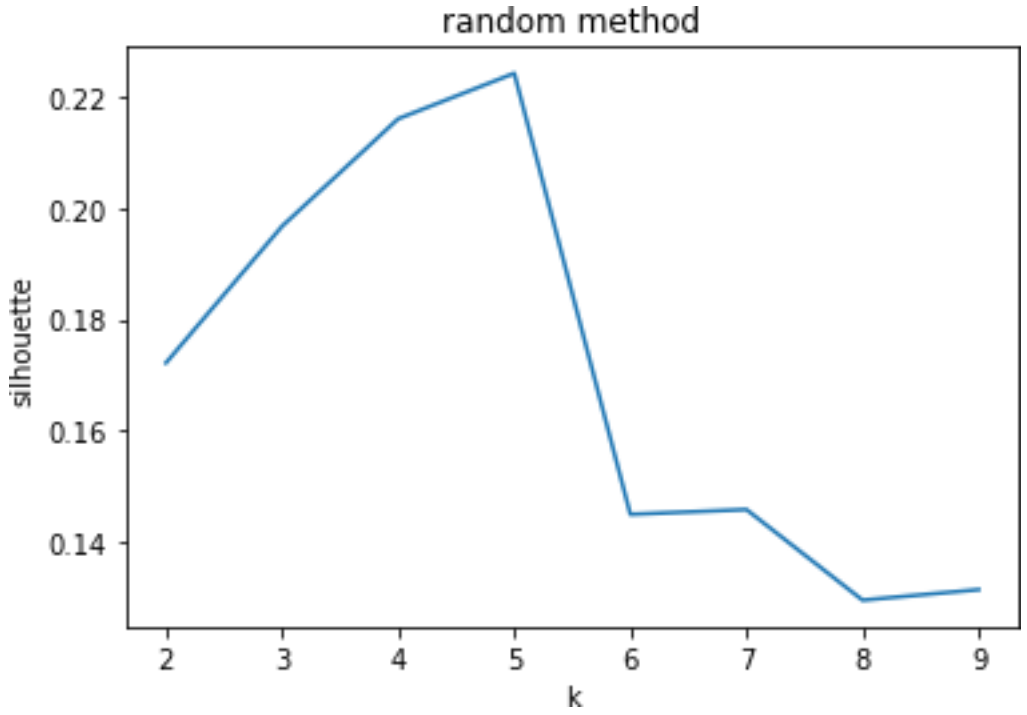
CMPT 459
Guangcheng Lin
301354550
Assignment 2

Task:

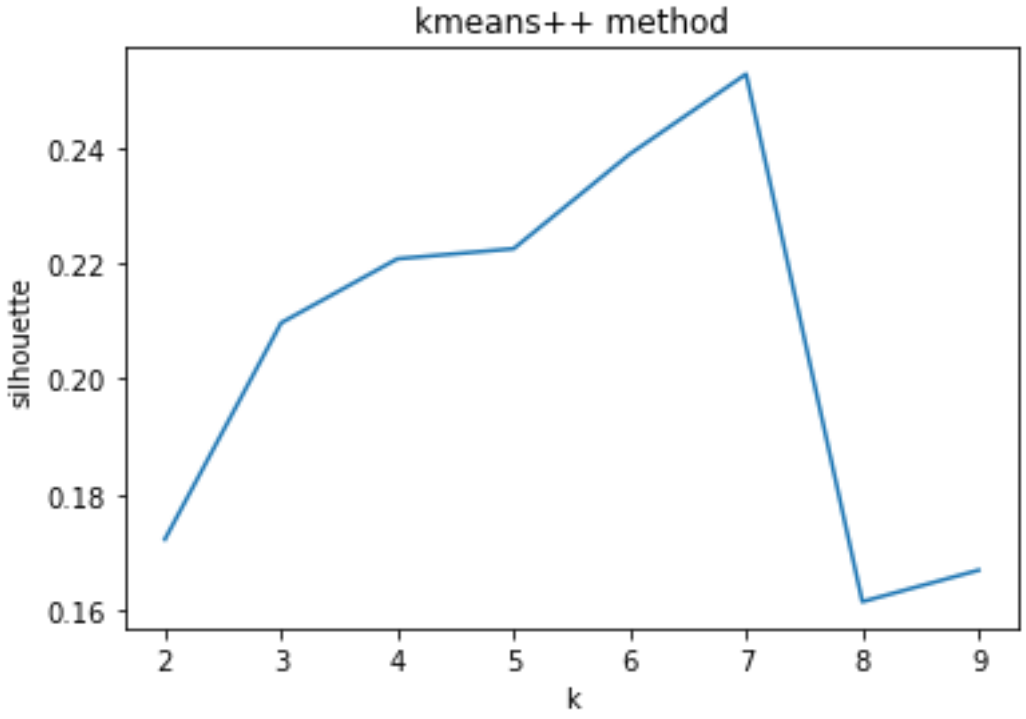
Q2.

