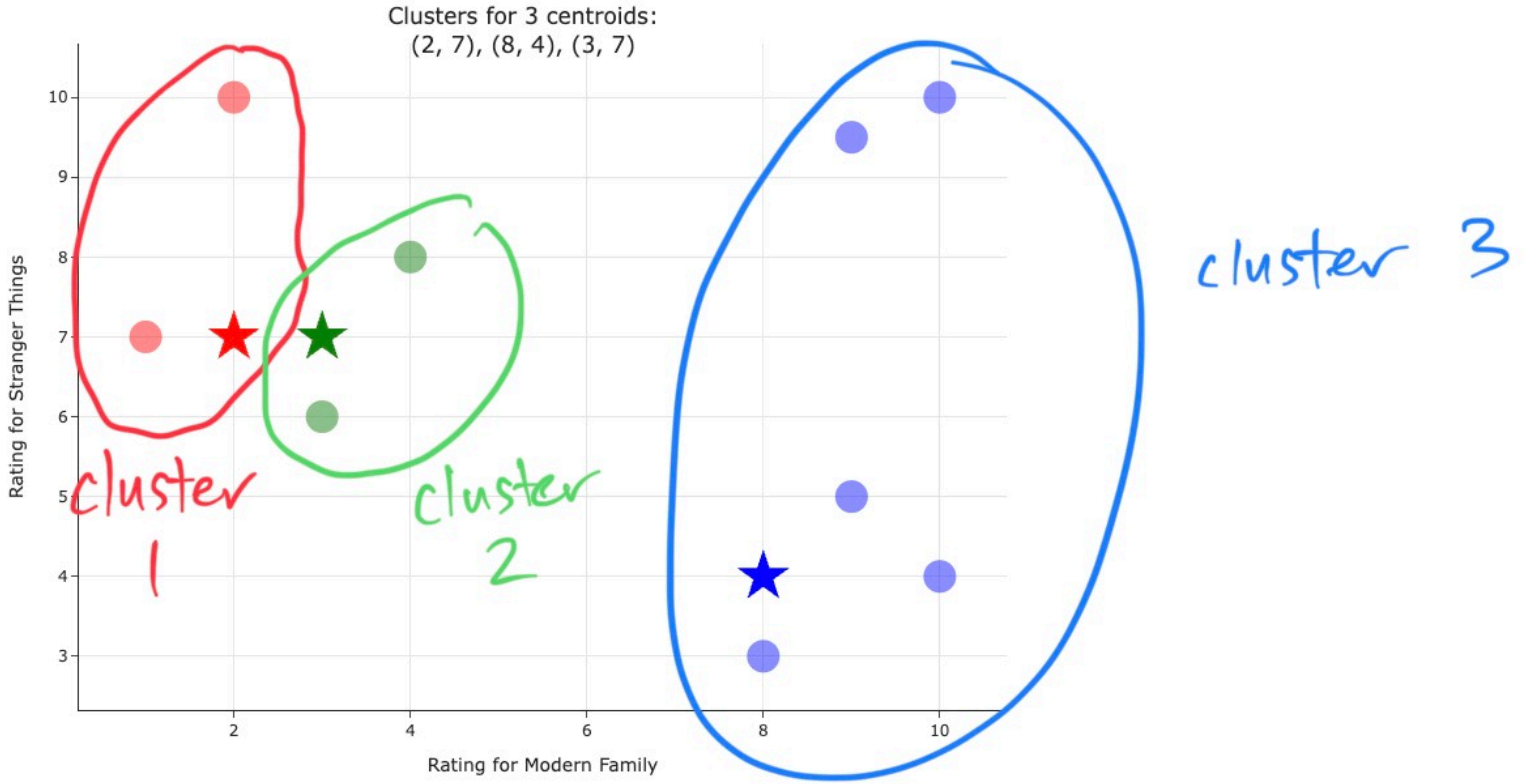
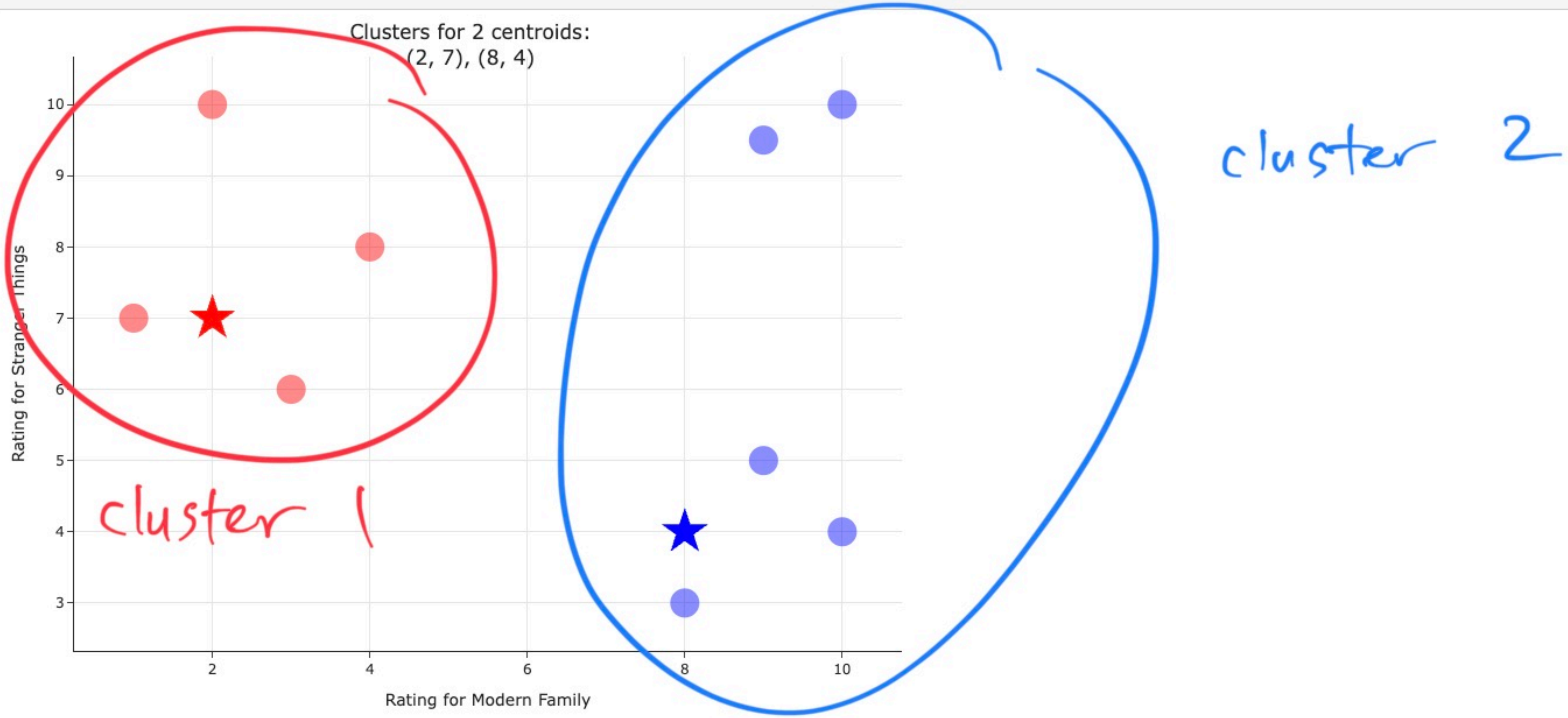
- But here, even though $k = 3$, the data are not colored "naturally"!

