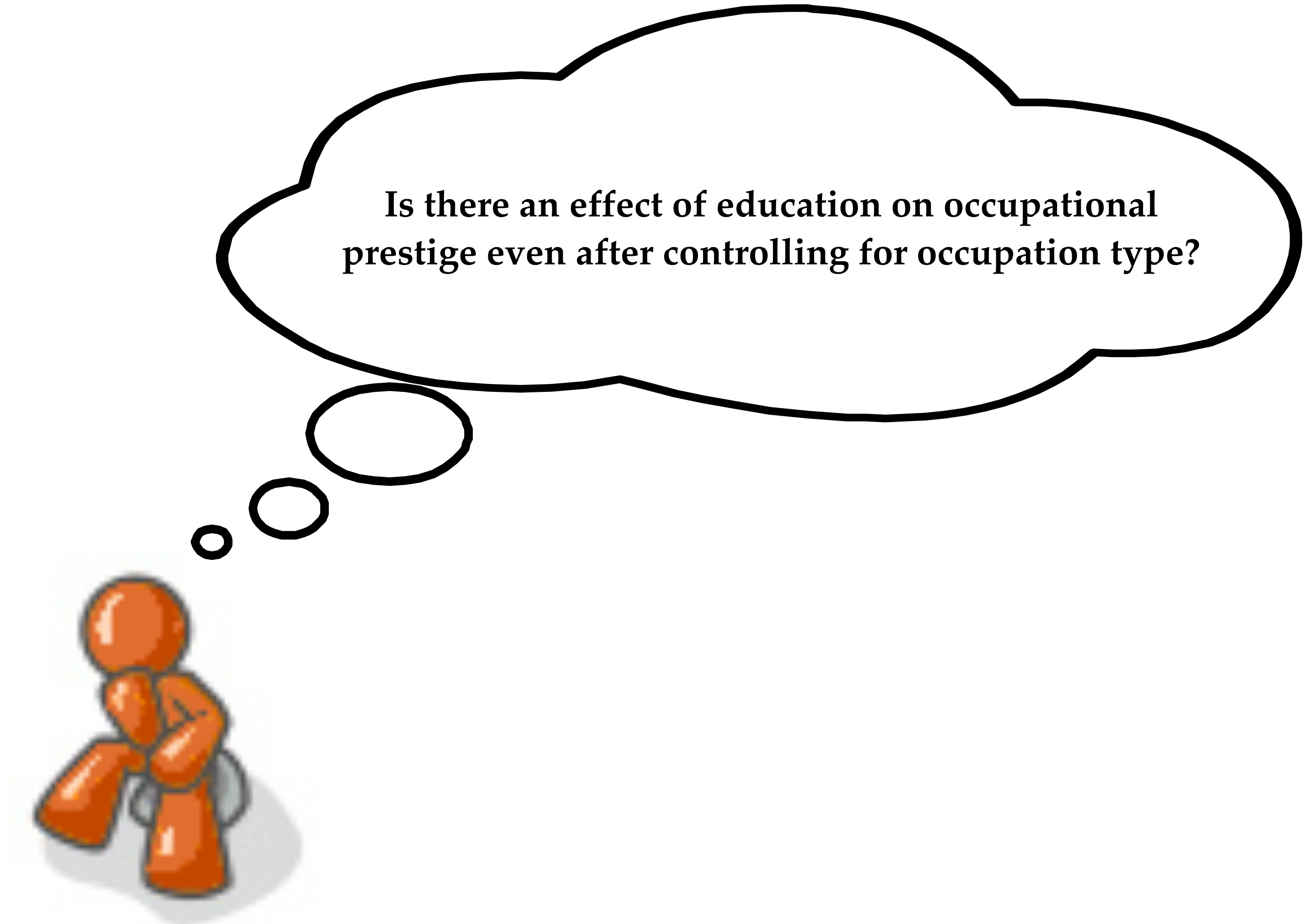
```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -10.732      3.677  -2.919  0.00434 **
education      5.361      0.332  16.148  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.103 on 100 degrees of freedom
Multiple R-squared:  0.7228,	Adjusted R-squared:   0.72
F-statistic: 260.8 on 1 and 100 DF,  p-value: < 2.2e-16
```

$$\widehat{\text{prestige}} = -10.7 + 5.4(\text{education})$$

Is there an effect of education on occupational prestige?

Yes. Each one-year difference in average education is positively associated with a 5.4-unit difference in Pineo-Porter occupational prestige score, on average. This effect is statistically reliable, $p < .001$.


Is there an effect of education on occupational prestige even after controlling for occupation type?

Occupational Type

Occupation type	***N***	***p***
Blue-collar	47	0.46
Non blue-collar	55	0.54

```
               prestige  education blue_collar
prestige      1.0000000  0.8501769  -0.6355115
education     0.8501769  1.0000000  -0.8038075
blue_collar  -0.6355115 -0.8038075   1.0000000
```

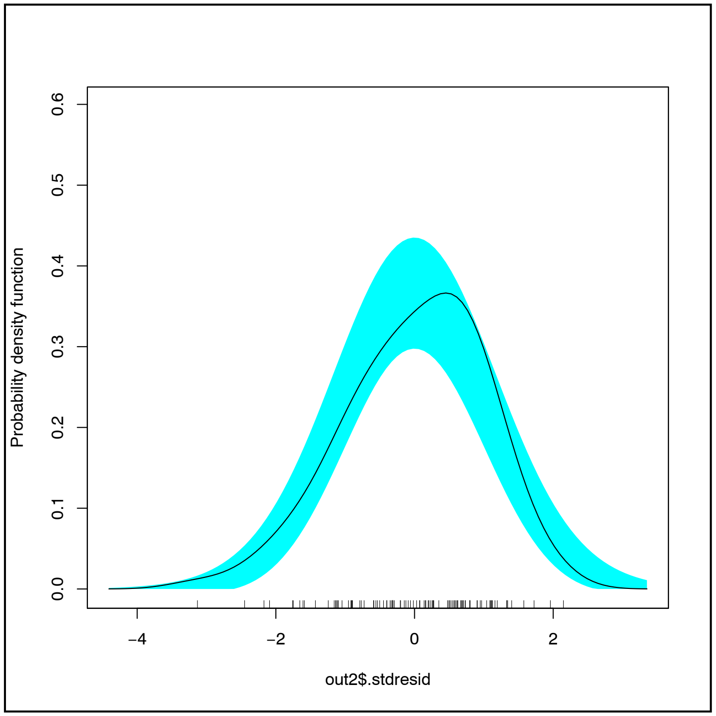
Having a blue-collar occupation is negatively associated with prestige (i.e., lower average prestige for blue-collar occupations)

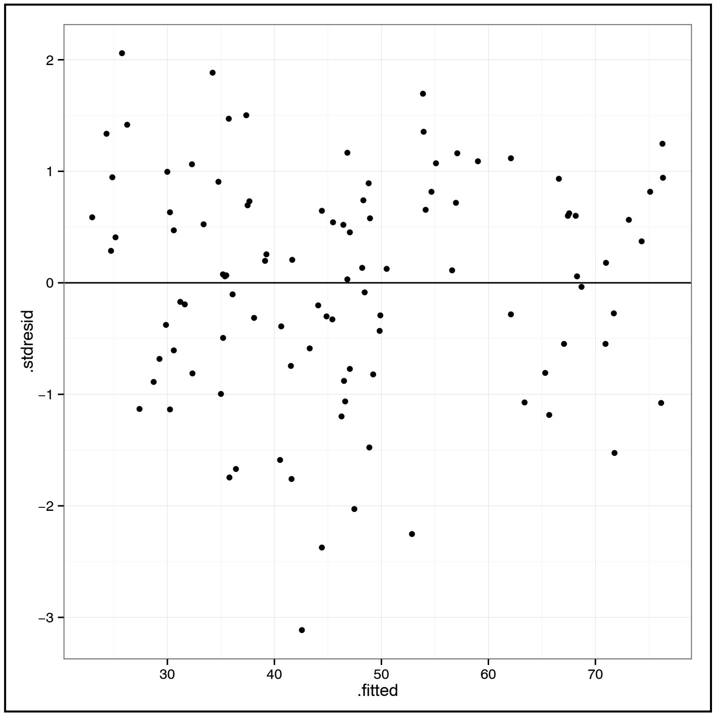
Having a blue-collar occupation is negatively associated with education (i.e., lower average education for blue-collar occupations)

Model

```
prestige ~ 1 + education + blue_collar
```

The main-effects model allows us to examine the effect of education on occupational prestige, controlling for occupation type.



