GitHub Wrapped

Logan O'Neal, Eric Vaughan, Caleb Fisher, Jacob King, Hayden Curl

Project Questions

Developer Spirit Animal

What type of developer am I?

Repo Licenses

How have the most common licenses on GitHub changed over time? What licenses do I tend to use?

Most Used Languages

How has the most used language on GitHub changed over time? How does my personal development journey compare?

GitHub Network Map

What would a map of all the individuals a user has worked with on GitHub look like?

Commit Messages

What are the most common contributions on GitHub? How do my personal contributions compare?

Developer Spirit Animal

- Uses metrics such as repo names, favorite languages, and recent contribution activity to form the GPT-4 prompt
- Relevant GitHub endpoints:
  - GET /user/repos
  - GET /repos/{owner}/{repo}/languages
  - GET /events

Developer Spirit Animal

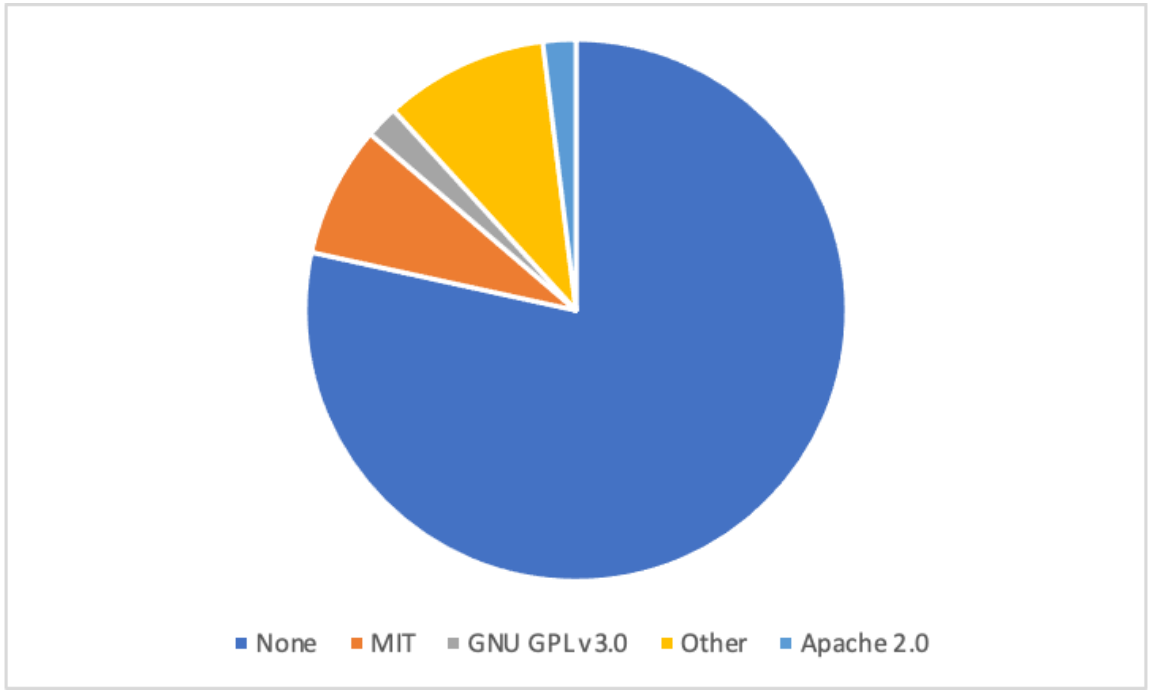
- Uses DALL-E 3 to generate an image based on developer spirit animal
- Personalizes the image based on GitHub profile information such as bio and profile README

Licenses

- Performed an analysis of all our group member's public repositories
- Accomplished by parsing data retrieved from GitHub API
- License data is organized into a structure that contains:
    - License Key
    - License Name
    - License URL
- License Type is determined by the License Name

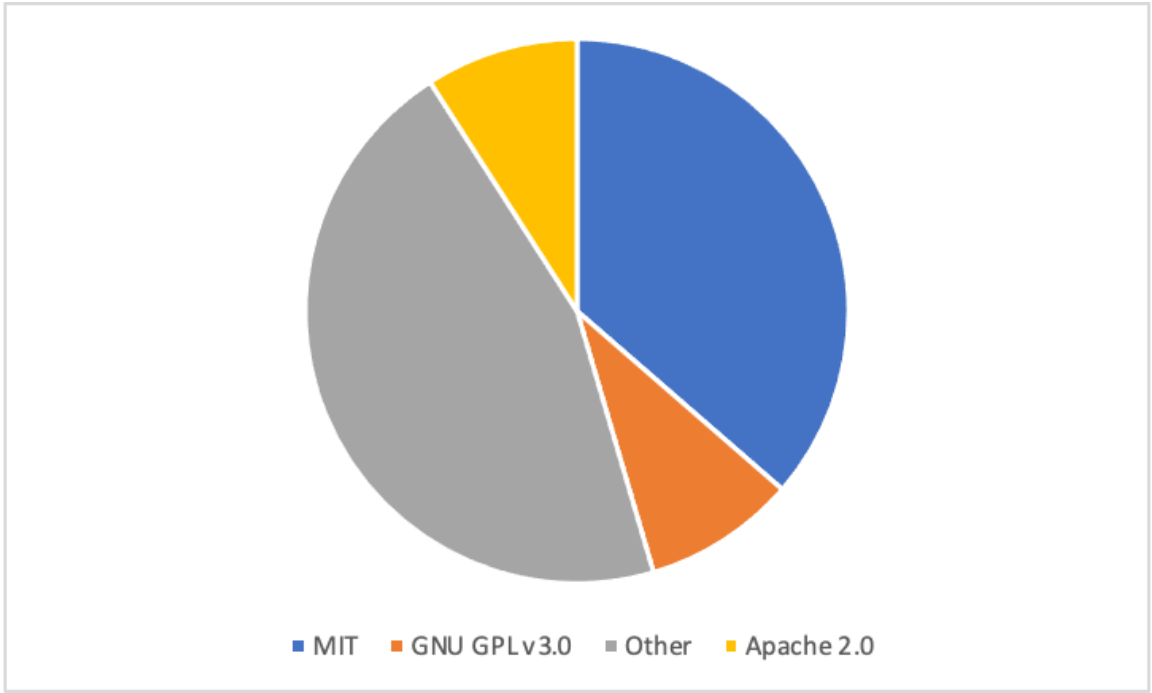
Licenses

- Out of 51 repositories analyzed:
- None (78.43%)
- MIT (7.84%)
- GNU GPL v3.0 (1.96%)
- Other (9.80%)
- Apache 2.0 (1.96%)

