Name: Vaibhav Panchanadam
Student Number: 101311133

Question 1:

A:
The mean of any given dataset can be calculated using the following formula:

$$\bar{x} = \frac{\sum_{i=0}^{n} x_i}{n}$$

Which essentially states that the mean is equal to the sum of all items in the dataset (represtned by x_i) divided by the total number of items in the dataset (represented by *n*). For the provided data, the mean would be:

$$\bar{x} = \frac{\sum_{i=1}^{30} x_i}{30} = \frac{\sum_{i=1}^{30} x_1 + x_1 ... + x_{30}}{30} = \frac{\sum_{i=1}^{30} 11.76 + 11.79 + ... + 12.53}{30}$$

Adding up the items in the numerator (per row),

$$\bar{x} = \frac{119.18 + 121.5 + 213.7}{30}$$

$$\bar{x} = \frac{364.38}{30}$$

$$\bar{x} = 12.146$$

Therefore, the mean speed of the runners is 12.146 seconds.

To calculate the variance of a given dataset, the following formula would be used:

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} (x_i)^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}\right]$$

For the given dataset, the formula would appear as follows (inner term of the numerator of the fraction being the numerator of the previous equation):

$$s^2 = \frac{1}{29}\left[\sum_{i=1}^{30} (x_i)^2 - \frac{(\sum_{i=1}^{30} x_i)^2}{30}\right]$$

$$s^2 = \frac{1}{29}\left[(1420.477 + 1476.2548 + 1530.2608) - \frac{(364.38)^2}{30}\right]$$

$$s^2 = \frac{1}{29}\left[(4426.9926) - \frac{(364.38)^2}{30}\right]$$

$$s^2 = \frac{1}{29}[(4426.9926) - \frac{(364.38)^2}{30}]$$
$$s^2 = \frac{1}{29}[(4426.9926) - 4425.75948]$$
$$s^2 = \frac{1}{29}[(4426.9926) - 4425.75948]$$
$$s^2 = \frac{1}{29}[1.23312]$$
$$s^2 = 0.042521$$

Therefore, the variance of the dataset of 30 runners and their times is (rounded to 6 decimal places) 0.042521.

B:

To calculate the median of a dataset, you use the formula below to find the position of the median (given that the dataset is sorted):

$$Median = \frac{n+1}{2} th\ term$$
$$Median = \frac{31}{2} th\ term$$
$$Median = 15.5 th\ term$$

Since the term is in between the 15th and 16th terms, you would take the average of these two terms, and this would represent the median. These would be:

$$15th\ term = 12.14,\ 16th\ term = 12.15$$

Therefore,

$$Median = \frac{12.14+12.15}{2}$$
$$Median = 12.145$$

The median of the dataset of runners and their times is 12.145.

The mean is greater than the median (12.146>12.145), however, due to the size of the difference being much smaller than the differences between values (being 0.001), the distribution can be assumed to be mound-shaped, with a singular large mound in the middle of the distribution.

C:

Consult the next page:

Stem	Leaf
117	6 9
118	2 7
119	4 7 8
120	1 1 3 5
121	0 1 2 4 5 8
122	0 2 3 5 5 8
123	2 4 7
124	0 6
125	0 3

Key:
117|6 = 11.76

D:
Based on the stem-leaf plot the graph has a very slight left-skew, however, the skew is generally negligible (due to the size of the values), making it once again mound-shaped.

E:
Firstly, using the previously found value for s^2, calculate the value of s:

$$s = \sqrt{s^2}$$
$$s = \sqrt{0.042521}$$
$$s = \sqrt{0.042521}$$
$$s = 0.206206$$

Now, from the mean, add and subtract the given times:

$$\bar{x} + 2s = 12.146 + 2(0.206206)$$

$$\bar{x} + 2s = 12.146 + 0.412412$$
$$\bar{x} + 2s = 12.5584$$

Similarly,

$$\bar{x} - 2s = 12.146 - 0.412412$$
$$\bar{x} - 2s = 12.146 - 0.412412$$
$$\bar{x} - 2s = 11.7336$$

Based on $\bar{x} + 2s = 12.5584$ and $\bar{x} - 2s = 11.7336$, we can determine that all the values within the dataset are contained within $x \pm 2s$.

F:

Based on the generally mound-shaped nature of the graph, this is fairly close to the empirical rule. According to the empirical rule, 95% of the data is contained within 2 standard deviations of the mean. In this instance, 100% of the data is contained instead, likely related to the very slight left-skew occurring in the graph. Therefore, overall, the graph does generally follow the empirical rule.

Question 2:

A:

Quantitative Continuous- This value is quantitative as it is a measure (given by the word "amount"), and is additionally continuous as it is possible to wait in smaller proportions than whole numbers of time.