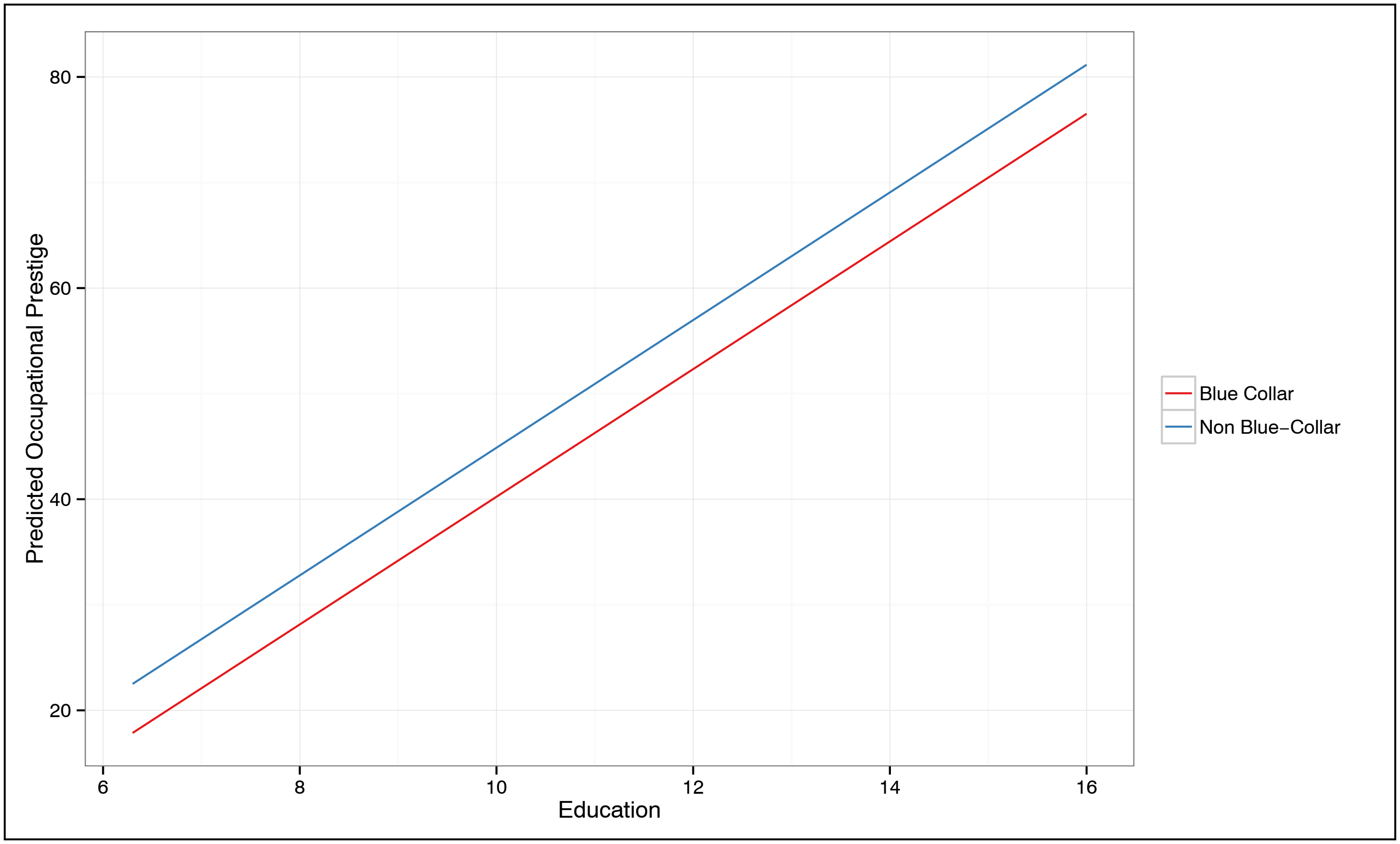


```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -20.2341     7.1746  -2.820   0.0058 **
education     6.0464     0.5543  10.908   <2e-16 ***
blue_collar   4.6455     3.0191   1.539   0.1271
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.042 on 99 degrees of freedom
Multiple R-squared:  0.7293, Adjusted R-squared:  0.7238
F-statistic: 133.3 on 2 and 99 DF,  p-value: < 2.2e-16
```

$$\widehat{\text{prestige}} = -20.2 + 6.0(\text{education}) + 4.6(\text{blue\_collar})$$

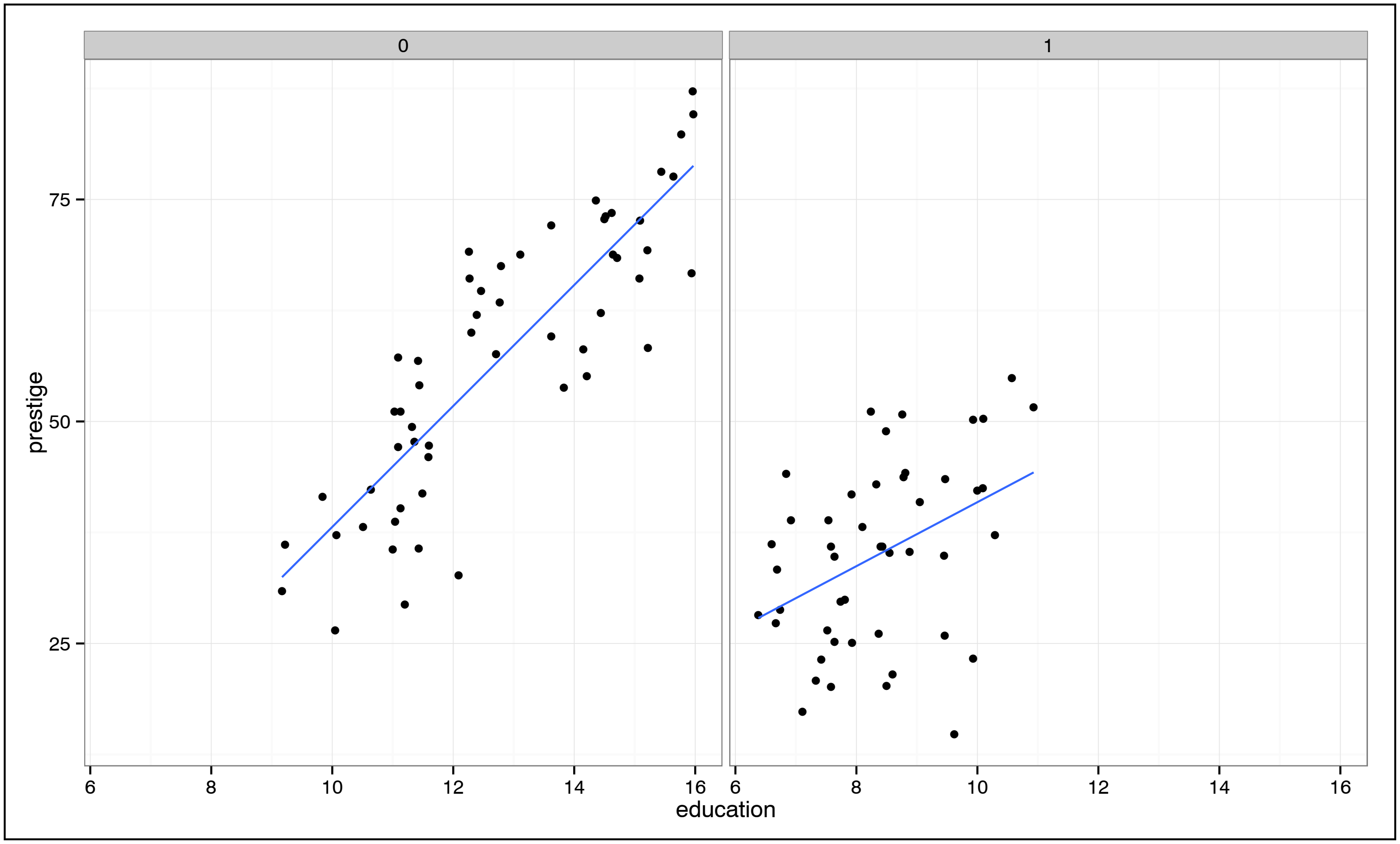
(For now, we will ignore the fact that the effect of occupational type (`blue_collar`) is not statistically reliable.)



