**Limitations of current analysis / opportunities for future work**

**Dataset limitations**

1. < 200 projects
2. Only npm
3. Not excluding forked projects
4. All in the same period - can only talk about correlation not causation

**Refining how we are calculating features - Reliability of the library/api**

1. The python libraries of readability and sentiment is general to all texts. Readme files have specific formats that affect the reliability of the scores
   1. Unable to delete the urls
   2. Code embedded in the readme
   3. Other languages (ex. Chinese, Russian)
   4. Pictures and badges (show up as urls)

**Future improvement of the measurement**

1. Clean up the readme
2. Maybe build a sentiment/readability interpretation model specifically for readme

**Additional features to consider**

1. Presence of women in previous time periods (if there is already a woman there, will that bring more women)
2. Sentiments of comments
3. **Core contributors of projects**
4. **Ratio of approved male pull requests to approved female pull requests**