Analysis of Gender Pay Gap for Web Developers

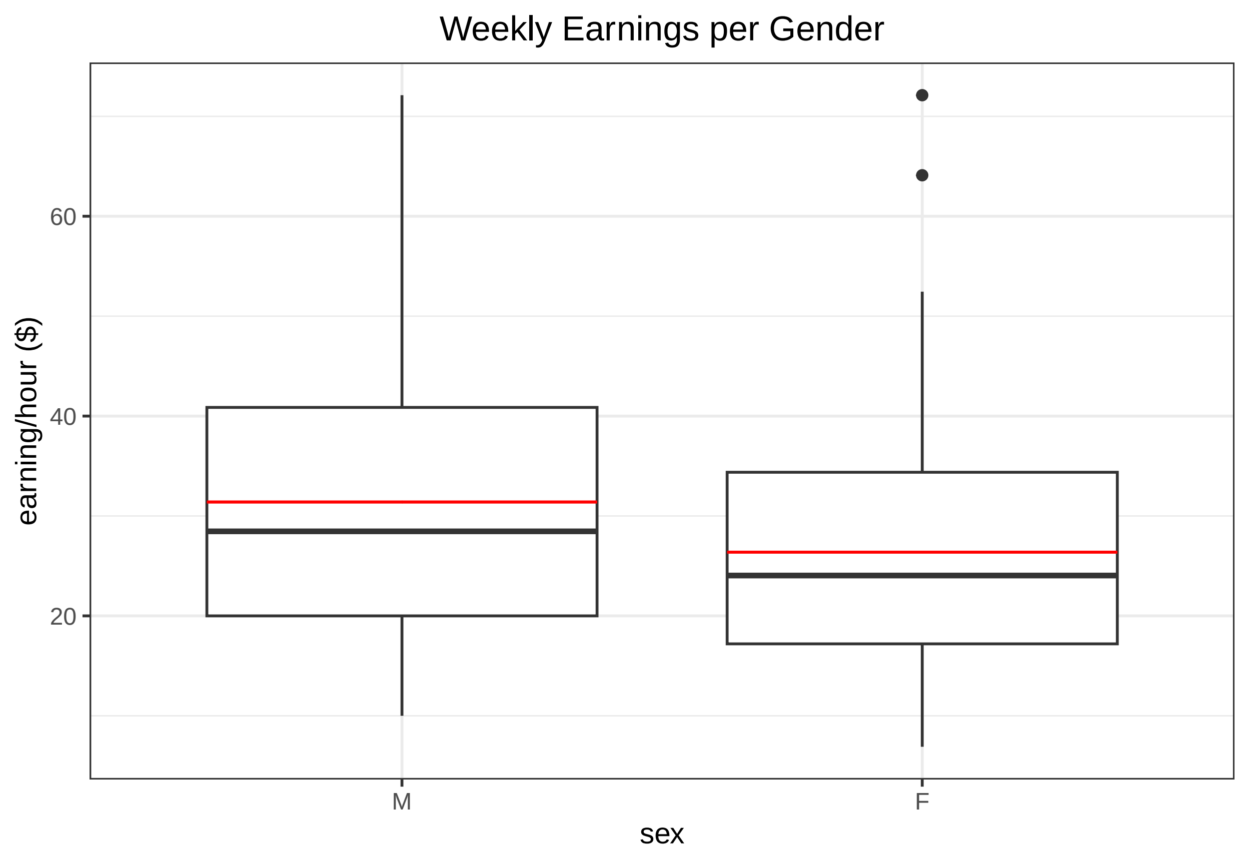
Akos Almasi & Tunay Tokmak

The Data

The data set consists of web developers. It has 216 observations. A feature is generated as weekly earnings as a division of earnwke and uhours columns.