In [5]:
```
util.visualize_centroids([(2, 7), (8, 4), (3, 7)])
```



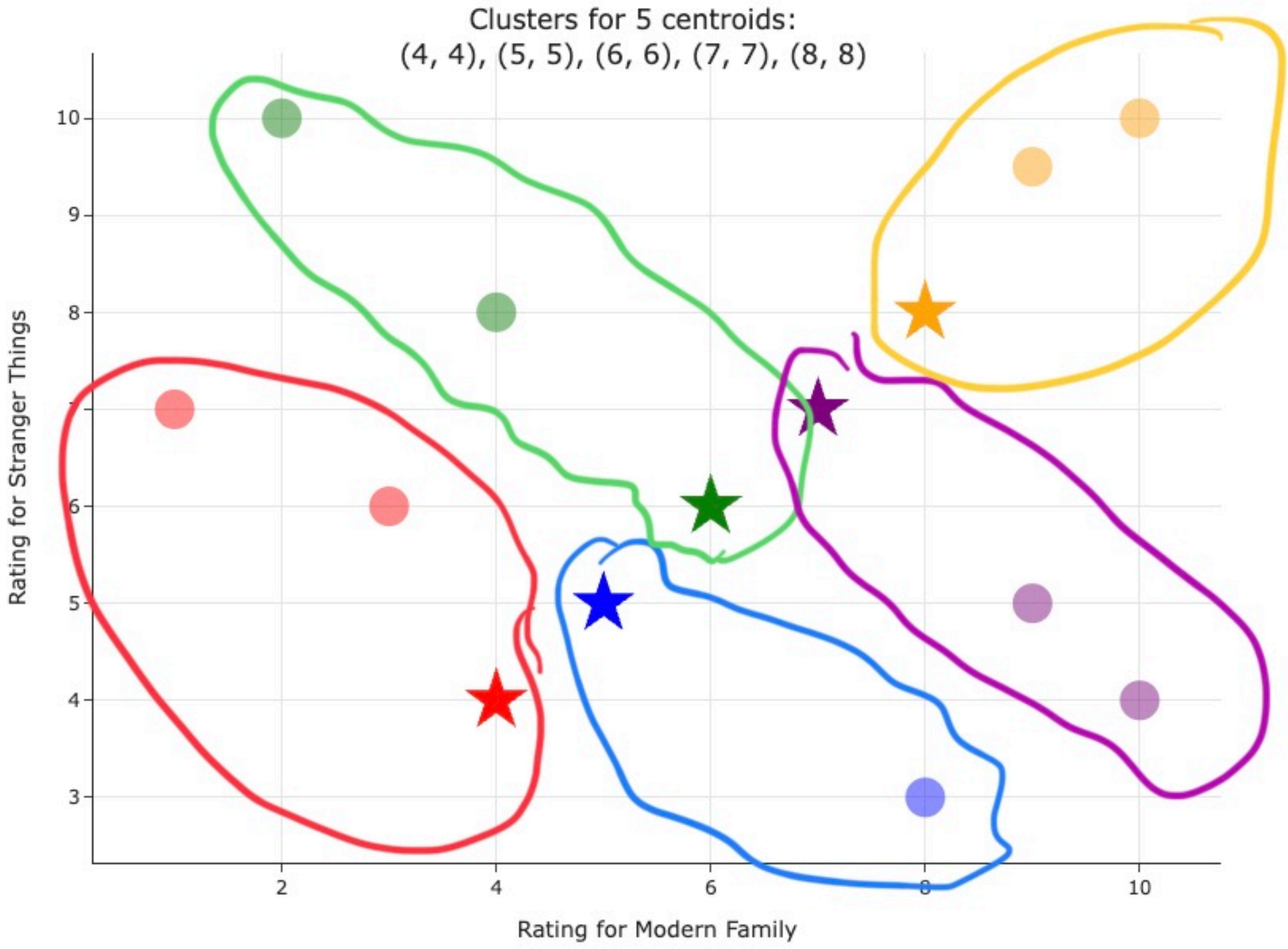
- Nothing is stopping us from setting $k = 2$, for instance!

In [6]:
```
util.visualize_centroids([(2, 7), (8, 4)])
```



- Or $k = 5$!

In [7]:
```
util.visualize_centroids([(4, 4), (5, 5), (6, 6), (7, 7), (8, 8)])
```



