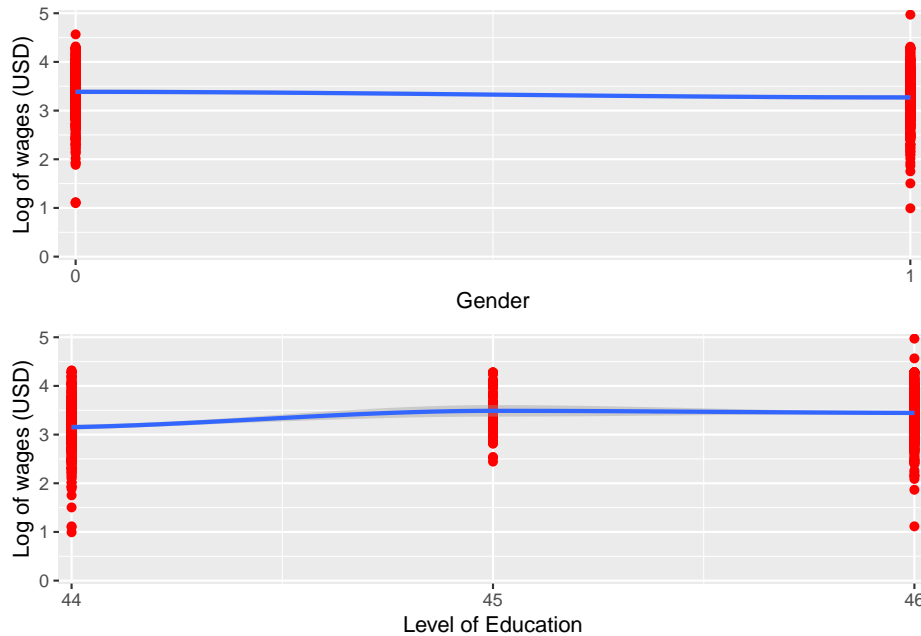


Appendix

Loess graphs



Based on the non-parametric regression lines, it seems that the association between gender and log wages is linear. For education and log wages, the overall picture is also linear, however, there is a change in the trend between Professional degrees and PhD, possibly because we have a smaller number of Professional degrees in our sample.

A linear spline model with knot at Professional degree might produce a better fit model, which could be useful for predicting wages. However, for the overall association of the data this is likely not significant as we show in multiple regression model with education as Categorical variable.

Additional findings from Data Summary Table

- Professional Degree holders earn more than PhDs. This could be due to low sample size.
- Masters degree holders still earn much less on avg.
- The distribution of hourly wages is right-skewed while log hourly wages are almost symmetric.

Further considerations for regression findings

In **Model 1** the intercept shows the log hourly wages of men which is hard to interpret in log unit (32 dollars per hour in the previous table).

For all models, logpoint differences are not exactly identical to percentage differences, here I ignored this for simplicity.

Since the y variable of all models match, their R-squared may be compared. While the last model has the highest R-squared, it is still pretty low at 0.081. However, we are not creating these models for predicting wages, therefore, this is a not major concern for us. If we wanted to build better prediction models we could: consider increasing variation in education or add a linear spline at professional level degrees.

All models use heteroskedastic robust formulas.