--- Data & File Descriptions ---

A) Raw data files

Raw data includes the data measured by measurement devices/sensors from the field/lab (either csv, xlsx, or txt), or the data filled in Excel spreadsheets in the field/lab by the field/lab user to record any experimental condition or measurement that could not be recorded by a device (mostly xlsx) in "Mahdavi & Siegel (2020) AS&T" (Phase 1). Due to data privacy, only a representative sample of the raw data (meaning one from each type) is shared (if shared).

See "Table 1.pdf", "Data Pipeline and Code Configuration.pdf", and "list_of_files.xlsx" files to understand more about raw data.

A sample of raw data is in the "Raw Data Files" folder.

B) Code files

Codes include all ".py" or ".ipynb" that read raw data or previously processed data to generate processed data, figures, or statistical analyses. There are four different code types used in this project:

- **Processing codes:** which process from raw data (stored in the "Processing Codes" folder).
- **Statistical preparation codes:** which processes from either raw or previously processed data to make a dataframe compatible for a statistical test (if a previously processed file is not) (stored in the "Stats & Calculation Codes" folder).
- **Calculation codes:** which present numeric or dataframes from the processed data (stored in the "Stats & Calculation Codes" folder).
- **Visualization codes:** which aim to plot figures. This may come with additional processing of data to make the dataframe compatible for plotting (stored in the "Plotting Codes" folder).

See "Data Pipeline and Code Configuration.pdf", and "list_of_files.xlsx" files to see the list of code blocks. (One generic code titled "notion_correction" is not shown in "Data Pipeline and Code Configuration.pdf" but it is called by many code blocks).

C) Processed data files

Processed data includes the data generated after processing in data pipelines in "Mahdavi & Siegel (2020) AS&T". See "Data Pipeline and Code Configuration.pdf", and "list_of_files.xlsx" files to check the processed data lists and blocks. "Data Pipeline & Code Configurations.jpg" illustrates the entire data pipeline processing that generates processed data using codes from the raw data or previously processed data.

A separate code block ("df_summary.ipynb") also presents a summary of some processed files. For data privacy purposes, not all the processed dataframes have been presented. Processed files are not shared as spreadsheet files either.

D) Plots

Plots include all the figures presented in "Mahdavi & Siegel (2020) AS&T" (Phase 1) from the processed data. See "Data Pipeline and Code Configuration.pdf", and "list_of_files.xlsx" for more information. The

plots aren't presented in separate files (e.g., jpg) but are illustrated in the same code file generating them (in Jupyter). See "Plotting Codes" for more information.

E) Other files

Any other file (mostly guidelines or descriptions) not classified above. A full list is available in "list_of_codes.xlsx" (sheet "other_summary_files") provided in the repository.