**Driving Data Cleaning Standard Operating Procedure**

Driving files and drives were programmed by Austin Zamarripa ([czamarr2@jhmi.edu](mailto:czamarr2@jhmi.edu)). For driving-specific questions please contact him at his listed email. The driving script needed to clean drives was created by Matt D. Novak and Austin Zamarripa – for questions, please contact ([czamarr2@jhmi.edu](mailto:czamarr2@jhmi.edu)).

Prior to using script, make sure you have the following items in place

1. Make sure you have R and R studio Downloaded to your computer
   1. Make sure you have the following packages downloaded:
      1. dplyr
      2. tidyr
      3. stringr
   2. If you don’t, the way to install will be in the script. To do so:
      1. Remove the “#” in front of #install.packages(PACKAGENAME)
      2. Hightlight “install.packages(PACKAGENAME)” and run
      3. When you have done this once, you can either 1) delete it, or 2) add the “#” back to prevent R from reading this line again
   3. Make sure you change the desktop to match your computers main file path
2. Copy the following “Driving Simulator Information” from the Flash Drive Memorex to your desktop.
3. This will give you this SOP, and the folders, and files needed for any future start up. The main files you need are in the “drive\_analyzer” folder.
   * 1. analysis\_summary
     2. block\_info
     3. data\_to\_analyze
     4. old\_data
        1. I recommend this folder for adding the old study data, and placing into the correct study identifiers (e.g., 2106 for 2106 data).

To clean driving data

1. When a person completes their study, download all of their .csv files to a USB. Do not download the plb files, you will only need the .csv files.
   1. plb are playback files and can mess up the script.
2. Make sure all files are appropriately named as “**Study\_Participant\_Session\_Timepoint\_DriveIdentifer**”. See notes below for possible issues this creates.
   1. Every issue I run into clean these data come from this mistake.
   2. “Study” is the name you call your study (e.g., 2106)
   3. “Participant” is their unique identifier (e.g., AZ13)
   4. “Session” is just the session they are on (e.g., Session 4, or simply “4”). For single sessions, or between subject designs, use 1. You will always need a value here
   5. “Timepoint” is the timepoint being assessed (e.g., BL or 1hr)
   6. “DriveIdentifer” is the actual drive you do (e.g., 4A). This tells the script how to clean your file and which specific drives. If this is wrong, it may not work and clean your file, and all the other files in the folder. **This is the most common error**!
3. Place all of the driving data in the “data\_to\_analyze” folder. The R script will pull these files and clean them. You can clean as many participants are you want.
4. When all of the files are in the correct folder, run the script. You do this by highlighting the whole script and pressing “Command + Return”.
   1. If this breaks, you will know immediately because no output file will be generated. If these files do not clean, contact Austin Zamarripa (czamarr2@jhmi.edu)
5. When this happens, a new file will be created in the analysis\_summary folder. This will be labeled as the date+Summary.csv. If you clean multiple times, it will save each version with the words “copy” behind each new time.
6. When you have run the Script, open the Drive Summary file and begin inspecting the outcomes. For the most part, nothing should ever change. You may see some NAs listed in some categories but those are okay to ignore.
   1. For ease, I recommend opening the file with excel to make the transfer to the master file easier.
   2. NAs are generated when participants did not meet the standards of that task. Most often this is seen with the amber dilemmas because a participant did not change their behaviors.
7. Copy all of the driving data, and add it to your master file as is.

Possible reasons for errors. This list will be evergrowing. Some solutions are not easy to fix.

* The files are not named correctly. Verify the individual files are labeled correctly with the format of Study\_Participant\_Session\_Timepoint\_Drive File.
* The participant performed undirected driving behaviors (e.g., turning when not prompted to). These can interfere with specific driving outputs with the R Script. I tried to make this script ignore these instances but I can only update it when I see unique behaviors that aren’t captured in the data-cleaning process. If this happens, contact Austin Zamarripa