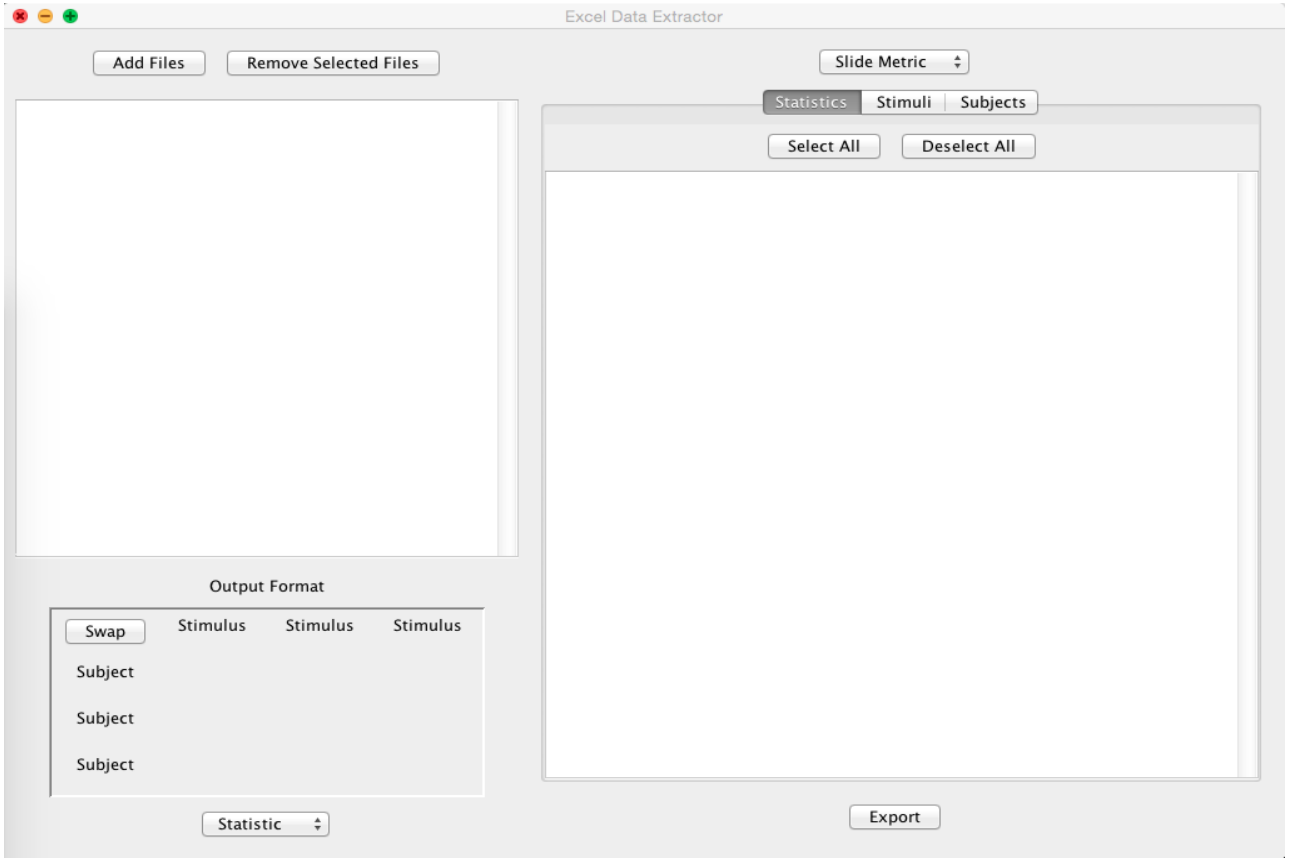
Excel Data Extractor

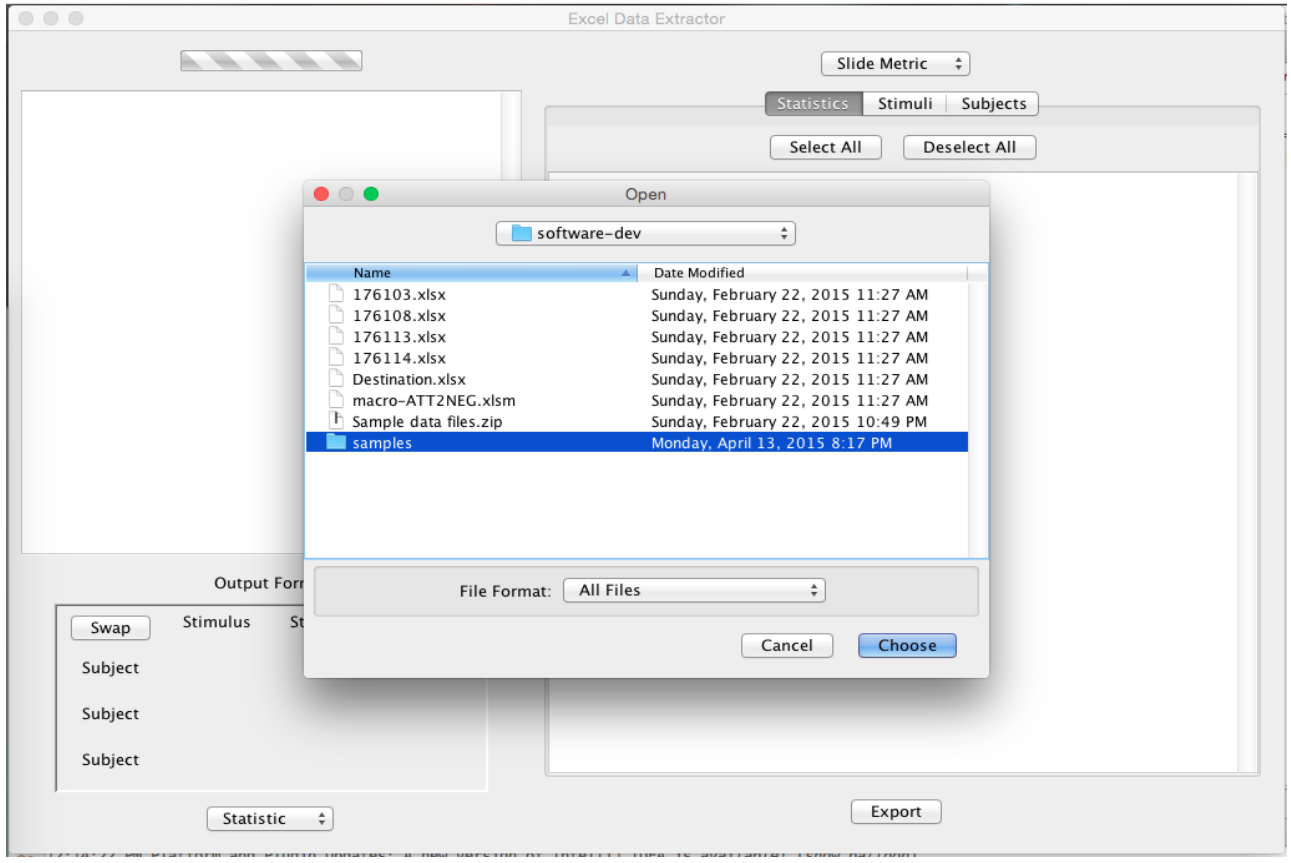
Welcome to Excel Data Extractor! This guide will demonstrate the capabilities of this software by walking through a sample use case. The guide is broken into four main parts - Upload, Selection, Export, and Data Assumptions. To begin, double click the application file (Excel-Data-Extractor.jar) to bring up the following window:

