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` on LumiNUS 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.

   Hgt_m: 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) Create separate histograms for male and female FEV, then obtain separate numerical summaries for males and female FEV. Comment on what you observe.

   (e) Create a scatterplot with height (in metres) on the x-axis and FEV on the y-axis.
   Hint: In R, use command `plot(x,y)` to create a scatter plot of `y` vs `x`, where vector `x` is the vector of values of the explanatory and `y` is the vector of values of the response.

   (f) Compute the correlation between FEV and height and comment on your results.
   Hint: In R, use command `cor(x,y)` to calculate the correlation value of the two vectors `x` and `y`.

2. Question of interest: Do male college students follow their school's teams more closely than female?

   The following data was collected in class on Monday morning, after a particularly exciting and important basketball game. The question asked was: Did you watch the game on TV last night?

	Whole Game	Part of the Game	None of It	Total
Male	10	12	4	26
Female	21	24	30	75
Total	31	36	34	101

   (a) What is the response variable and explanatory variable in this study?

   (b) For each gender, find the percentage of watching game (all, part and none).

   (c) State the pairs of percentages for comparison. Plot a bar plot which helps to compare the percentages found.

   (d) Is it fair to say that males were more likely to watch the game than females?

   (e) Can the conclusion above be applied to the population of all students?