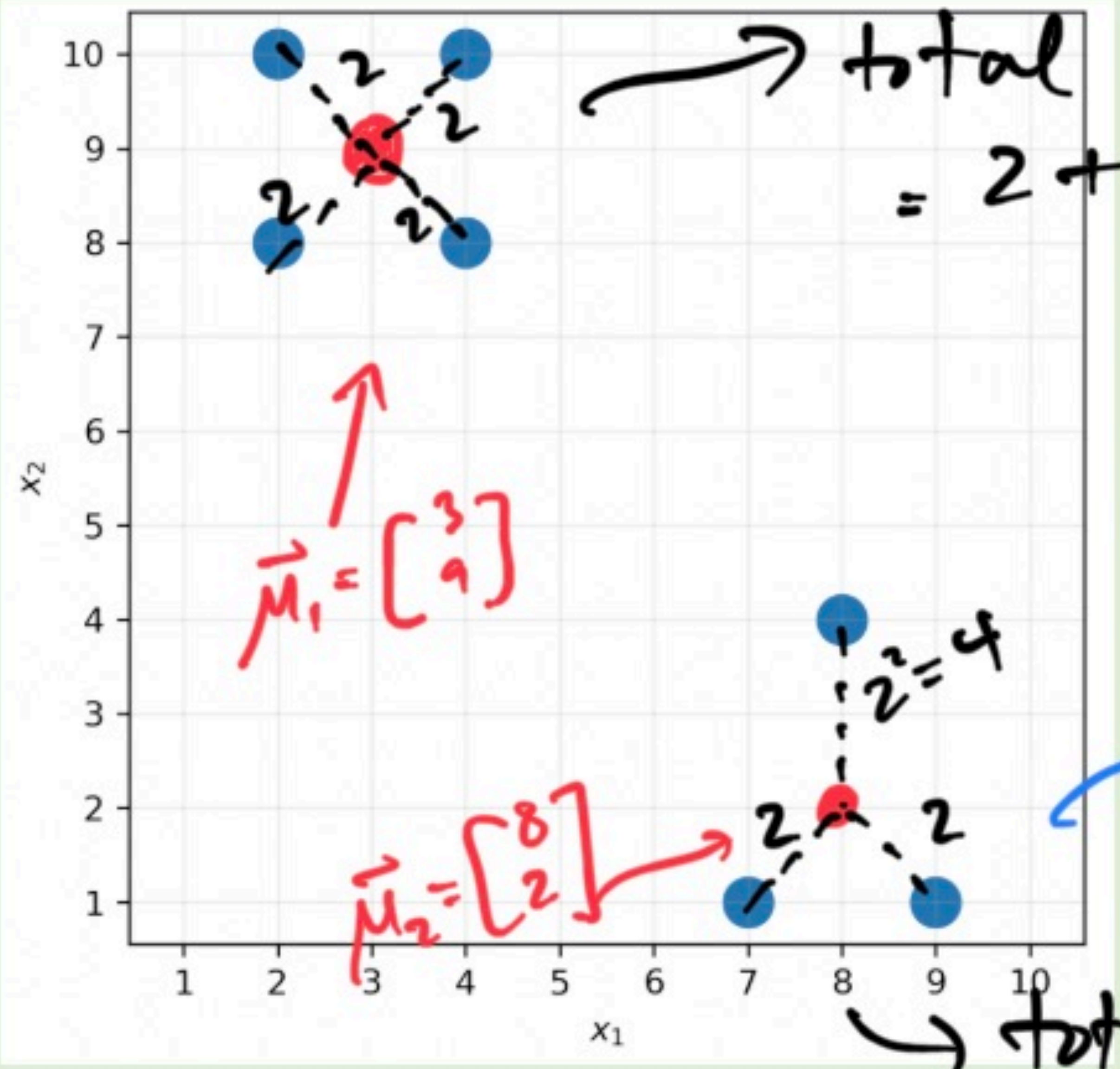
Activity

Recall, inertia is defined as follows:

$$I(\vec{\mu}_1, \vec{\mu}_2, \ldots, \vec{\mu}_k) = \text{total squared distance of each point } \vec{x}_i \text{ to its closest centroid } \vec{\mu}_j$$

inertia $= 8 + 8 = \boxed{16}$

Suppose we arrange the dataset below into $k = 2$ clusters. What is the **minimum possible inertia**?



total sq distance to $\vec{\mu}_1$

$= 2+2+2+2 = \boxed{8}$

x : 7, 8, 9

$\rightarrow \bar{x} = 8$

y: 1, 1, 4

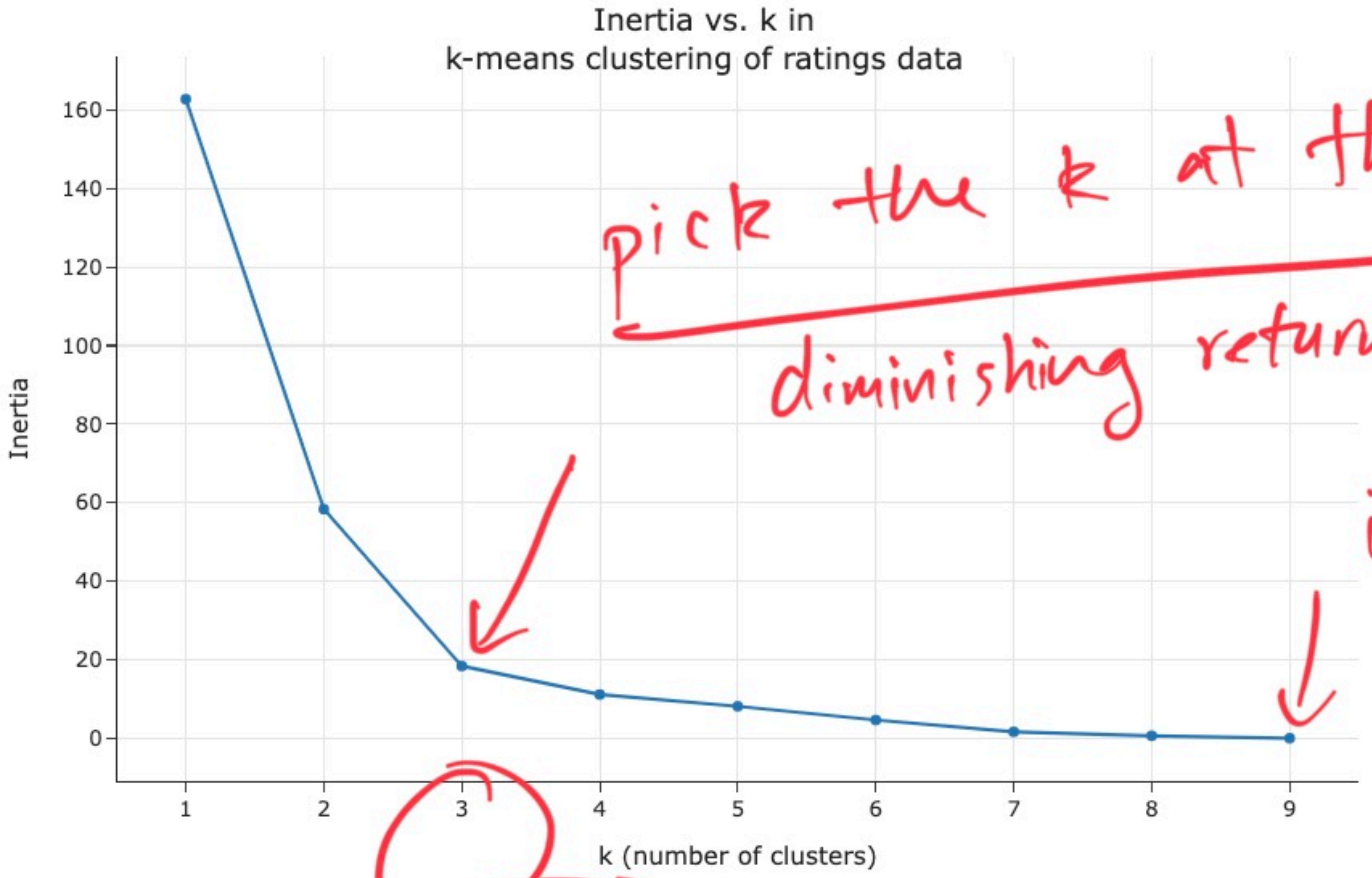
$\rightarrow \bar{y} = 2$

total sq distance to $\vec{\mu}_2$

$= 2+2+4 = \boxed{8}$

- For several different values of k, let's compute the inertia of the resulting clustering, using the scatter plot from the previous slide.

In [29]:
```
util.show_elbow()
```



pick the k at the elbow!

