ChiPy Mentorship Blog

**How I got started**

On September 6th, 2017, I stepped into a small auditorium at 12PM, and I began to set up. The auditorium was newly furnished, with sterile, white chairs, and gray cushions. Most of the seating was in place already; I just had to move a few pieces and make room for snacks near the floor-to-ceiling glass windows at the back.

I had expected 25 people to attend, but as the morning wore on, I realized that I may need more seating. Fortunately, though it was last minute, the auditorium could be expanded to seat 50, instead of the customary 25. I registered my reservation, and started to expand the room, placing chairs, and checking my A/V setup.

Then, finally, at 1PM, I began to present the work that I had been performing over the last six years. At 1PM, I started defending my dissertation.

In many ways, this marked the end of one journey, and the beginning of another. After successfully finishing the dissertation, my graduate school experience was winding down: I had some revisions to work on and projects to wrap up, but the dissertation defense was complete. This left room for exploration, and at the end of the dissertation, I had one big question to explore “what’s next?”

At the time, my answer to this question was simply “Data”.

Over the next several months, I tried to answer that question in more detail. I searched in many different places for an answer – I asked myself that question, I read online forums, I reflected on my experiences and skills, and, probably most important for me, I chatted with friends.

For me, this part of the journey was slow. There was no “aha” moment. No flash of brilliance where I realized what my future/career had in store for me. Instead, it was more like listening to a song over and over. In the beginning, it’s just exciting and enjoyable. Then, at some point, you’ve heard it enough times to know it. And sometimes, after listening enough times, you can pick up on new things in the song – maybe it’s an instrument that you hadn’t focused on, or a beat that you missed. Something hiding in plain sight.

By continuing to explore the idea of data analysis, I slowly understood that I wanted to focus on developing professional data analysis skills, and that I wanted to work on my coding. Conversation after conversation, article after article, I noticed that quantitative analysis was a theme I enjoyed building on, and a skill that I felt I could be really good at.

After realizing that I wanted to develop as a quantitative professional, the next step, forming a path, was much easier. At this point, I focused on improving my technical skills using the tools of the trade: Python and R. And that’s when my roommate forwarded an email that led me to ChiPy mentorship program!

**The ChiPy mentorship program.**

I’ve always had a strong belief in the power of mentorship and apprenticeship. For me, learning in a small team, and teaching in a small team is a powerful way to improve, and a powerful way to give back to the community.

The ChiPy program takes this concept and applies it to the world of Python coding. The experience is meant to provide a forum for two Pythonistas to share an experience and learn from each other (probably more mentee learning from mentor, but us mentees can try to help our mentors learn too!). The goals of the mentorship are largely set by the mentor and mentee, however, there is a formal requirement for the program: at the end of several months, the mentee must develop a Python-based project to showcase the skills they’ve learned during the program.

**Goals for the mentorship program.**

After chatting with my mentor, my project will be driven by a few specific goals.

1. Main goal: Build solid data science skills:
   1. Manipulating data frames
   2. Implementing statistical models and machine learning models
   3. Organizing a data science project
2. Secondary goal: Improve interactive visualization skills
   1. D3 or Bokeh

**My proposed project: Analysis of Chicago Public Schools (CPS) data**

To achieve the project goals, I proposed to analyze data from Chicago Public School system.

Every year, CPS publishes data on the number of teachers employed at each school, as well as teacher salaries. CPS also publishes data on school performance, whether its graduation rates, standardized test performance, or even attendance rates.

I want to analyze this dataset and build a data science project around the question: “What impacts CPS school performance?”

To start, I plan on analyzing school funding and staffing to understand what effect these two variables have on school performance. Following that, I’d like to integrate other sources of information to see if they are linked with school performance (geographic information about the city of Chicago, such as crime rates, incomes, employment etc).