B:

Qualitative- This is qualitative as it utilizes the category of brand of cell phones to classify the data, rather than a specific number (for example, the *number* of cell phones one has).

C:

Quantitative Discrete- This value is quantitative as it refers to the number of courses that a student takes, and is discrete as (barring the possibility of a student dropping a course) it is not possible to take half a course.

D:

Qualitative- This is qualitative as the measurement of intensity can vary from person to person. Additionally, the measurements of pain are categories rather than numbers.

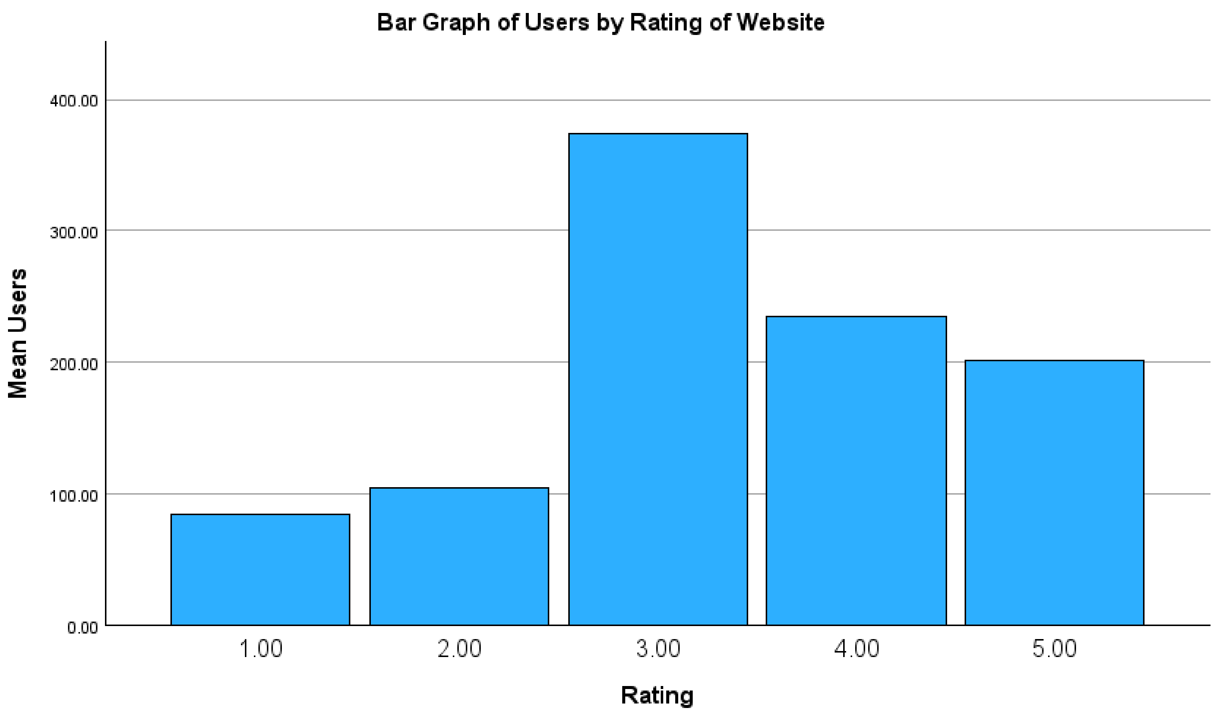
E:

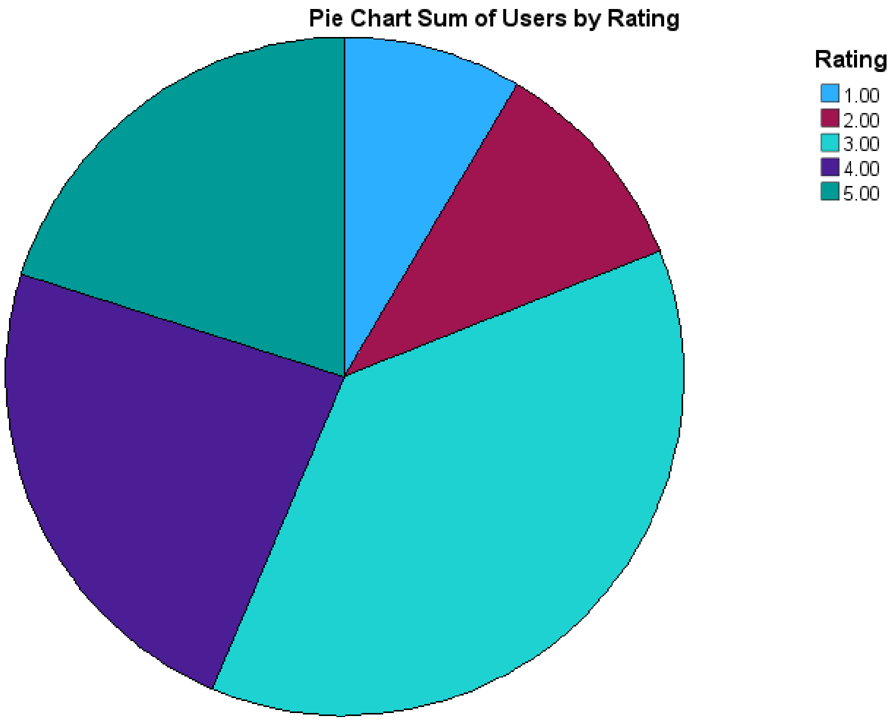
Quantitative Discrete- This value is quantitative as is refers to the *number* of patients a doctor treats, and is discrete because it is not possible for a doctor to treat half a patient (only whole-number measurements are possible)

