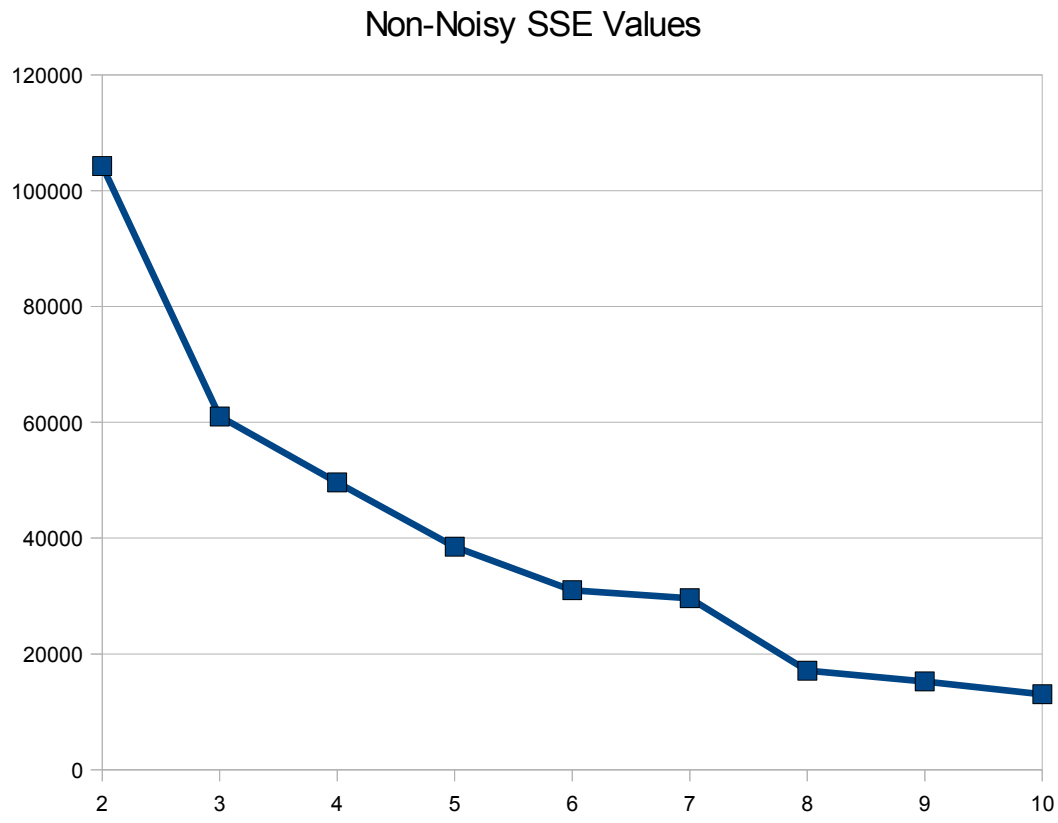
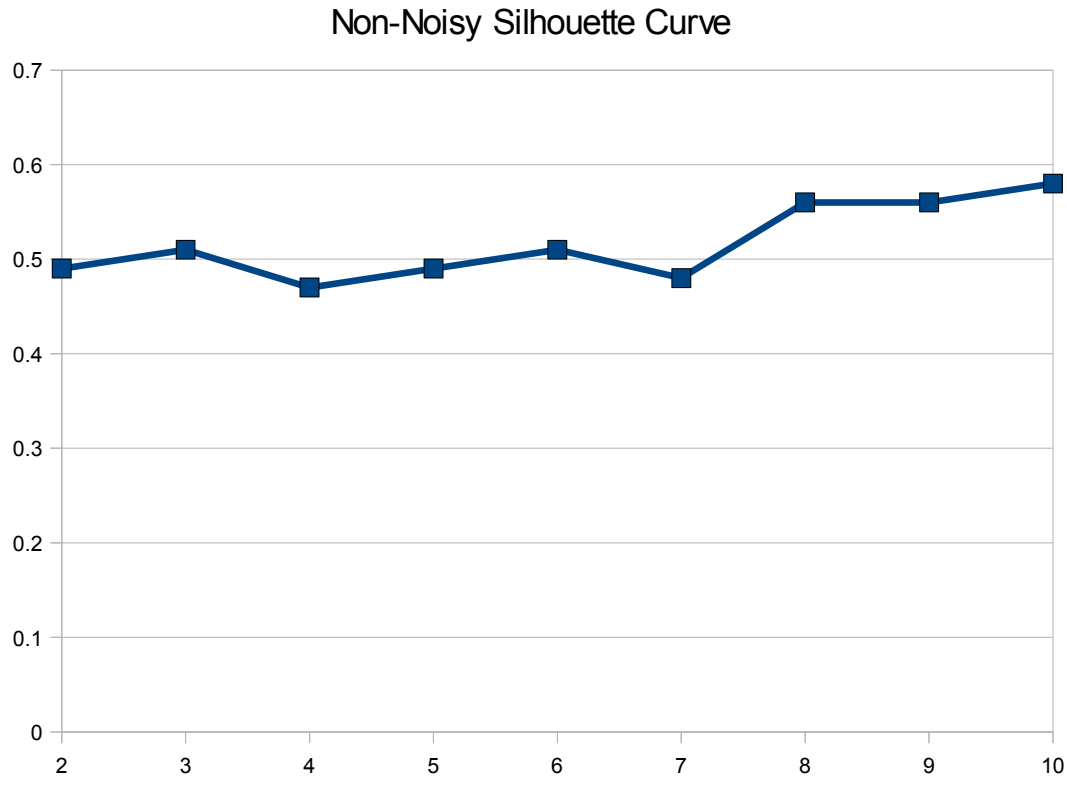
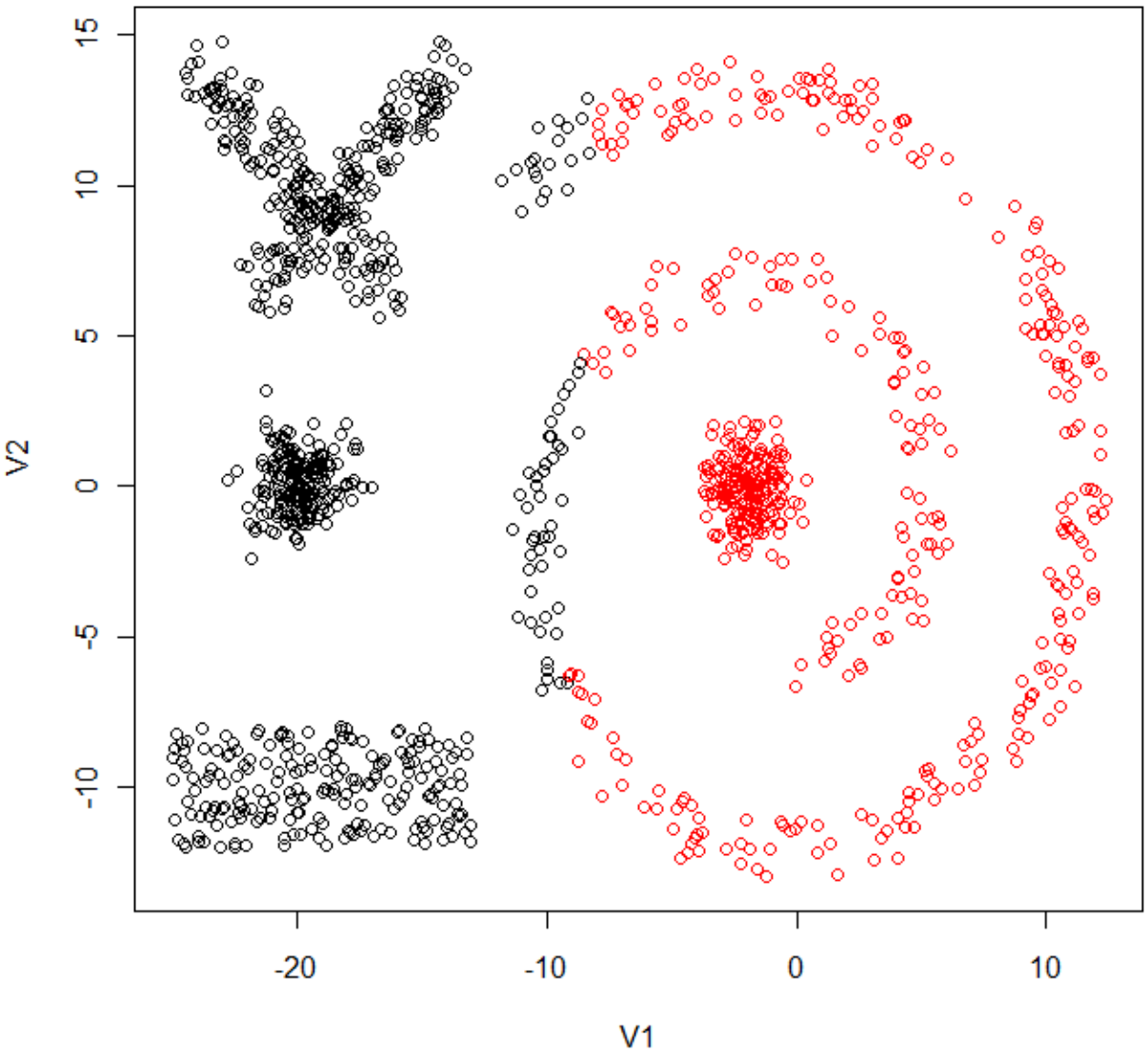
1.


