CCSS Summer Workshops Types of Questions.

#These are the types of questions that I plan to ask for ALL sessions(Stata, R, Python). If a question is specific to only 1 software. Mention that.

#Include estimated steps and time for each question.

* Recognize all .txt/.csv files inside a folder. Import each one, subset, save, combine them all with append. Repeat this process for a similar dataset that contains demographic information. Vertical merge. Then merge the 2 together to bring over more information.
* Merging datasets horizontal. Unique id both datasets. Duplicates 1 to many. Duplicates take max of duplicate. User create id variable by combining other variables. Merge by rows when applicable.
* Log file and script files. Opening and running. Minor change to get script working. Quick replication. View log files to make sure stuff is working properly.
* Regressions. All conditions prior to running. Linear and logistic. Analyzing output.
* Ttest and other tests comparing stats by 2 groups.
* Value labels for neater output. In accordance with one above.
* Importing data using codebook and fixed formatting.
* Quick replication. Follow readme, make minor changes. Get code running.
* Simple find and replace. More and more layers. If else conditions . Mention undo function/procedures/actions
* Substring to get part of string. Use that part to extract what is desired. Splitting string by specific character.
* Installing and using packages
* Taking code from help documentation, links online and get running on personal machine
* Creating results, saving code that generate results and code to export results so can send to other researchers. This is how replicate, generates following results. Basic output tables. Export to familiar format.
* Lag variable. Find and replace, cumulative sum by group.
* Check for duplicates/outliers. Make note of them. Delete them. Missing values too(string and numeric). Hard code and auto saved results.
* Measuring change over time. Statistically and visually. Between groups too.
* Variable transformations. Re-running regressions with transformations applied.
* Subsetting on import. Target specific areas to reduce size of dataset to speed up processes.
* Text and sentiment analysis python and R only
* Repeated measures in loops obvisously. Basic loops though.
* Matrices. Mata. Stata only.
* Foundation of flow. Examine type of data. Format of data. Sample working code, apply to personal use case.
* Many of above can be combined to make a more complicated problem. Or slightly tweekd to make a similar problem but not exact.