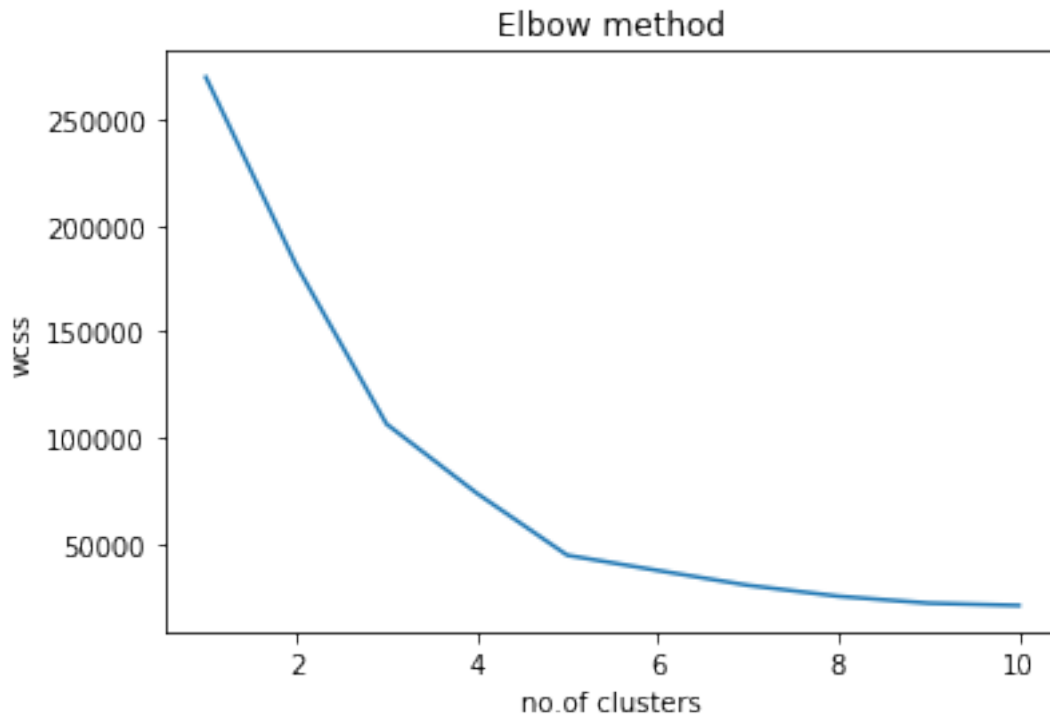# K-Means

September 27, 2018

```python
In [15]: import pandas as pd
         import matplotlib.pyplot as plt
         import numpy as np

In [16]: dataset = pd.read_csv('/media/coea/A4F698A8F6987BEC/A-Z/K_Means/K_Means/Mall_Customers.
         x = dataset.iloc[:,[3,4]].values

In [17]: from sklearn.cluster import KMeans

In [18]: wcss = []
         for i in range(1,11):
             kmeans = KMeans(n_clusters=i,init='k-means++',max_iter=300,n_init=10,random_state=0
             kmeans.fit(x)
             wcss.append(kmeans.inertia_)
         plt.plot(range(1,11),wcss)
         plt.title('Elbow method')
         plt.xlabel('no.of clusters')
         plt.ylabel('wcss')
         plt.show()
```
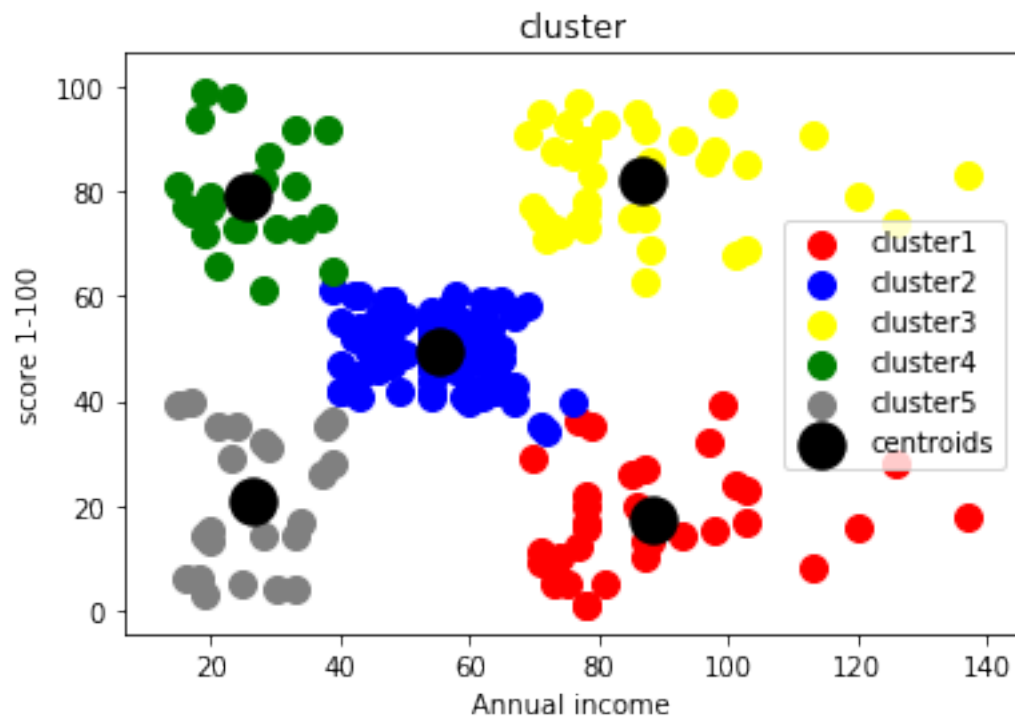
Elbow method

```python
In [19]: kmeans = KMeans(n_clusters=5,init='k-means++',max_iter=300,n_init=10,random_state=0)
         y_kmeans = kmeans.fit_predict(x)

In [20]: plt.scatter(x[y_kmeans==0,0],x[y_kmeans==0,1],s=100,c='red',label='cluster1')
         plt.scatter(x[y_kmeans==1,0],x[y_kmeans==1,1],s=100,c='blue',label='cluster2')
         plt.scatter(x[y_kmeans==2,0],x[y_kmeans==2,1],s=100,c='yellow',label='cluster3')
         plt.scatter(x[y_kmeans==3,0],x[y_kmeans==3,1],s=100,c='green',label='cluster4')
         plt.scatter(x[y_kmeans==4,0],x[y_kmeans==4,1],s=100,c='gray',label='cluster5')
         plt.scatter(kmeans.cluster_centers_[:,0],kmeans.cluster_centers_[:,1],s=300,c='black',l
         plt.title('cluster')
         plt.xlabel('Annual income')
         plt.ylabel('score 1-100')
         plt.legend()
         plt.show()
```

cluster

In [21]:

In [ ]: