

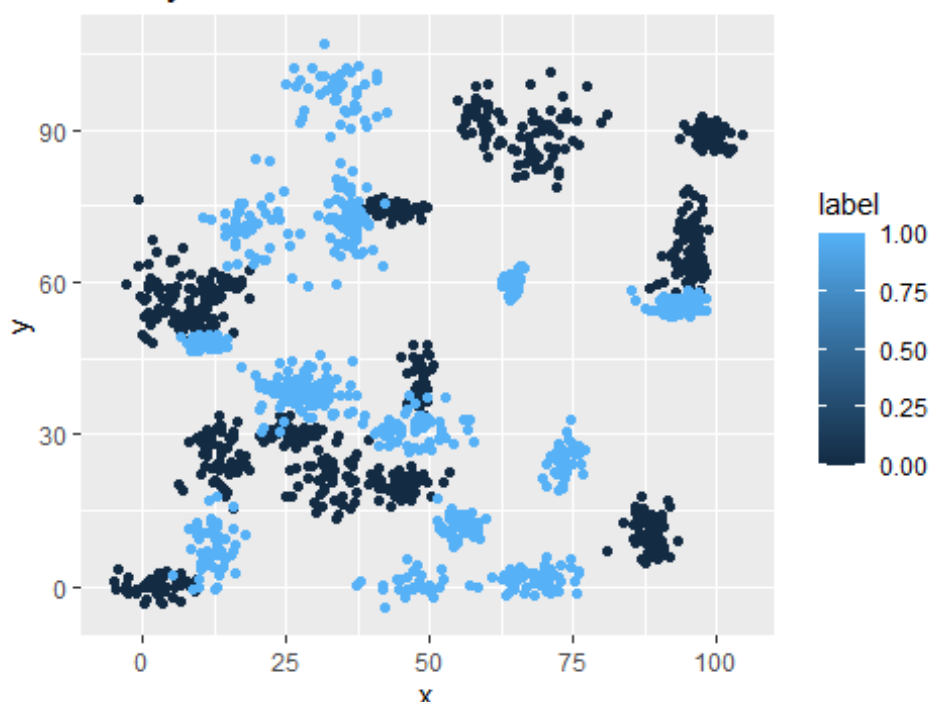
10.2 Clustering

Shaquiel Pashtunyar

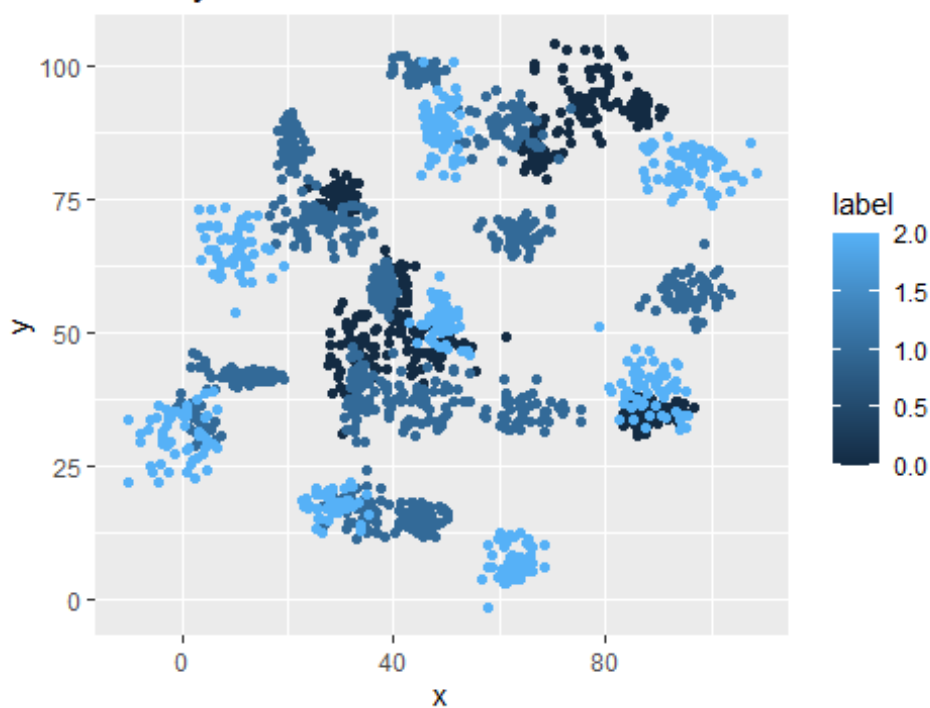
2022-08-11

Scatterplots of both data sets

Binary Data



Trianry Data

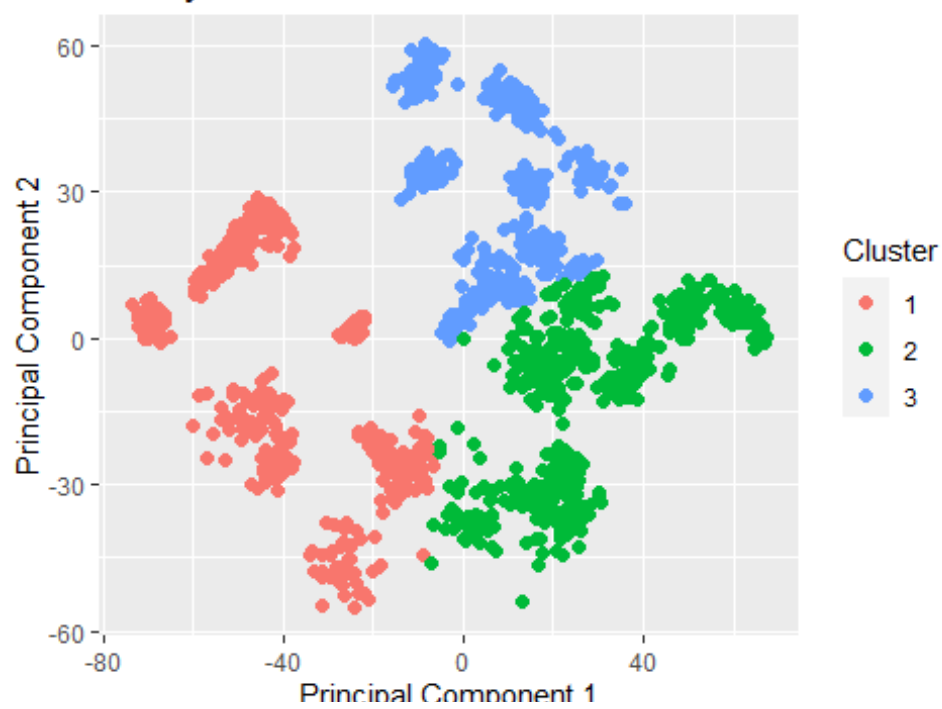


Fit a k nearest neighbors' model for each dataset for $k=3$, $k=5$, $k=10$, $k=15$, $k=20$, and $k=25$. Compute the accuracy of the resulting models for each value of k .

