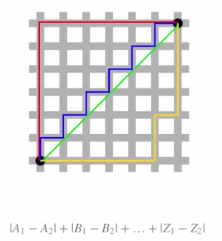
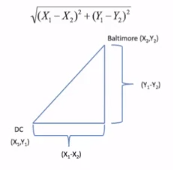
Clustering

# Hierarchical Clustering

**Method to explore data**

1. Find closest two things
2. Put them together
3. Find next closest
   1. Produces a tree that shows ordering of how close they are together
4. Chose a merging approach?



Define close?

**Continuous – Euclidian distance**

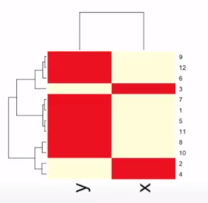
* Straight line, geometry

Continuous – Correlation similarity

Binary – Manhattan distance

* Follow the actual steps

# How to cluster

1. Calculate distance from each, pair-wise
   1. dist(dataframe)
   2. creates a matrix of pair-wise distances from each other
2. hclust tree
   1. hClustering <- hclust(distmatrix)
   2. plot(hClustering
3. heatmap function
   1. heatmap(dataMatrix)

# K-Means Clustering

**A partitioning approach**

1. Fix a number of clusters
2. Get “centroids” of each cluster
3. Assign things to closest centroid
4. Recalculate centroids

**Requires:**

1. Defined distance metric
2. Number of clusters
3. Initial guess to centroids

**Example**