Link to forked repository

https://github.com/SnellJ2/cs472project/tree/mine_repository/repo_mining

Testing: In order to collect the data from the /roobeer/ project on github I wrote a few scripts to collect, analyze and visualize data. Below is a quick summary of the scripts used.

Jonas-Snell_CollectFiles.py - Uses github api to collect commit data, which is parsed to get only the source files that are also programming files.

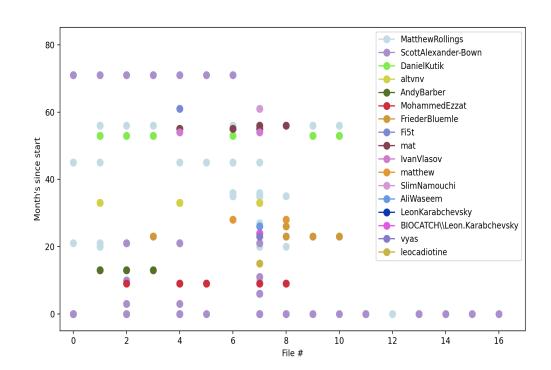
ionas authorsFileTouches.pv - Cleans up the data from CollectFiles into a new .csv file.

MatthewRollings,2021-05-24T07:35:11Z,rootbeerlib/src/main/java/com/scottyab/rootbeer/RootBeerNative.java ScottAlexander-Bown,2021-05-08T14:18:37Z,rootbeerlib/src/main/java/com/scottyab/rootbeer/RootBeerNative.java

<u>jonas_scatterplot.py</u> - Takes the data from authorsFileTouches and plots it visually. Now we can interpret the data based on authors.

Data Interpretation:

From this data we can see visually who committed to one of the 17 source files and what week they committed to it. For example "matthew" committed to file 3 around the 25th week and files 6 and 8 around the 30th week. From that we can extract that matthew committed 3 changes total.



On the other hand **Matthew Rollings** committed around **20 changes**.

Summary:

Though it's a bit cluttered due to the amount of contributors, it's possible to match color to name and get a rough idea who had a high and low amount of commits. Of course this can't be taken as the full story on who is productive to the project, but it is a good place to start.