The main-effects model makes an assumption that the effect of education is exactly the same for the different levels of occupation type.

For both blue-collar and non blue-collar occupations, the effect of education on occupational prestige is that a one-year difference in average education is positively associated with a 6.0-unit difference in Pineo-Porter occupational prestige score, on average. This effect is statistically reliable, $p < .001$.

Do the data suggest that the effect of education on occupational prestige is the same for all levels of occupation type?



The data suggest that the effect of education on occupational prestige may vary for the different levels of occupation type.

Another way of stating this, is that the data suggest an interaction between education and occupation type.

Still another way of stating this, is that the effect of education on occupational prestige *depends on* occupation type.

While the data suggest an interaction between education and occupation type, we don't know whether there is an interaction in the population, or whether the interaction we are seeing in the data is due to sampling error.

To examine this, we create an interaction term and then include that term in the model *along with* the constituent main-effects (the main-effects that we use initially to create the interaction term).

Interaction terms are created by computing the product of main-effects

$$(\text{education}) \times (\text{blue\_collar})$$

Create interaction term

```
Prestige$educ_bc = Prestige$education * Prestige$blue_collar
```

```
                 occupation prestige education blue_collar income educ_bc
1   government administrators     68.8     13.11           0  12351    0.00
2            general managers     69.1     12.26           0  25879    0.00
3                 accountants     63.4     12.77           0   9271    0.00
4         purchasing officers     56.8     11.42           0   8865    0.00
5                    chemists     73.5     14.62           0   8403    0.00
...                       ...      ...       ...         ...    ...     ...
98                bus drivers     35.9      7.58           1   5562    7.58
99               taxi drivers     25.1      7.93           1   4224    7.93
100              longshoremen     26.1      8.37           1   4753    8.37
101               typesetters     42.2     10.00           1   6462   10.00
102               bookbinders     35.2      8.55           1   3617    8.55
```

Fit model that includes constituent main-effects and interaction as predictors

Model

```
prestige ~ 1 + education + blue_collar + educ_bc
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -29.9907     7.9775  -3.759  0.00029 ***
education     6.8113     0.6185  11.013  < 2e-16 ***
blue_collar  34.8711    12.2897   2.837  0.00553 **
educ_bc      -3.2083     1.2666  -2.533  0.01290 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.804 on 98 degrees of freedom
Multiple R-squared:  0.7459,	Adjusted R-squared:  0.7381
F-statistic:  95.9 on 3 and 98 DF,  p-value: < 2.2e-16
```

$$\widehat{\text{prestige}} = -30.0 + 6.8(\text{education}) + 34.9(\text{blue\_collar}) - 3.2(\text{education})(\text{blue\_collar})$$

The key predictor in this model is the interaction term

$$H_0 : \beta_{\text{Interaction}} = 0$$

In this model, this term is statistically reliable, $B = -3.2$, $p = .013$.

This suggests that the interaction we saw in the data is (likely) not due to sampling error.

This indicates that the effect of education on occupational prestige *depends on* occupation type.

This implies that we should no longer write/speak about the effects of education without considering occupation type....

Nor should we speak/write about the effects of occupation type without considering the level of education.

To understand what the different effects in the model are, we will examine the more general interaction model

$$\text{Prestige} = \beta_0 + \beta_1(\text{Education}) + \beta_2(\text{blue\_collar}) + \beta_3(\text{Education})(\text{blue\_collar}) + \epsilon$$

The `blue_collar` predictor takes the value of either 0 or 1

`blue_collar` = 0

$$\text{Prestige} = \beta_0 + \beta_1(\text{Education}) + \beta_2(0) + \beta_3(\text{Education})(0) + \epsilon$$

$$\text{Prestige} = \beta_0 + \beta_1(\text{Education}) + \epsilon$$

`blue_collar` = 1

$$\text{Prestige} = \beta_0 + \beta_1(\text{Education}) + \beta_2(1) + \beta_3(\text{Education})(1) + \epsilon$$

$$\text{Prestige} = \beta_0 + \beta_1(\text{Education}) + \beta_2 + \beta_3(\text{Education}) + \epsilon$$

$$\text{Prestige} = [\beta_0 + \beta_2] + [\beta_1 + \beta_3]\,(\text{Education}) + \epsilon$$

`blue_collar` $= 0$

$$\text{Prestige} = \beta_0 + \beta_1(\text{Education}) + \epsilon$$

`blue_collar` $= 1$

$$\text{Prestige} = [\beta_0 + \beta_2] + [\beta_1 + \beta_3]\,(\text{Education}) + \epsilon$$

$\hat{\beta}_0$ The estimated intercept is the average value of prestige when education level = 0 for non blue-collar occupations

$\hat{\beta}_2$ The estimated effect of blue_collar is the difference in intercepts between blue-collar and non blue-collar occupations
(It is the estimated difference in the average prestige when education = 0 between blue-collar and non blue-collar occupations)

$\hat{\beta}_1$ The estimated effect of education is the effect of education for non blue-collar occupations

$\hat{\beta}_3$ The estimated effect of the interaction is the difference in the effect of education between blue-collar and non blue-collar occupations
(It is the estimated difference in slopes)

