

Using the next data points and the Euclidean equation to calculate the distance matrix. Moreover, it's important to highlight the notation where X represents the Sepal Width, and Y represents the Sepal Length.

	Data points	
A1	3	4.9
A2	3.2	4.7
A3	3.1	4.6
A4	3.6	5
A5	3.9	5.4
A6	3.4	4.6

$$\sqrt{(X(\text{centroid}) - X(\text{data point}))^2 + (Y(\text{centroid}) - Y(\text{data point}))^2}$$

(The Euclidean distance equation)

To create the distance matrix using the euclidean distance: **Tiny Hint: I have rounded the outputs to 4 digits.**

We need to calculate the interaction for each data point towards the centroids.

Point (A1,A1) =	0
Point (A1,A2) =	0.2828
Point (A1,A3) =	0.3162
Point (A1,A4) =	0.6083
Point (A1,A5) =	1.0296
Point (A1,A6) =	0.5

Point (A2,A2) =	0
Point (A2,A3) =	0.1414
Point (A2,A4) =	0.5
Point (A2,A5) =	0.9899
Point (A2,A6) =	0.2236

Point (A3,A3) =	0
Point (A3,A4) =	0.6403
Point (A3,A5) =	1.1314
Point (A3,A6) =	0.3

Point (A4,A4) =	0
Point (A4,A5) =	0.5
Point (A4,A6) =	0.4472

Point (A5,A5) =	0
Point (A5,A6) =	0.9434

Point (A6,A6) =	0
-----------------	---

(Let's build our distance matrix using these outputs)

The original Distance Matrix

	A1	A2	A3	A4	A5	A6
A1	0					
A2	0.2828	0				
A3	0.3162	0.1414	0			
A4	0.6083	0.5	0.6403	0		
A5	1.0296	0.9899	1.1314	0.5	0	
A6	0.5	0.2236	0.3	0.4472	0.9434	0

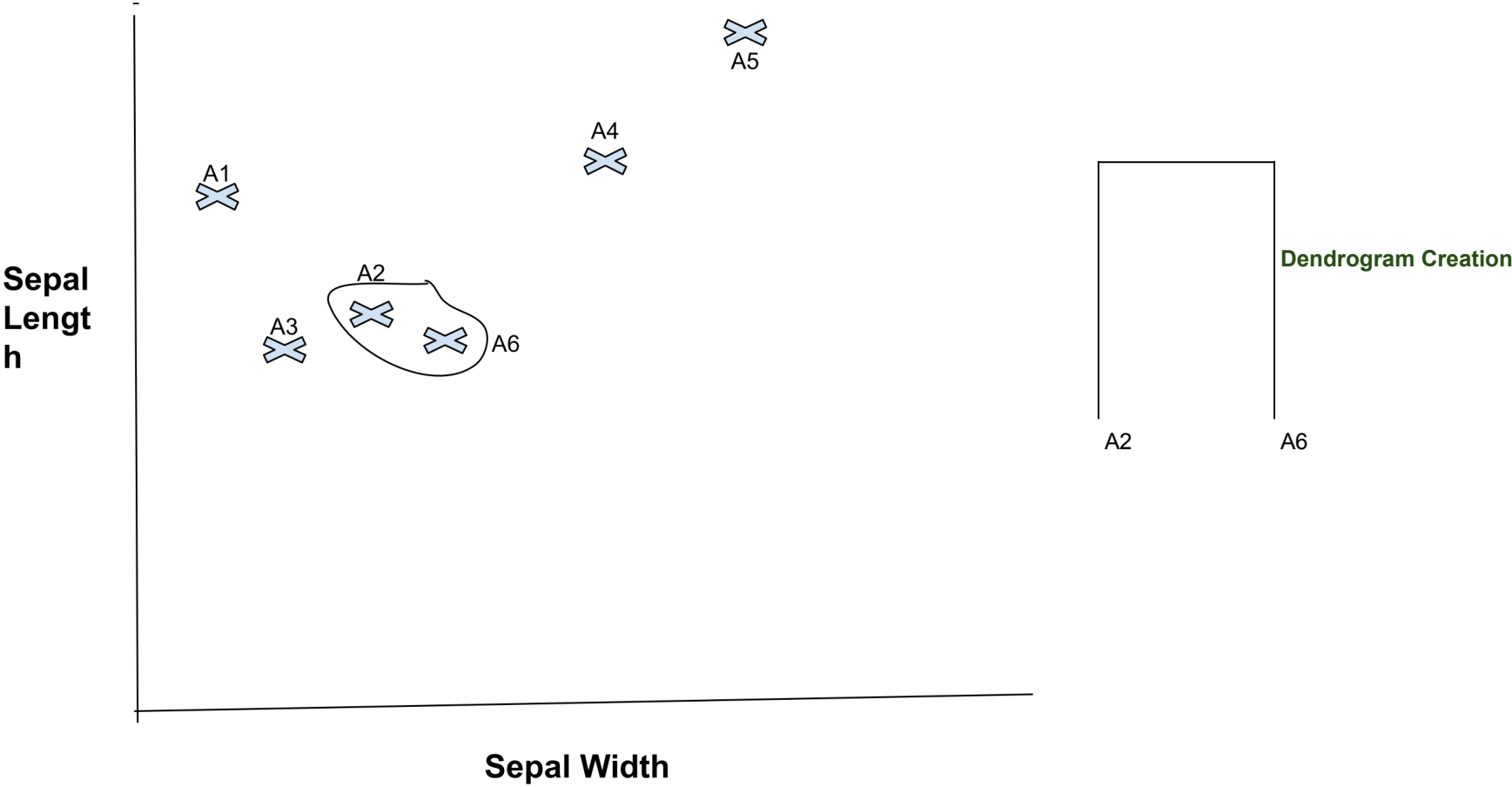
After getting the distance matrix I will use it to build up the dendrogram and scatter charts manually

