**Plan – my workflow**

**Get timestamps in milliseconds for the Pupil lab data**

**Merge raw data**

* A strategy will be to loop over each row in the manually annotated Boris data and for each timestamp find the nearest timestamp in the pupil lab data.
* A way to do this is by subtracting the annotated timestamp from all timestamp in the Pupil data and find the row with the lowest divergence; subtract all information from the other columns in this row.
* Loop over every single one.

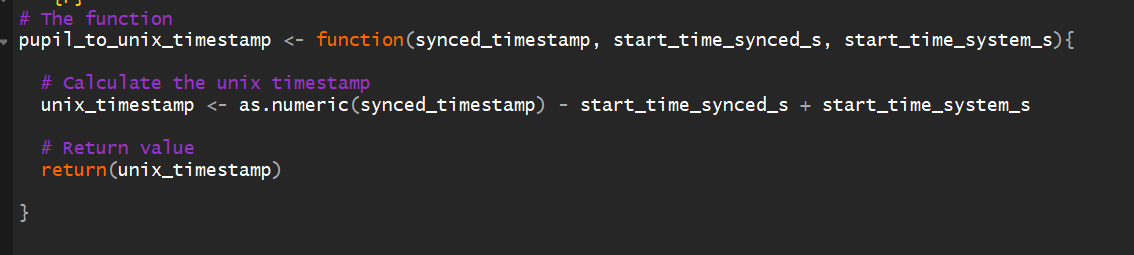
**Visualizations and potential modelling**

**Questions for meeting 22/04**

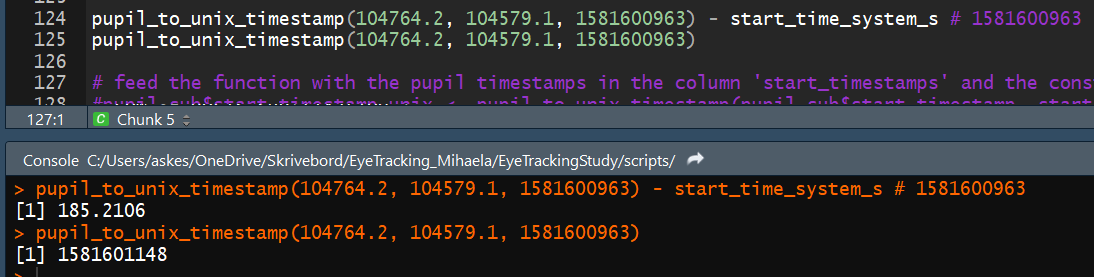
* What columns are necessary?
* The time durations for each block are often longer than the time manually annotated with fixations. How come? Did the eyetracker keep recording?
* Is it correctly understood that I should just extract one row of from the pupil data per fixation annotated in the boris data? That is what I’ve currently done.

**Questions for Mihaela.**

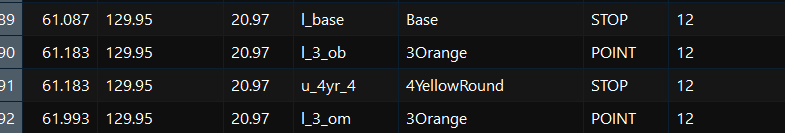
* The function to convert from pupil to UNIX is giving me a hard time. I’ve based it exactly on the one from the documentation



* It says one should add start\_time\_system\_s which is a weird insanely high number from the json file. Then the output is as well a weird insanely high number. If we just don’t, it makes way more sense – and we will get something that seems like timestamps from approx. 0-185.21 (in this case)



There are some rows in boris with identical timestamp but different information meaning that they will merge with the same row from pupil data



* I don’t understand how one can be looking at 3Orange and 4YellowRound at the same time? Is that on purpose?

There are also a few cases where the timestamp difference between to rows in boris is so small, that they will be assigned the same fixation from the pupil data. I don’t really know what to do about that.

**Meeting 05/05**

* Behaviour column contains u l and ACC BC which is all behaviour and not related to the object looked. Only the L. Thus create code to extract only rows with an l. probably solved with some regular expressions
* Look into the documentation again with calculating timestamps (links in email the 5th of may)
* What start timestamp should I extract? Might not be a big difference but could be relevant.

**To do**

* Start recording stamp – from the json file or the csv file? I have made both functions
* We still have the issue that some of the pupil data is attached more than once. I Could do something like: if duration + timestamp > the next timestamp, then drop column – because it means that it did not belong to that one.
* I have made a function to select only behavior “l”
* Create a good readme file
* Create some plots and visualisations similar to what has been done in the little preprint