Lab 2: Univariate Data Exploration

1) A social researcher in a particular city wishes to obtain information on the number of children in households that receive welfare support. A random sample of 400 households is selected from the city welfare rolls. A check on welfare recipient data provides the number of children in each household.
   a) Identify the population of measurements that is of interest to the researcher.
   b) Identify the sample.
   c) What characteristics of the population are of interest to the researcher?

2) The nine measurements that follow are furnace temperatures recorded on successive batches in a semiconductor manufacturing process (units are F):

   953   955   948   951   949   954   950   959

   a) Calculate the sample mean and sample standard deviation.
   b) Find the sample median of the data.
   c) How much could the largest temperature measurement increase without changing the sample median?

3) A survey on knee injuries recorded the following data on type of injury (A=meniscal tear, B=MCL tear, C=ACL tear, D=patella dislocation, E=PCL tear), the data is given in injury.csv
   a) Construct a bar chart for this data.
   b) Construct a pie chart for this data.

4) A small part for an automobile rear-view mirror was produced on two different punch presses. In order to describe the distribution of the weights of those parts, a random sample was selected, and each piece was weighed in grams, resulting in the data set weight.csv
   a) What is the average weight?
   b) Find the variance of the weights.
   c) Find the five-number summaries of this dataset.
   d) Compute the interquartile range of the weights.
   e) Draw a boxplot of the data with x-axis and y-axis labelled.
   f) Draw a histogram of the data with x-axis and y-axis labelled.
   g) Comment on your boxplot and histogram.

5) In an experiment designed to study the behaviour of certain individual cells that had been exposed to beryllium, the interdivision times (IDTs) of cells were deter-mined for a large number of cells both in exposed (treatment) and unexposed (control) conditions. The data is given in IDT.csv
   a) Construct a histogram of this data.
   b) Calculate $\log_{10}(x)$ for each observation.
   c) Construct a histogram of the transformed data.
   d) What is the effect of the transformation?