

1. Consider the data below. What is  $m$ ? What is  $p$ ? Draw a graph showing these points.

$$\mathbf{X} = \begin{bmatrix} 3 & 2 \\ 2 & 2 \\ 4 & -1 \end{bmatrix}$$

2. Compute the mean of these points, and plot it on your graph.
3. Compute the total sum of squares for this dataset.
4. Let's say we have  $k = 2$  clusters. What are the means of these clusters? Use the means to compute the within-cluster sum of squares for  $k = 2$ .
5. What if we have  $k = 3$  clusters? What is the within-cluster sum of squares?

6. **Elbow Plot:** Finally, compute the proportion of explained variance for  $k = 1, 2, 3$ , and draw a graph showing this pattern ( $k$  on the  $x$ -axis and proportion of explained variance on the  $y$ -axis).