Grouping Get Connected students

Kenny Yue

1. Build a matrix of distances between all addresses.
2. Starting with a random address, sort the matrix by ascending distance to that address
3. Removing the addresses from the matrix as you go along, place the addresses closest to the start address into a new group until you reach your criteria (size of group or max distance).
4. Once a group is full, choose another random address and resort the matrix by distance to that address
5. Continue like this until all addresses are taken out of the matrix.

**Clustering in R**

## Data Preparation

To perform a cluster analysis in R, generally, the data should be prepared as follows:

1. Rows are observations (individuals) and columns are variables
2. Any missing value in the data must be removed or estimated.
3. The data must be standardized (i.e., scaled) to make variables comparable. Recall that, standardization consists of transforming the variables such that they have mean zero and standard deviation one.

**GOAL:**

* Creating automation by the correct ratio of male to female.
* Manage the grouping to be as close as possible.