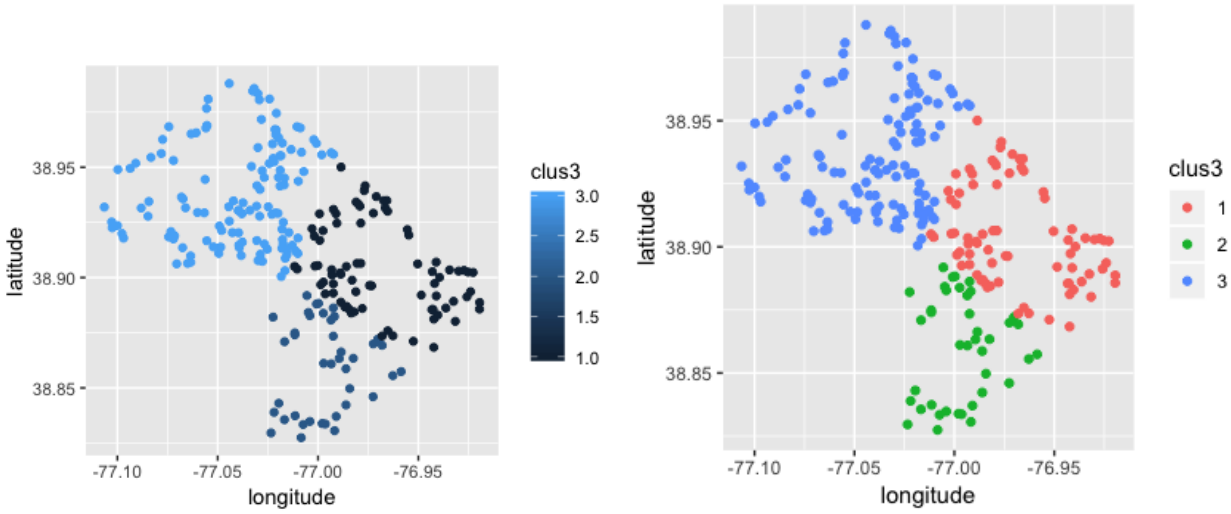
Analytics of Business Intelligence - Fall 2019, Practice Set # 5

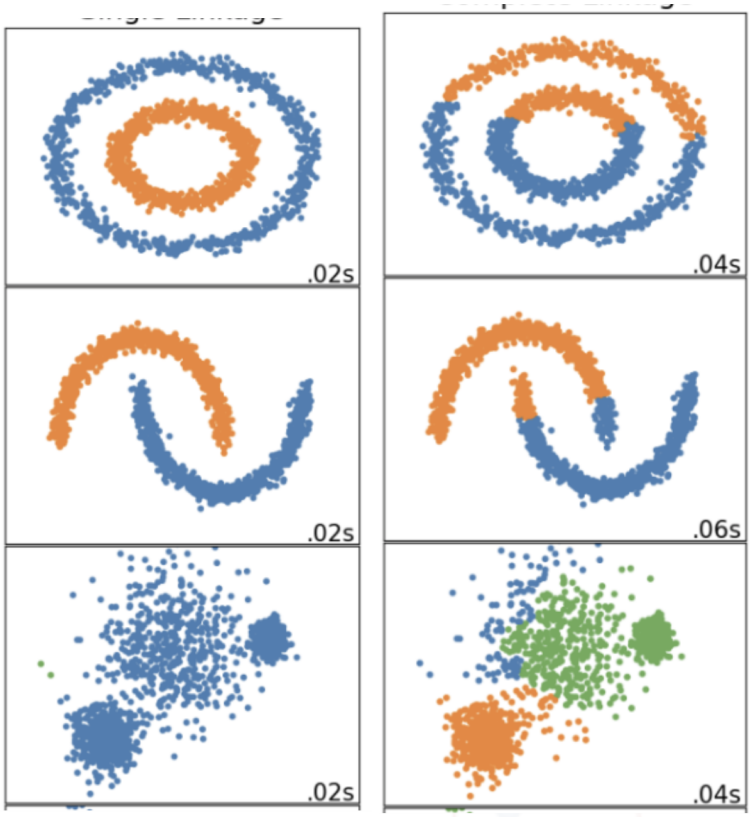
1. **There are 2 clusters. Cluster 1 has 2 data points, (4,10) and (1,8), and Cluster 2 has 3 data points, (8,5), (7,4), and (9,2). Calculate the error rate.**
2. **(Code) Given the table food, that has 2 columns, price and pounds. Calculate the elbow graph from 2 clusters to 10. The units of price are in dollars and the units of pounds are in 100's.**
3. **(Code) Given the code below, change the graph from the first plot to the second one.**




```
1  ggplot(bike,aes(x=longitude,y=latitude,color=clus3)) + geom_point()
```

4. **(Code) The table dt has 5 columns, (station, k1, k2, k3, k4). The column station has the number of the bike station, k1 has the distance from each bike station to that kiosk, k2 has the distance from each bike station to that kiosk, etc. Calculate the average distance between each bike station and 3rd closest kiosk.**

5. **Given the 5 data points, (4,10) and (1,8), (8,5), (7,4), and (9,2). Calculate the 5 levels of clusters (5 to 1), using the single linkage method, complete linkage method, average method, and centroid method.**
6. **Given the clusters below, which vertical plots were made using the single linkage method vs the complete linkage method?**



7. Write a simple R program that mimics the kmeans() algorithm.