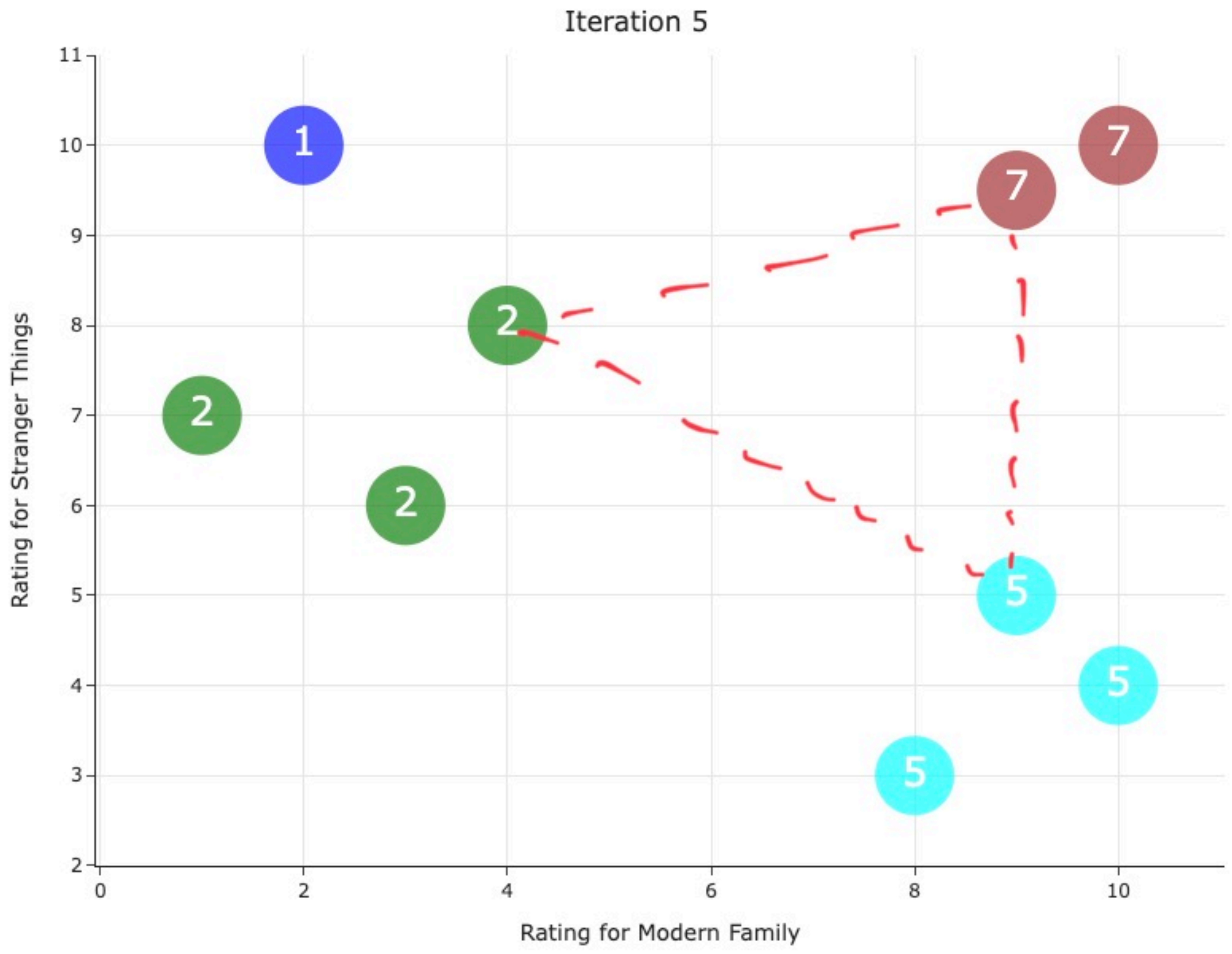
diminishing returns after the elbow

inertia = 0 because each point is its own centroid

k=3

http://localhost:8888/nbclassic/notebooks/lectures/lec26/lec26-blank.ipynb#/slide-53-0

In [25]:
```
util.color_ratings(title='Iteration 5', labels=[2, 1, 2, 2, 5, 5, 5, 7, 7])
```



whichever distance is smallest, we'd merge that pair if we wanted $k=2$.

- And finally, we merge **cluster 2** and **cluster 1**.
- If we just want $k = 3$ clusters, we stop here! If we wanted $k = 2$ clusters, we'd then merge the two closest clusters, based on the single linkage criterion.