* how the data were collected
  + Yelp Dataset Challenge
  + python JSON to CSV converter
* the goal:
  + What do you hope to learn?
    - How users tastes influence (or are influenced by) their friendship network
  + Why does it matter to you?
  + Why might it matter to others?
    - Businesses could target particularly influential/central individuals
    - If users have friends with very similar tastes, then ostensibly we could recommend friends based on combination of tastes and network structure
* some sanity checks and visualization are useful to gain a basic understanding of the data.
  + It is often necessary to clean or filter the data to deal with problematic observations—e.g., missing data, extreme outliers, etc.
  + State explicitly what design choices and assumptions you are making in this cleaning.
    - Only looking at restaurants.
    - Getting rid of users who have rated fewer than 10 restaurants
* Next, specify the modeling task—e.g., regression, classification, clustering, dimensionality reduction, recommendation system, etc.—along with the model(s) you’re considering.
  + Dimensionality reduction/recommendation/clustering/network analysis
* State explicitly why this learning task is a natural way to frame the goal. In most cases this should include a loss function that quantitatively balances model fit and model complexity, along with an algorithm for optimizing this loss function.
* Clearly define success metrics which quantify performance—e.g. accuracy, confusion matrix, ROC, etc.—and evaluate these metrics on both training and test data to assess fit and generalization.
* Be sure to address the issue of complexity control (a.k.a. “model selection”), as discussed extensively in class.
* Discuss the practical aspects of your project, including the scalability and computational complexity of the storage and runtime for the methods used.