**How can we choose a good K for K-means clustering**

https://www.quora.com/How-can-we-choose-a-good-K-for-K-means-clustering

You can choose the number of clusters by visually inspecting your data points, but you will soon realize that there is a lot of ambiguity in this process for all except the simplest data sets. This is not always bad, because you are doing unsupervised learning and there's some inherent subjectivity in the labeling process. Here, having previous experience with that particular problem or something similar will help you choose the right value.

If you want some hint about the number of clusters that you should use, you can apply the *Elbow method*:

First of all, compute the sum of squared error (SSE) for some values of k (for example 2, 4, 6, 8, etc.). The SSE is defined as the sum of the squared distance between each member of the cluster and its centroid. Mathematically:

SSE=∑Ki=1∑x∈cidist(x,ci)2SSE=∑i=1K∑x∈cidist(x,ci)2

 If you plot k against the SSE, you will see that *the error decreases as k gets larger*; this is because when the number of clusters increases, they should be smaller, so distortion is also smaller. The idea of the elbow method is to choose the k at which the SSE decreases abruptly. This produces an "elbow effect" in the graph, as you can see in the following picture:



In this case, k=6k=6 is the value that the Elbow method has selected. Take into account that the Elbow method is an heuristic and, as such, it may or may not work well in your particular case. Sometimes, there are more than one elbow, or no elbow at all. In those situations you usually end up calculating the best k by evaluating how well k-means performs in the context of the particular clustering problem you are trying to solve.

93.3k Views · [View 241 Upvoters](https://www.quora.com/api/mobile_expanded_voter_list?key=basya9FB35U&type=answer) ·