Questions to ask tomorrow:

- Manuel verification is 384 pages – is there any way to get a smaller sample that is representative. I don't mind doing the work, but 384 might take me a week. As such, my strategy would be to write the first draft of everything but the results / conclusion and abstract.
- With the results – I am not sure what else I could do.
  - There is an enormous prevalence of apt-get, it overshadows the rest really. Wondering if I could leave it out.
  - I don't have evolutionary data, but I have readings from the datasets separately and was wondering if I could maybe do something with that? Like have 2 readings with show some sort of evolution from Github as one dataset is from 2007 to 2019 and the other is from 2022. But feels like a period of files vs a snapshot … not really sure if it is that good evolutionary. Additionally, I would only have two datapoints.
  - The results are not super good for the temporary files, except for one. Same goes for the bloaters, there is actually not a lot … was wondering what I should do with this. I thought the churn smell would be much more prevalent, but hey, at least I'm detecting it.
- How did Camilo scrape the results from StackOverflow, is there some kind of criteria that he used?
- What do you think of the title?
- Going forward, how does the feedback work?