Proposal notes

When reviewing CVE List vulnerabilities, I looked at JavaScript and Java vulnerablities. Many of the JavaScript vulnerabilities pertain to issues with various browsers. I also explored Java issues for comparison and this was less the case.

Hypotheses

H_{1a}: GitHub projects will share dependencies and some of those dependencies will be out of date, leading to a security compromise

Findings

Out of date dependencies almost always relate to abandoned projects, even within the Apache foundation, rather than currently maintained projects with security issues. Many of the out-of-date dependencies pertained to personal projects with one developer and a small number of commits or large projects which are no longer maintained.

Initial analysis arrives at the obvious, using projects that do not have any recent commits are likely to have security vulnerabilities.

There is occasionally a pattern where a CV item will reference previous CV items. We can use this and other NLP to pull out general industry trends in terms of security vulnerabilities. This is very much subject to making assumptions about reporting consistency and incentives.

Further research

1. Validate the abandoned projects security risk finding on a larger scale
2. See if certain projects or classes of projects have reoccuring types of vulnerabilities popping up.

Issue (1) could be a real concern because it can be very hard for a developer to figure out if a project is currently being maintained on github when they're using several dependencies to try and get things done.

Issue (2) may give us a more general sense of risk pertaining to certain technologies.

Solving Issue (2) may be more feasible in the short-term because it requires collecting less data, although issue (1) may be of more interest.