Non-Noisy SSE Values
120000
100000
80000
60000
40000
20000
0
2
3
4
5
6
7
8
9
10

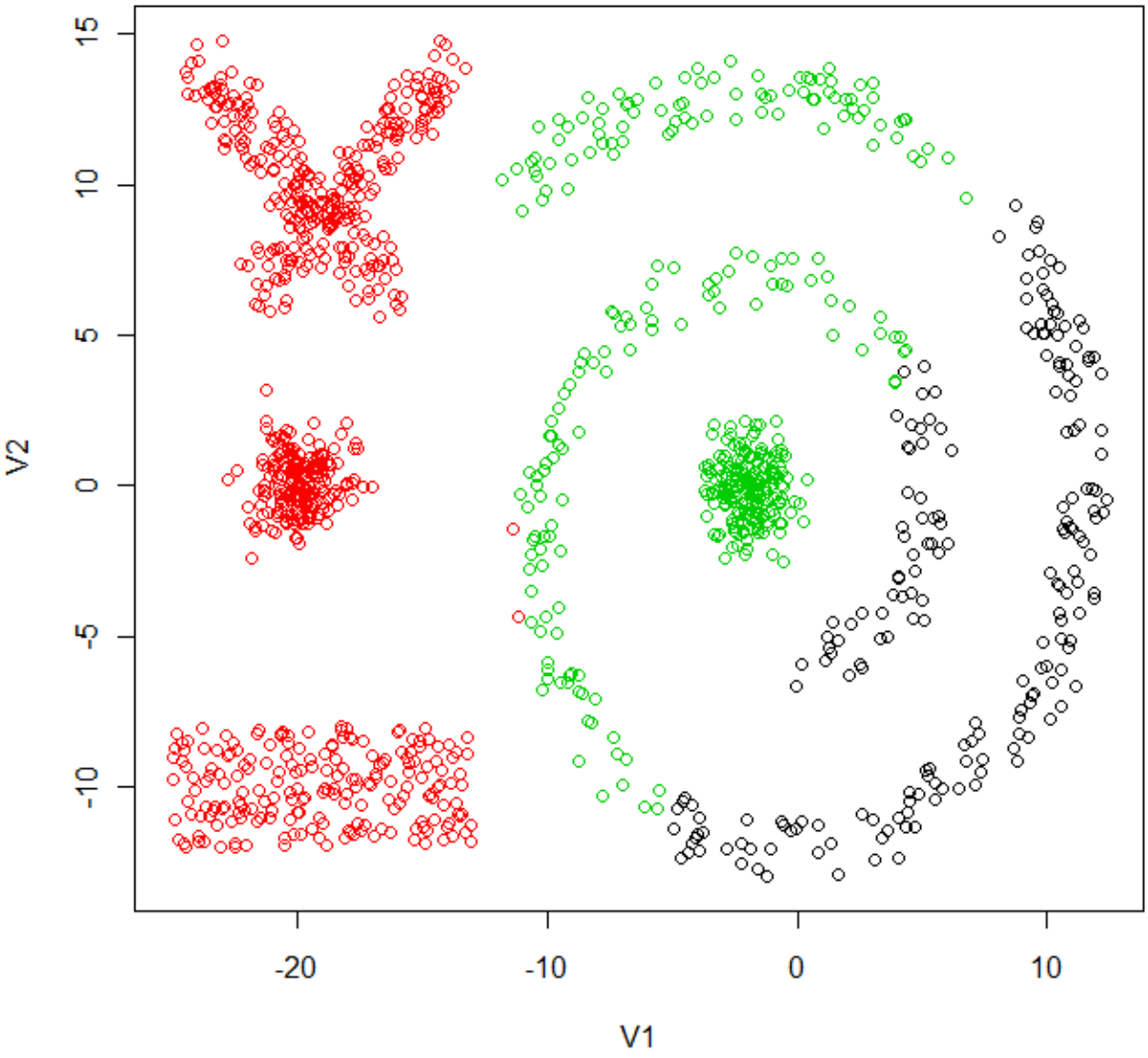

Non-Noisy Silhouette Curve
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
2
3
4
5
6
7
8
9
10

Plots for Non-Noisy Data Clusterings:

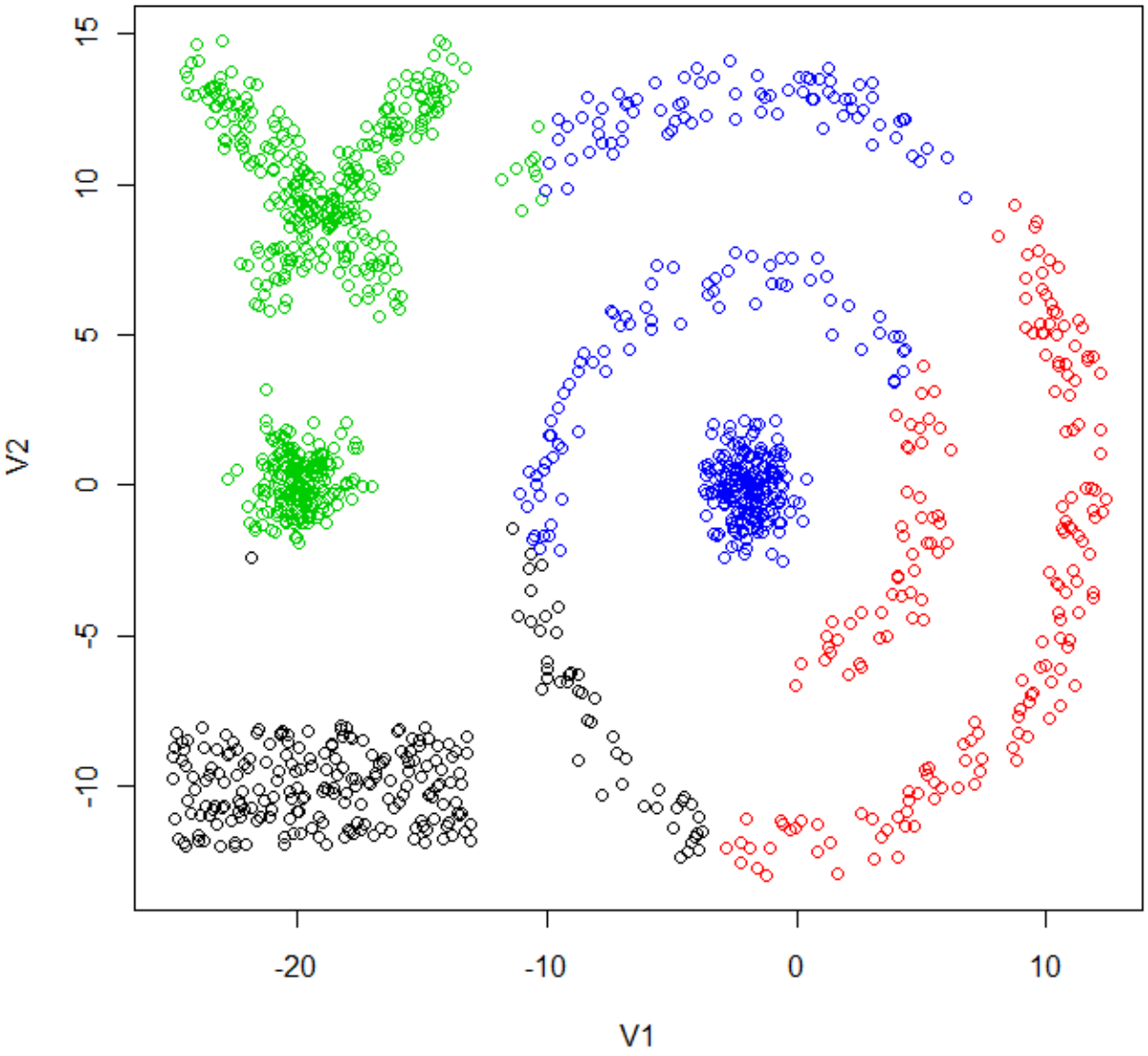
k = 2



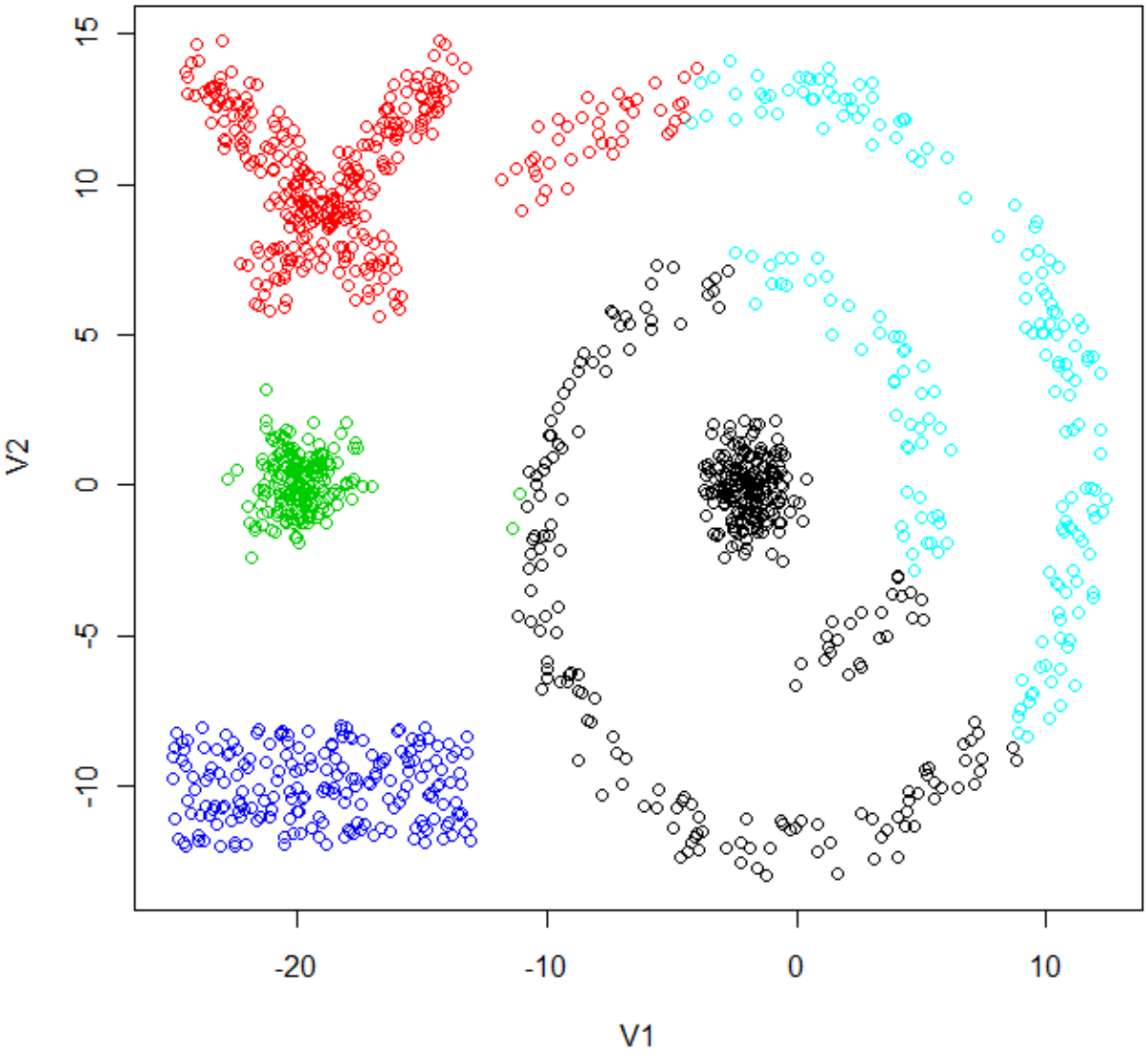
k = 3



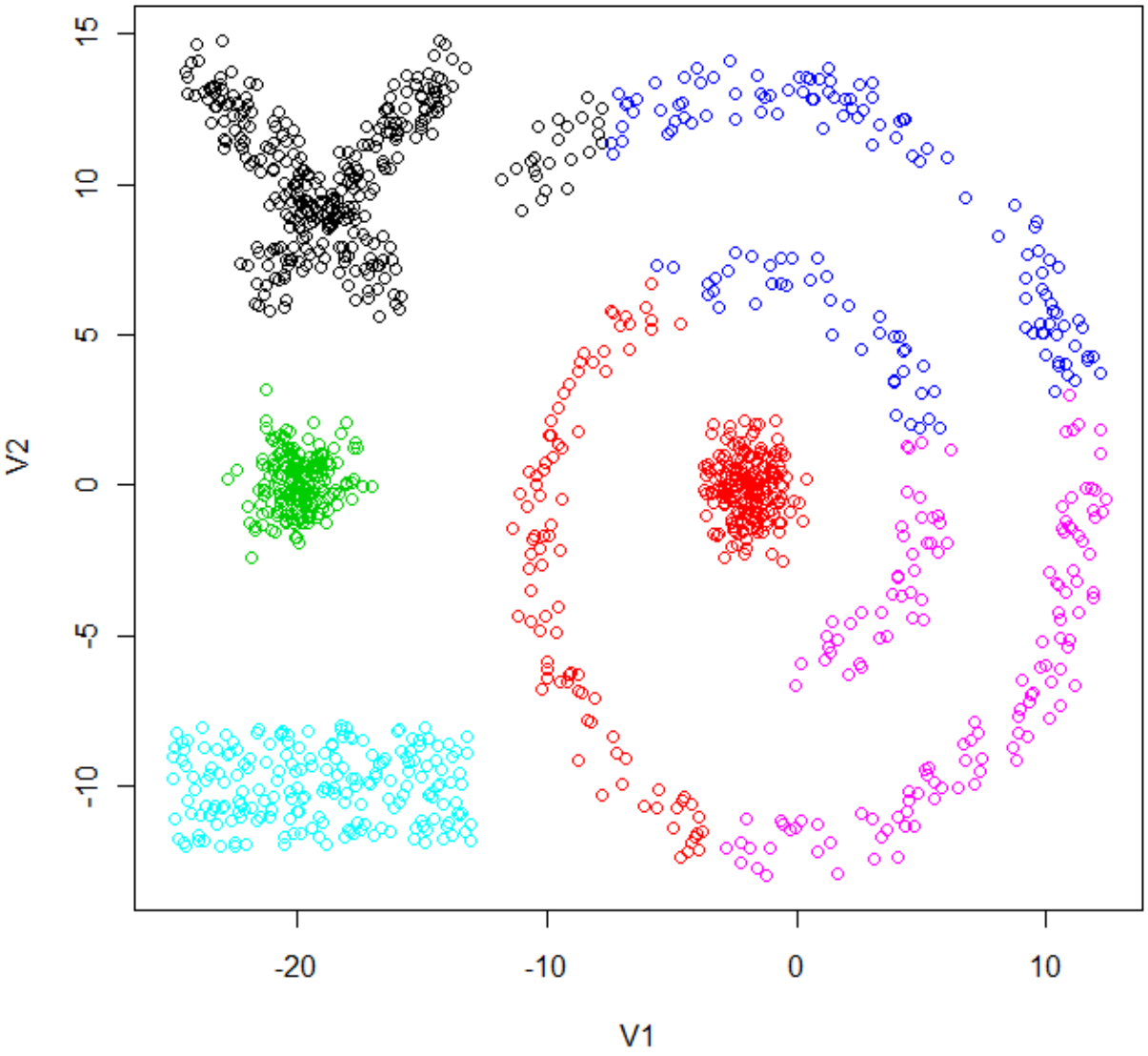
k = 4



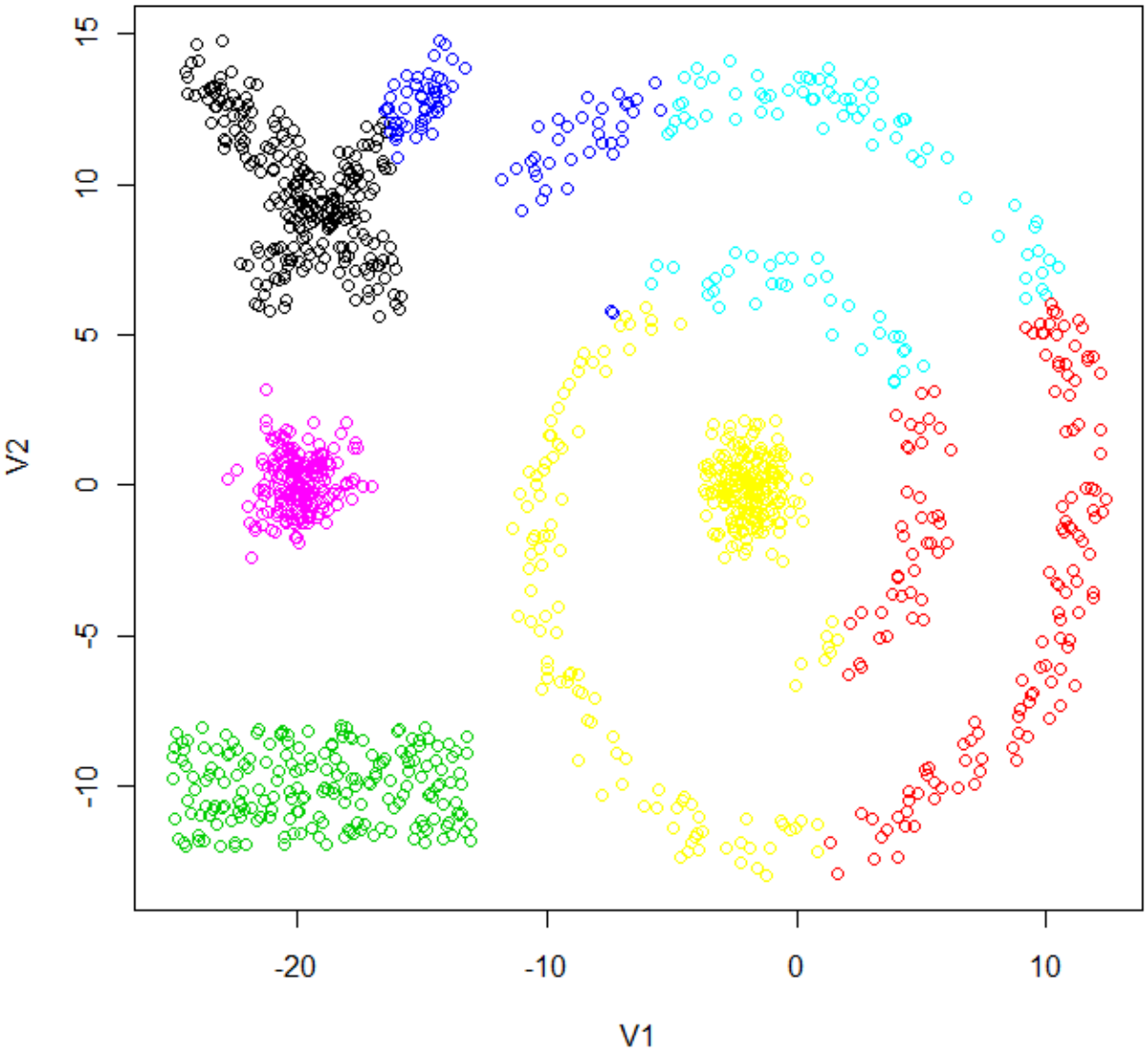
k = 5



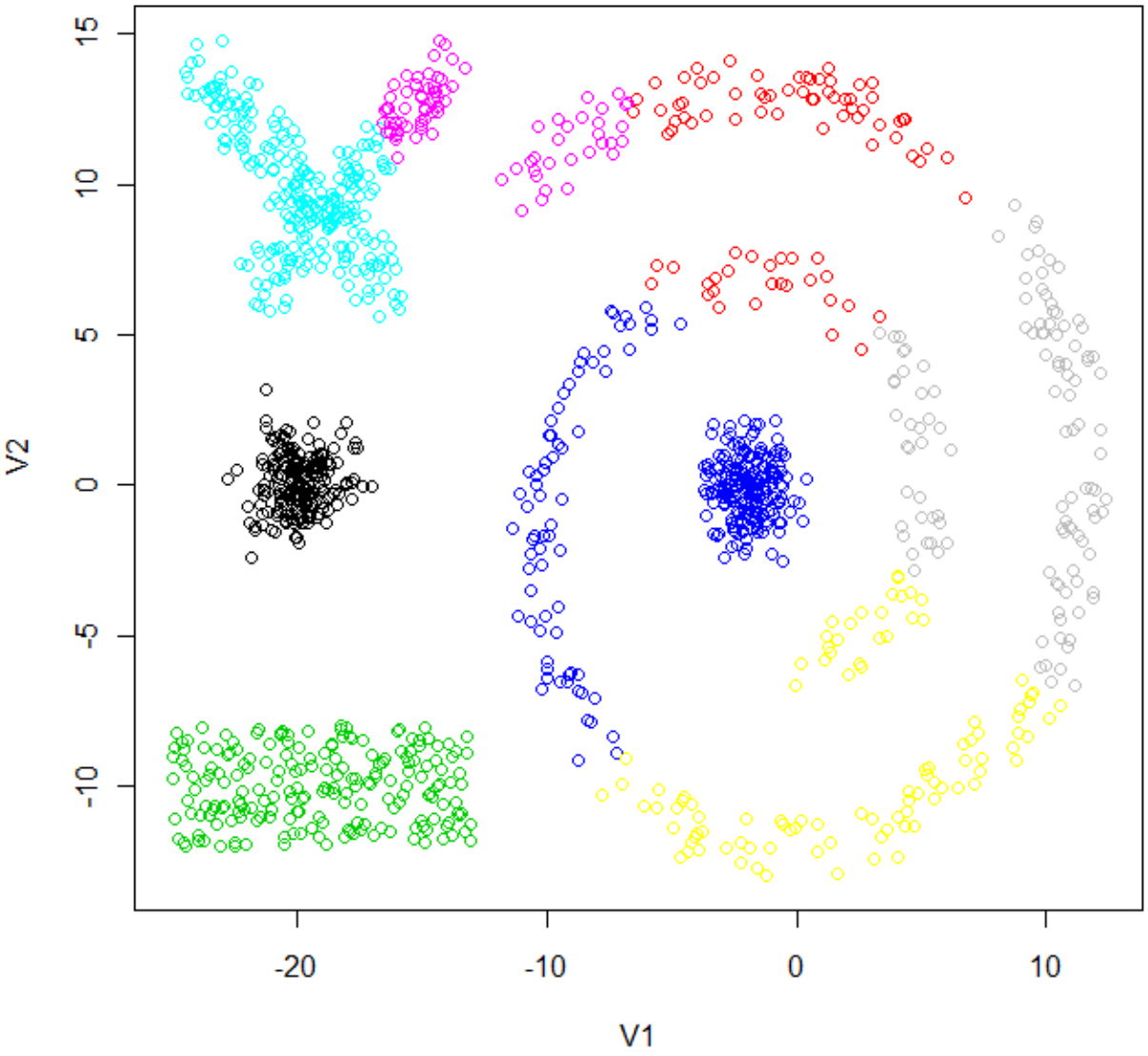
