

## Tutorial 2

1. Forced Expiratory Volume (FEV) is an index of pulmonary function that measures the volume of air expelled after 1 second of constant effort. The dataset `FEV.csv` contains measurements for 654 children aged 3 to 19 years of age. The purpose of the data collection was to study how FEV is affected by certain other variables. The variables that we shall work with are

**Age:** Age in years.

**FEV:** FEV measurement.

**Hgt:** Height in inches.

**height:** Height in meters

**Sex:** 0 = female, 1 = male.

**Smoking status:** 0 = current non-smoker, 1 = current smoker.

- (a) What is the response variable in this study?
- (b) Create a histogram of FEV and comment on it.
- (c) Create a boxplot of FEV and identify how many outliers there are. Investigate your data and comment on these outliers.
- (d) Generally, is the sample of FEV normally distributed?
- (e) Create separate histograms for male and female FEV, then obtain separate numerical summaries for males and female FEV. Comment on what you observe.
- (f) Create a scatterplot with height (in metres) on the  $x$ -axis and FEV on the  $y$ -axis.
- (g) Compute the correlation between FEV and height and comment on your results.