BS 806

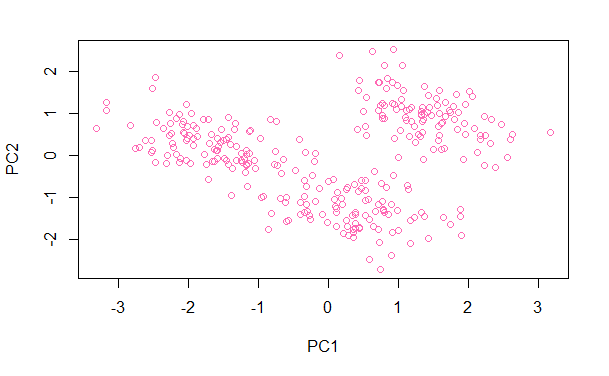
Homework 10

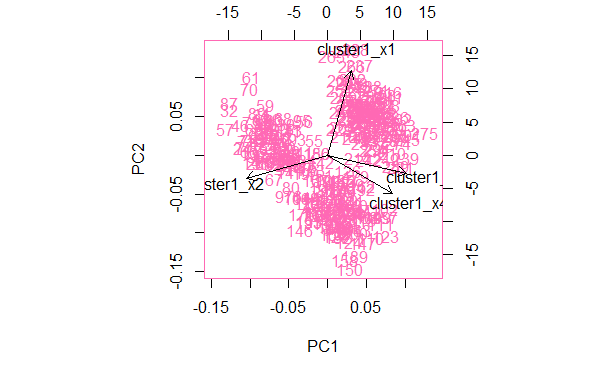
Irene Hsueh

**Question 1**

The standard deviations of the variables were quite close to each other, so there was no need for standardization of the data.

**Question 2**

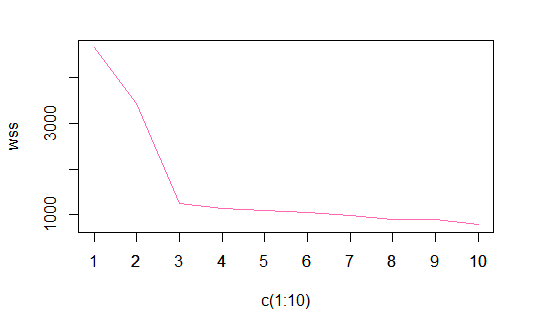




The plot of the first two principal components seem to suggest that there are 3 clusters in the data. X1 forms one cluster, X2 forms the second cluster, and X3 and X4 form the third cluster.

**Question 3**

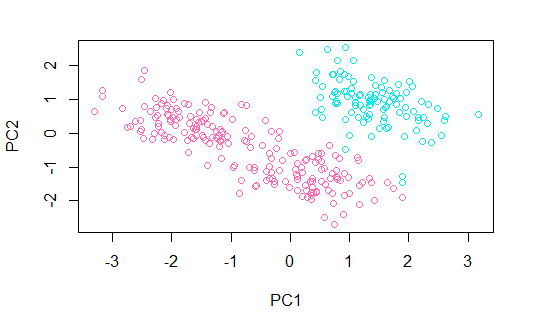
Part A



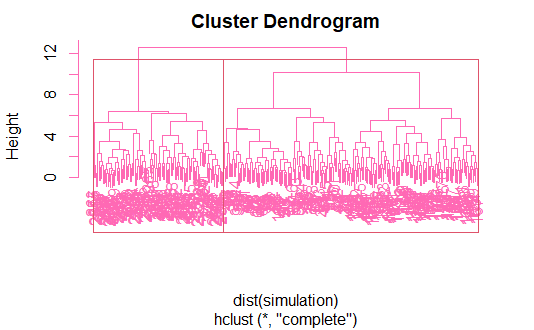
The graph of WSS for different number of clusters suggest that there is most likely 3 clusters.

Part B

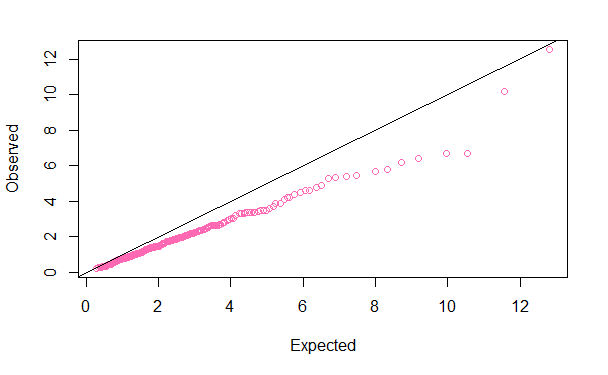
Part C

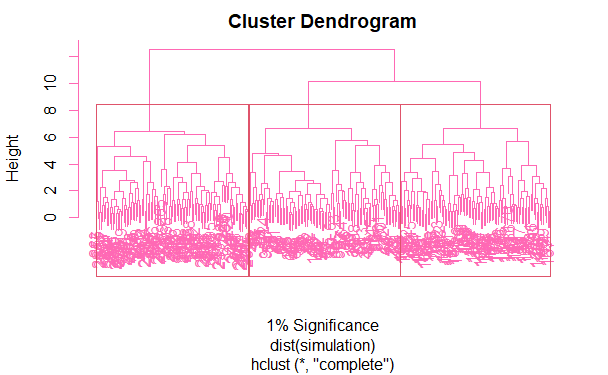


**Question 4**



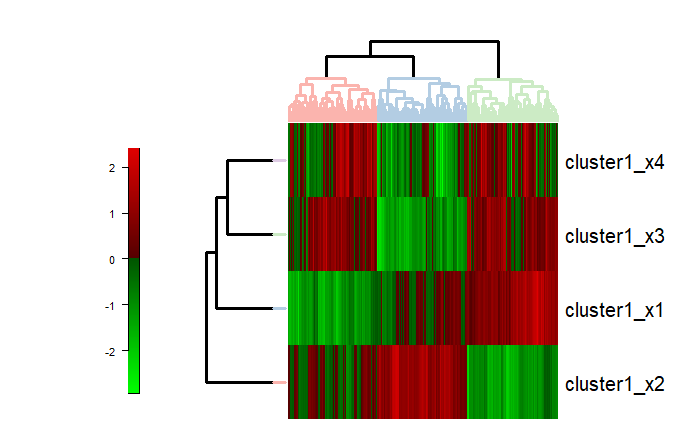
**Question 5**





The Q-Q plot shows significant departure from the reference line, which suggests there are clusters. At 1% significance, cutting at a height of 9.976784 gives 3 clusters.

**Question 6**



There are 3 significant clusters at the 1% significance level

**Question 7**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | Hierarchical Clustering | | |
| **1** | **2** | **3** |
| K-Means Clustering | **1** | 100 | 0 | 0 |
| **2** | 0 | 99 | 0 |
| **3** | 0 | 0 | 101 |

Comparing k-means clustering and hierarchical clustering, all 300 observations have consistent clustering sorting.