Exercise 3.4 (Group exercise) - Solution

a. In order to plot the Tukey box plot five main features are required as shown in Table C.8 in the lecture notes. These are:

- the lower quartile

- the lower adjacent value

- the median

- the upper adjacent value

- the upper quartile

Consider the data of resistivity in direction 1. Based on Equation C.10 from the lecture notes a value v is required such that

$$v = nQ_v + Q_v$$

Therefore for the lower quartile (i.e. the 0.25 quartile) it is:

$$v = 25 \cdot 0.25 + 0.25 = 6.5$$

v has a non integer value. The value is splitted to its integer part $k = 6$ and the fractional part $p = 0.5$

Therefore the lower quartile is calculated as:

$$x_v^o = (1 - p)x_6^o + px_{6+1}^o = (1 - 0.5) \cdot 24.7 + 0.5 \cdot 25.3 = 25 \text{ kOhm}$$

In the same way for the upper quartile (the 0.75 quantile) it is:

$$v = 25 \cdot 0.75 + 0.75 = 19.5$$

v has a non integer value. The value is splitted to its integer part $k = 19$ and the fractional part $p = 0.5$.

Therefore the upper quartile is calculated as:

$$x_v^o = (1 - p)x_{19}^o + px_{19+1}^o = (1 - 0.5) \cdot 28.3 + 0.5 \cdot 28.7 = 28.5 \text{ kOhm}$$

In order to calculate the median it is:

$$v = 25 \cdot 0.5 + 0.5 = 13$$

The 13th value of the data set is equal to 26.9 kOhm.

To evaluate the adjacent values the interquartile range is required:

$$r = Q_{0.75} - Q_{0.25} = 28.5 - 25 = 3.5$$

The lower adjacent value is the smallest observation that is greater than or equal to the lower quartile minus *1.5r.* It is:

$$Q_{0.25} - 1.5r = 25 - 1.5 \cdot 3.5 = 19.75 \text{ kOhm}$$

Thus from Table 3.1.1 the lower adjacent value is 20.2 kOhm.

In the same way the upper adjacent value is found as:

$Q_{0.75} + 1.5r = 28.5 + 1.5 \cdot 3.5 = 33.75$ kOhm

Therefore from Table 3.4.1 the upper adjacent value is a value less than or equal to 33.75 kOhm, that is 29.9 kOhm. Table 3.4.2 summarizes the above features of both data sets.

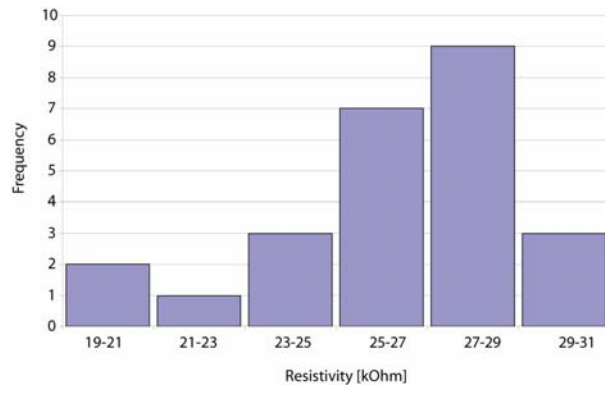|	Direction 1	Direction 2
Lower adjacent value	20.2	3.8
Lower Quartile	25.0	8.65
Median	26.9	11.1
Upper Quartile	28.5	14.2
Upper adjacent value	29.9	17.8
Outside values	-	23.4

Table 3.4.2: Descriptive statistics.

The Tukey box plots are shown in Figure 3.4.1.
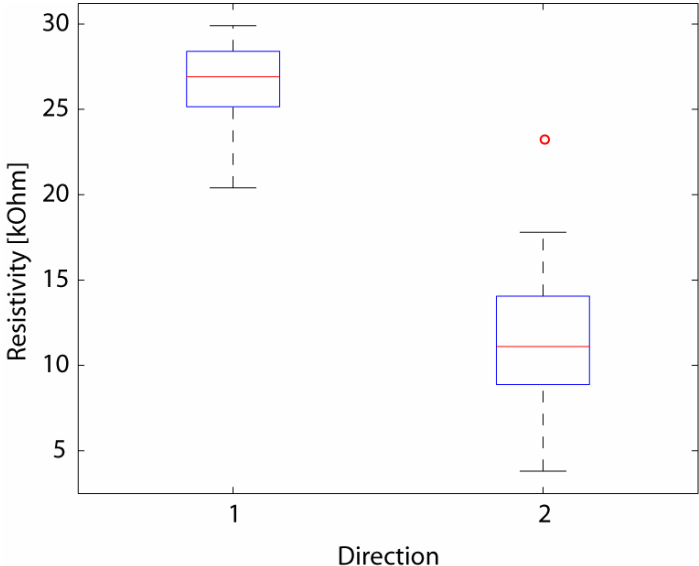
Figure 3.4.1:   **Tukey box plot for direction 1 and 2.**

b. Both are left skewed, although the data of direction 2 are less skewed.

c. It is proposed that the number of intervals k is: $k = 1 + 3.3 \cdot \log(n)$, where n is the number of data. By substituting $n = 25$, k is obtained as: $k = 1 + 3.3 \cdot \log(25) \approx 6$. From the histogram in Figure 3.4.2, it is seen that the distribution is skewed to the left.

Figure 3.4.2: **Histogram.**