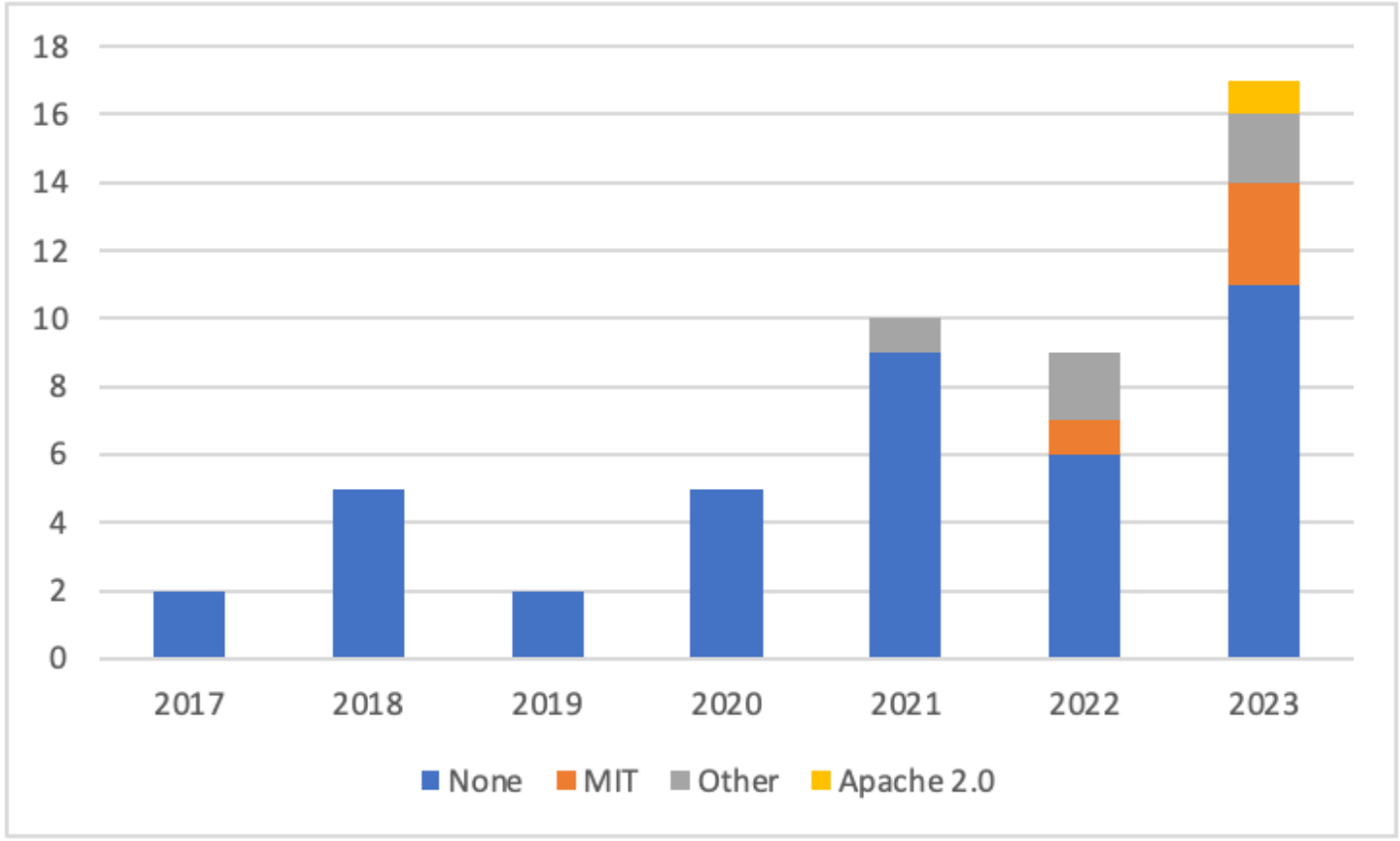


Licenses

- Out of the 11 Licensed Repositories:
- MIT (36.36%)
- GNU GPL v3.0 (9.09%)
- Other (45.45%)
- Apache 2.0 (9.09%)