k = 6



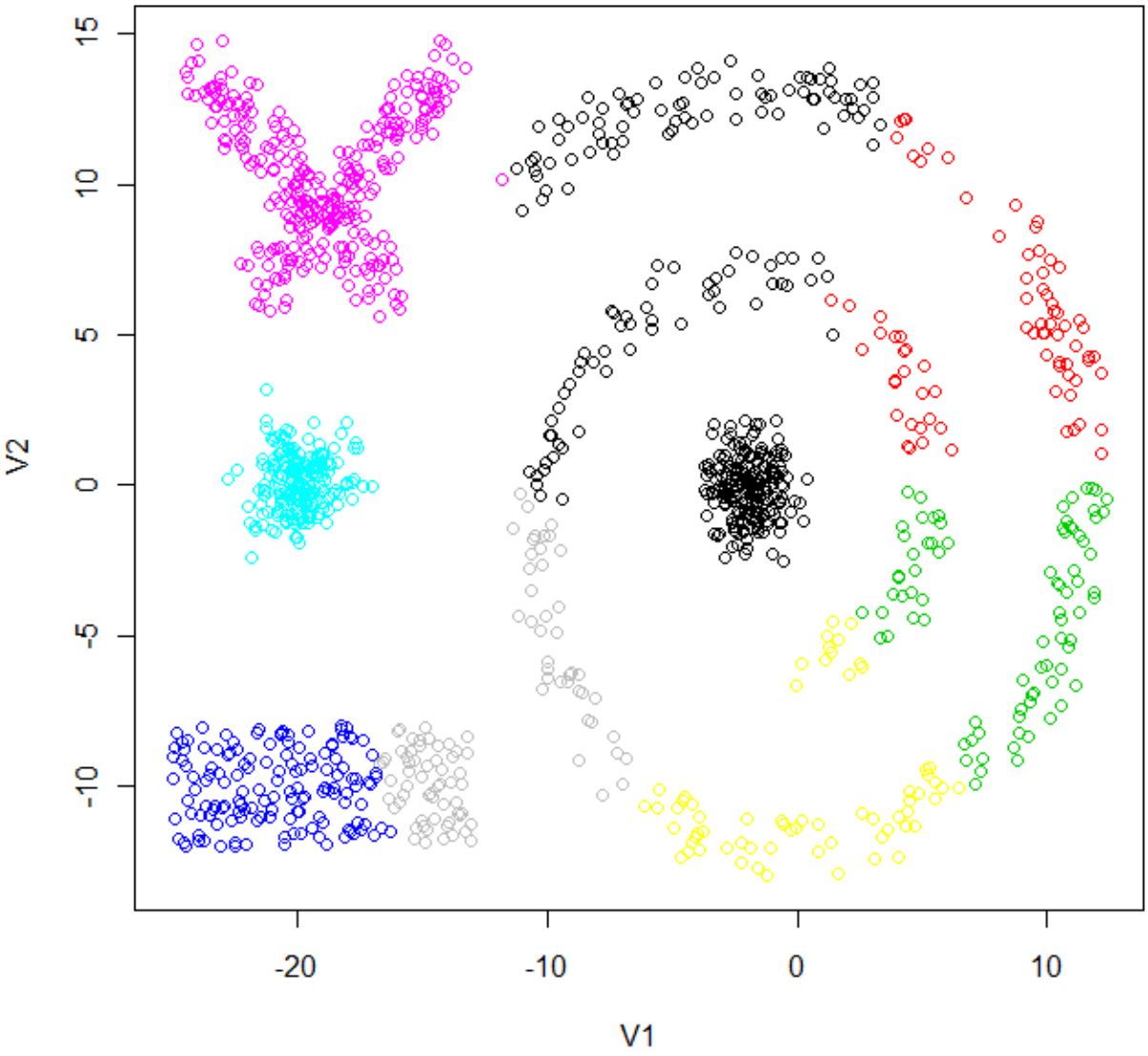
k = 7



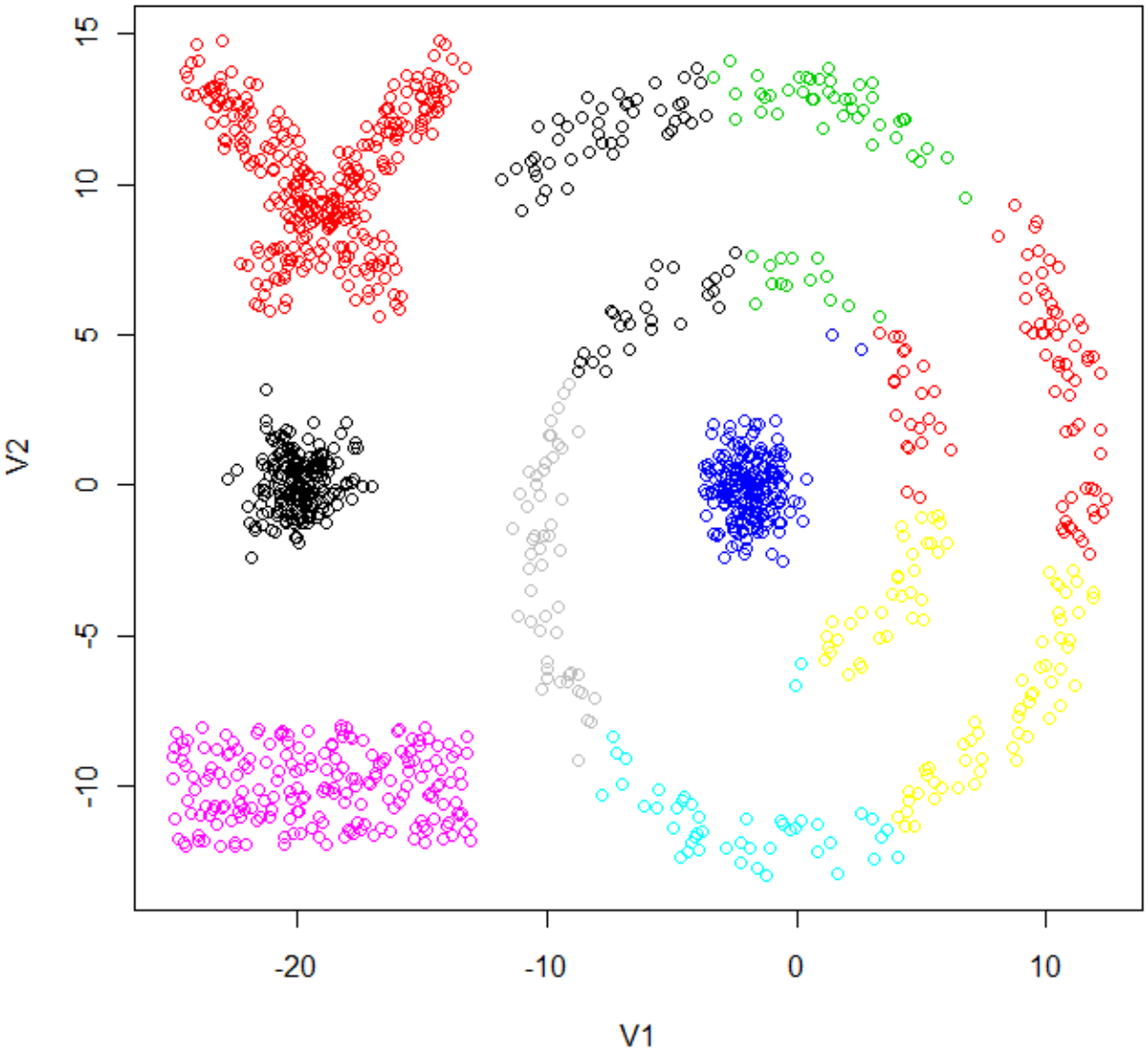
k = 8



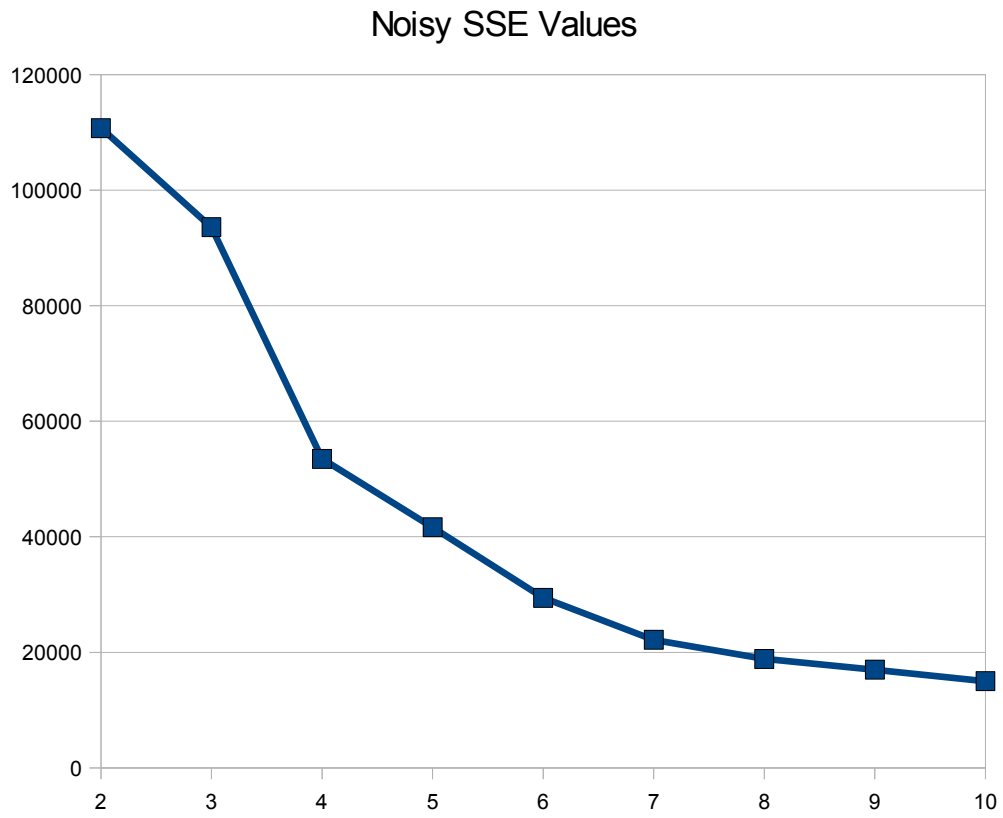
k = 9

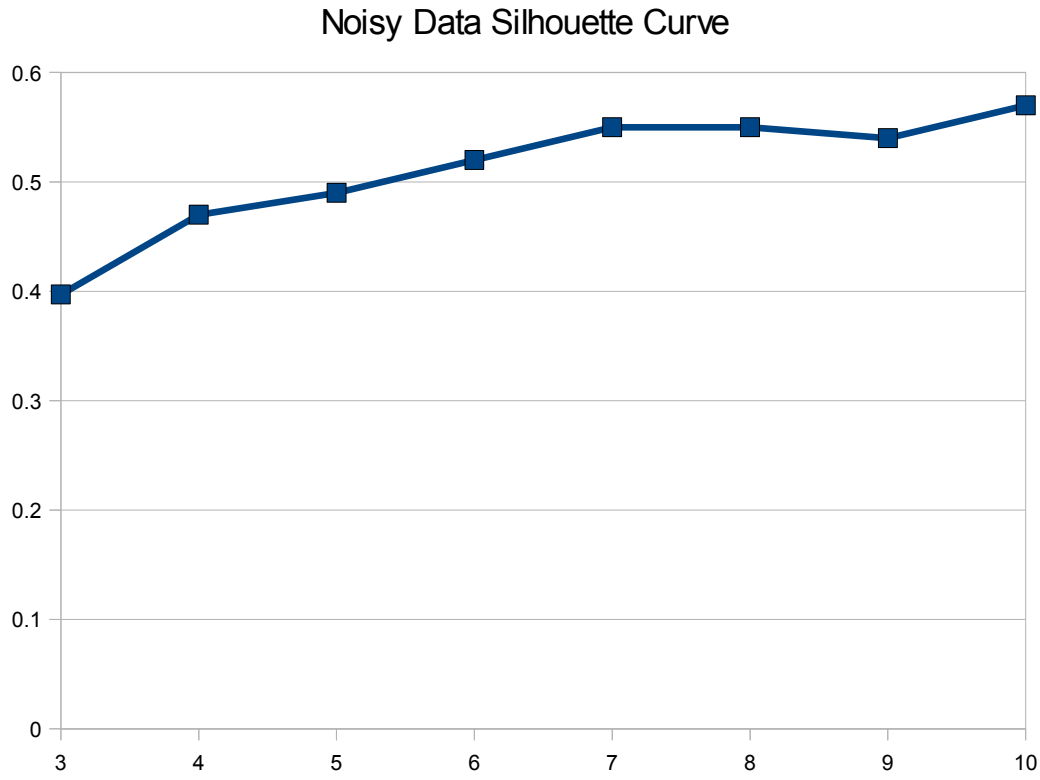


k = 10



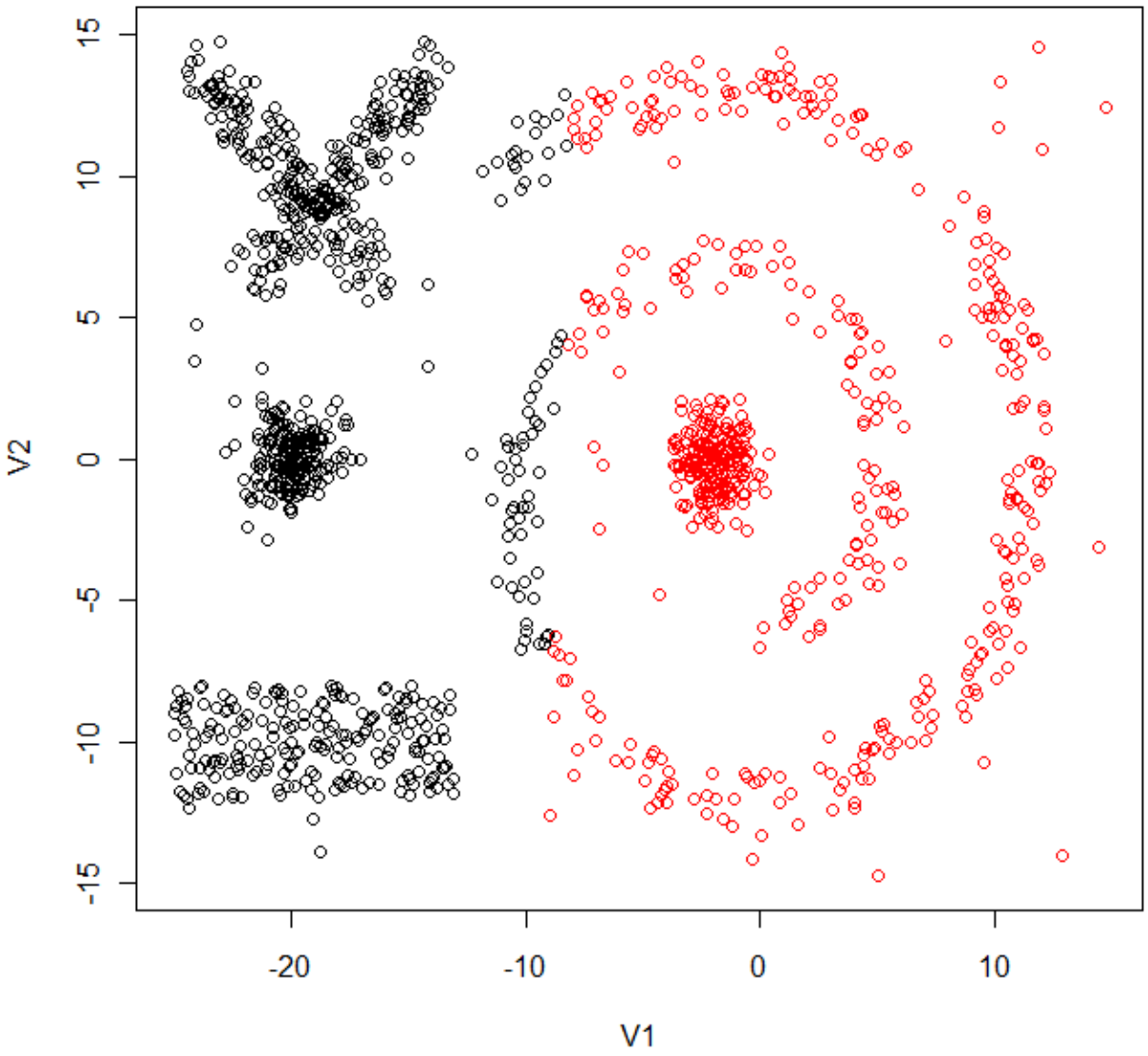
2.


Noisy SSE Values
120000
100000
80000
60000
40000
20000
0
2
3
4
5
6
7
8
9
10

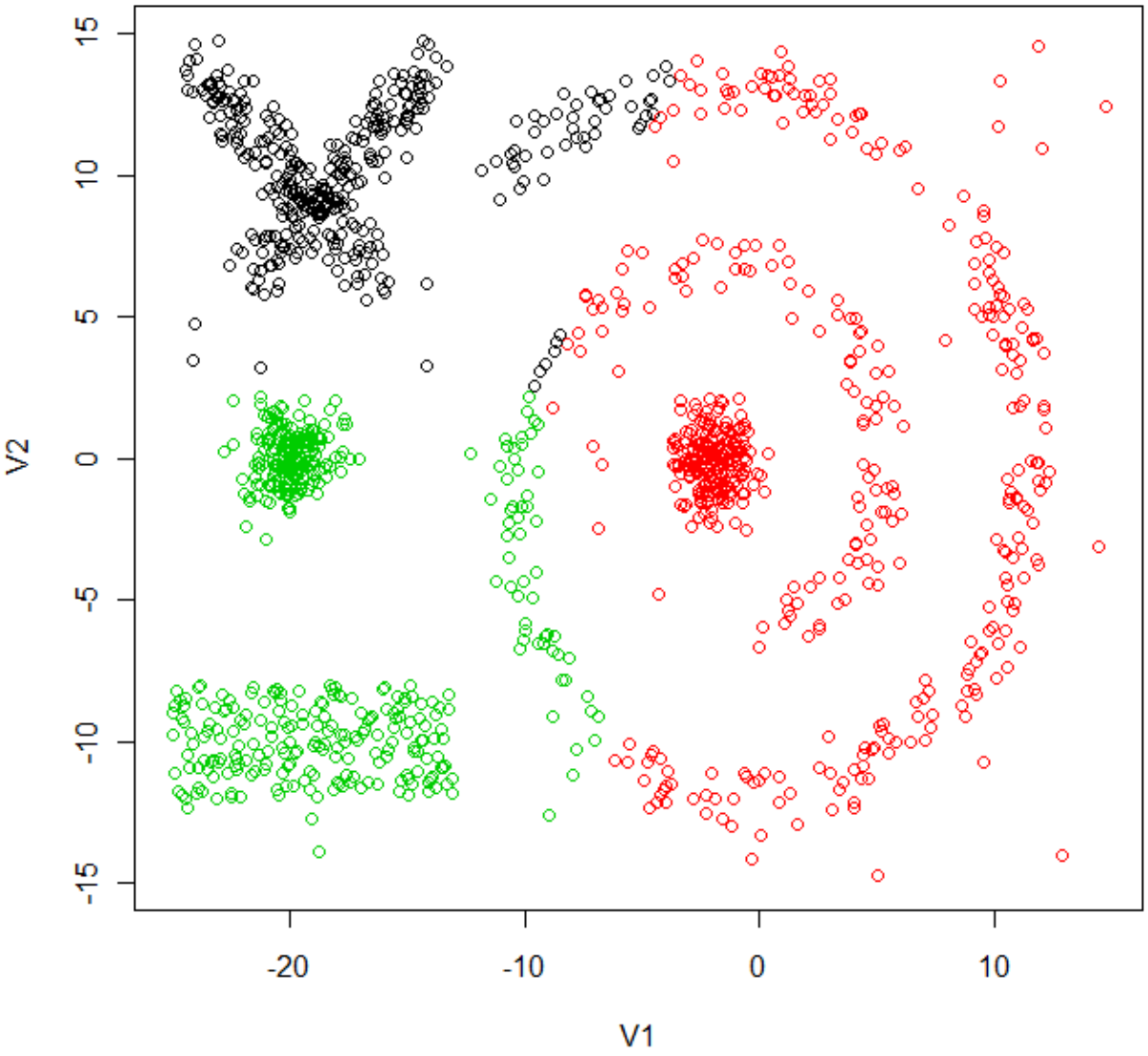

Noisy Data Silhouette Curve
0.6
0.5
0.4
0.3
0.2
0.1
0
3
4
5
6
7
8
9
10

Plots for Noisy Data Clusterings:

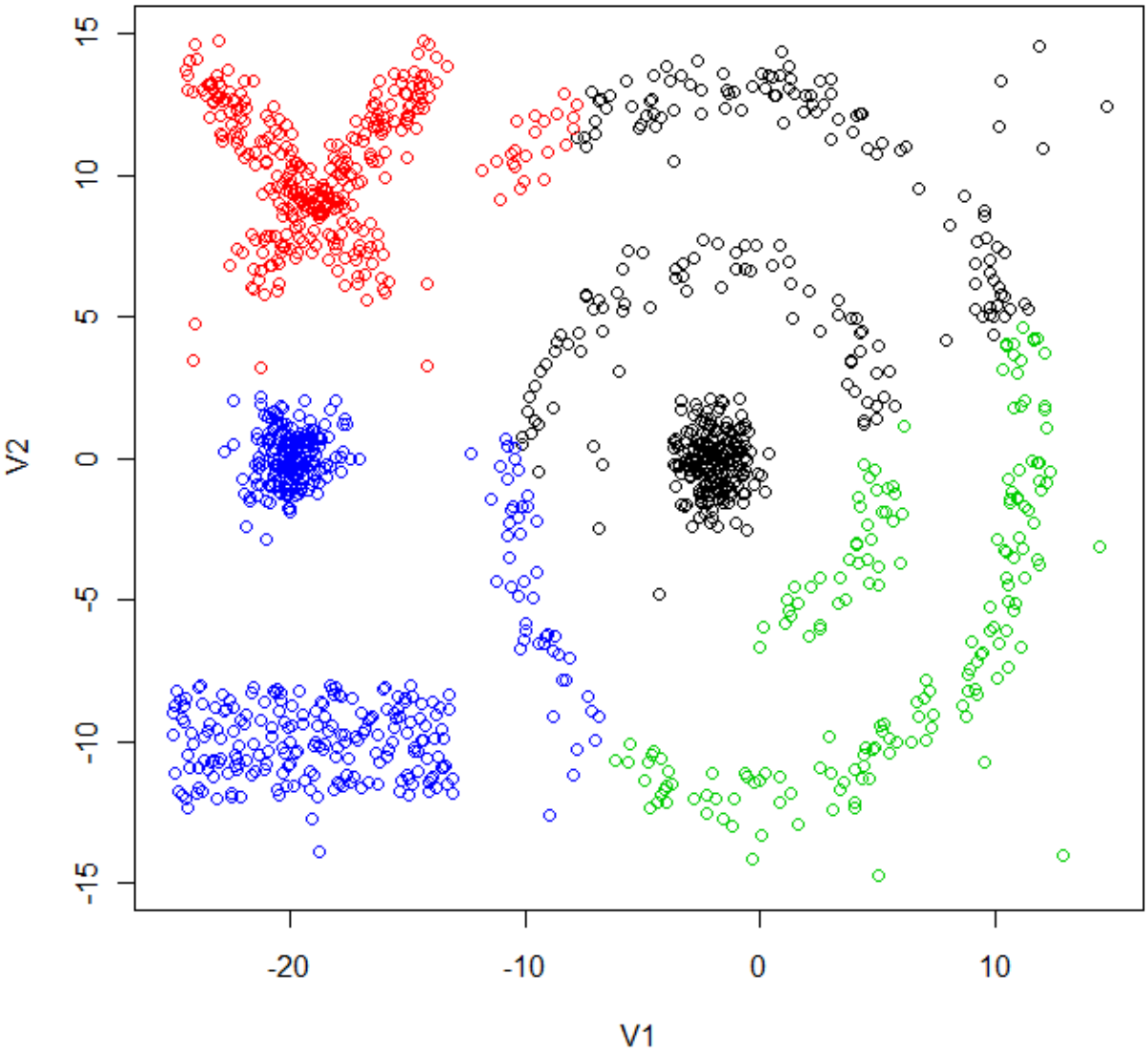
k = 2



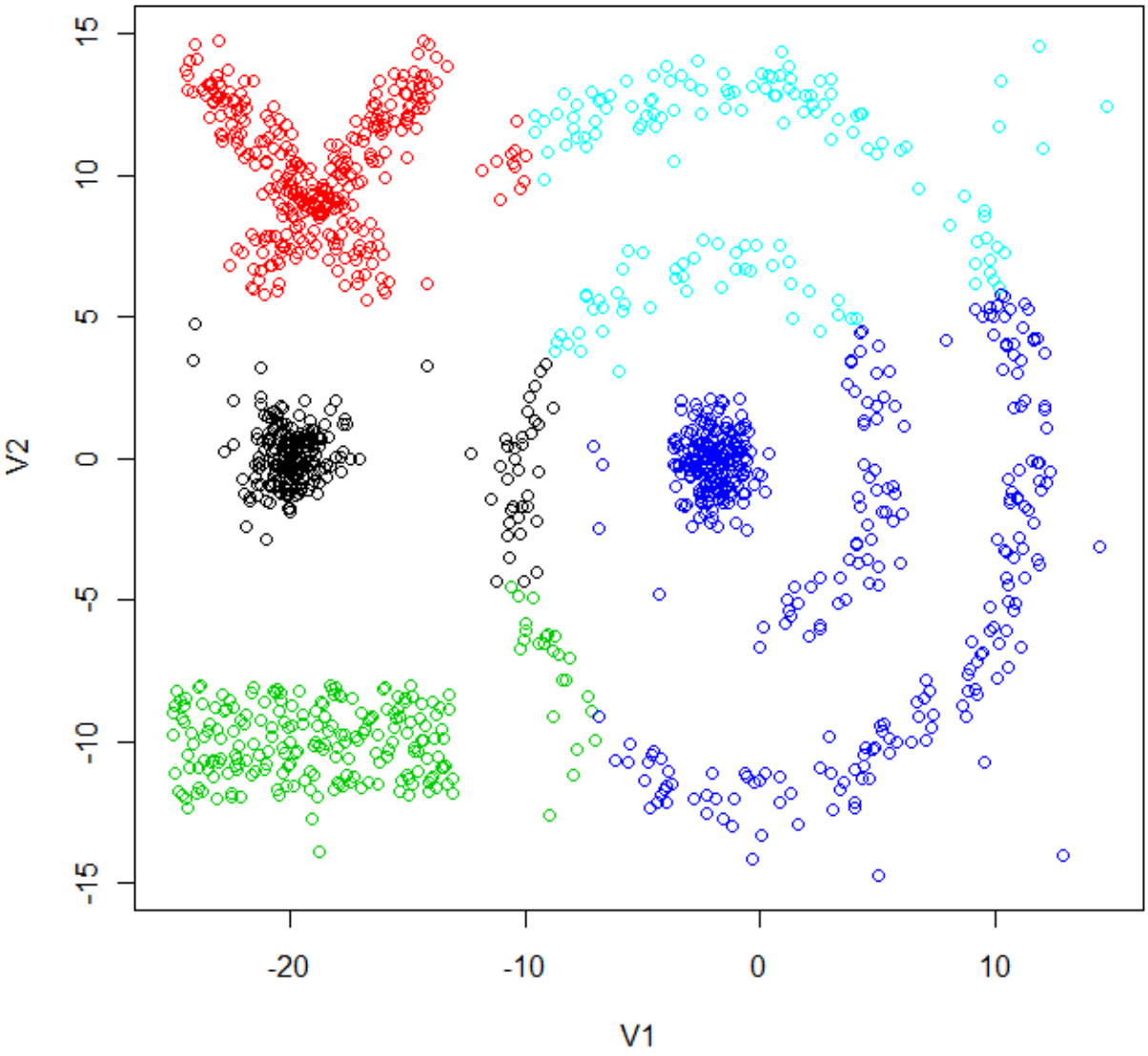
k = 3



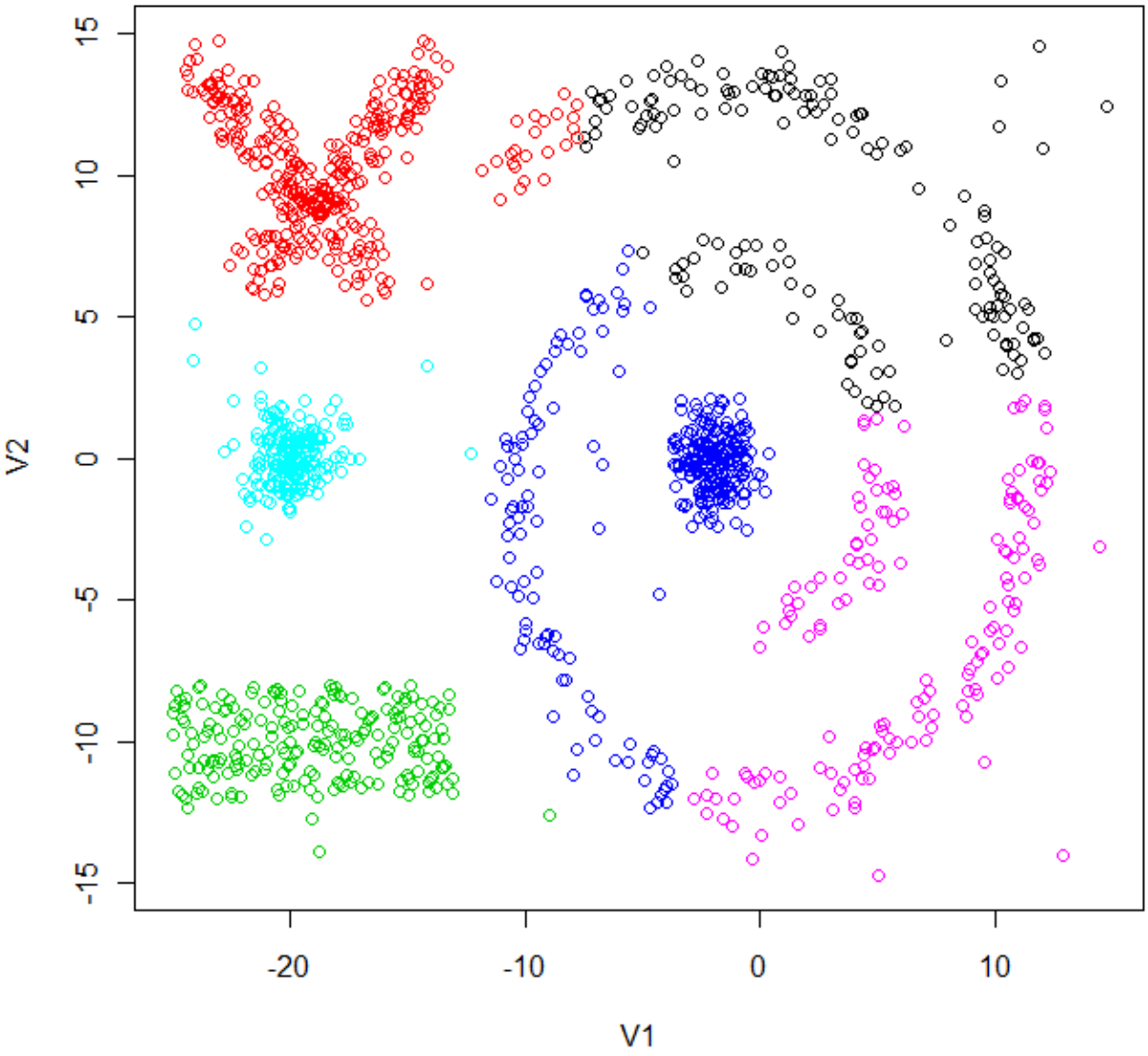
k = 4



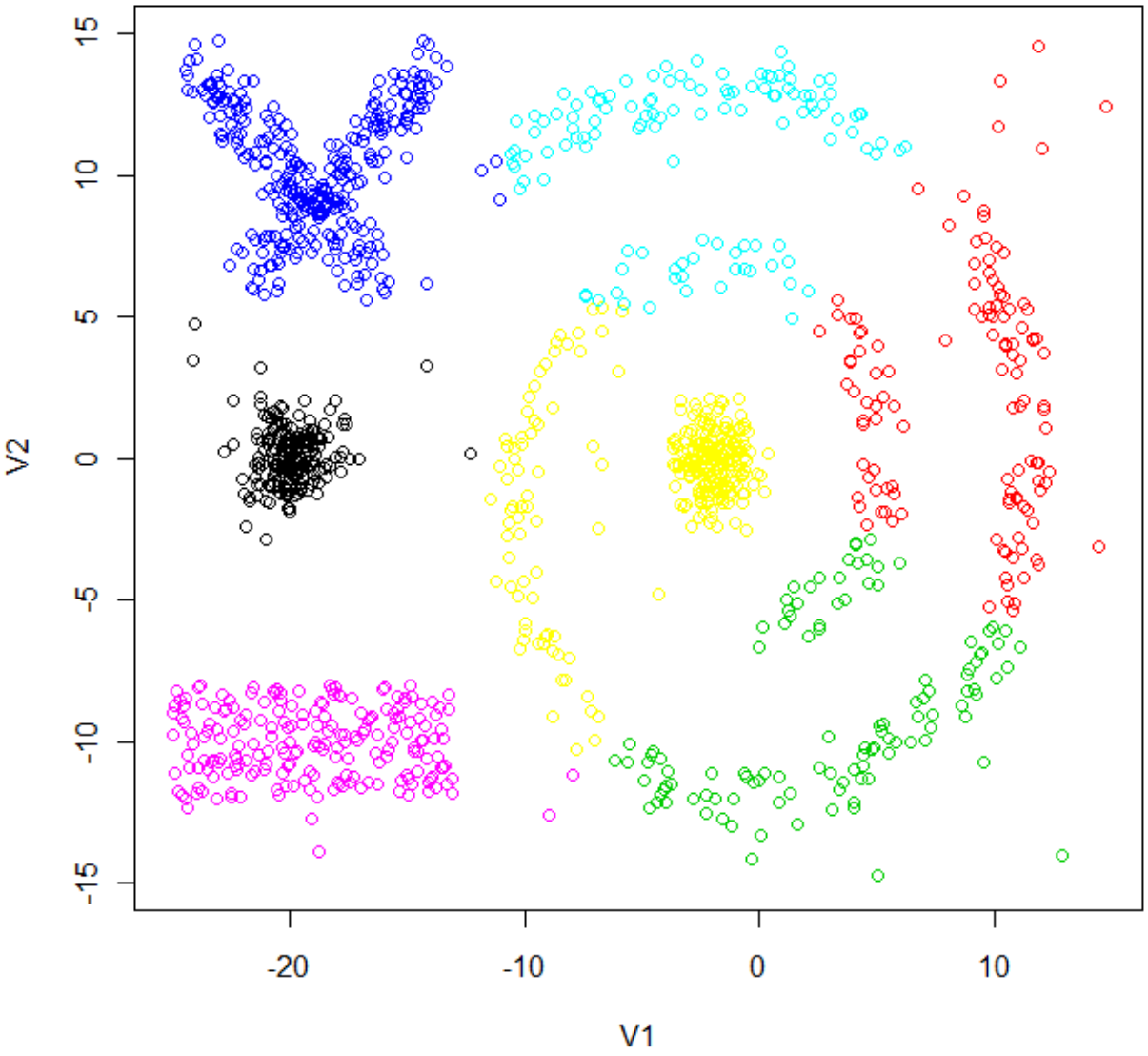
k = 5



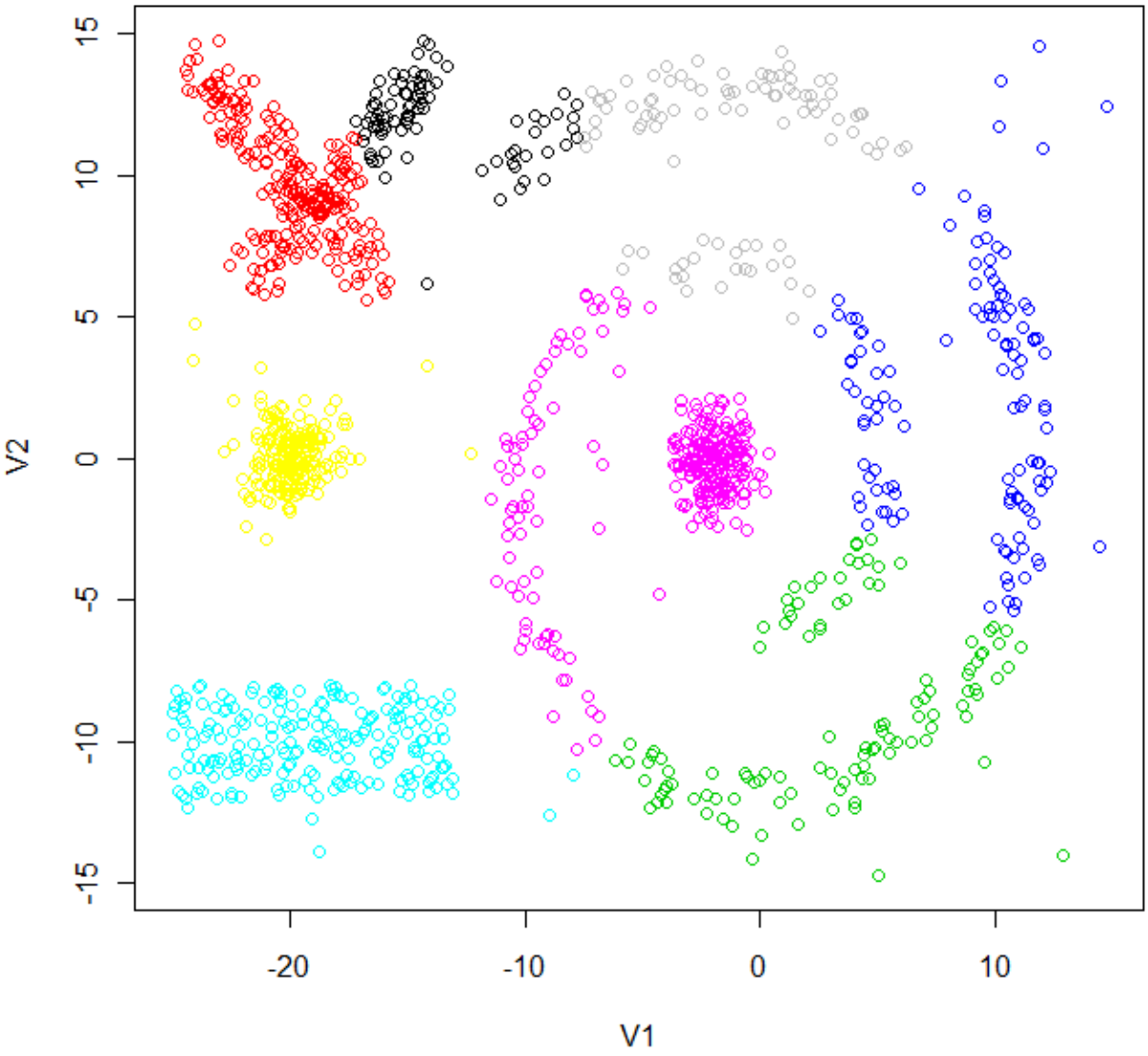
k = 6



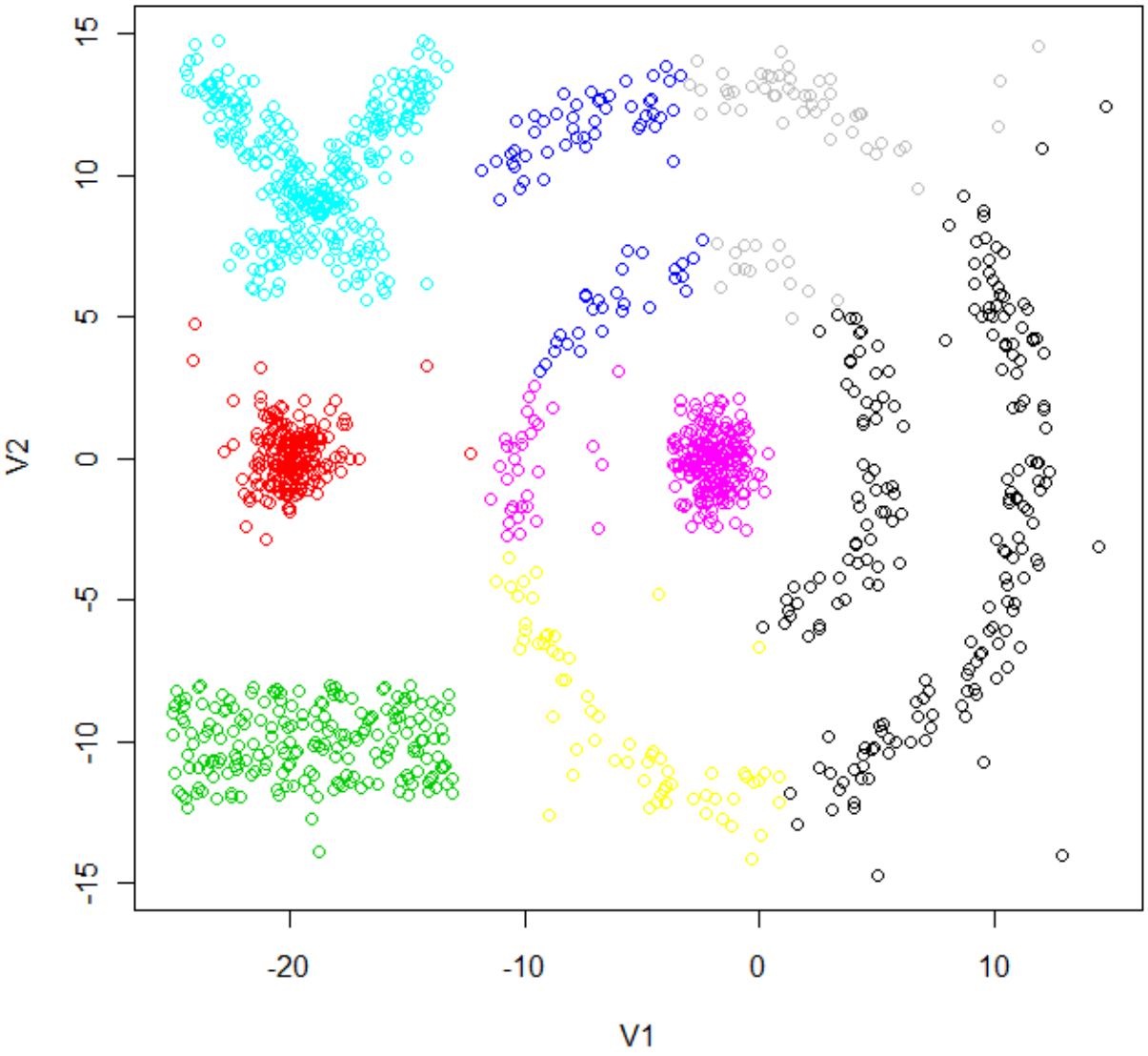
k = 7



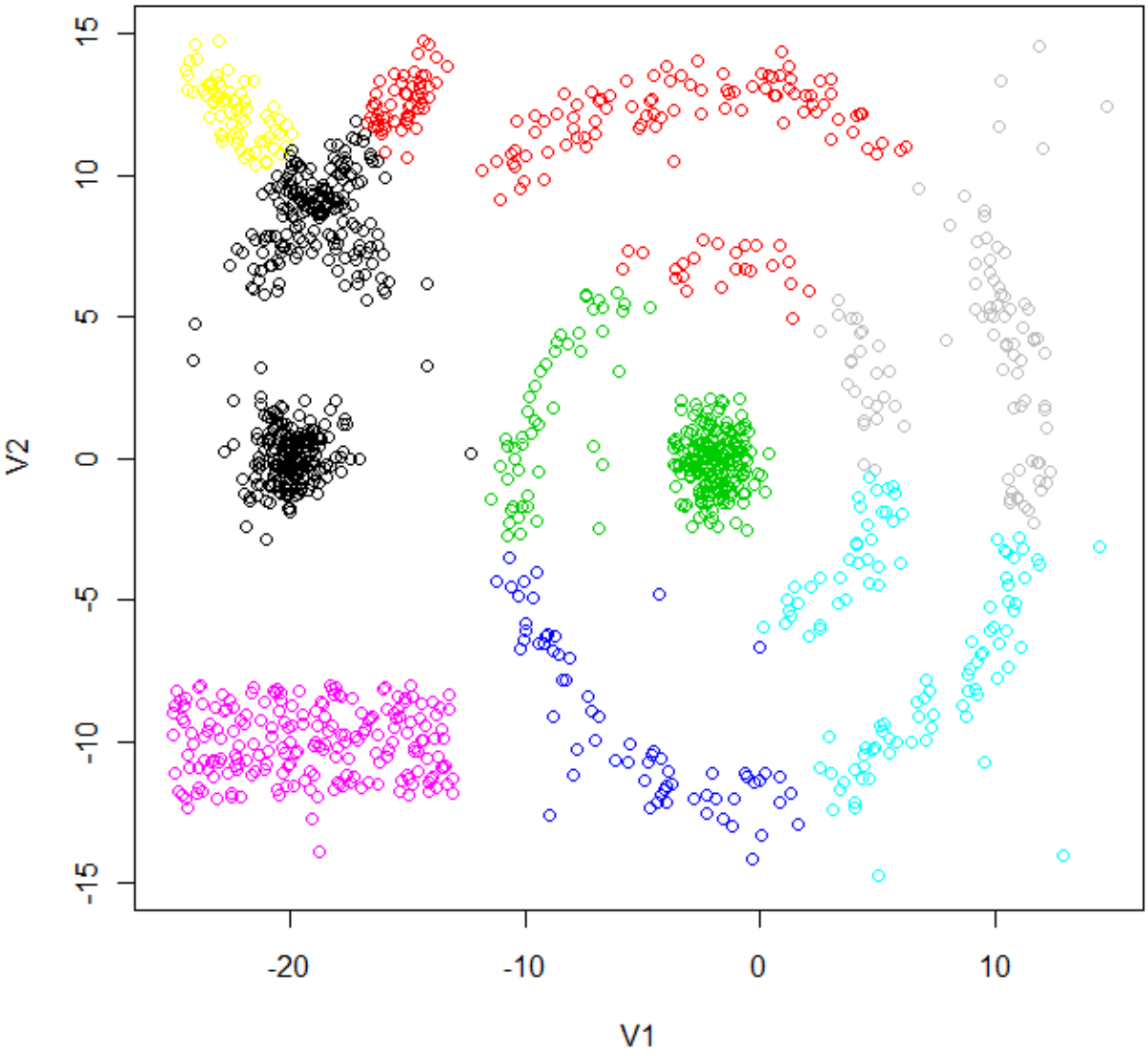
k = 8



k = 9



k = 10



3. Judging from both the SSE curves and the Silhouette curves, I'd have to say that k=10 is the most correct clustering.

4.

Non-Noisy Cophenetic Correlation

	Single	Complete	Average	Ward
Cophenetic Correaltion	0.66	0.77	0.78	0.72