F:
Qualitative- This is qualitative as, similar to the previous qualitative example, the measure of "length of assignment" is not consistent among who you ask. Additionally, sorting the data this way sorts it into categories rather than numbers, making it qualitative.

Question 3:

Using SPSS, the graphs (bar and pie) are attached below (images exported from SPSS, document linked):





Question 4:

A:

Using SPSS, the quartiles are attached below (being the "weighted average" section, or the first row):

Percentiles

		Percentiles 25	50	75
Weighted Average (Definition 1)	Reaction	.4075	.4800	.5325
Tukey's Hinges	Reaction	.4100	.4800	.5300

B:

The formula for the position of the *p*th percentile is as follows:

$$Position = \frac{p}{100}(n + 1)$$

The third quartile is calculated using the 75th percentile (or rather, p=75). Substituting the values of the dataset into the formula:

$$Position = \frac{75}{100}(50 + 1)$$

$$Position = \frac{3}{4}(51)$$

$$Position\ =\ 38.25$$

Therefore, the value is at the 38.25th position.

To calculate the actual value, you would need to round *p* down to the nearest whole number (meaning 38). Then, you'd have to use the difference between *p* and *rounded p*, and multiply this by the difference of the nearest whole numbers (meaning *rounded p*, and *rounded p*+1).

$$3rd\ Quartile\ =\ 38th\ Position\ +\ 0.25(39th\ Position\ -\ 38th\ Position)$$

$$3rd\ Quartile\ =\ 0.53\ +\ 0.25(0.54\ -\ 0.53)$$

$$3rd\ Quartile\ =\ 0.53\ +\ 0.25(0.01)$$

$$3rd\ Quartile\ =\ 0.5325$$

Therefore, the third quartile of the given dataset is 0.5325.

C:

[drawing on next page, calculations done here]