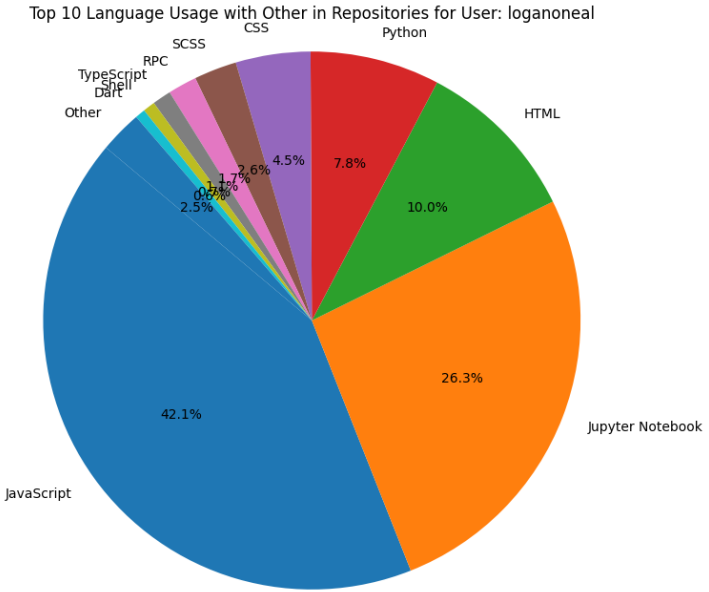
Licenses Over Time

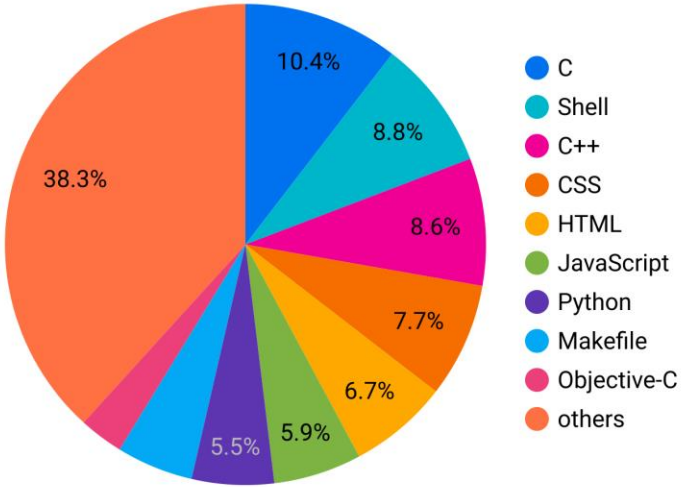


Language Popularity

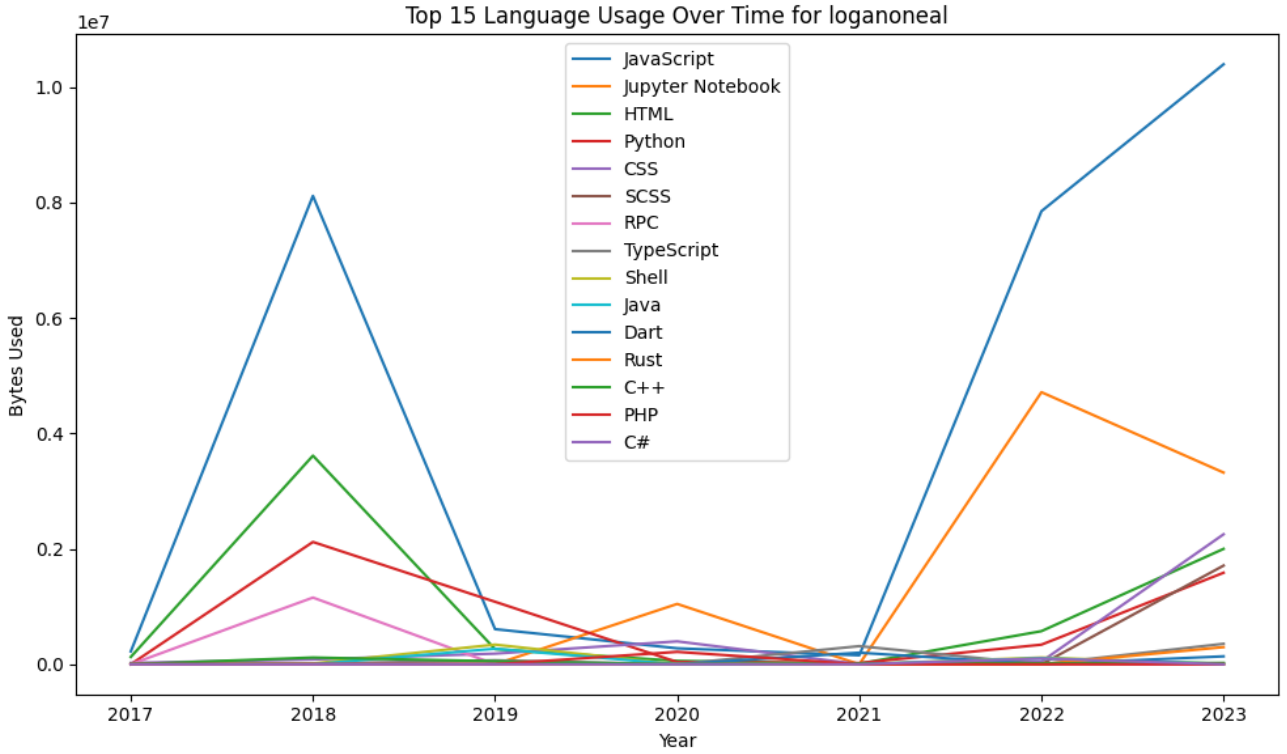
- Contrast the language distribution in a user's repositories with the global Github trend.
- Analyze a user's commit history across different languages in comparison to the entirety of Github.
- Utilized Bigquery to structure data from across Github and Google Data Platform for modeling