Thus, by testing whether the interaction term is zero is equivalent to testing whether the slopes (effect of education) is the same for all levels of another predictor.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -29.9907     7.9775  -3.759  0.00029 ***
education     6.8113     0.6185  11.013  < 2e-16 ***
blue_collar  34.8711    12.2897   2.837  0.00553 ** 
educ_bc      -3.2083     1.2666  -2.533  0.01290 *  
```

$\hat{\beta}_0 = -30.0$ The average value of prestige when education level = 0 for non blue-collar occupations is estimated to be –30.0 (extrapolation)

$\hat{\beta}_2 = 34.9$ The estimated difference in the average prestige when education = 0 between blue-collar and non blue-collar occupations is 34.9 (on average, when education = 0, blue-collar occupations have a prestige that is 34.9 units higher than non blue-collar occupations.

$\hat{\beta}_1 = 6.8$ The estimated effect of education for non blue-collar occupations is 6.8 (each one-unit difference in education is positively associated with a 6.8-unit difference in prestige, on average, for non blue-collar occupations.

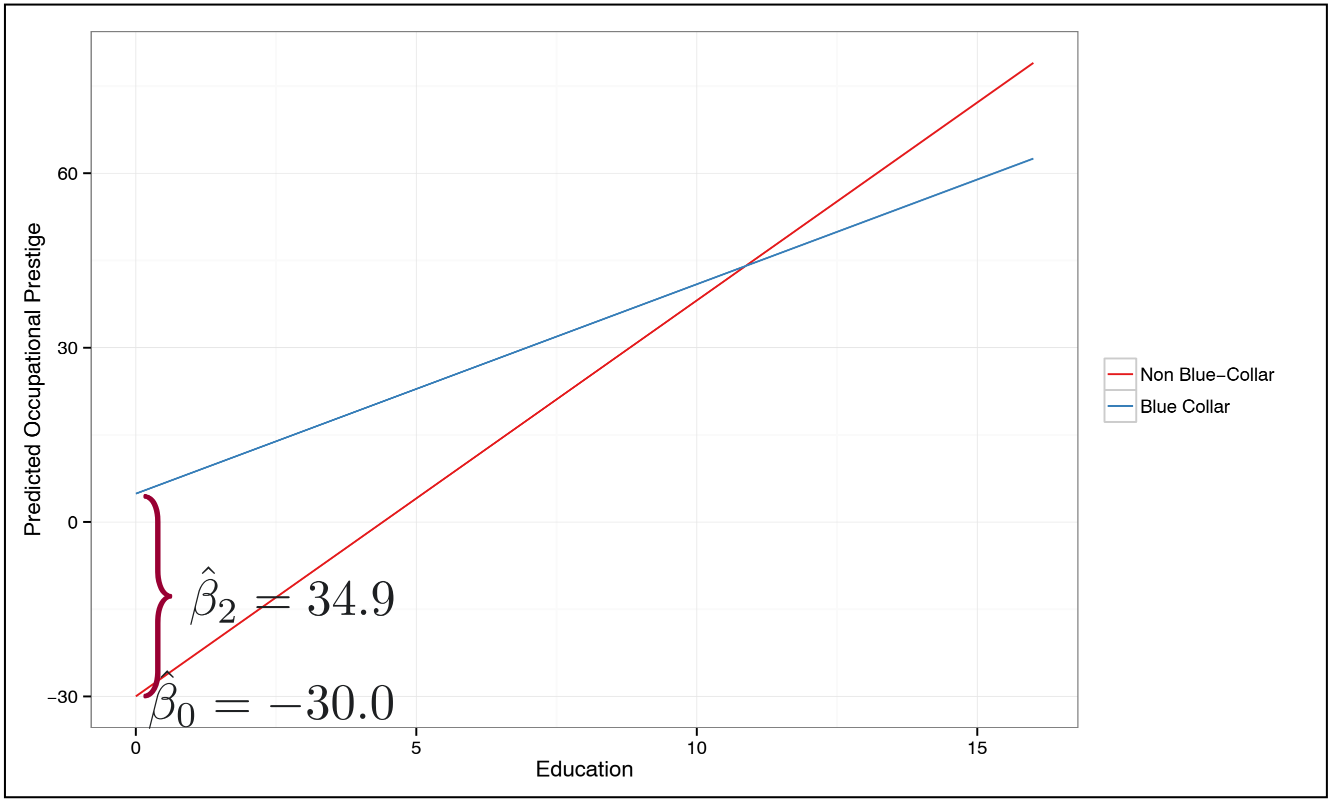
$\hat{\beta}_3 = -3.2$ The estimated difference in the effect of education between blue-collar and non blue-collar occupations is –3.2 (the effect of education on occupational prestige is 3.2 units lower for blue-collar occupations than for non blue-collar occupations blue_collar = 0

$$\hat{\text{Prestige}} = -30.0 + 6.8(\text{Education})$$

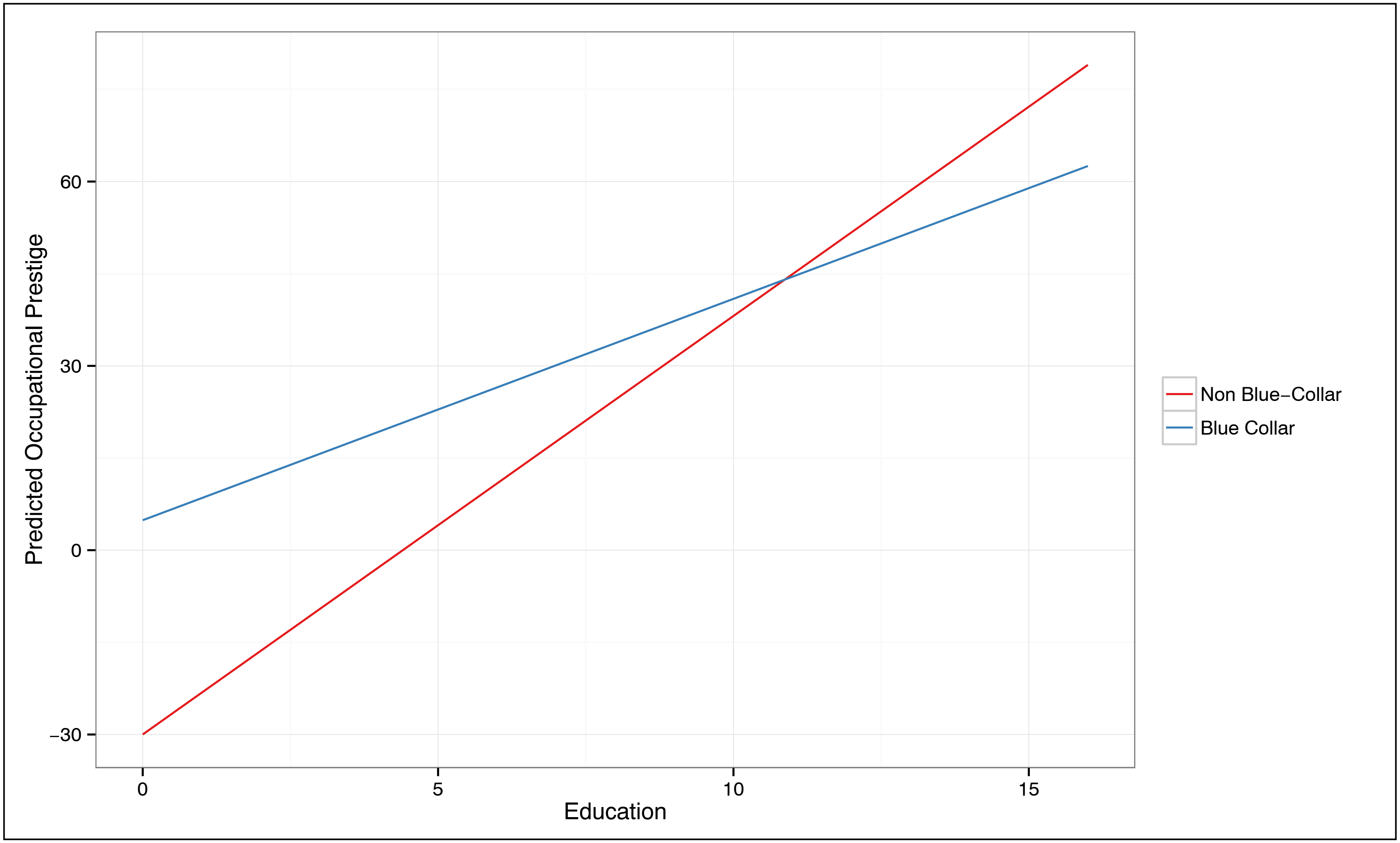
blue_collar = 1

$$\hat{\text{Prestige}} = [-30.0 + 34.9] + [6.8 + -3.2]\,(\text{Education})$$

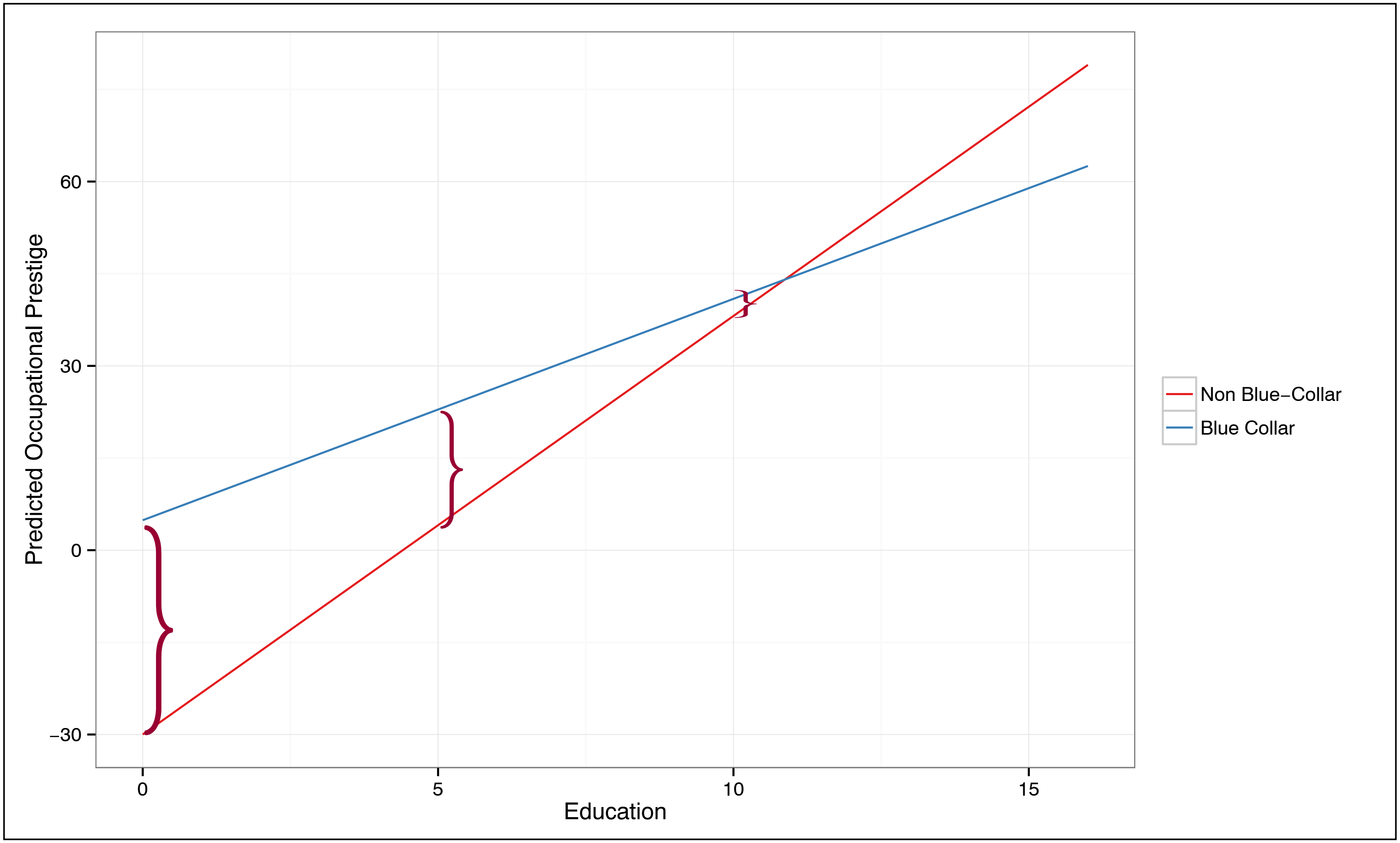
$$\hat{\text{Prestige}} = 4.9 + 3.6(\text{Education})$$



The effect of education varies across levels of occupation type.
(The slopes of the lines are different.)

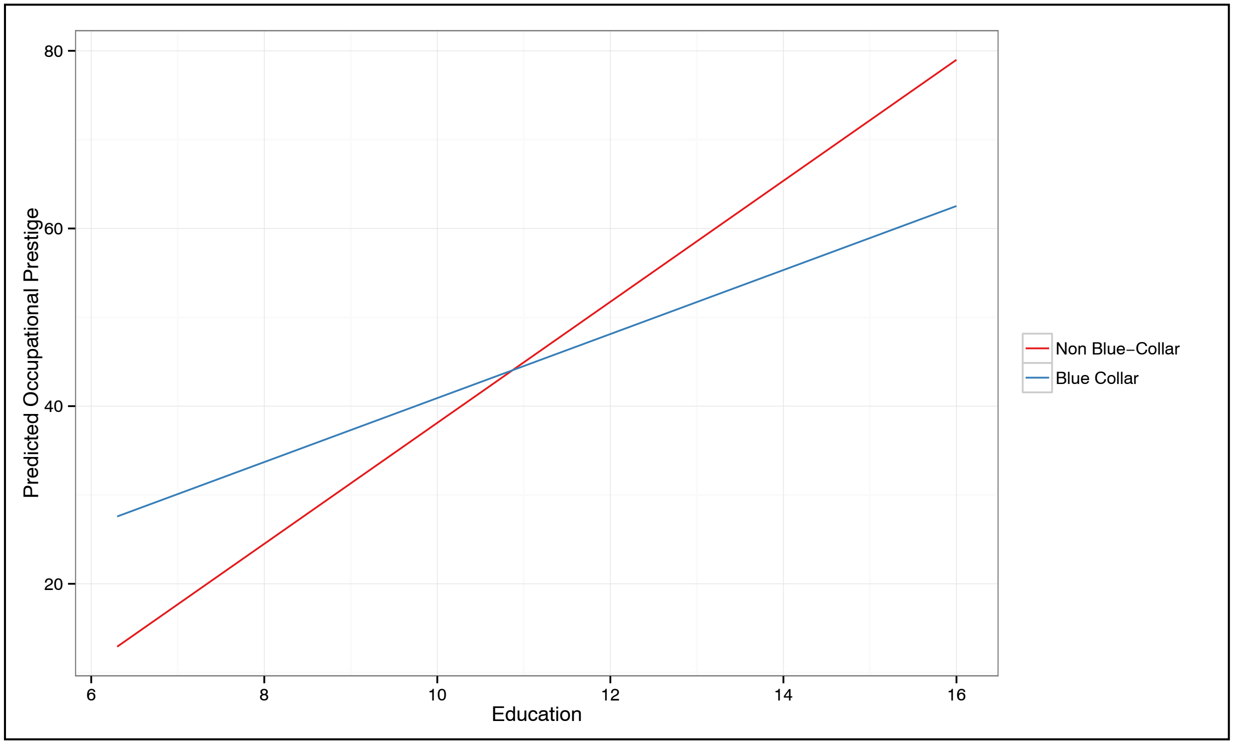


The effect of occupation type varies across levels of education.
