Tutorial 3 On-site Questions

A dataset on house selling price was randomly collected [1], `house_selling_prices_FL.csv`. It's our interest to model how y = selling price (dollar) is dependent on x = the size of the house (square feet). A simple linear regression model (y regress on x) was fitted, called Model 1.

The given data has another variable, NW, which specifies if a house is in the part of the town considered less desirable (NW = 0).

1. Derive the correlation between x and y.

2. Derive a scatter plot of y against x. Give your comments on the association of y and x.

3. Derive R^2 of Model 1. Verify that $\sqrt{R^2} = |cor(y, x)|$. In which situation we can have $\sqrt{R^2} = cor(y, x)$?

4. Form a model (called Model 2) which has two regressors (x and NW). Report the coefficient of variable NW in Model 2. Interpret it.

5. Estimate and report the price of a house where its size is 4000 square feet and is located at the more desirable part of the town.

---

[1] *Statistics: The Art and Science of Learning from Data*, 4th, Agresti, Franklin, Klingenberg