Show your working in all calculations.

3. a) Suppose you are given the data in Figure 2

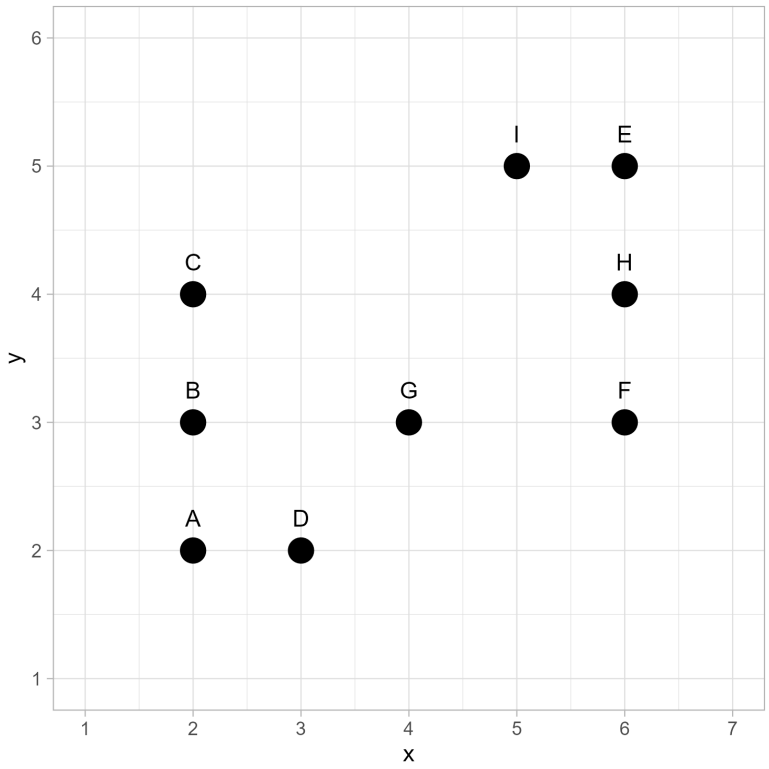

Figure 2: Data for Question 3.

You are asked to use the simple k-means clustering to group this data, with the following specifications:

- Use $k = 2$ with initial centroids $c_1 = (4, 1)$ and $c_2 = (5, 1)$;
- Use the *Manhattan distance* to calculate all dissimilarities.

Perform TWO iterations of the k-means algorithm as defined above. You should detail your process and calculations [8%], and report: (i) the final position of the centroids [4%], and (ii) the final attribution of points to each cluster [4%]. You can use the letter identifiers from figure 2 to show which points are allocated to each cluster.

b) Choose either DBSCAN or agglomerative hierarchical clustering (pick only one). Provide a short explanation of how your chosen method works. Include the general idea behind the method, and a sketch of the main steps (or a pseudocode) of the method [8%]. Indicate at least one advantage and one drawback of your selected method [6%].

(Total: 30% of the exam marks)

END OF ASSESSMENT