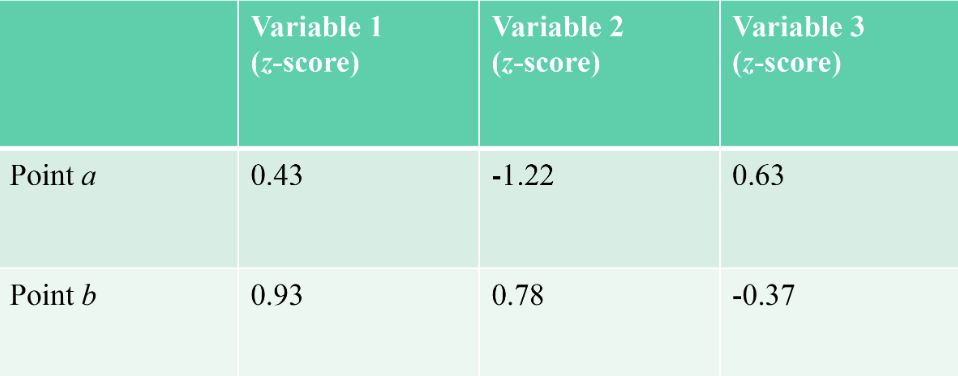
Question 3

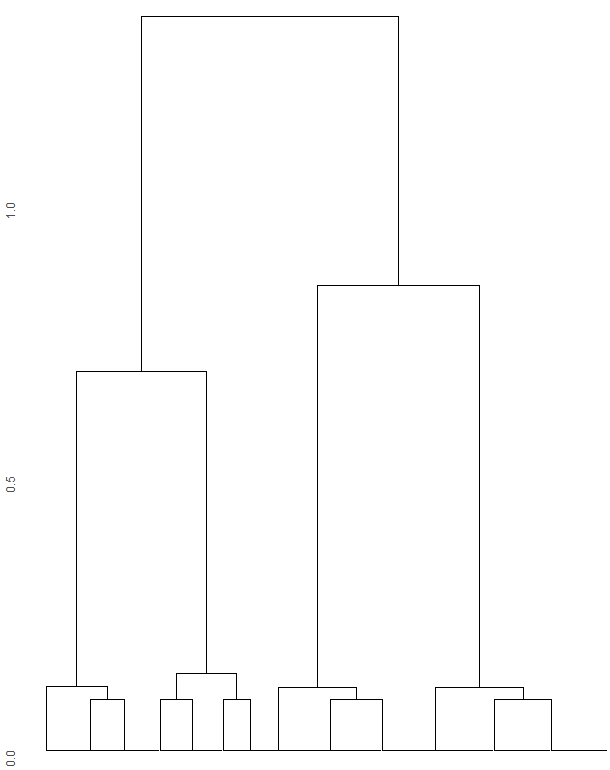
Please review the following chart, and respond to the question.



What is the (normalized) Euclidean distance between point a and point b?

**Question 6**

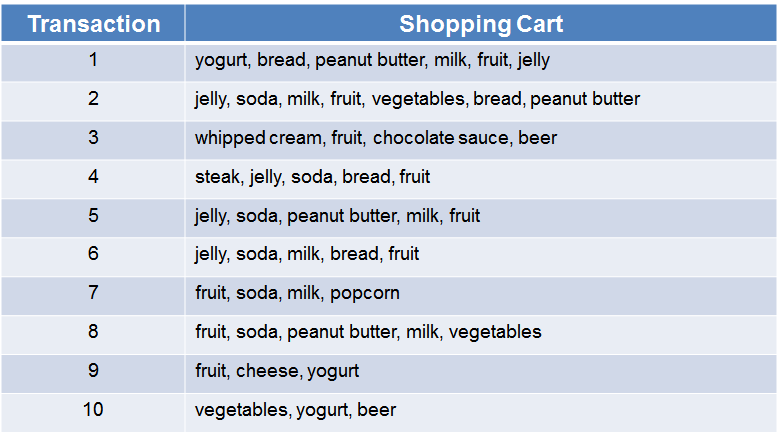
Based on the dendrogram shown, would it be most appropriate to split the data set into three, four, or five clusters?



* Three
* Four
* Five

**Question 7**

Based on the image below, please answer the following questions.



What is the support of {peanut butter, jelly}?



What is the confidence of the association rule: "if peanut butter, then jelly"?

Question 8

The HardwareStore.csv file contains data on 6495 transactions at a chain of hardware stores. Each column represents a different item sold by the store. Each row is a transaction; a 1 indicates that the customer bought that item, and a 0 that the customer did not buy the item. Import the data into R, convert the values to binary, and create a set of association rules with a minimum support of 0.005 and a minimum confidence of 0.1.

What is the confidence of the rule: “if Saw, then Screws”?



What is the highest lift ratio in the set of rules created?

Question 9

If we have a data set of car accidents that includes the following two binary variables:

* **SUV**(was an SUV involved?)
* **Major damage** (did the $ amount exceed 10,000?)

Consider the rule: “if **SUV**, then **major damage**”

If we work for an insurance company and are trying to determine the rate adjustment for insuring SUVs, which association rules metric(s) should we look at? Why?