* To read up on
  + Multidimensional scaling
  + Topic models/analysis
  + Correspondence analysis
    - patrick’s job talk analyzing faculty profiles
* Cluster analysis of therapists
  + Using treatment orientation data
    - Then analyze what the clusters refer to what
    - compare profile text…word cloud, sentiment analysis
  + Using profile text
* Network analysis of therapeutic techniques (by co-occurrence)
  + This also gives you cluster analysis
  + Can correlate with: fee, number of issues treated, number of orientations used, years of experience

Story:

PsychologyToday contains perhaps the most extensive public, searchable directory of mental health care providers in the US. (how many?). If a person anywhere in the US were to search Google for “thereapist [*city*]”, PsychologyToday’s therapist directory woud most likely be the first search result. As such, there is good reason to think it is also the most widely used web-based directory for finding therapists. Once on the site, people can filter results by several criteria – the problem the patient is seeking help for, the treatment modality/orientation desired, the desired sex/gender of the provider, list of accepted insurances, and others.

From here, users of the directory may view profiles of providers that meet their criteria. Each profile provides a variety of information, such as a provider’s title, degrees, years of experience, fee, issues treated, and treatment orientations/modalities. Providers may also display a brief statement (a few paragraphs) describing their approach, goals, experience, and any other information they want to share about themselves and their practice.

People using the site may read through many profiles before .

Given that the majority of people using this site probably have little familiarity with the particulars of mental health care, particularly issues such as treatment orientations.

A few classes of questions.

The first class of questions concerns what information the average user uses when determining which therapists to contact. The average user most likely filters search results by a number of concrete criteria that they are familiar with: how much they want to pay, whether they need the provider to accept a particular insurance, provider location, and the issue(s) they want their provider to have experience treating. Once users have limited results based on the criteria they require, they are most likely left making a decision based on the remaining available information:

* Provider’s personal statement
* Treatment orientations / approaches
* Years of experience

What information can be gleaned from these remaining factors? Where does the greatest variance lie? Which of these aspects provides the most useful information? Is there a way to simplify this process for users who don’t know what to look for? Another way to think about this problem is to assume that people generally don’t know how to judge and evaluate ‘fit’ of therapists based on technical information like treatment orientation, and are left choosing a therapists based on how their profile makes them feel and aligns with hwo they think their problem needs to be fixed.

Another approach to this issue is user-blind. Ignoring how people probably most likely use the site, another class of questions focuses on describing

1. Do some descriptives to see the big words that are used in peoples profiles
2. Conduct dimension reduction techniques to see what major dimensiosn profiles vary on (e.g., positivity/negativity in sentiment, importance of insight, focus on non-judgmentalism/safety/acceptance/warmth [Rogerian themes], )
   1. Could do this with atheoretically – like Mark’s wine analysis. Simply take the most common words, provide counts of them, and then look for clusters based on words counts of each of these
   2. Or could do it theoretically – create a few variables based on some theory I have (insight variable, Rogerian variable, skills/tools/ability, time/past/history, work, fix, difficulty/challenge/stress)
3. Network analysis on treatment orientations
   1. This would be cool, but requires significant cleaning of data