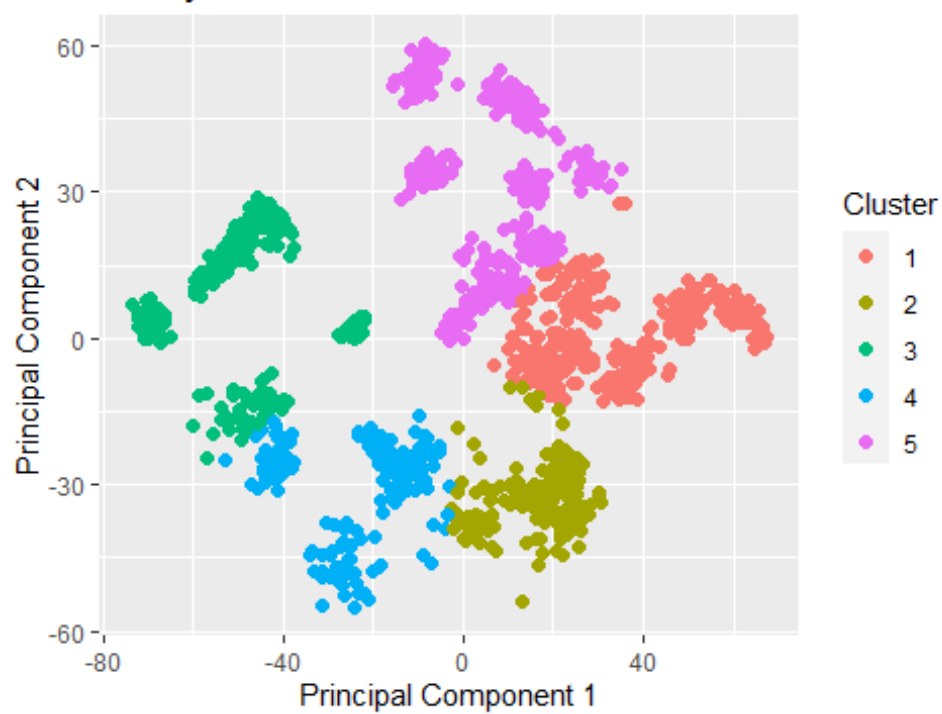
Plot the results in a graph where the x-axis is the different values of k and the y-axis is the accuracy of the model.

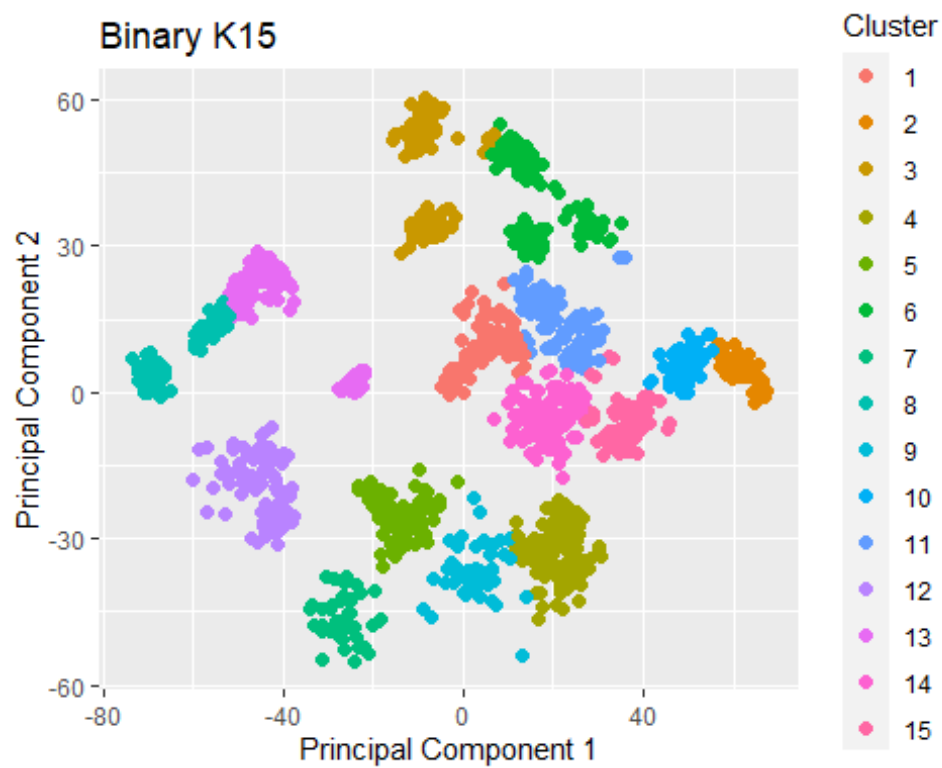
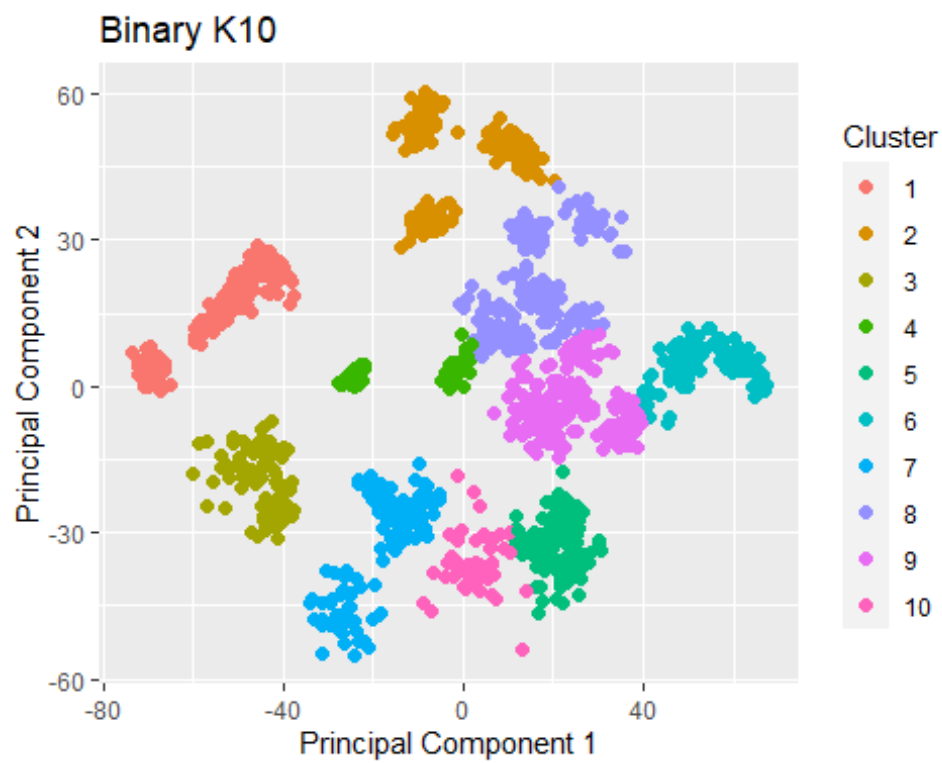
Binary Plots

