The goal of this activity is to assess your approach, consideration and processing of some fairly simple data. Do not take more than 4 hours, and feel free to leave portions open-ended and/or undocumented. We will review and you will have an opportunity to clarify all parts.

Lastly, don’t be too critical of code style or performance. Think of this as a quick proof-of-concept and not a production ready process!

Part 1:

The [GHTorrent](https://ghtorrent.org/) project has been scraping Github commit data for a while now. Downloads of their data from MySQL or MongoDB is available here - <https://ghtorrent.org/downloads.html>

Grab the most recent database dump from the link above and run some basic statistics on the data, showing the following:

* Top 5 repos by commit activity (# of commits)
* Top 5 most frequent words used in the commit messages
* Both of the above sliced by year, i.e. “Top 5 repos for 2019”

Hint: Some code to get you started on this data is here - <https://github.com/src-d/datasets/tree/master/CommitMessages>, and some more to get you thinking about approach here - <https://gist.github.com/paulmillr/2657075>

Part 2:

Continuing on what you played with in Part 1, but now for more up-to-date statistics, you will pull in some data from the Github API [https://docs.github.com/en/rest/reference/repos#statistics](https://docs.github.com/en/rest/reference/repos" \l "statistics)

Using the repository for Vue.js(<https://github.com/vuejs/vue>), get the following:

* All of 2020’s commit activity ([https://docs.github.com/en/rest/reference/repos#get-the-last-year-of-commit-activity](https://docs.github.com/en/rest/reference/repos" \l "get-the-last-year-of-commit-activity))
* Most active contributor in terms of commits ([https://docs.github.com/en/rest/reference/repos#get-all-contributor-commit-activity](https://docs.github.com/en/rest/reference/repos" \l "get-all-contributor-commit-activity)

For both parts, it is up to the individual to determine how they’d like to display the data. It should be presented in human-readable form but does not need to be “pretty”.

Please package all your code and outputs into a zip file once complete. You may also commit it to a GitHub repo if that’s more convenient.