User Language Distribution Vs Github Overall

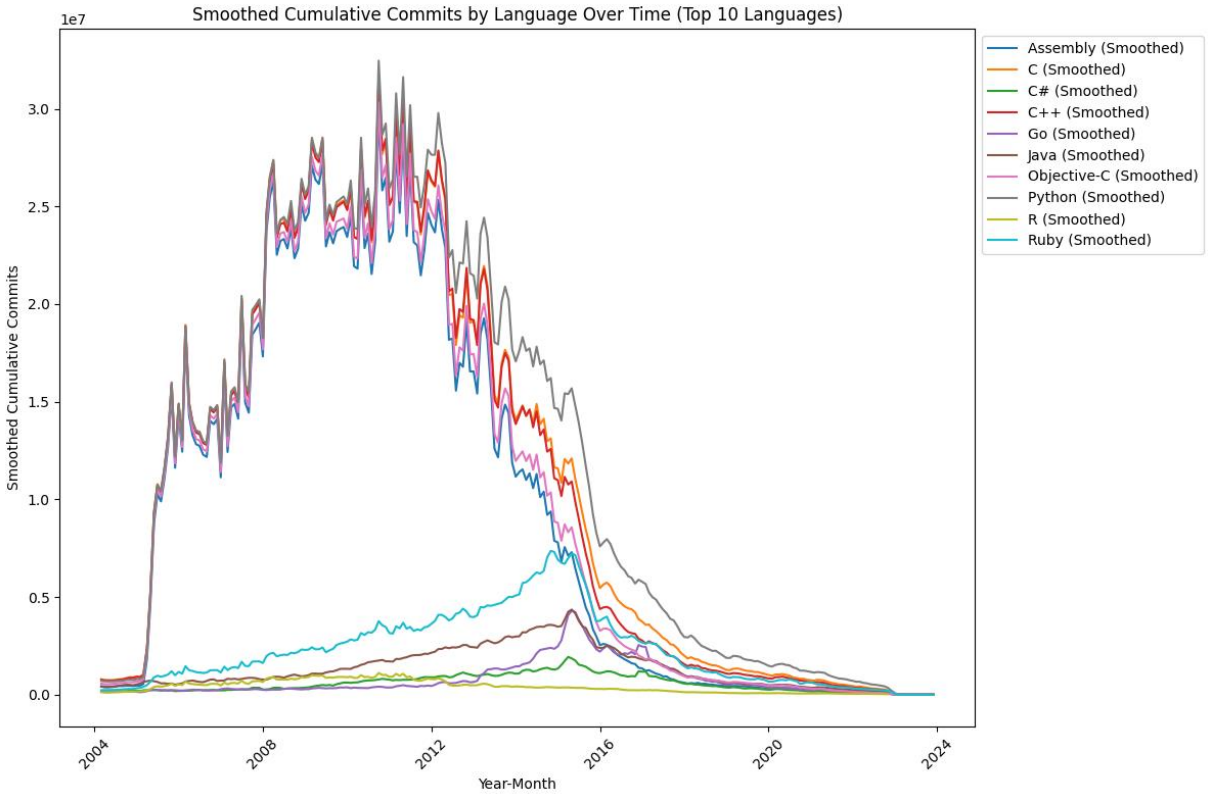




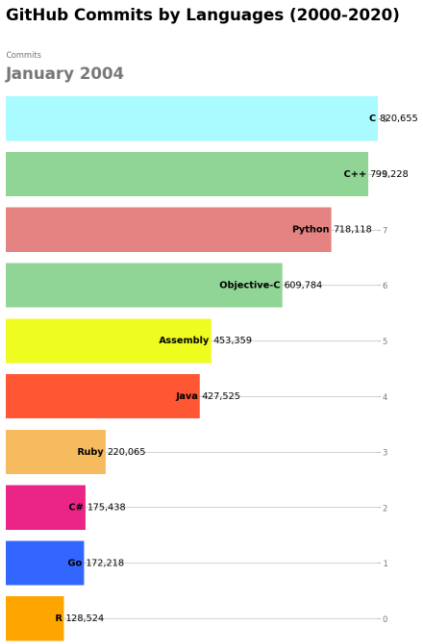
User Language Trend

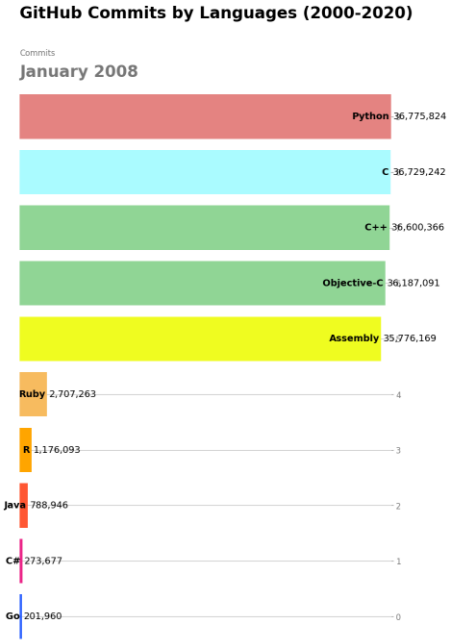


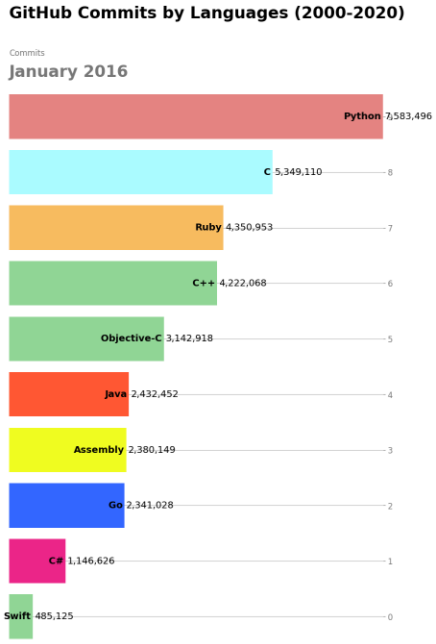
Github Language Trends

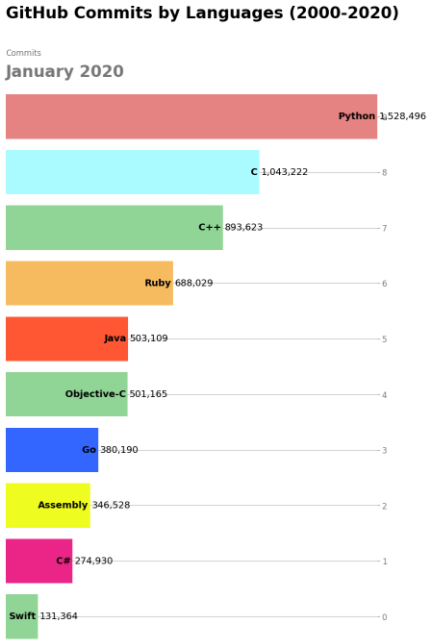


