Processing Log Data

Below are the two research questions related to user interaction:

RQ3 What kinds of search interface features are being used during the initial search when using x versus the baseline search interface?
RQ4. What kinds of search interface features are being used during the resumed search, when using x versus the baseline search interface?

Initially, I had the raw JSON data of user logs which contained everything they did for both sessions in one object. I pre-processed the data and programmatically cleared all system-level logs (e.g. stop/start events, network events, etc), and all the practise session logs. Later, for each user, I had two lists of interaction logs for their two sessions containing only their activities for the task.

From that data, I programmatically counted frequencies of each interaction for every user for both of these sessions and generated a data frame. Later I exported these two data-frames in CSV. The CSV files also have columns for each interaction's parent classification category from Wilson's model.

There is a predefined list of possible interaction codes along with Wilson's classification category. So, even if we update any of the categories (e.g. changing "input" to "information" for every query input), the generated CSV files will have the correct classification without any change in the code. Removing a particular interaction will be also done in the same way with no code change. Until this point, I had done everything on python.

There will be two more CSV files when I will use my code to process Baseline logs. My next goal is to load these programmatically generated CSV files in R, group them and plot them in an appropriate visualization. I am attaching the two generated CSV files with this email.