IQR = 3rd Quartile-1st Quartile

IQR = 0.5325-0.4075

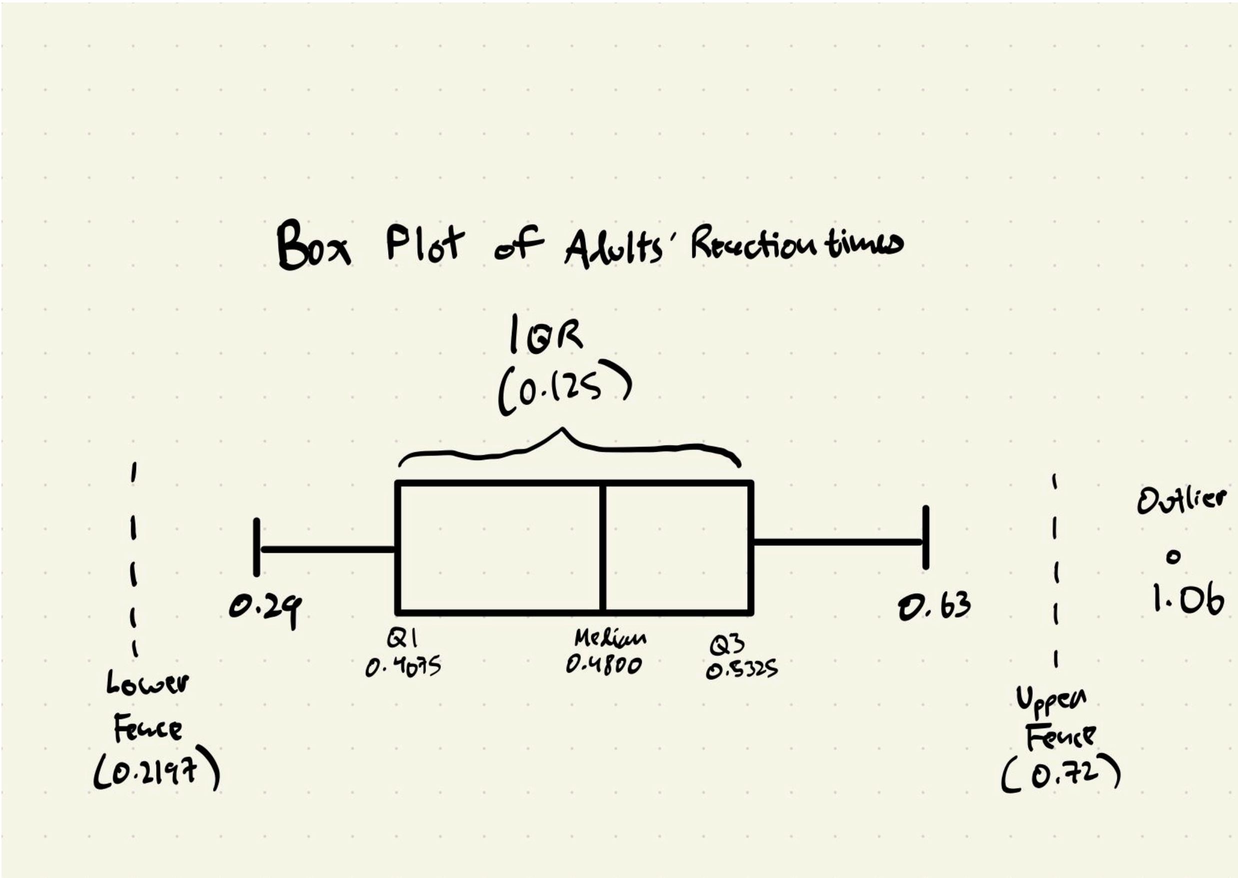
IQR = 0.125

Upper fence = 0.5325 + 1.5(0.125)

Upper fence = 0.72

Lower fence = 0.4075-1.5(0.125)

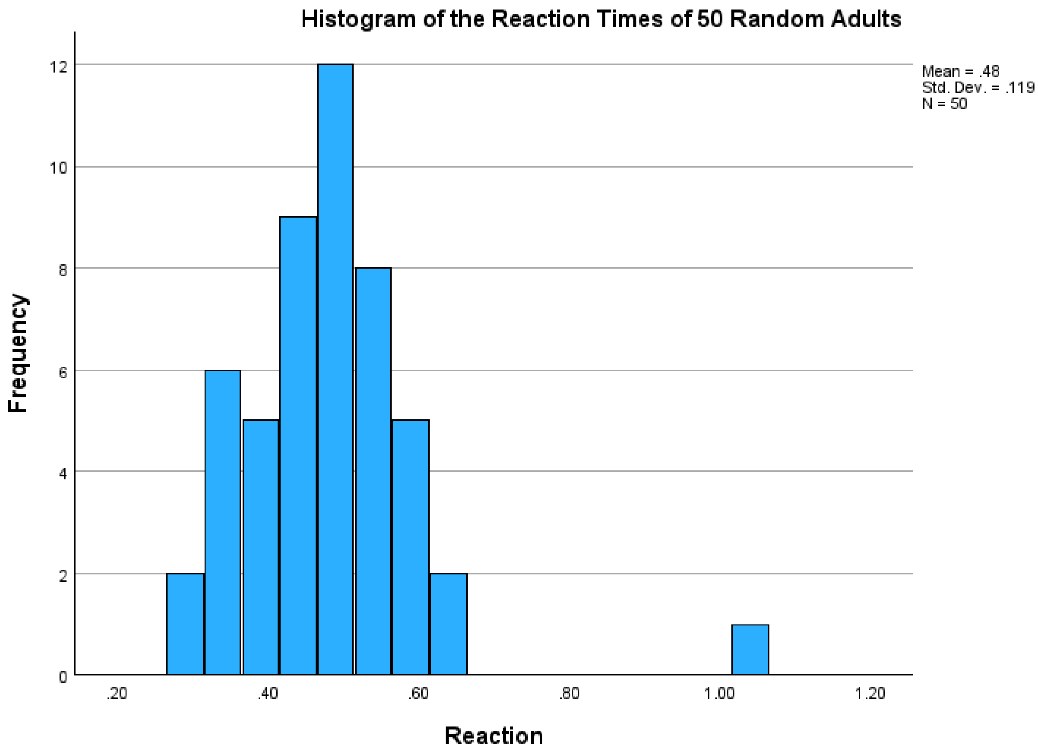
Lower fence = 0.2197



(drawn by hand, digitally, on a tablet)

D:

Displayed below is the histogram of the dataset in the question:



Based on the histogram, one can conclude that the graph is generally mound-shaped (no skew), with a slight left skew. However, it must be noted that there is an outlier present in the graph, being the value of 1.06.