In random method, the best k is 5.



Q3.

The best k is 7.



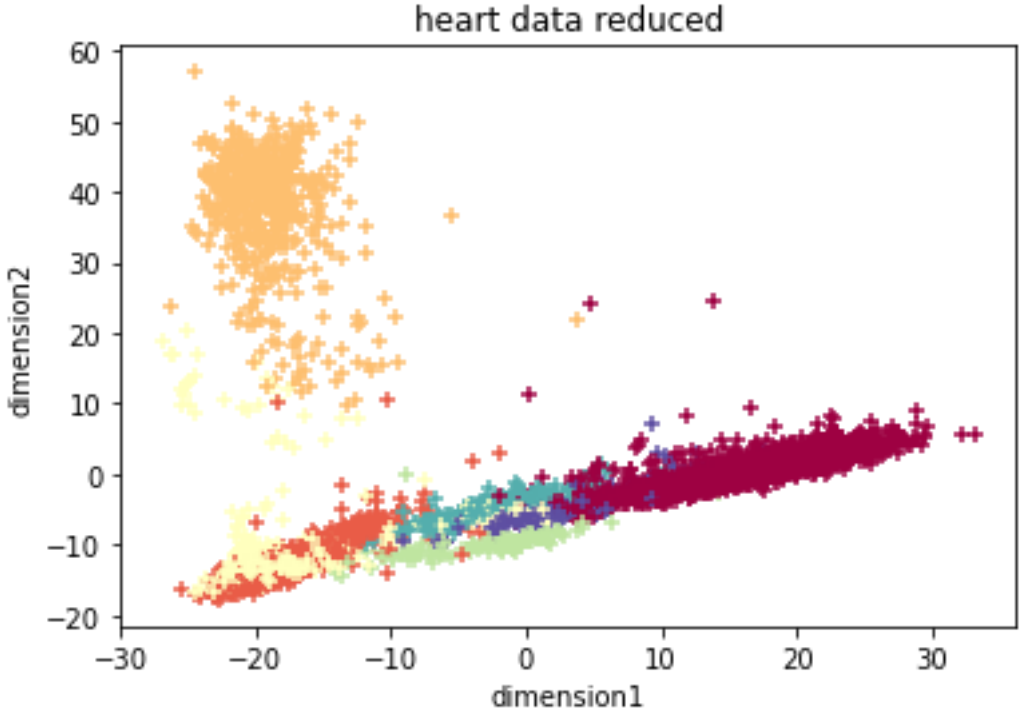
Conclusion:

As we all know, one weakness of the K-means algorithmis that it is sensitive to the initialization of the centroids.

In this experiment, we can find that kmeans++ method is more robust. Comparing to kmeans, it requires less time and has better efficiency to finish clustering. On the other hand, kmeans++ has better robust such that the data it draws will not appear to be very variable, and the results will be relatively stable. So that kmeans++ has better performance.

Q4.

Visualize_cluster graph plotted below (while the best value of k is 7):


heart data reduced
dimension2
dimension1
60
50
40
30
20
10
0
−10
−20
−30
−20
−10
0
10
20
30