Tutorial 3

1. Consider dataset `gasoline.csv`, which includes the gasoline mileage performance for 32 different automobiles. The data frame contains the following columns:

   y: response variable, the gasoline mileage performance (miles/gallon).

   $x8$: overall length of vehicle (inch).

   $x9$: width of vehicle (inch)

   $x10$: weight of vehicle (lb)

   $x11$: type of transmission (1=automatic, 0=manual).

   The answers for questions below should be derived by R.

   (a) Produce QQ plot of the response and comment if the response variable is approximately normally distributed.

   (b) Plot a histogram and a density plot overlay on the histogram for the response variable. Produce the boxplot of the response. Give your comment about these plots.

   (c) In the given dataset, how many vehicles are the automatic transmission type and what is the range (min, max) of the gasoline mileage performance for these vehicles?

   (d) In the given dataset, how many vehicles are the manual transmission type and what is the range (min, max) of the gasoline mileage performance for these vehicles?

   (e) Produce boxplots for the response variable classified by type of transmission.

   (f) Produce a scatter plot for the gasoline mileage performance vs the overall length of vehicle. Give you comment.

   (g) Produce a scatter plot for the gasoline mileage performance vs the weight of vehicle, classified by the type of vehicle, then add a legend for the plot.

2. Derive answers for all questions above by Python.