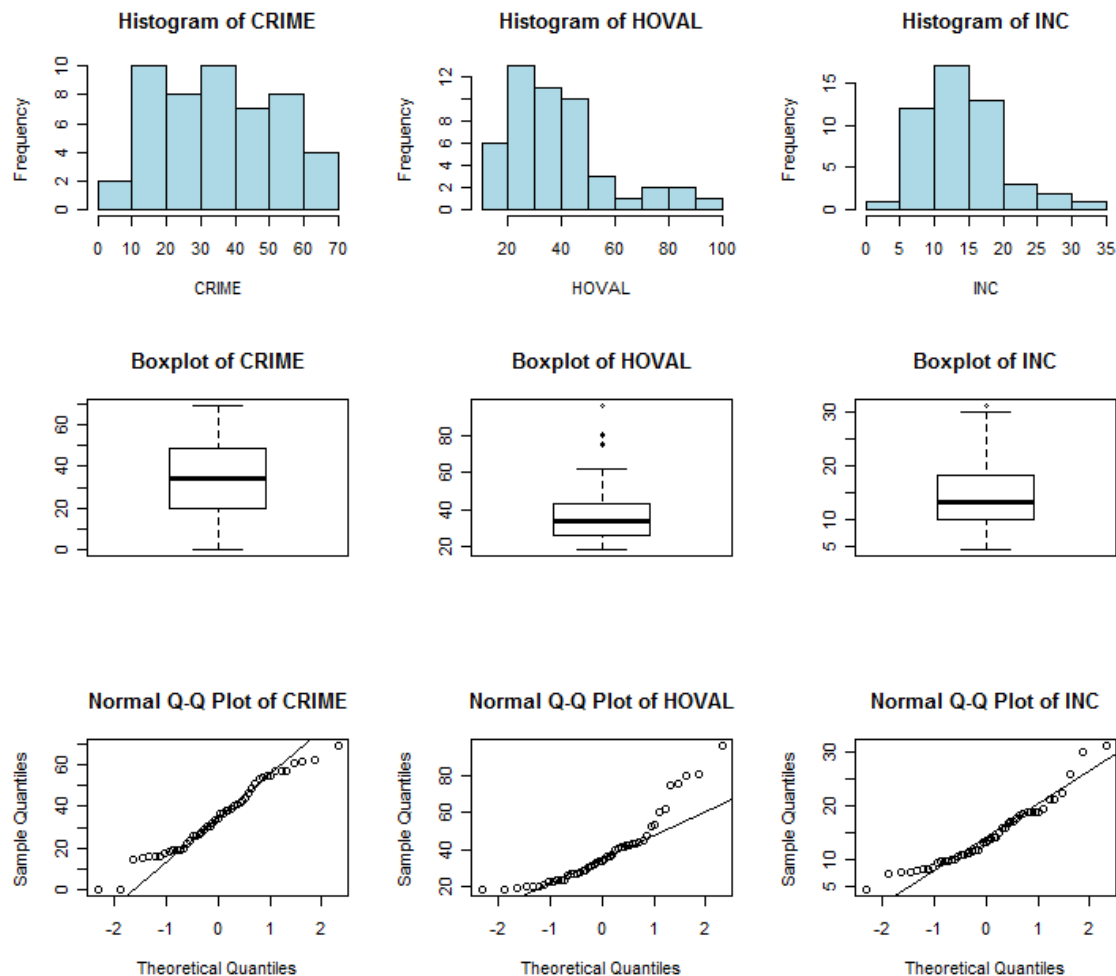


Solutions to Lab 1 - Introduction to R

1. The histogram, boxplot, and QQ plot for CRIME, HOVAL, and INC are shown below.

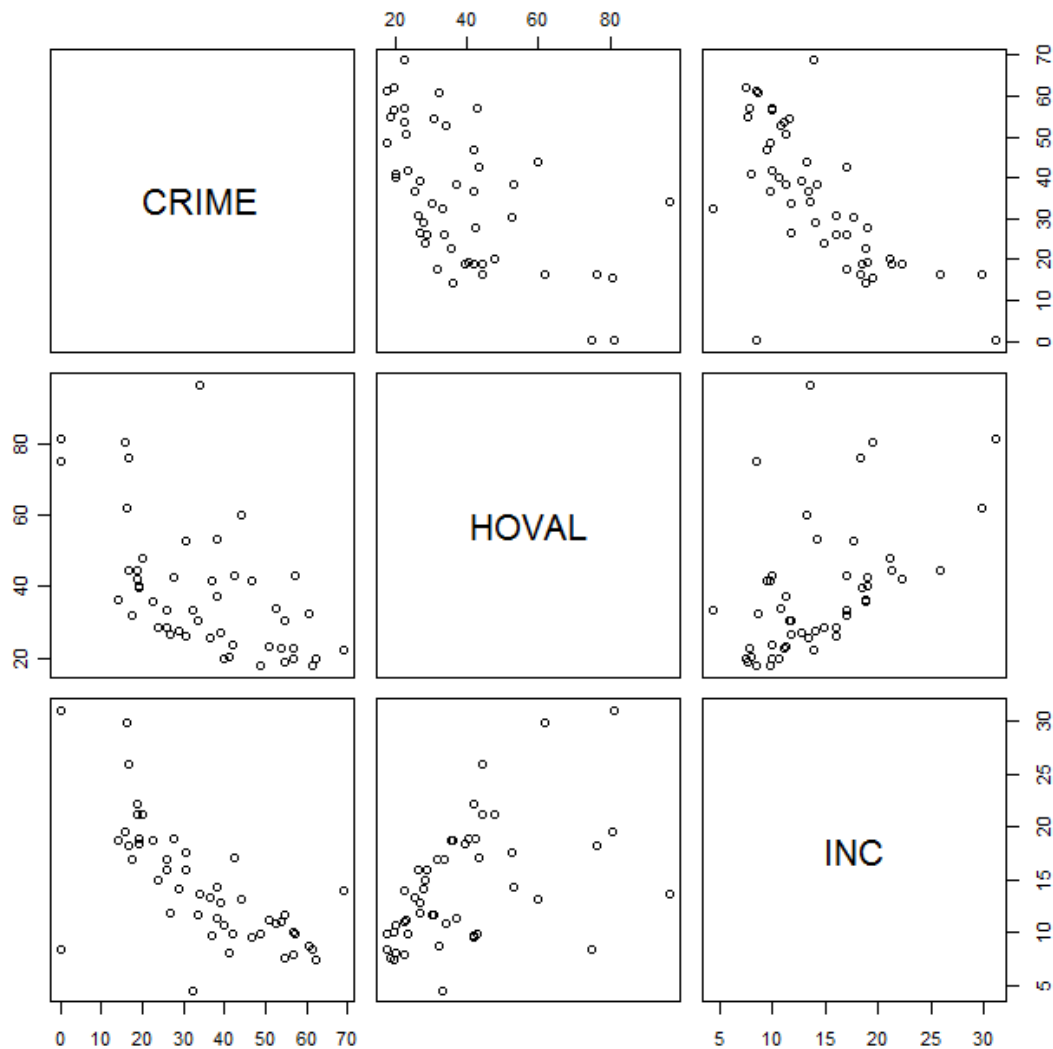


CRIME variable: the roughly normal shape of the histogram, the symmetry of the boxplot with no outliers, and the moderate linearity of the points suggest a roughly normal distribution.

HOVAL variable: the positively skewed shape of the histogram, slight asymmetry of the boxplot with several outliers in extremely high values, and outliers with extreme values in relation to the qqplot line suggest a distribution that is positively skewed.

INC variable: the positively skewed shape of the histogram, slight asymmetry of the boxplot, and moderate linearity of the points suggest a distribution that is slightly positively skewed.