-----> Kindly note that I am focusing on building the next calculations moving from the bottom of the distance matrix to upper side

	A1	A2	A3	A4	A5	A6
A6	0.5	0.2236	0.3	0.4472	0.9434	0

**Interpretation:**  
 Since I have A2 = 0.2236 and A6 = 0, which the both data points have the smallest values in the original distance matrix. Then we will merge these two points and the new updated distance matrix is:

	A1	A2	A3	A4	A5	A6
A1	0					
A2	0.2828	0				
A3	0.3162	0.1414	0			
A4	0.6083	0.5	0.6403	0		
A5	1.0296	0.9899	1.1314	0.5	0	
A6	Point (A1)	Point (A2,A6)	Point (A3)	Point (A4)	Point (A5)	



Interpretation:

This new cluster will be the baseline that help me during the computation.

Point (A2,A6) = 0.22

Moving to the points (A1,A3), let’s compute the single linkage of the Euclidean distance original matrix data points as following:

The Euclidean distance between the point (A2, A6) & the point (A1,A3) can be calculated as:

**Notation: D in the following equations stands for The Euclidean Distance**

$$\text{Min}( \text{Distance} (A2, A6), \text{Distance} (A1, A3))$$

$$\min(\text{Distance}(A2, A1), \text{Distance}(A1, A3))$$

When we break down this equation into more data points it will be far easier, we do that because we don't have the values of the distance (A1,A3) on the original distance matrix.

$$\min(D(A2, A1), D(A2, A3), D(A6, A1), D(A6, A3))$$

Using the original matrix to generate those numbers, we will have:

$$\min(0.29, 0.14, 0.5, 0.3)$$

**In conclusion:**

Since the smallest value is equal 0.14, then this new value is less than the Distance (A2,A6) = 0.22, we need to update the Euclidean distance matrix one more time as following:

	A1	A2	A3	A4	A5
A1	0				
A2	0.2828	0			
A3	0.3162	0.1414	0		
A4	0.6083	0.5	0.6403	0	
A5	1.0296	0.9899	1.1314	0.5	0

Point (A2,A6)