Linear Regression Exercise

Tom Madsen

June 29, 2018

Let's return to the Economist dataset we explored in the Graphics Workshop. We'll use our knowledge of linear regression to explore the relationship between HDI (an index of human development) and CPI (an index of perceived corruption).

1. What types of variables are present in this dataset? Examine the dataset for categorical and numerical variables

2. *Using ggplot*, create histograms of the continuous variables. What do you notice about their distributions?

3. *Using ggplot*, create a scatter plot of HDI vs. CPI. Is the relationship linear? If not, investigate a transformation of one or the other.

4. Fit a linear regression to explain HDI using CPI. Practice interpreting the regression coefficients. Do they both have meaningful interpretations?

5. *Using ggplot*, lay the regression line over the scatterplot. Does the regression line do a good job of summarizing the relationship?

6. Create a histogram and a q-q plot of the residuals. Are the assumptions of linear regression met? If not, how might we adjust our analysis to meet them?