(The distance between the lines is different.)



Disordinal Interaction

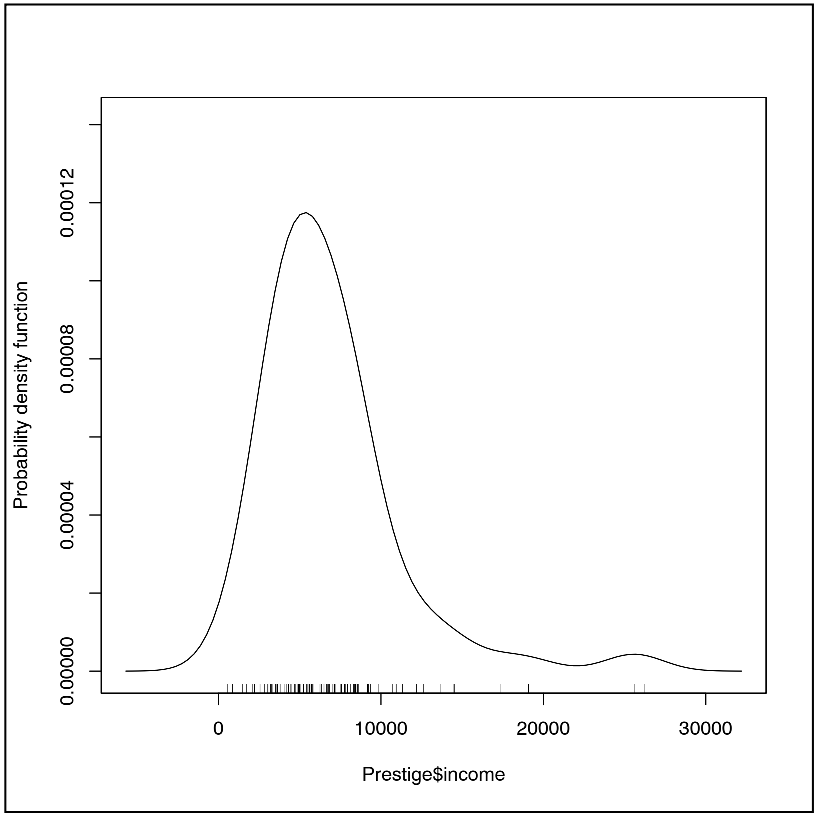
Disordinal interactions indicate a predictor has one kind of an effect in one level of the other predictor and a different kind of effect in a different level.

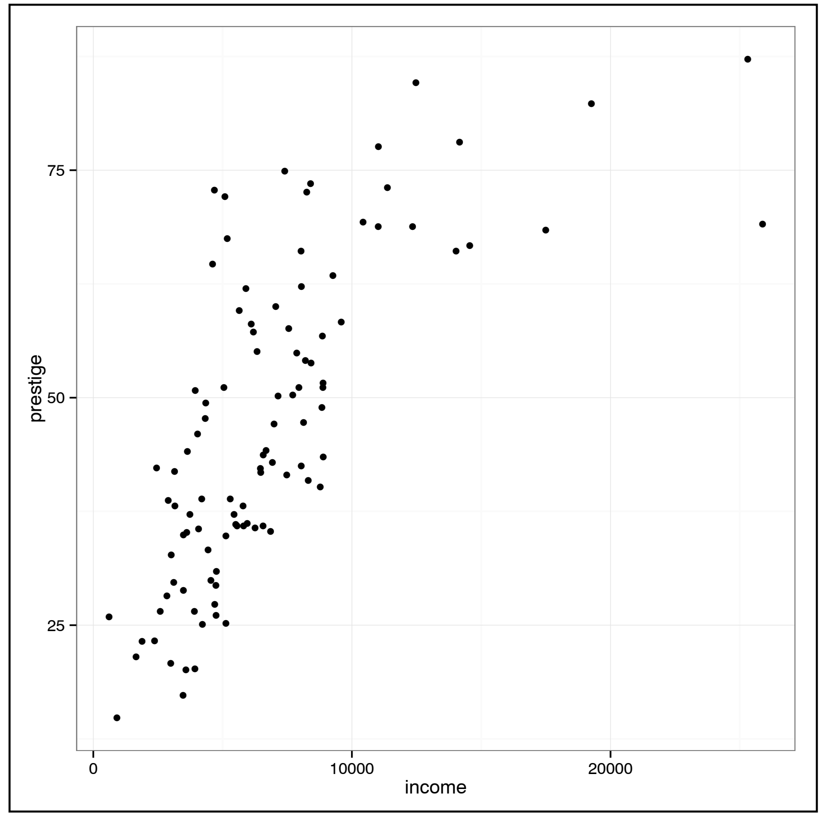


Blue-collar occupations have a higher estimated average occupational prestige than non blue-collar occupations for some education levels (≤11) and lower estimated average occupational prestige than non blue-collar occupations for other levels of education (>11)

Is there *still* an interacion effect between education and occupation type on occupational prestige after controlling for income level?

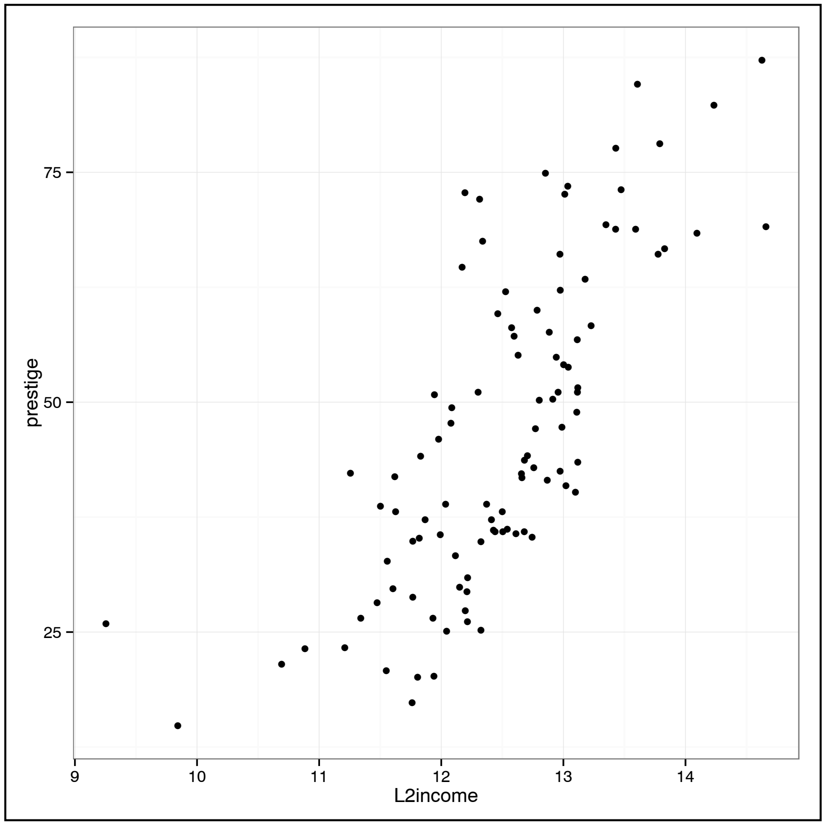
Income





	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
1	1	102	6797.9	4245.92	5930.5	6161.49	3060.83	611	25879	25268	2.13	6.29	420.41

Log Transform Income



	prestige	education	blue_collar	L2income
prestige	1.0000000	0.8501769	-0.6355115	0.7410561
education	0.8501769	1.0000000	-0.8038075	0.5481051
blue_collar	-0.6355115	-0.8038075	1.0000000	-0.3696937
L2income	0.7410561	0.5481051	-0.3696937	1.0000000

Model

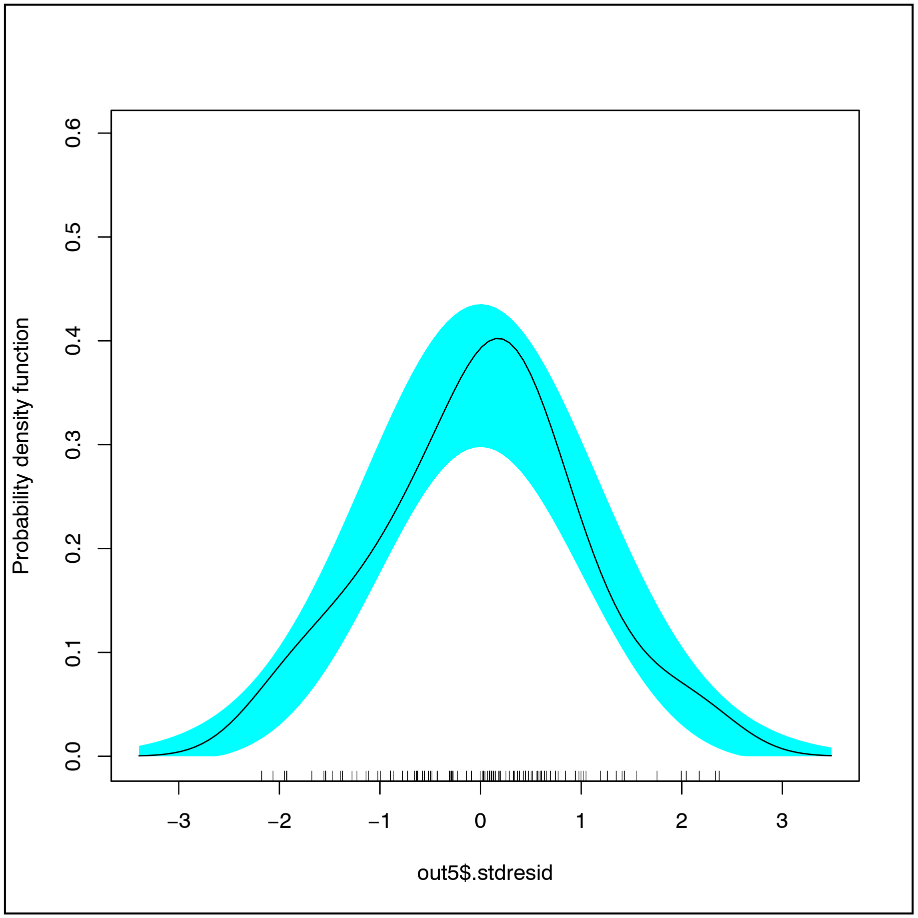
```
prestige ~ 1 + education + blue_collar + L2income + education:blue_collar
```

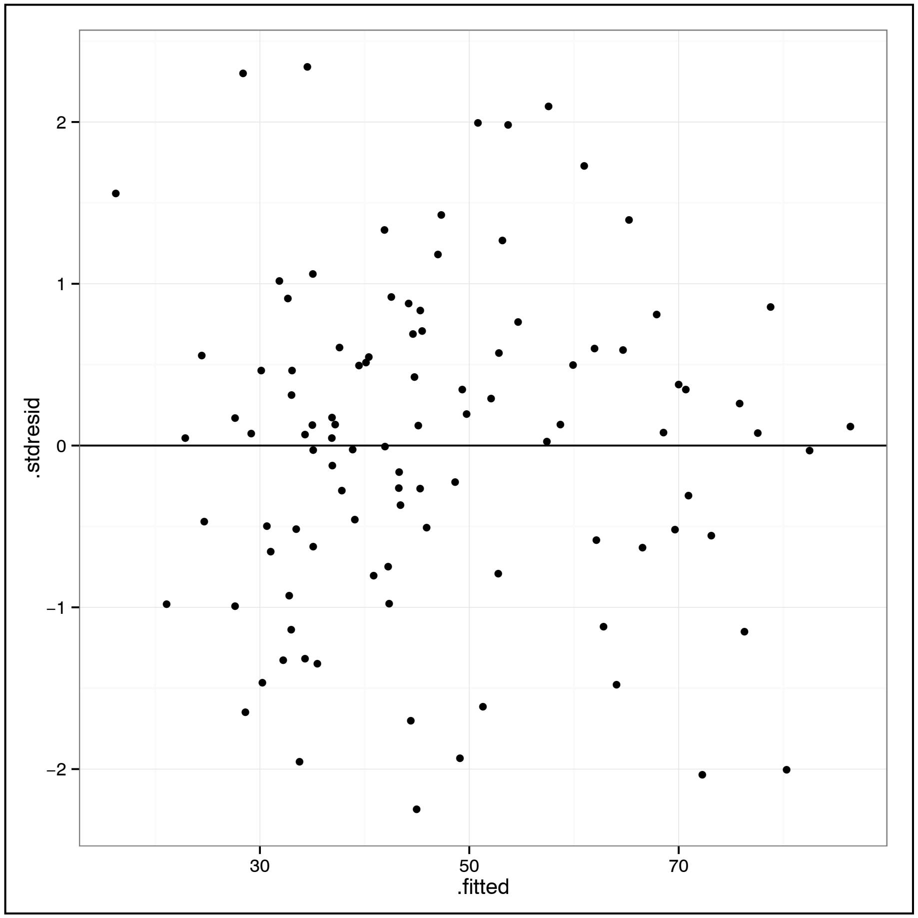
```
Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)           -101.5757    11.4677  -8.858 3.92e-14 ***
education                4.8884     0.5567   8.781 5.73e-14 ***
blue_collar             22.1974     9.9667   2.227   0.0282 *
L2income                 7.5179     1.0010   7.510 2.90e-11 ***
education:blue_collar   -2.1353     1.0224  -2.088   0.0394 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.037 on 97 degrees of freedom
Multiple R-squared:  0.8393, Adjusted R-squared:  0.8327
F-statistic: 126.7 on 4 and 97 DF,  p-value: < 2.2e-16
```

