**Q1. Find the feature selection function to convert an image into a vector**

I will use 2-layer spatial color histogram to select the features from the images.

I will scale the colors(R, G, B) by dividing their value with 16, then count the value of each color to put them in different bins.

I tried using kmeans on the dataset with k varied from 1 to 30, here is their total within-cluster sum of squares, with seed = 1994:

### Q2. Divide the images into k clusters using kmeans.

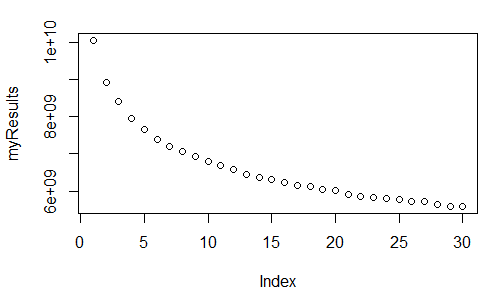
The smaller the sum is, the better as it is a measure of error.

Figure 1Sum of squares varied by k

The appropriate number of clusters (k) should be around 7 to 15, as the result’s change isn’t significant with k > 15.

So I will choose k =13, just based on preference.

### Q3. Rendering the result in HTML page

The code will make a HTML file named “Output” that will display the clusters and the images in each clusters.