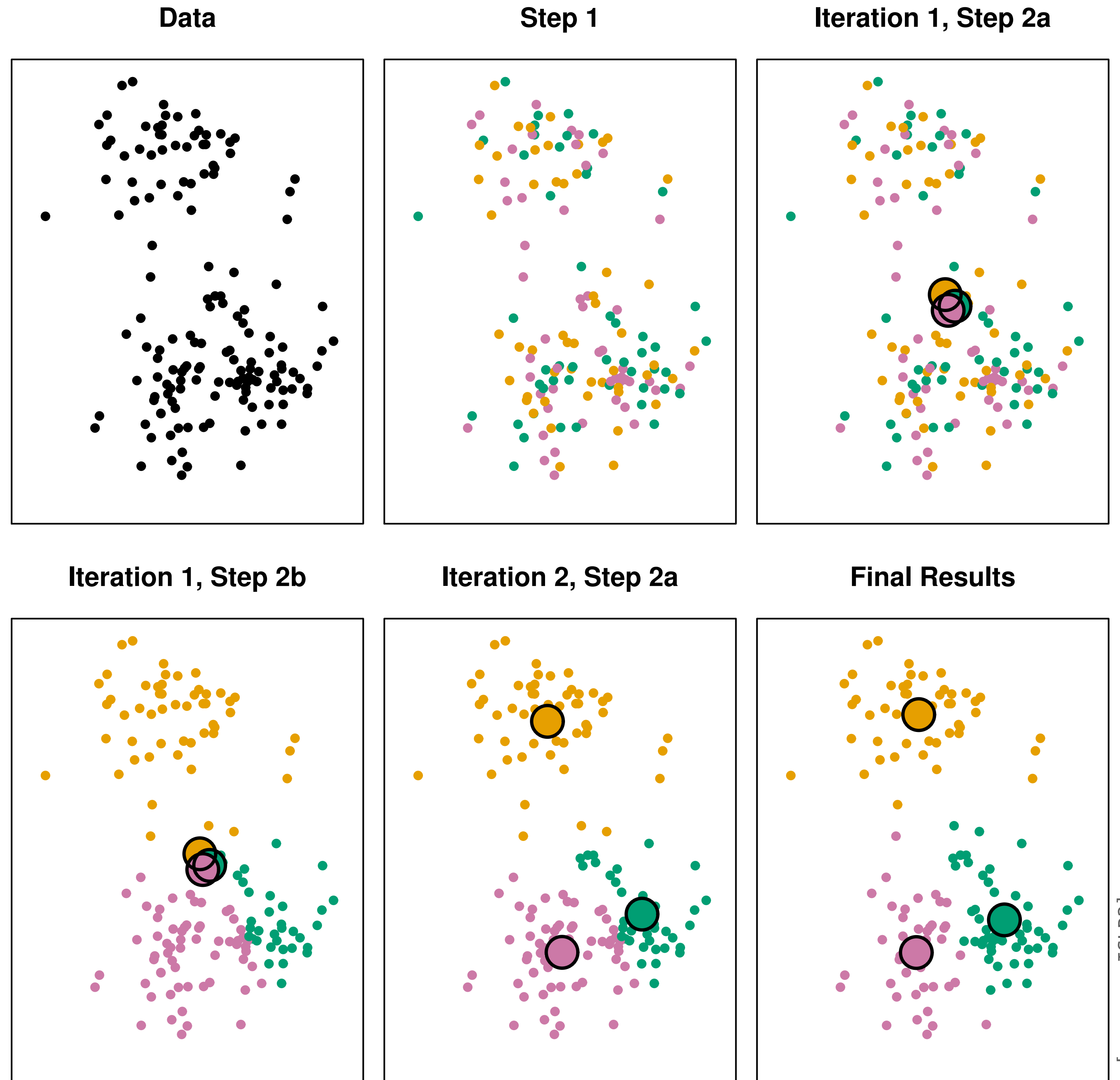


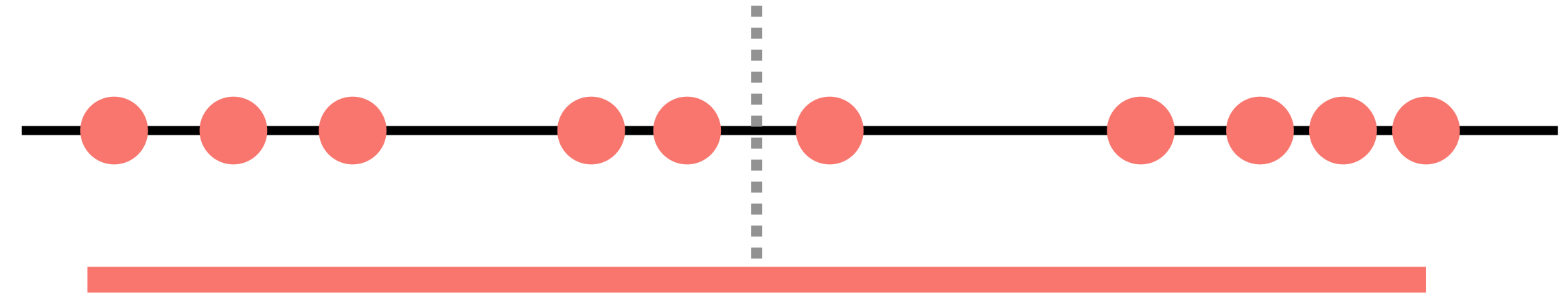
# K-Means: Algorithm

1. Choose **k** random points as cluster centers
2. For each data point, assign it the cluster whose centroid is the closest
3. Using these assignments, recalculate the centers
4. Reiterate from step (2) until **convergence**:
  - cluster membership does not change
  - center only changes very very little

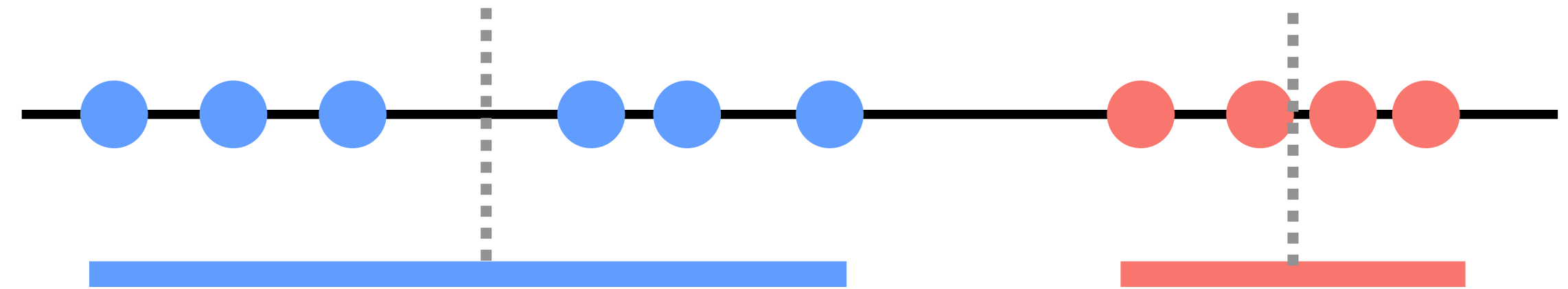


# How to decide on K

- Start with  $K=1$ : compute total variation



- Then  $K=2$ : compute total variation



- Then  $K=3$ : compute total variation

•  
•  
•