Controlling for differences in income, there is still a statistically reliable interaction between education and occupation type on occupational prestige, $p = .039$.

Residuals


Probability density function
0.0
0.1
0.2
0.3
0.4
0.5
0.6
−3
−2
−1
0
1
2
3
out5$.stdresid


.stdresid
2
1
0
−1
−2
30
50
70
.fitted

$$\hat{\text{Prestige}} = -101.6 + 4.9(\text{Education}) + 22.2(\text{blue\_collar}) + 7.5(\text{L2income}) - 2.1(\text{Education})(\text{blue\_collar})$$

The average value of prestige when education level = 0 for non blue-collar occupations is estimated to be –101.6 (extrapolation), controlling for income.

The estimated difference in the average prestige when education = 0 between blue-collar and non blue-collar occupations is 22.2 (on average, when education = 0, blue-collar occupations have a prestige that is 22.2 units higher than non blue-collar occupations), controlling for income.

The estimated effect of education for non blue-collar occupations is 4.9 (each one-unit difference in education is positively associated with a 4.9-unit difference in prestige, on average, for non blue-collar occupations), controlling for income.

The estimated difference in the effect of education between blue-collar and non blue-collar occupations is –2.1 (the effect of education on occupational prestige is 2.1 units lower for blue-collar occupations than for non blue-collar occupations), controlling for income

The estimated effect of income is 7.5 (each two-fold difference in income is positively associated with a 7.5-unit difference in prestige, on average, controlling for education and occupation type.

Understanding the Model

To better understand and interpret the different effects in the model, we will plot the fitted values.

We know we want to display the effect of education on the x-axis, and the effect of occupation type (blue_collar) using different lines.

To display the effect of income, we also need to use different lines.