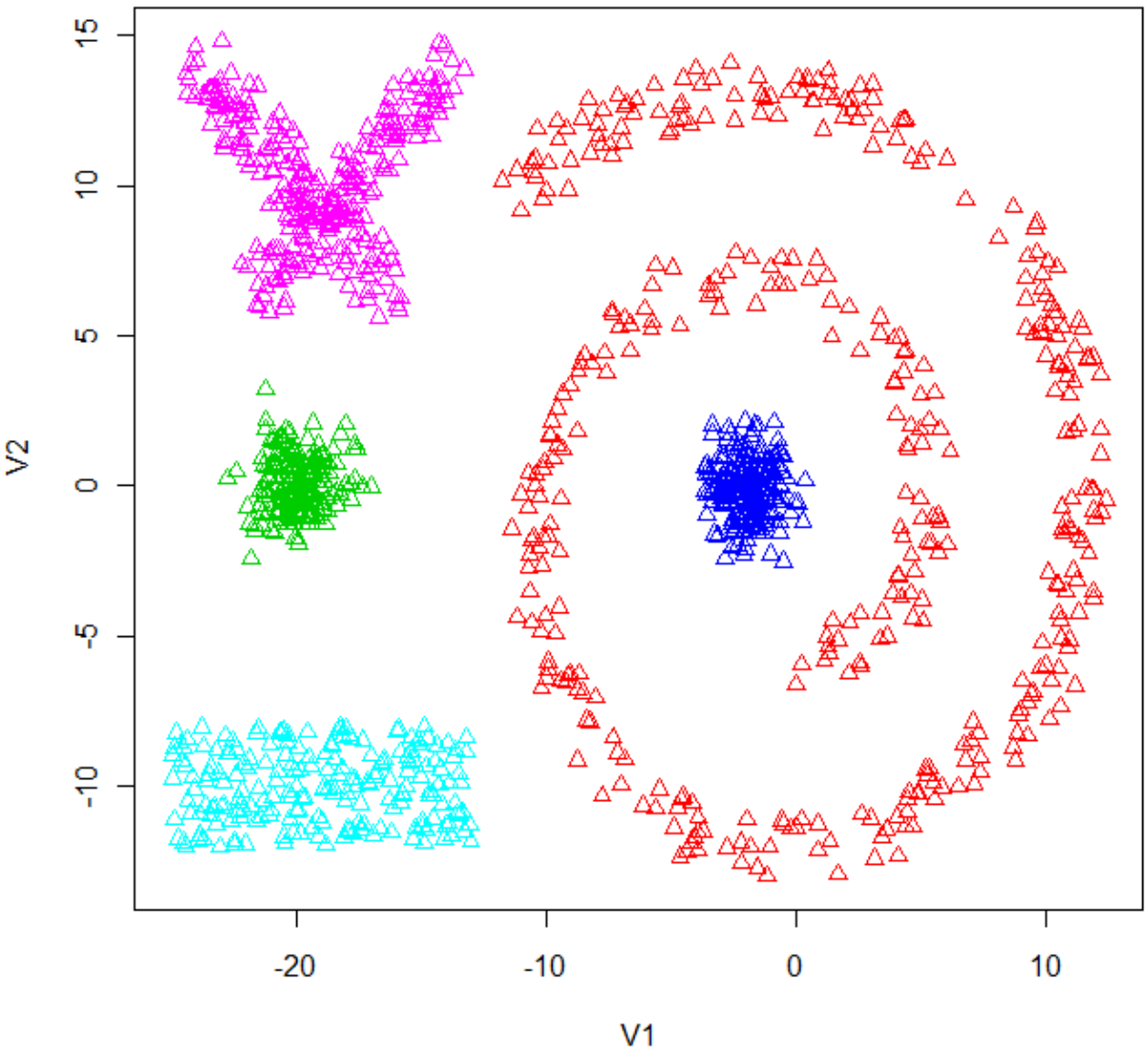
5.

Noisy Cophenetic Correlation

	Single	Complete	Average	Ward
Cophenetic Correaltion	0.65	0.72	0.77	0.71

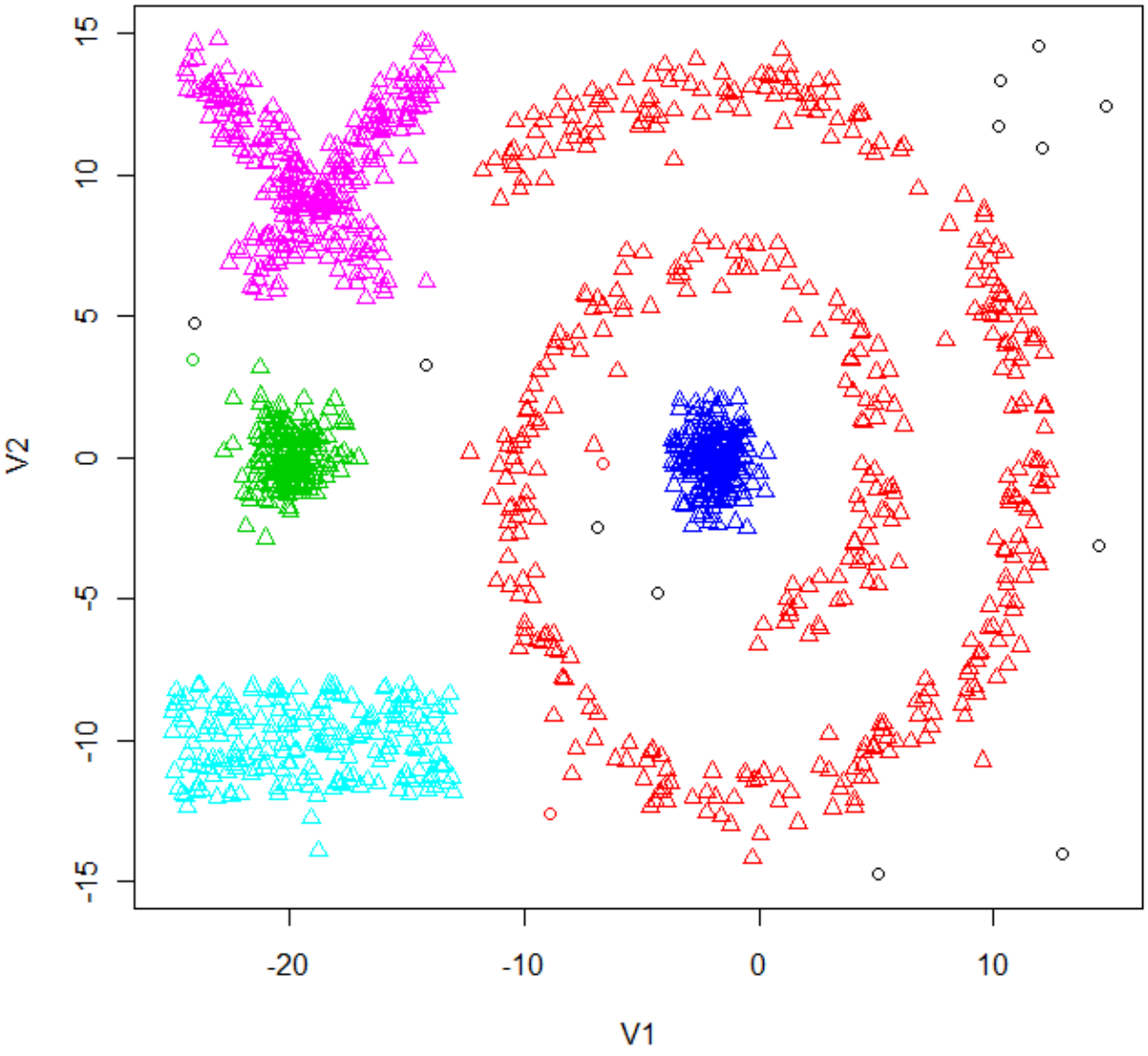
6. The best settings of linkage are the same for questions 4 and 5. In this case, the best linkage setting is "Average Linkage" in both cases.

7. Non-Noisy Data: Eps = 2.5, Minpts = 4



a) As minpts increases, DBScan identifies more of the datapoints as noise. It starts with the sparser sections; however, if minpts is high enough then all points will be classified as noise.
b)As eps increases, the size of clusters will increase and the number of clusters will decrease. If eps is high enough, then all points will be placed in the same cluster.

8. Noisy Data: Eps = 2.5, Minpts = 4



a) As minpts increases, the number of outliers increasees. With a very small minpts value, every noise point is classified into a cluster. If a very high number is chosen, then every point is classified as an outlier.
b) As eps increases, the size of the cluster increases. This is because eps represents the reachability distance of the DBScan algorithm. If you increase eps too high, the entire dataset will be placed in the same cluster.