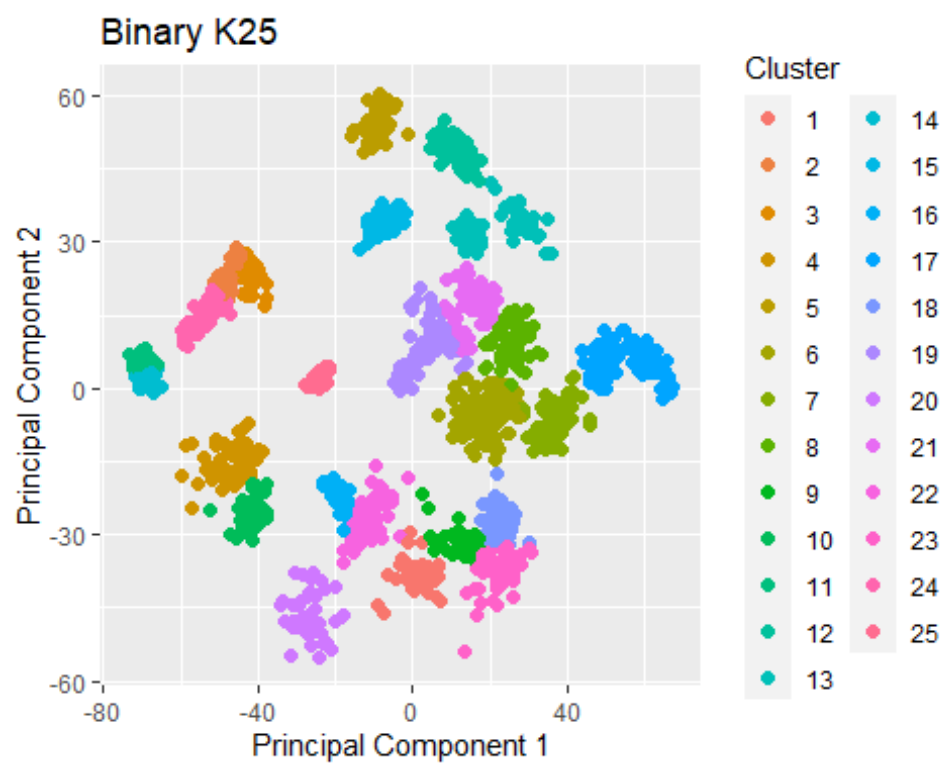
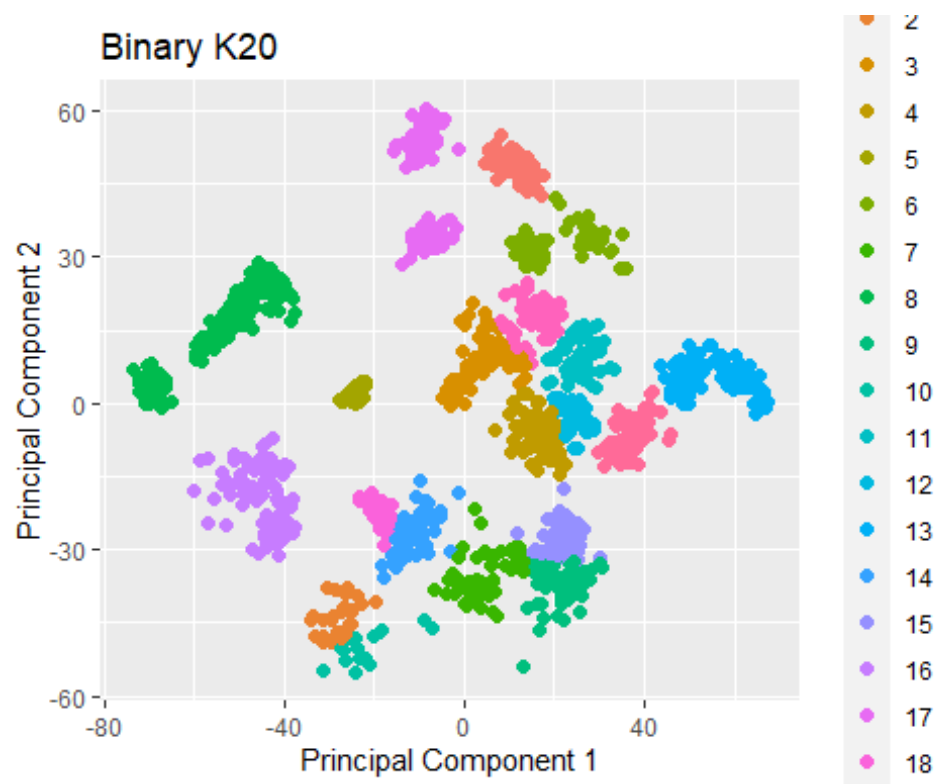
Binary K3

