* Q2.2 – Suppose that data for analysis include the attribute *age*. The *age* values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.
  + A) What is the *mean* of the data? What is the *median*?
    - Mean = 29.96
    - Median = 25
  + B) What is the *mode* of the data? Comment on the data’s modality (i.e., bimodal, trimodal, etc.)
    - Mode = 25 and 35
    - Data is bimodal because it has two modes (25 and 35), which both appear four times in the data set.
  + C) What is the *midrange* of the data?
    - Midrange = 41.5
  + D) Can you find (roughly) the first quartile (Q1) and the third quartile (Q3) of the data?
    - First quartile (Q1) = 20
    - Third quartile (Q3) = 35
  + E) Give the *five-number summary* of the data.
    - Minimum = 13
    - First Quartile (Q1) = 20
    - Median = 25
    - Third Quartile (Q3) = 35
    - Maximum = 70
  + F) Show a *boxplot* of the data (see image below):



* + G) How is a *quantile-quantile plot* different from a *quantile plot*?
    - In a quantile plot, the individual data points in a univariate distribution are plotted against their calculated quantile value. However, a quantile-quantile plot will plot data points from one univariate distribution against the corresponding quantile value data points of a separate univariate distribution. The quantile-quantile plot allows the user to see if there is any shift or trends between two separate distributions.
* Q2.5 Briefly outline how to compute the dissimilarity between objects described by the following:
  + A) Nominal attributes:
    - Nominal attributes values are typically symbols or names of things, which make them categorical. To find the dissimilarity between two nominal attributes, one needs to calculate the ratio of mismatches by subtracting the total number of matches from the total number of attributes and dividing by the total number of attributes.
  + B) Asymmetric binary attributes
    - In an asymmetric binary attribute, it is decided that one outcome is more important than the other. In these cases, the negative matches (both objects are “0”) are ignored as unimportant, and the mismatches are summed up and divided by the total comparisons (minus unimportant matches).
  + C) Numeric attributes
    - Numeric attributes dissimilarity can be calculated by several different methods. The first method is Euclidean, which calculates the straight-line distance between two objects. Another method is the Manhattan method, which calculates the distance between two points as you would move through city blocks (X blocks over + Y blocks up/down). Supremum method is also used for numeric attributes. This method reviews all the attributes describing objects and calculates the attribute with the maximum distance between objects.