```
myData = expand.grid(
    education = seq(from = 0, to = 16, by = 0.1),
    blue_collar = c(0, 1),
    L2income = c(9, 15)
    )
```

All predictors displayed as different lines (everything except for the predictor on the x-axis) you should pick meaningful values for.

After setting up the data set, use the `predict()` function.

```
preds = predict(lm.5, newdata = myData)
```

Bind the predictions to the prediction data

```
myData = cbind(myData, preds)
```

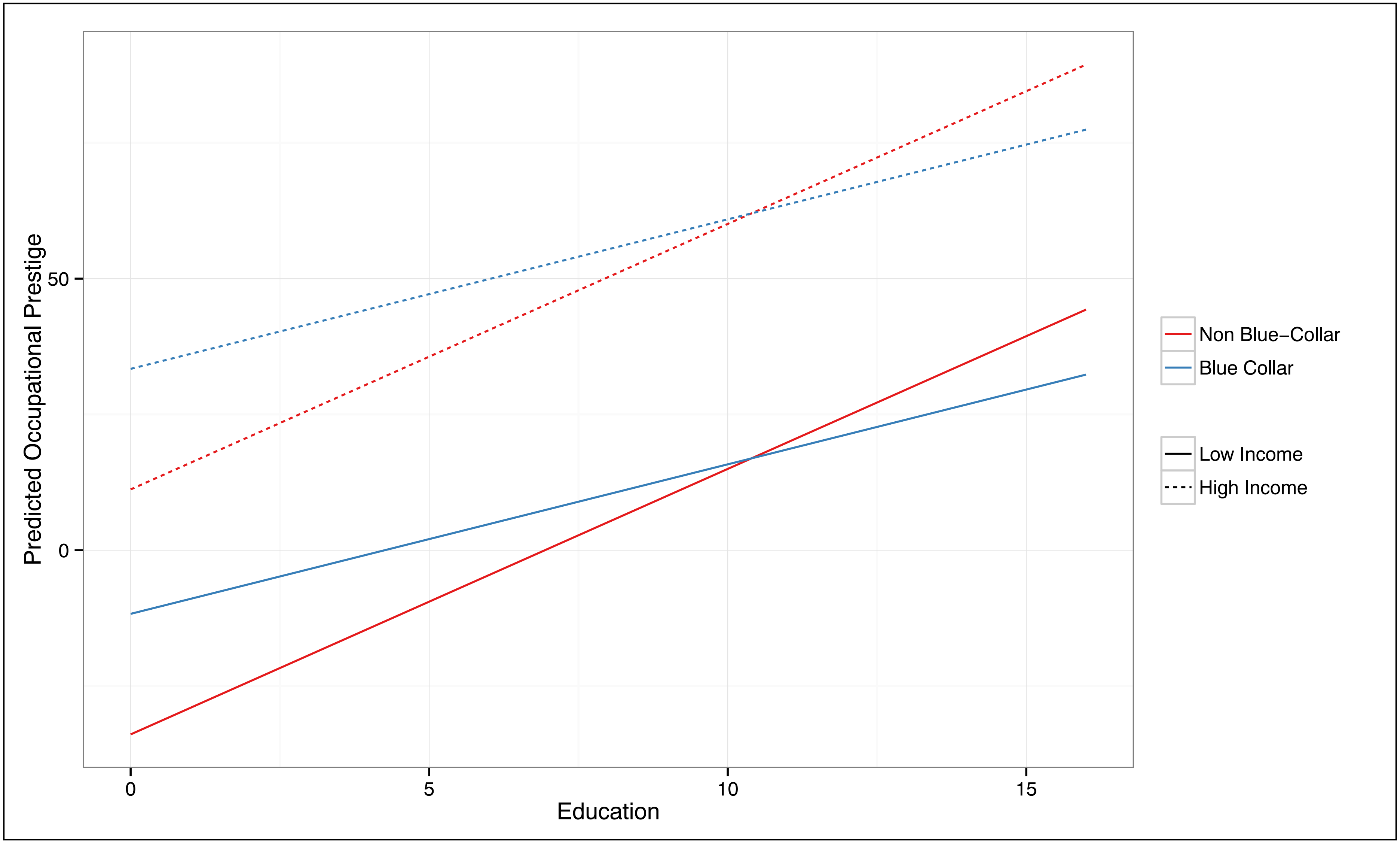
```
myData$blue_collar = factor(myData$blue_collar,
    levels = c(0, 1),
    labels = c("Non Blue-Collar", "Blue Collar")
    )
```

Change all predictors whose effects are displayed through different lines to factors.