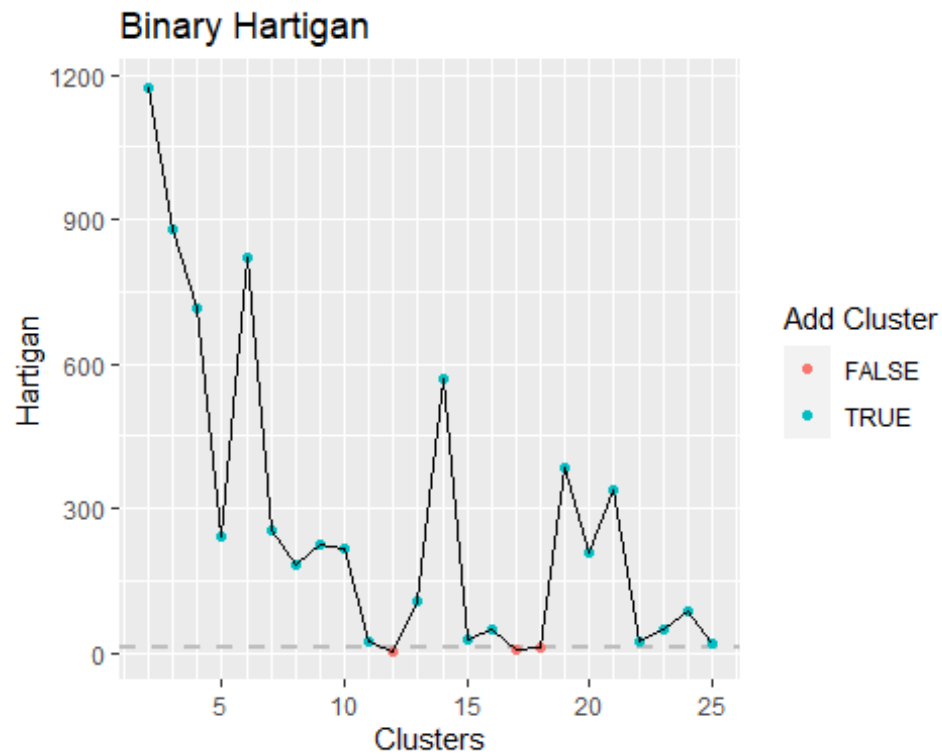


Binary K5



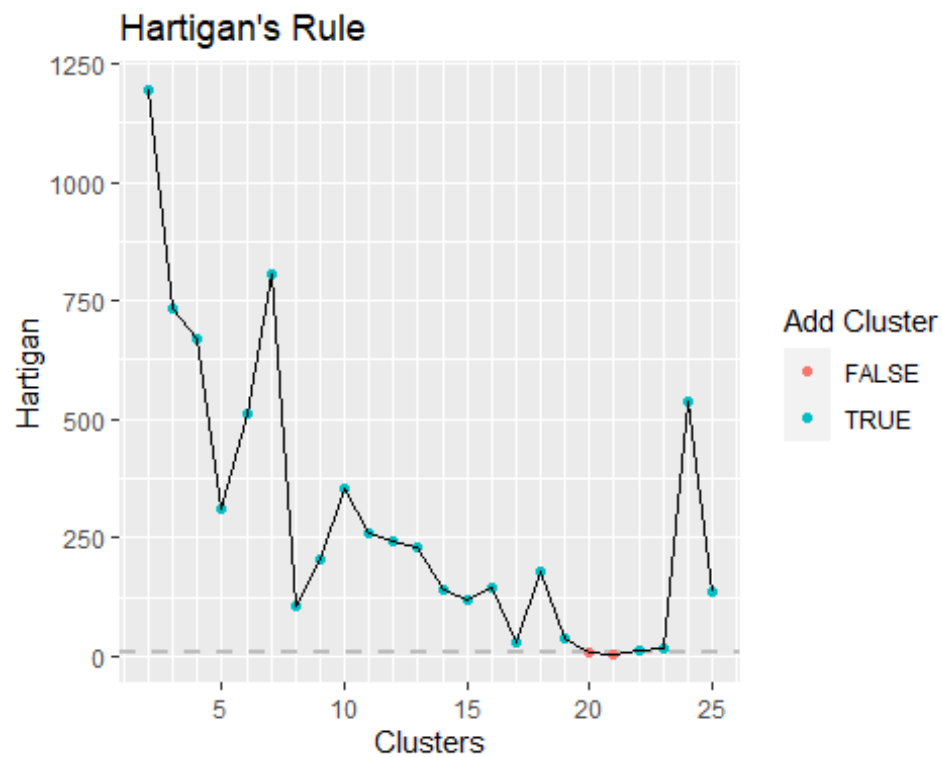






The best number of clusters to use here is 12

Trinary Hartigan

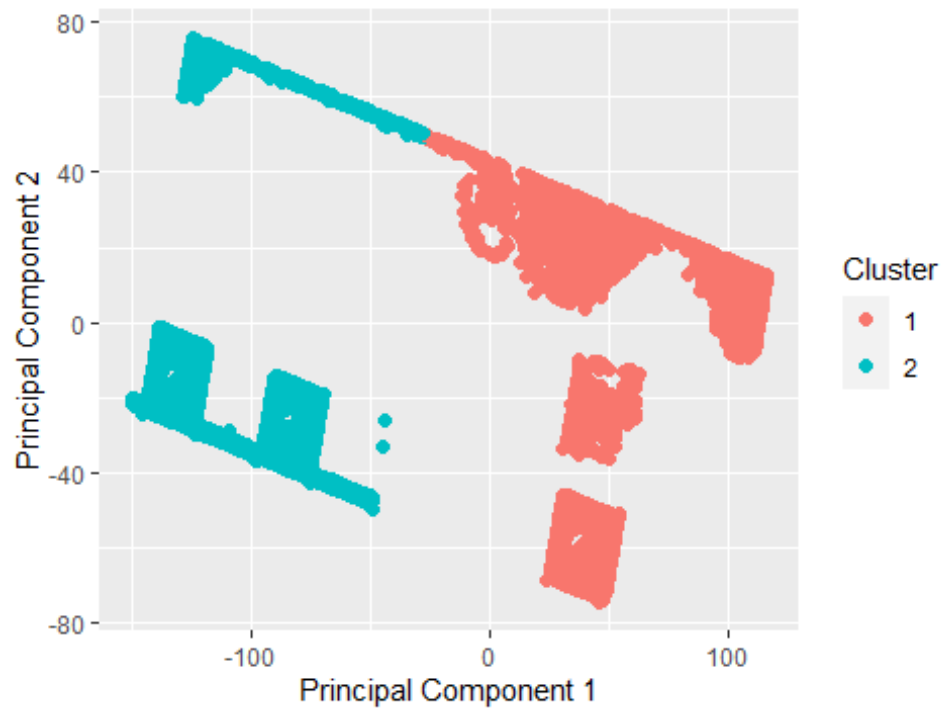


20 is the best fit here

