

HW5 - Intro Quantitative Analysis

In this assignment you will practice basic quantitative data analysis methods.

You will download and mine the 2017 Open Source Survey results from Zenodo:
<https://zenodo.org/record/806811> (https://zenodo.org/record/806811#.W7PD0y_Mz1L)

The survey, led by GitHub, "collected responses from 5,500 randomly sampled respondents sourced from over 3,800 open source repositories on GitHub.com, and over 500 responses from a non-random sample of communities that work on other platforms. The results are an open data set about the attitudes, experiences, and backgrounds of those who use, build, and maintain open source software." See <http://opensourcesurvey.org/2017/>

You can find more information on the survey design and instrument [here](https://github.com/github/open-source-survey) (<https://github.com/github/open-source-survey>). See also R. Stuart Geiger's [Summary Analysis of the 2017 GitHub Open Source Survey](https://arxiv.org/abs/1706.02777) (<https://arxiv.org/abs/1706.02777>)" presenting frequency counts, proportions, and frequency or proportion bar plots for every question asked in the survey."

Formulate two research question about participating in open source development, motivate them in 1-2 paragraphs with a few citations to relevant literature, and answer them using a quantitative analysis of data. Go beyond the basic frequency counts from R. Stuart Geiger's paper and focus your research questions on correlations, regressions, or descriptive breakouts between subgroups. You could use any language you are familiar with.

Submit a short 1-2page report describing the research questions and motivation, data cleaning and statistical analysis steps, and (briefly) results.

Grading will be based primarily on: 1) quality of research questions asked; 2) appropriateness of the statistics used for the research questions; 3) execution.