Lab 4: Univariate and Bivariate Data Exploration

1) A study was done on a sample of first year students at a university. The study measured the students resting pulse rate, height and weight as well as noting their gender.
   The data is available in pulse.csv.
   a) Construct a histogram of the resting pulse rate of all the students. Comment on what the frequency distribution indicates about the distribution of students' resting pulse rate.
   b) Determine the five-number summary for the resting pulse rate of males and females separately.
   c) Draw a comparative box-plot of resting pulse rate as a function of gender, ensuring that all axes are well labelled.
   d) Comment on any similarities and differences (as shown in the comparative box-plot) in the resting pulse rates of males and females.
   e) Find the mean and standard deviation of the resting pulse rate of each of the samples of males and females.
   f) Compare the mean with the median of each gender. What do these comparisons indicate about the distributions?
   g) Construct a suitable graph to determine if there is an association between gender and smoker.

2) A sample of blood pressure for a group of adult aged 24 to 50 was collected and the data is available in bp.csv. These subjects are divided into 3 groups based on the systolic and diastolic blood pressure values:

   NORMAL – both the systolic blood pressure is within its normal range (between 90 and 139) and the diastolic blood pressure is within its normal range (between 60 and 89).
   HIGH – if either measurements of blood pressure are above their normal upper limits.
   LOW – if either measurements are below their normal lower limits.

   a) What is the distribution of age in this dataset?
   b) Construct a suitable graph to determine if there is an association between systolic blood pressure and bmi.
   c) Construct a suitable graph to determine if there is an association between blood pressure group and bmi.
   d) Construct a suitable graph to determine if there is an association between systolic blood pressure and diastolic blood pressure.
   e) Compute the sample correlation coefficient for systolic blood pressure and bmi. Does it confirm your impression from your plot in part (a)?
   f) Compute the sample correlation coefficient for systolic blood pressure and diastolic blood pressure. Does it confirm your impression from your plot in part (d)?