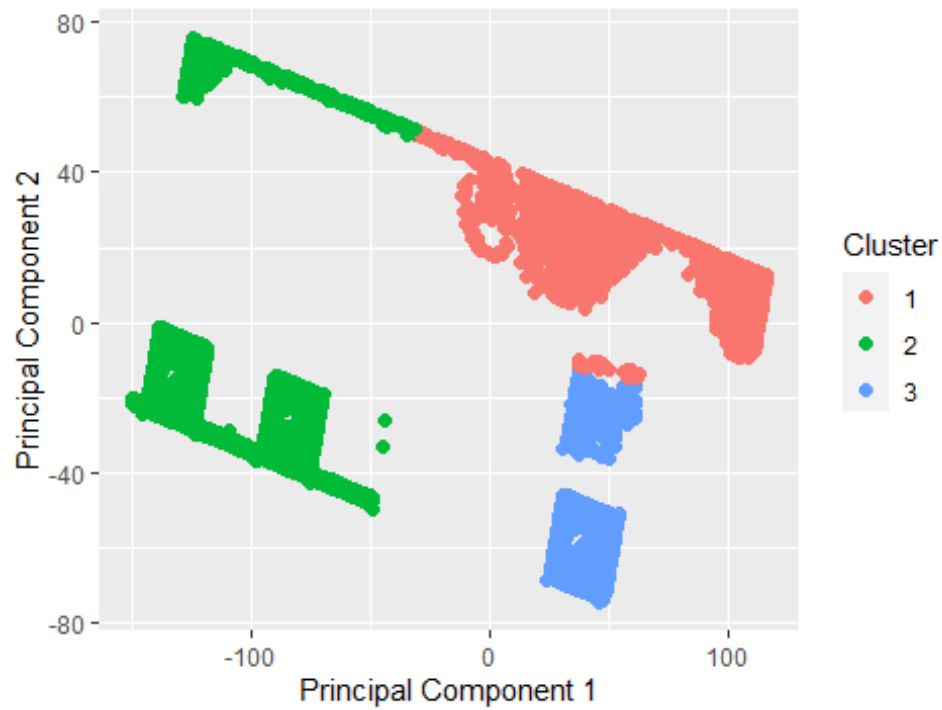
Q1 Just by looking at the two scatterplots, we can see that the data doesn't really show a linear model, so a linear classifier would not be better. We can also see that each color clusters together in different regions.