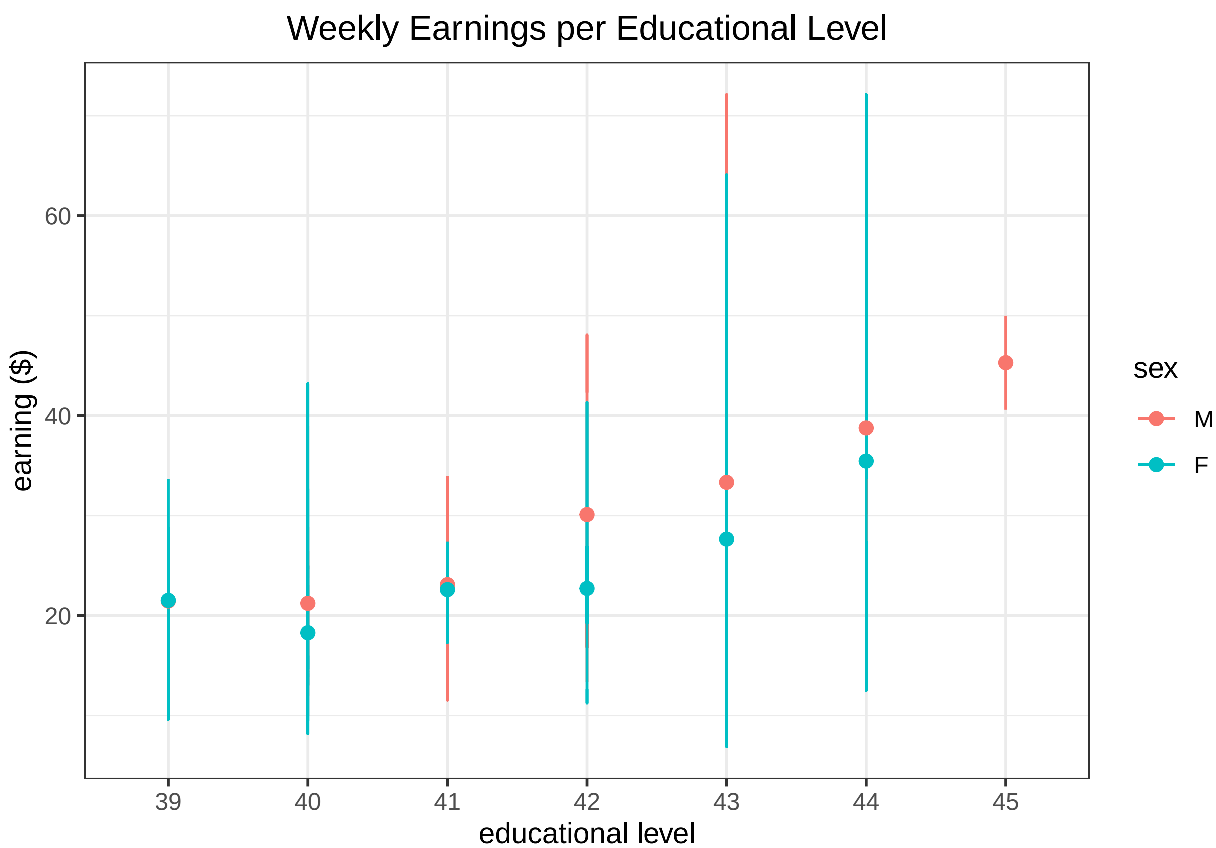
Unconditional Gender Gap

The chart clearly demonstrates that men earn higher than women on average as web developers. The mean for men is 31.39838 USD when it is 26.37298 USD for women.



Gender Gap Variance with Educational Level

The graph demonstrates that the average salary for both sexes increase as the educational level of the web developers increases. However, it also demonstrates that the average salary for women is below the average salary of men. So, there is a gender gap.



Linear Models

There are two different linear models. The first model takes solely the sex categorical variable into account, while the second model considers both sex and grade92 categorical variables. Thus, we can observe the correlation between educational level and income as well as gender and income.

The R squared values which can be observed in the summary table illustrates that the simple linear regression with only sex variable explains only 3% of the variation while the second model, multiple linear regresion with added grade92 variables, explains 20.8% of the variation in the data. That shows that the educational level is more correlated with the earning/hour than sex categorical variable.

It is seen that the intercept for the first model is 31.40, while the slope coefficient is -5.03 . The slope coefficient means that if the sex is female, independent variable set to 1, this will decrease the dependent variable weekly earning by -5.03.

The second model's intercept is 23.57, while the slope coefficients for independent sex and grade variables are -1.64,0.92,5.08,9.32,15.68,21.73. respectively.

The slope coefficients demonstrates the effect of each estimator variables' on the weekly earning. Sex can be either 0 or 1 and only one grade variable can be 1 which will increase or decrease the value of dependent variable by the slope coefficient.

Table 1: The Coefficient Estimates and Confidence Intervals of Models

	Model 1 / Est.	Model 1 / 0.5 %	Model 1 / 99.5 %	Model 2 / Est.	Model 2 / 0.5 %	Model 2 / 99.5 %
(Intercept)	31.40*	28.03	34.77	23.57*	16.52	30.62
sexF	-5.03*	-9.85	-0.20	-4.69*	-9.33	-0.05
grade9240				-1.64	-9.72	6.43
grade9241				0.92	-8.44	10.28
grade9242				5.08	-5.65	15.81
grade9243				9.32*	1.70	16.95
grade9244				15.68*	5.20	26.15
grade9245				21.73*	10.45	33.01
Num.Obs.	216			216		
R2	0.032			0.176		
RMSE	13.69			12.63		
Std.Errors	Heteroskedasticity-robust			Heteroskedasticity-robust		

Note: ^^ * $p < 0.01$

Statistical Inference

First model is significant for both intercept and slope coefficient at a p .01. The 99% confidence intervals are [28.03,34.77], [-9.85,-0.20] respectively.

The second model is significant for intercept, and sex, grade9243,grade9244,grade9245 at a p value .01. The 99% confidence intervals are [16.52,30.62], [-9.33,-0.05], [1.70,16.95], [5.20,26.15], [10.45, 33.01] respectively.

The rest of the variables are not statistically significant. This means the model needs to be improved by adding either more dependent variables or having a different modelling approach.