Cluster analysis
Hierarchical Clustering
define when one thing is closer to each other and when things are further apart, . Group things together.
organize data, start with the data, start lumping them together and gradually lead to a cluster.

Make a merge point and find closest things, and make a tree like structure,. Use different distance metrics.

Keeping finding and putting things together to get a tree sort of style.

Need a distance metric. Examples of Distance metric: euclidian distance, Correlation Similarity, Binary Distance or the manhattan distance.

Creates a tree called a dendrogram

dist(data frame) - get the distance between all the points and see how far the various points are from each other r.

heatmap() function very useful to see matrix data ,

K-Means Clustering

Not deterministic, run it a couple of different types to get a stable view of the data.

First define what does it mean for things to be closer to one another. Same distance metrics as above.

Way of partitioning points to a a certain number of groups.

Each group Will have a centroid. Iterate back and forth to figure out where the centroids are to produce a final estimate of where the centroids are how each can be used.

Continually recalculate and form new clusters to assign points to each of the clusters and assign the points to the clustering locations.

Kmeans() = (data frame, centers=3) . ,

names(means) - get the vector of where the clusters are ,

Use heat map to run kmeans to create a heat map with the clustered datapoint in the matrix. Useful for high dimensional data to look at matrix data in many ways to get different visualtions.

Dimension Reduction

Adding a pattern to the data to make a slight shift to all the data to get a better look at the step patters.

Trying to reorganize data to find patterns. Trying to plot different row and column means. Reorder data to get the row and column means to see how the data is divided.

Taking advantage of the matrix structure of the data to find things that are related and correlated to each other. Trying to create a reduced set of variables to explain much of the variability in your data.

Trying to find a lower-rank matrix and then find a data compression to get a smaller data representation,.

PCA - Principal components
SVD - Singular Value Decomposition