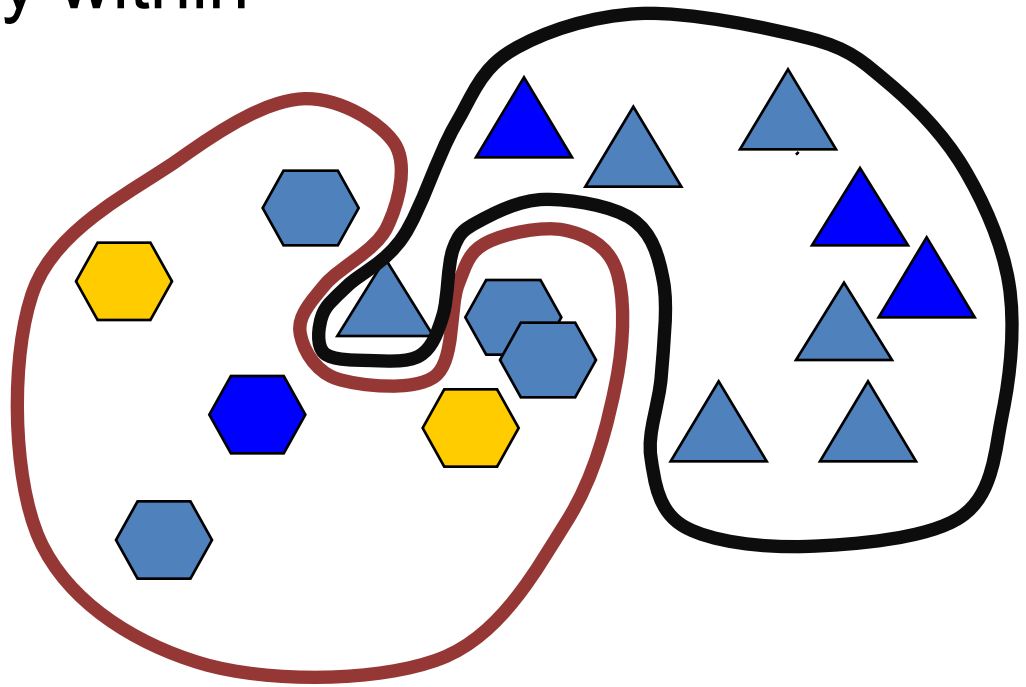
How Many Ks in K-Means?

Evaluating Your Segmentation Scheme

Two acid tests:

- Is there heterogeneity between segments?
- Is there homogeneity within segments?



No Segmentation

The original uploader was Brandon Dilbeck at English Wikipedia (Transferred from en.wikipedia to Commons) [CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0)], via Wikimedia Commons

Two Segments

The original uploader was Brandon Dilbeck at English Wikipedia (Transferred from en.wikipedia to Commons) [CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0)], via Wikimedia Commons

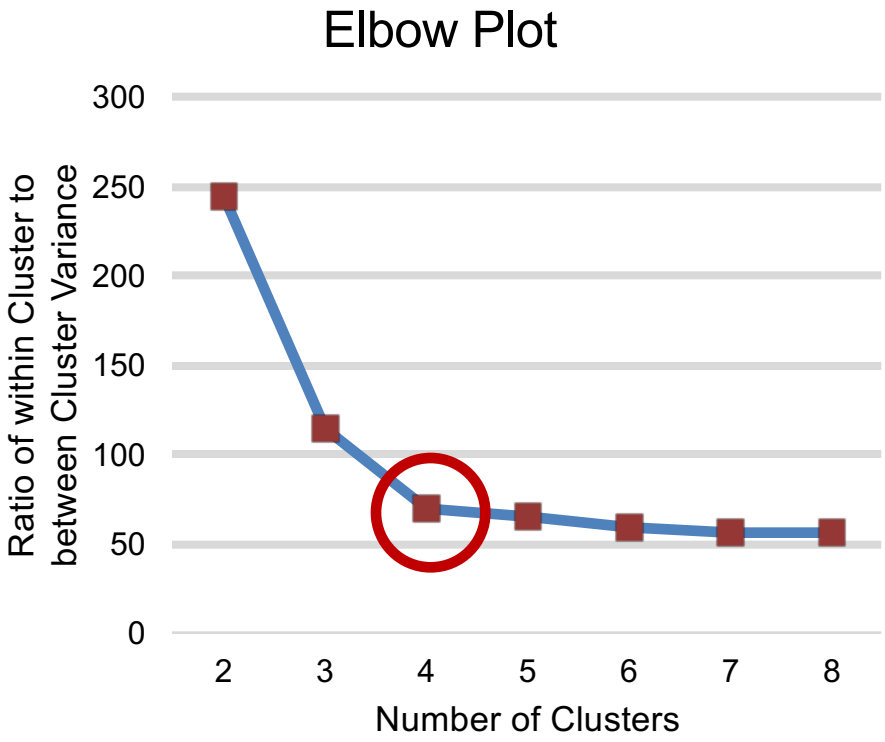
Images Credit: Mike Mozart, https://www.flickr.com/photos/jeepersmedia/

Ideal Segmentation

The original uploader was Brandon Dilbeck at English Wikipedia (Transferred from en.wikipedia to Commons) [CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0)], via Wikimedia Commons

Images Credit: Mike Mozart, https://www.flickr.com/photos/jeepersmedia/

By brandi sims (originally posted to Flickr as jelly belly) [CC BY 2.0 (http://creativecommons.org/licenses/by/2.0)], via Wikimedia Commons

Number of Clusters: Elbow Plot

- Ratio of within to between cluster variance is plotted against the number of clusters
- The point at which a sharp bend occurs indicates the number of clusters