Overall Language Plot



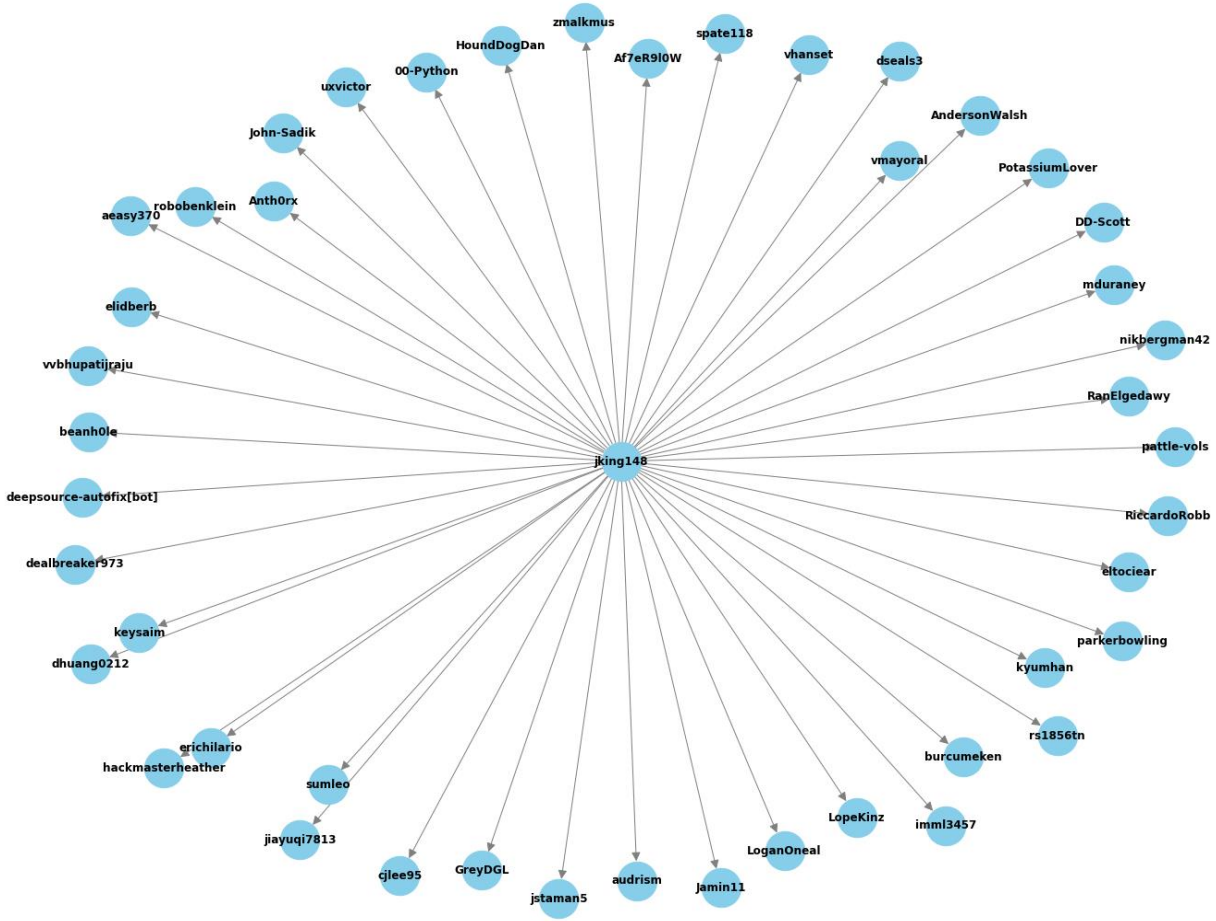






User Network Map

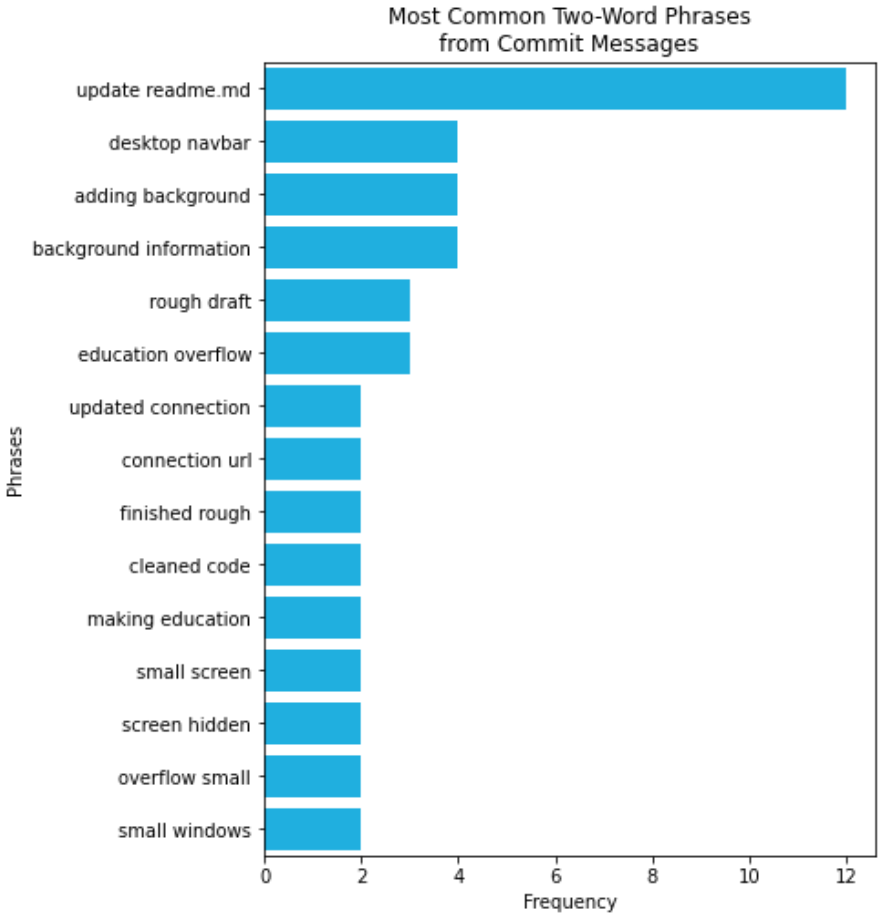
- A user's network map is comprised by using the Github API end point api.github.com/repos/{user}/{repo_name}/contributors.
- This will grab all contributors that have worked on the same repos as the user.
- To do this you will need first parse all repos the user has contributed to and then for each repo find all other contributors.
- By doing this you will have a list of all the users that your user has contributed with.