```
myData$L2income = factor(myData$L2income,
    levels = c(9, 15),
    labels = c("Low Income", "High Income")
    )
```

Now, plot

```
ggplot(data = myData, aes(x = education, y = preds,
                          color = blue_collar, linetype = L2income)) +
    geom_line() +
    xlab("Education") +
    ylab("Predicted Occupational Prestige") +
    scale_color_brewer(name = "", palette = "Set1") +
    scale_linetype(name = "") +
    theme_bw()
```



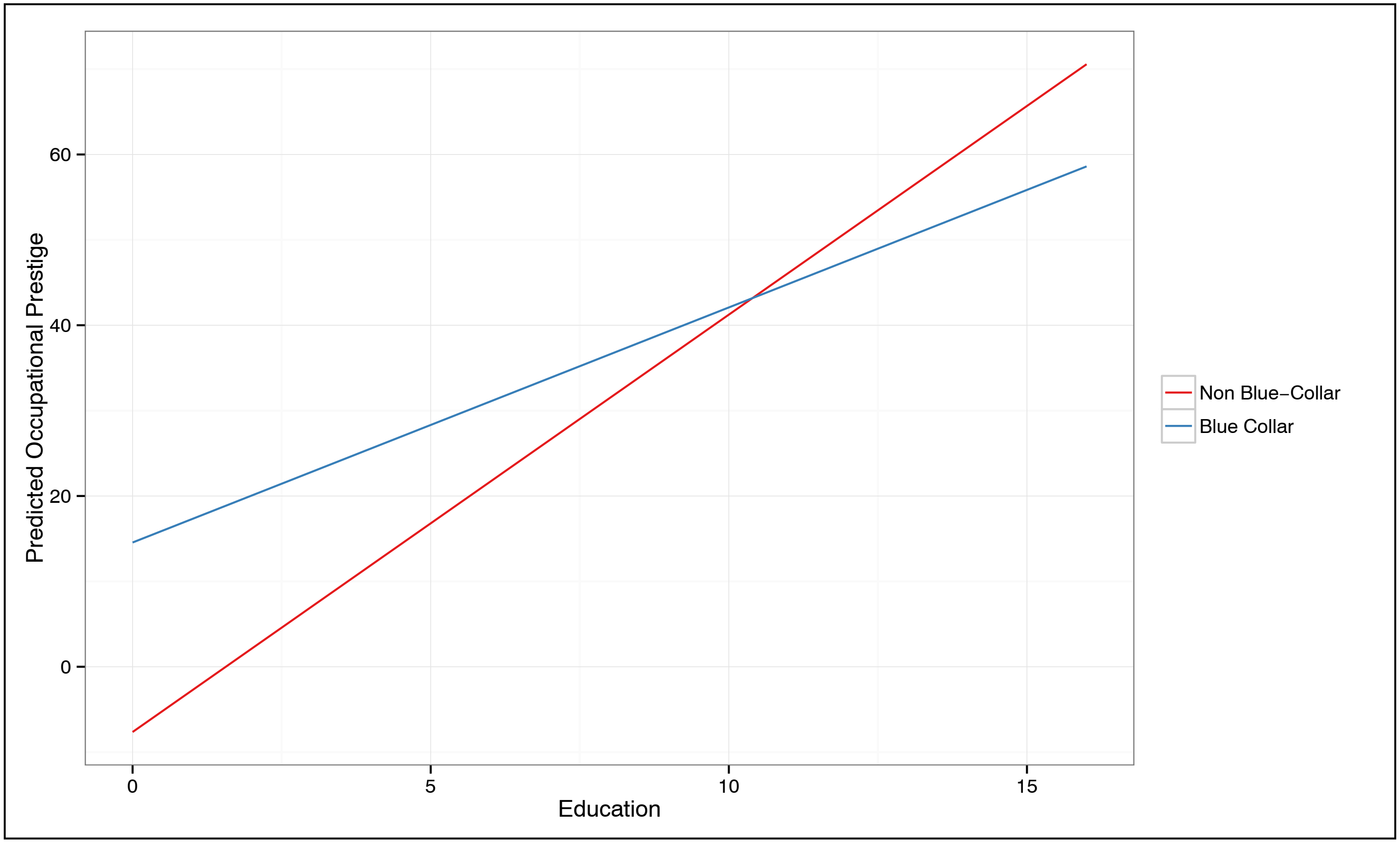
The interaction plot is the same for occupations that have low average incomes and occupations that have high average incomes (L2income is a main-effect), and occupations with high average incomes have a higher occupational prestige, on average, than those with a low average income (controlling for occupation type and education).

Control out the Effect of Income

To control out the effect of a predictor, we can select the mean value as our meaningful value for the predictor.

```
> mean(Prestige$L2income)
[1] 12.49447
```

```
myData = expand.grid(
    education = seq(from = 0, to = 16, by = 0.1),
    blue_collar = c(0, 1),
    L2income = 12.49447
    )
```



The interaction shown between education and occupational type on occupational prestige is the same. The predicted values represent the predicted occupational prestige for the average value of L2income. (Any caption should include *controlling for the effect of income*.)

OLS Regression Models Predicting Occupational Prestige Based On N = 102 Occupations.

	Estimate (SE)			
Predictor	Model A	Model B	Model C	Model D
Education	5.36^{***} (0.33)	6.04^{***} (0.55)	6.81^{***} (0.62)	4.89^{***} (0.56)
Occupation type†		4.65 (3.02)	34.87^{**} (12.28)	22.20^{*} (9.97)
Income				7.52^{***} (1.00)
Education x Occupation type			6.04^{***} (0.55)	-2.14^{*} (1.02)
(Intercept)	-10.73^{**} (3.68)	-20.23^{**} (7.17)	-3.21^{*} (1.27)	-101.58^{***} (11.47)
Model Summaries				
R^2	0.723	0.729	0.746	0.839
Adjusted R^2	0.720	0.723	0.738	0.832

Note. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

†Occupation type was dummy coded and the reference group is non blue-collar occupations