# An Introduction to Psychology Today Therapist Data

In this inaugural blog post, I want to introduce the motivation and data for a data science project I've been working on.

For the last 6 years, I have been working on a PhD in Harvard's [Clinical Science](https://psychology.fas.harvard.edu/clinical-psychology) program (part of the Psychology department). As part of my training, I have become familiar with the research base (or lack thereof) for various psychological therapies, as well as with the more general nuances and particulars in the field of mental health treatment.

Trying to put my expertise to good use, I have often provided guidance to friends and relatives seeking qualified therapists. And on many occasions I have done the leg work of searching for and evaluating potentional therapists.

Much of this time is spent on Psychology Today's (PT) [Find a Therapist](https://therapists.psychologytoday.com/rms/?tr=Hdr\_SubBrand) directory, the most extensive online public directory of mental-healthcare providers in the United States. If you were to search Google for a [therapist in your city](http://google.com/search?q=Boston+Ma+therapists), PT's therapist directory would most likely be the first (non-Ad) search result. Unless you live in an very rural area, there are likely to be 100s of therapists within 20 miles of you that have a profile on PT. As such, it is probably the most widely used method of locating potential therapists ([aside from asking friends and family](https://www.psychologytoday.com/blog/freudian-sip/201102/how-find-the-best-therapist-you)).

For a relatively small monthly fee, mental health providers (typically in private practice) can take advantage of PT's online visibility and reach by creating a profile in their directory. Profiles are highly structured. Providers can include their title (e.g., psychologist, counselor), degree(s), years of experience, fee, issues treated, treatment orientations/modalities, and brief open-ended description of their approach, goals, and experience. When searching for a therapist, potential consumers (or 'patients', but I'll stick with 'consumers' because much of the therapist-finding process is conducted in the mindset of a consumer) can use search filters to display only those profiles that meet certain criteria.

In using PT's directory, a number of things hit me:

1. Finding a therapist is HARD. Possibly for some of the following reasons.

2. Within a particular category (e.g., providers who treat relationship difficulties), lots of providers sound the same. Just for kicks, here are a few quotes from different therapists on a single search results page:

- I provide a safe, compassionate environment that is supportive and free from judgment.

- I provide a safe, non-judgmental environment to help you gain a deeper understanding of yourself and your life experiences.

- I provide a supportive, safe, nonjudgmental space to share your feelings and address aspects of yourself and your life that you'd like to change or enhance.

- I believe that most of us can greatly benefit from a safe, trusting and collaborative therapeutic relationship.

3. There are a lot of providers peddling pseudo-scientific, or downright non-scientific, treatments. (Snarky aside: 'Eclectic therapy' is just another way of saying 'I do whatever I want. I may or may not follow established research or best practices'). Can you imagine a physician saying this and getting away with it?

4. The mental health field is highly unregulated. As long as you don't use one of a few protected titles (e.g., psychologist, licensed [anything], psychiatrist), you can call yourself pretty much anything else, and can provide almost any psycho-social service/treatment you want. And those consumers who are not trained in the mental health field are none the wiser.

On the heels of these insights, I thought that a more thorough investigation of PT profiles would be a fascinating way to answer a variety of questions about mental health treatment providers and the field more generally. Because it is a national database, PT profiles can also be compared across regions.

For the remainder of this post, I will describe my data collection process and present some general descriptive information about the profiles, along with some commentary/analysis. In future posts I will use the data to address more pointed questions.

A few classes of questions.

-----------

## Data collection methods

All the code and data for this project can be found in my [github repository](https://github.com/stevenfelix/PsychologyToday). All code is written in Python 2.7.

### Sampling

From each of the 50 US states, 200 profiles were semi-randomly selected. To do this, the following process was followed for each provider: First, a random zip code was selected from a specific state (e.g., Kansas). PT was then queried using this zip code. Because PT appears to randomly shuffle its search results across identical queries (i.e., if you search 02139 twice, the order of therapists will be different), the first profile in the results was selected. This process was repeated >=200 times for each state (enough times that 200 different profiles were selected).

### Profile scraping

From each selected profile, I scraped the following information (if provided):

- provider name

- title (e.g., Licensed Clinical Social Worker)

- degree(s)

- years of experience (if a range provided, I rounded down)

- school

- year graduated

- licence (# and issuing state)

- fee (if range provided, average taken)

- accepts insurance (yes/no only)

- city, state, zip

- list of specialities

- list of issues treated

- list of "mental health" problems treated (this turns out to be useless)

- treatment orientation

- treatment modality

## What are the most common provider specialties, treatment orientations, and issues? Does it vary by region?

[plot]

--------------------

### Limitations of these data

- Bias: Because this is an opt-in, for-fee service, there is certainly a selection bias in any sample of profiles taken from PT. Profiles are typically created by providers who have a private practice. This means that providers who work out of hospitals or other large treatment centers/organization (e.g., the VA, Kaiser, community clinics) are less likely to have a PT profile. It is also likely that these providers may have a different set of characteristics than those in private practice (perhaps shorter average duration of treatment, more likely to take insurance, more likely to have certain types of degrees and training, more likely to treat more severe forms of psychopathology). As such, PT profile data cannot be used to understand the overall mental health field in the US. It is best thought of as an overview of the private-practice field (and even this is a bit biased, since there are certain clusters of private-practictioners who are less likely to have a PT profile).

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