Q4

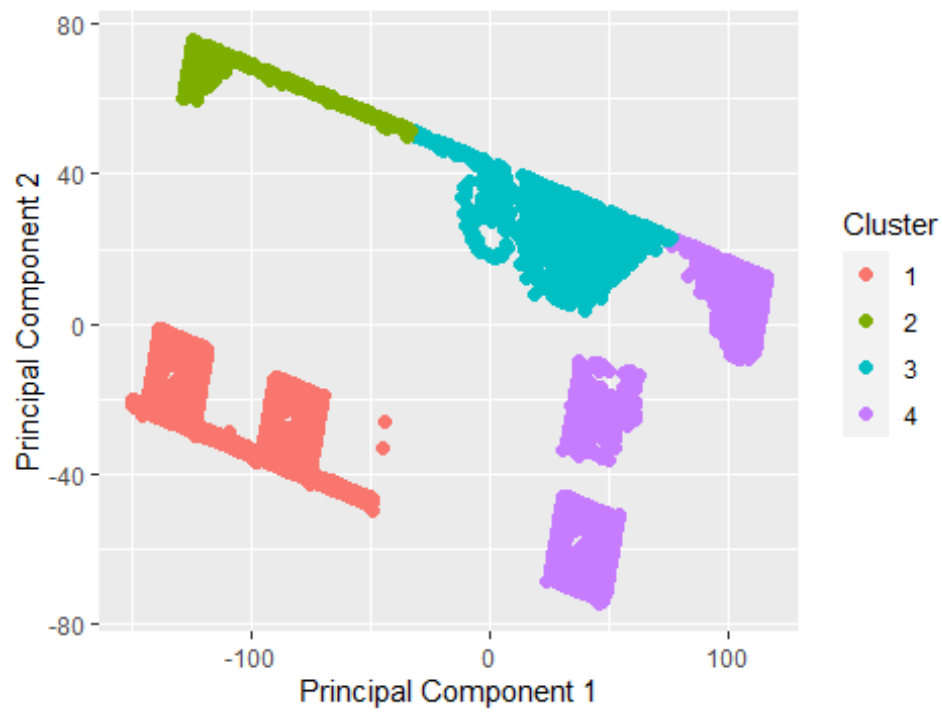
K-Means Results



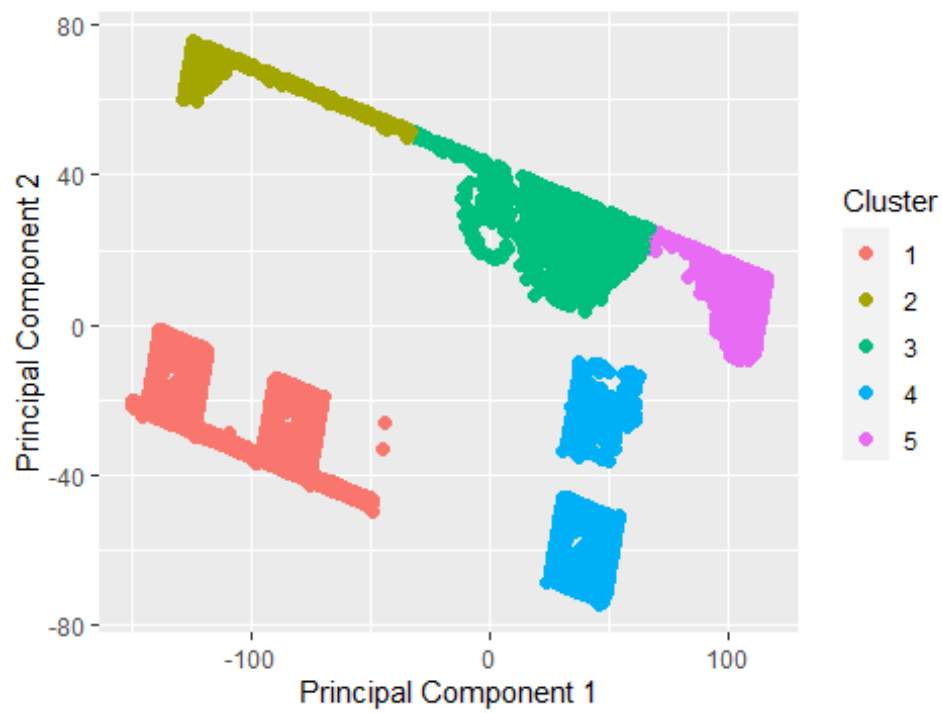
K-Means Results



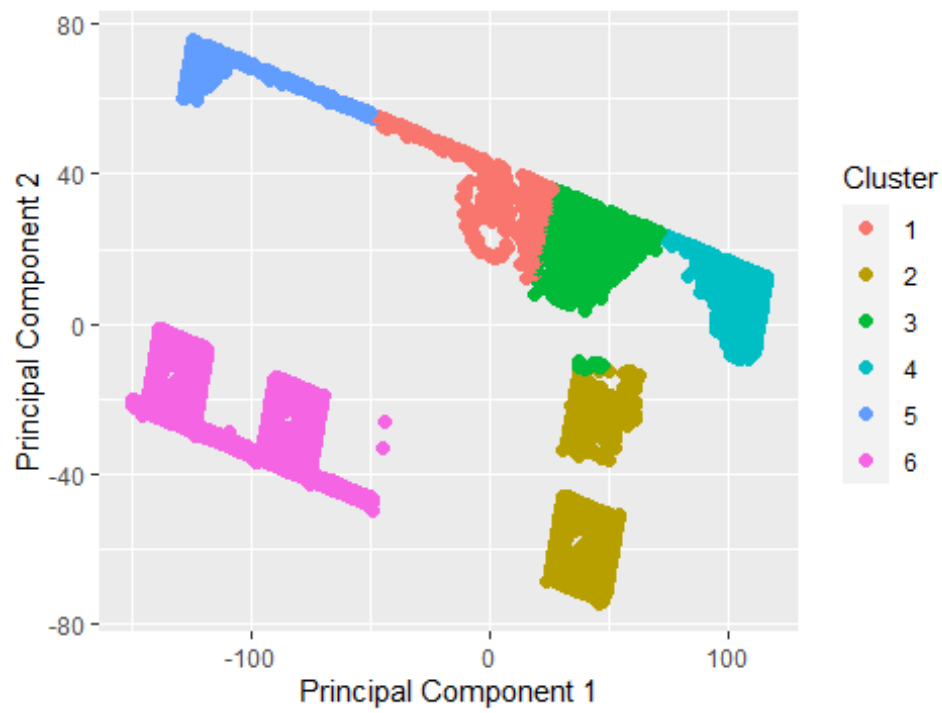
K-Means Results



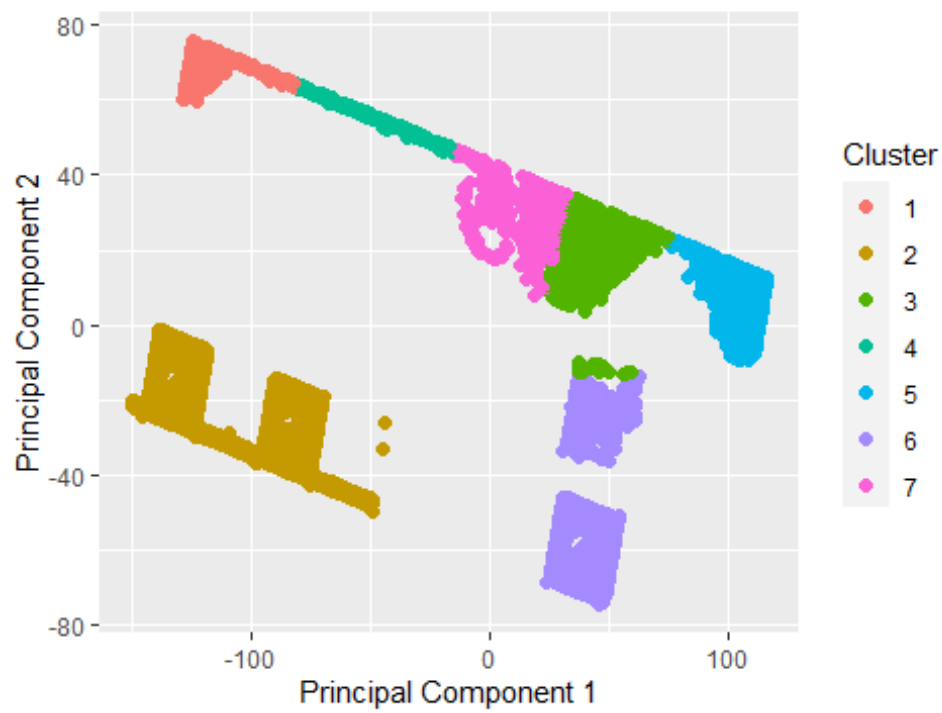
K-Means Results