2. The scatterplot matrix for CRIME, HOVAL, and INC variables are shown below.



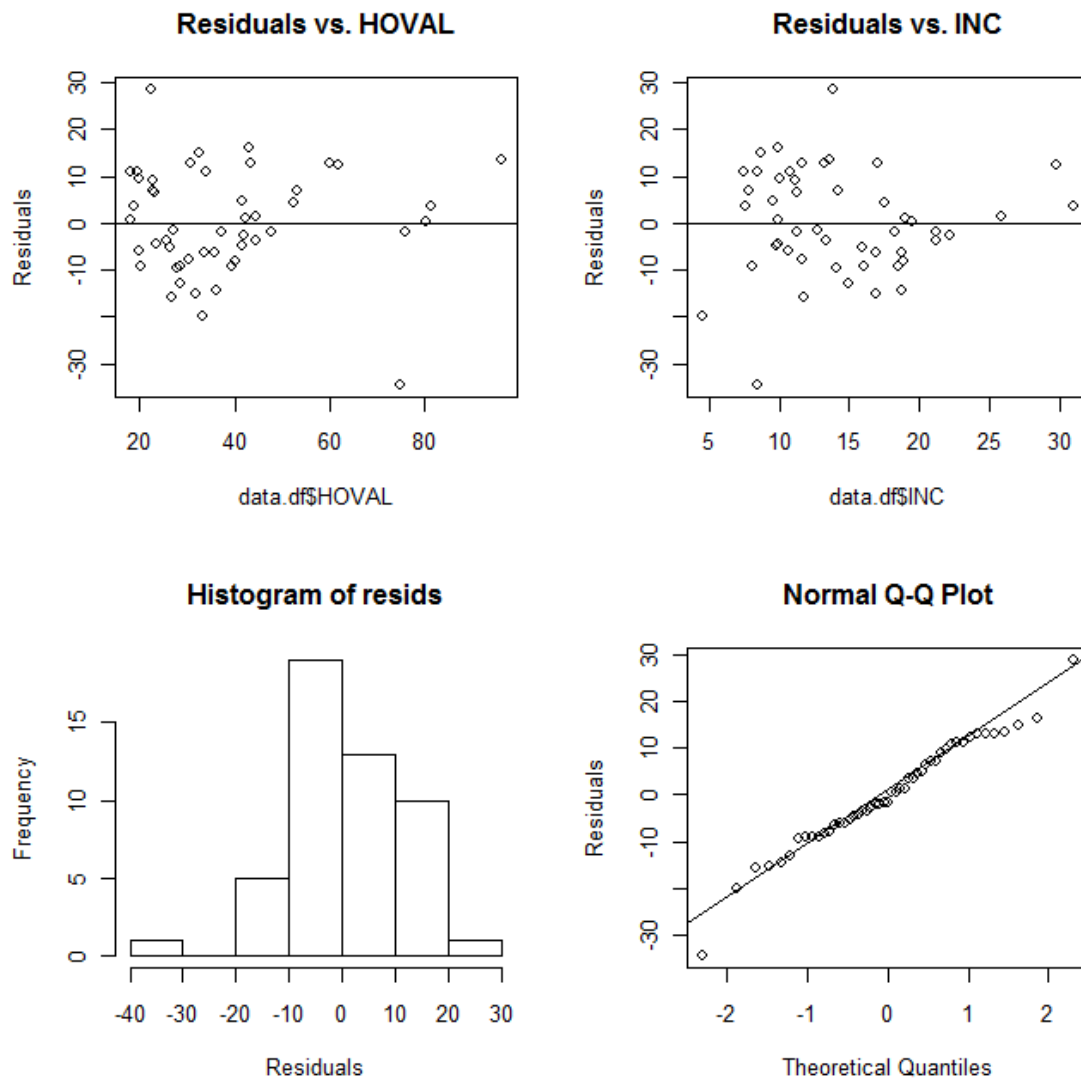
CRIME seems to have a negative relationship with HOVAL and INC; as housing value and household income increase, the crime rate decreases. The strength of the linear relationship appears quite strong, with some points dispersed.

HOVAL and INC are positively associated; as HOVAL increases, INC increases and vice versa. The strength of the linear relationship seems fairly strong, with more dispersed points.

3. The fitted equation is shown below.

$$\text{CRIME} = 68.6190 - 0.2739 \cdot \text{HOVAL} - 1.5973 \cdot \text{INC}$$

4. The plots are shown below.



Histogram of residuals shows that residuals are negatively skewed from normality, and QQplot shows that there are several outliers. From “Residuals vs. HOVAL”, we can see that the variances of residuals varies a lot with different HOVAL values, thus the assumption of constant variance is systematically violated. From “Residuals vs. INC”, we can observe that the variance of residuals seem to be more constant for different INC values than that of “Residuals vs. HOVAL”. To sum up, assumptions of linear regression are not fitted well.