zmalkmus
HoundDogDan
Af7eR9l0W
spate118
vhanset
dseals3
00-Python
uxvictor
AndersonWalsh
John-Sadik
vmayoral
PotassiumLover
Anth0rx
robobenklein
aeasy370
DD-Scott
mduraney
elidberb
nikbergman42
vvbhupatijraju
RanElgedawy
beanh0le
pattle-vols
jking148
deepsource-autofix[bot]
RiccardoRobb
dealbreaker973
eltociear
keysaim
parkerbowling
dhuang0212
kyumhan
rs1856tn
erichilario
burcumeken
hackmasterheather
sumleo
LopeKinz
imml3457
jiayuqi7813
LoganOneal
cjlee95
GreyDGL
audrism
Jamin11
jstaman5

Commit Message Analysis

- Overall goal is to try to summarize the most common topics being worked on and actions being taken.
- Analyzed a single user's commit messages across all their public repos.
- Analyzed commit messages from several users across GitHub and compared that to the single user's commit messages.
- Utilized BigQuery to collect and structure the data for analyzing the commit messages from several users.

Single User

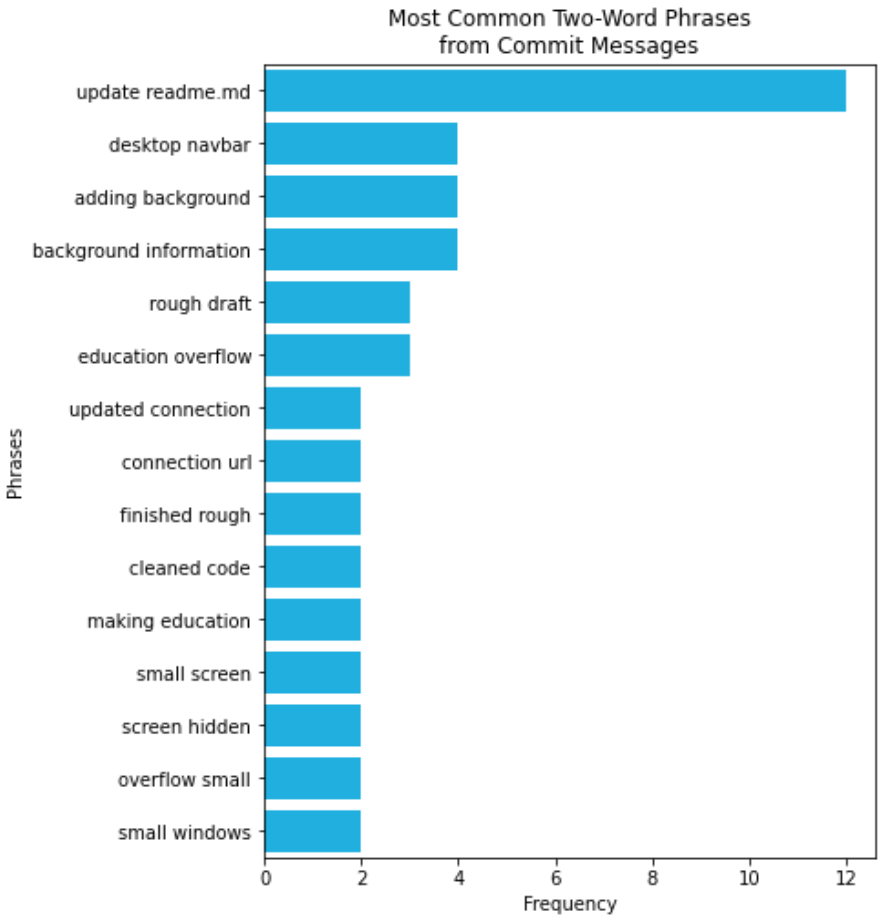


- Looked at consecutive two-word phrases.
- Example:

"Single User Commit Analysis"

"Single User"
"User Commit"
"Commit Analysis"

Single User



Several Users