Team Info

1. Members: Qinzhi Peng, Xinyan(Hathaway) Liu, Yujie(Johnny) Ye, Zhankai Ye

Outline

1. **Topic:** We will write a vignette for a [record linkage package](https://cran.r-project.org/web/packages/RecordLinkage/RecordLinkage.pdf) and explain this to students in the class.
2. **Task Assignment:** We will introduce functions related to supervised classification and weight-based deduplication of the record linkage package.
   1. Weight-based deduplication: Qinzhi Peng, Xinyan(Hathaway) Liu
      1. Using RLdata500 or RLdata1000 to generate record pairs
      2. Using mean of EM Algorithm to calculate weight
      3. Use emClassify(), which is one of the classification method in RecordLinkage package with the record pairs to generate to perform deduplication on data
      4. The result will be a classification table along with the following metrics:
         1. Alpha error
         2. Beta error
         3. accuracy
   2. Supervised classification: Yujie(Johnny) Ye, Zhankai Ye
      1. Generate Data
         1. Adjust the matches/non-matches sizes
      2. Training
         1. The package provides six methods to train (rpart, bagging, svm, ada, nnet, bumping)
         2. We can also check if there are any ways to adjust hyperparameters
      3. Classification
      4. Results  
         