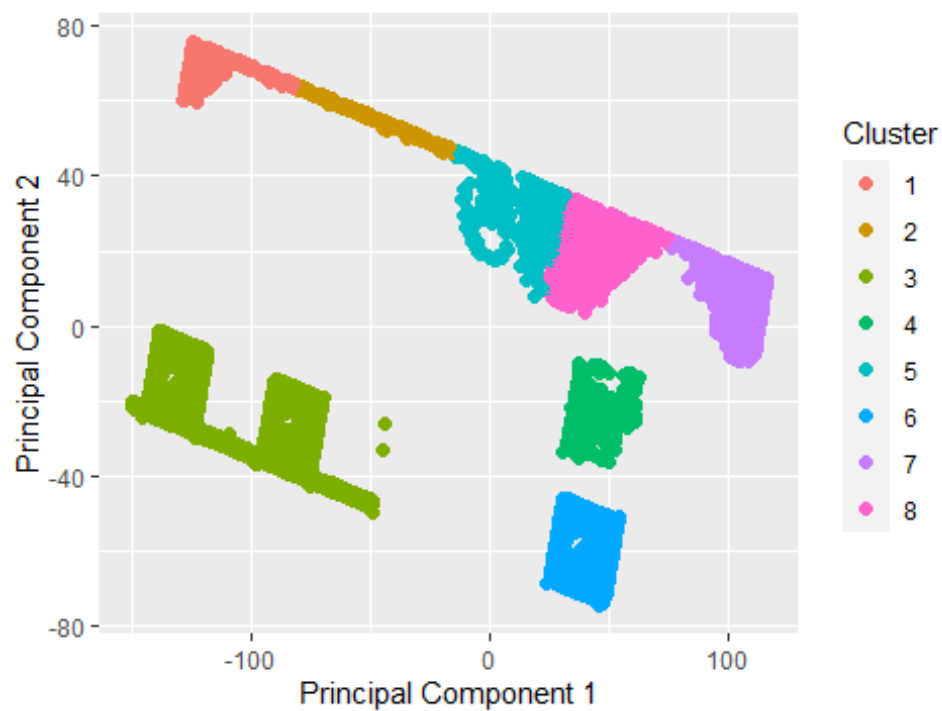
K-Means Results



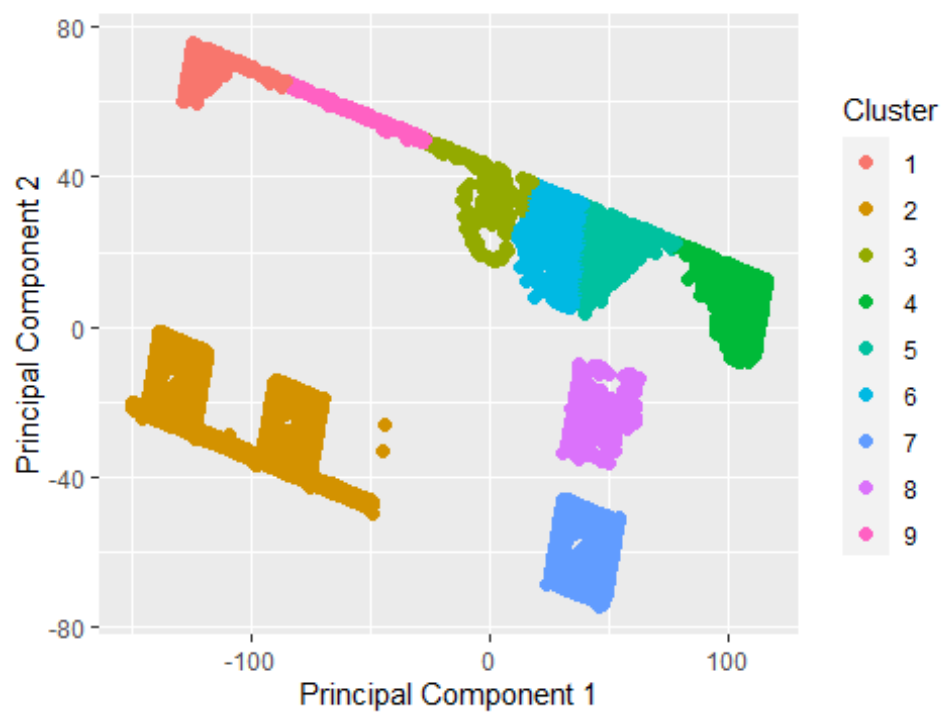
K-Means Results



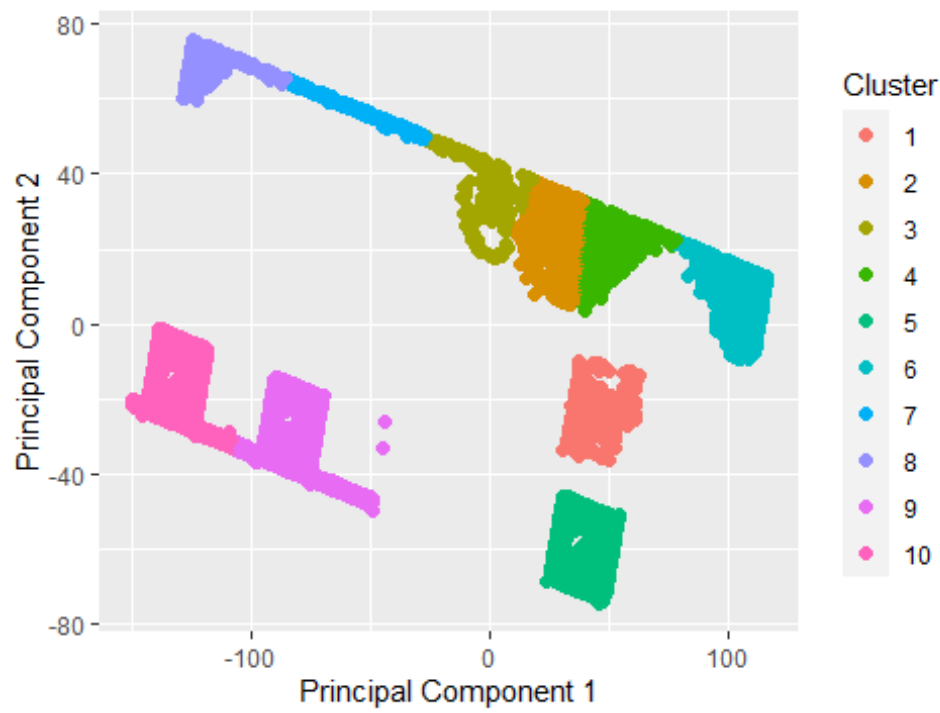
K-Means Results



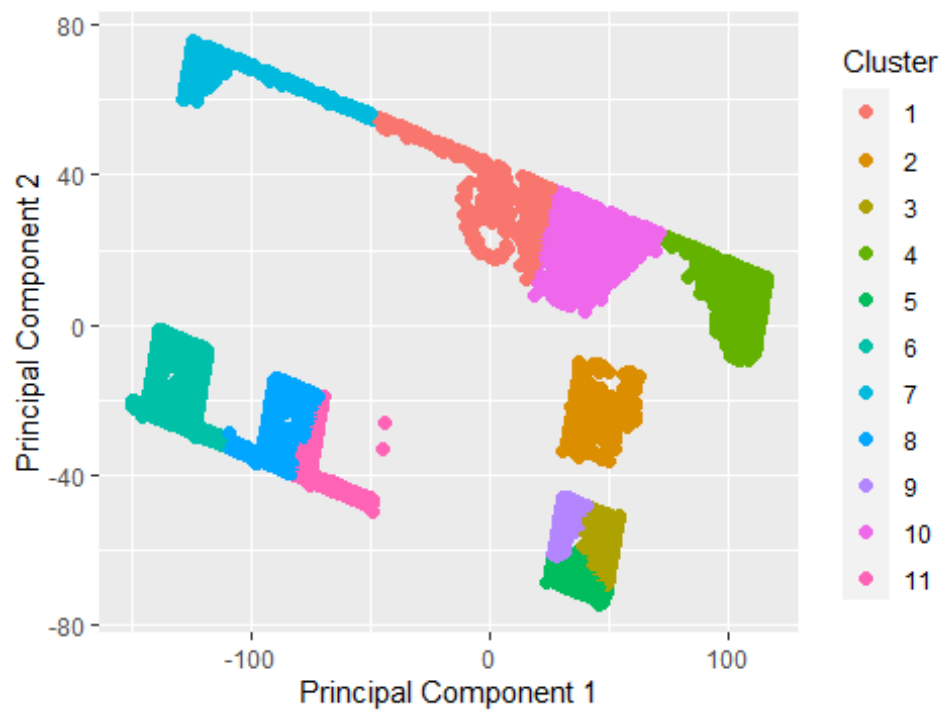
K-Means Results

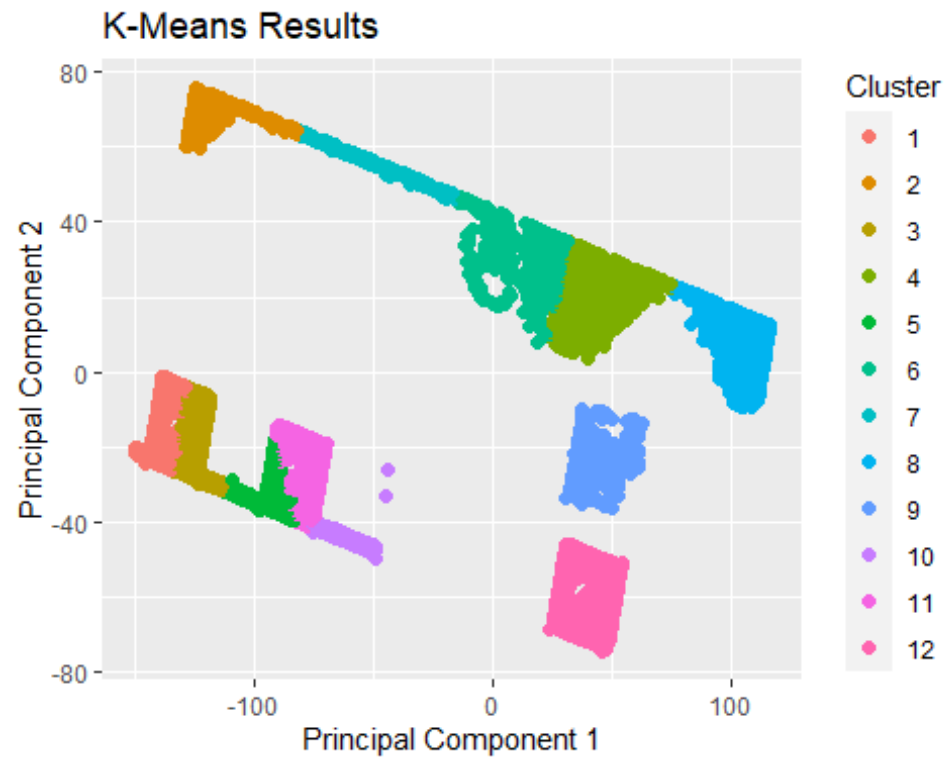


K-Means Results



K-Means Results





```
## [1] 6067.522
```

At this point I do not know how to use the euclidean function exactly. I see it is telling me the distance of points, but how i get that number per cluster is unknown