

## Scrum 1

- What have I worked on?

Last week, our client let us collect the data about NH and Rhode Island. We think it is difficult to crawl data from the website for two reasons: one is the data is pdf, the other is there is no clear definition for the headnote and text in the data. We talked with our client again and decided to crawl the data in 2018. Our work this week is crawling the pdf data about NH and Rhode Island in 2018. We used pypdf2 package. Not every case is described in the same format. We have to classify them according to different formats. The extracted code of each format is different.

- What will I be working on next?

The next goal is to identify which part of the data is headnote and which part is text. We try to distinguish between headnote and text. We will use our data to discuss with our client again.

- Have I run into any issues? Do I need help?

The reason why we have to distinguish the headnote and text is that our analysis focuses on the criminal cases. In M.A's cases, via keywords we can select the criminal cases. But we read some of the data about NH and Rhode Island, and found it's not clear whether the case is a criminal case or not.

- Have I talked to the client recently? When are we meeting with them next?

We talked to the client on 02/28/2020

We will have a meeting after spring week