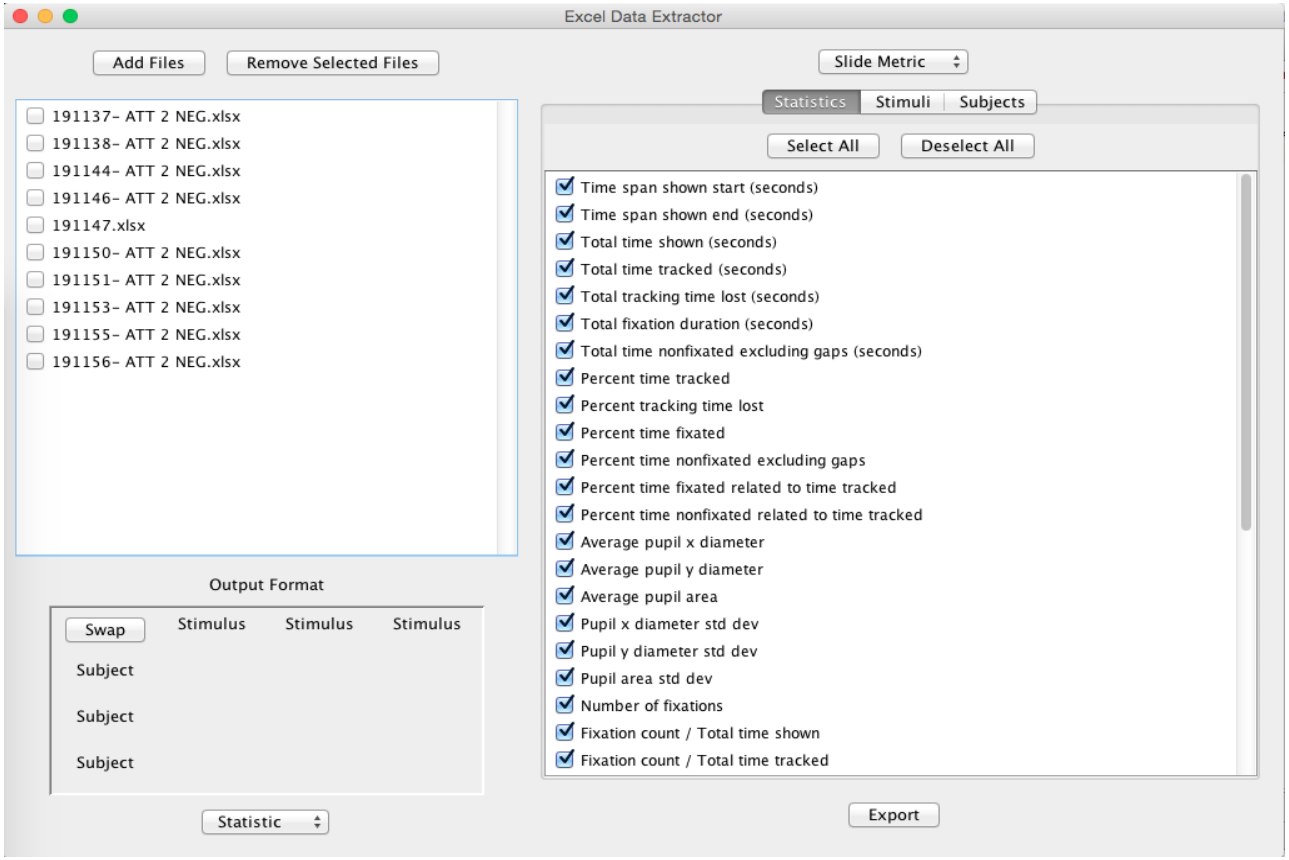


Upload

To upload files for analysis, click the “add files” button to bring up the file chooser:

Excel Data Extractor supports the upload of single files, multiple files, or directories. Only .xlsx files will be included; all other file formats will be ignored. Additionally, the excel files must be of a specific format to ensure the success of the application. See the Data Assumptions section for more information.
Navigate to the file or set of files you would like to upload and click “choose.” A loading bar will appear in the window, and will disappear when file upload is complete and all extracted data has been added to the window.





Selection

Once files have been uploaded, Excel Data Extractor allows for robust configuration of its output file to match the user's needs. At any point, a user can add additional files for analysis, or remove accidental inclusions by selecting the file and clicking "remove selected files."

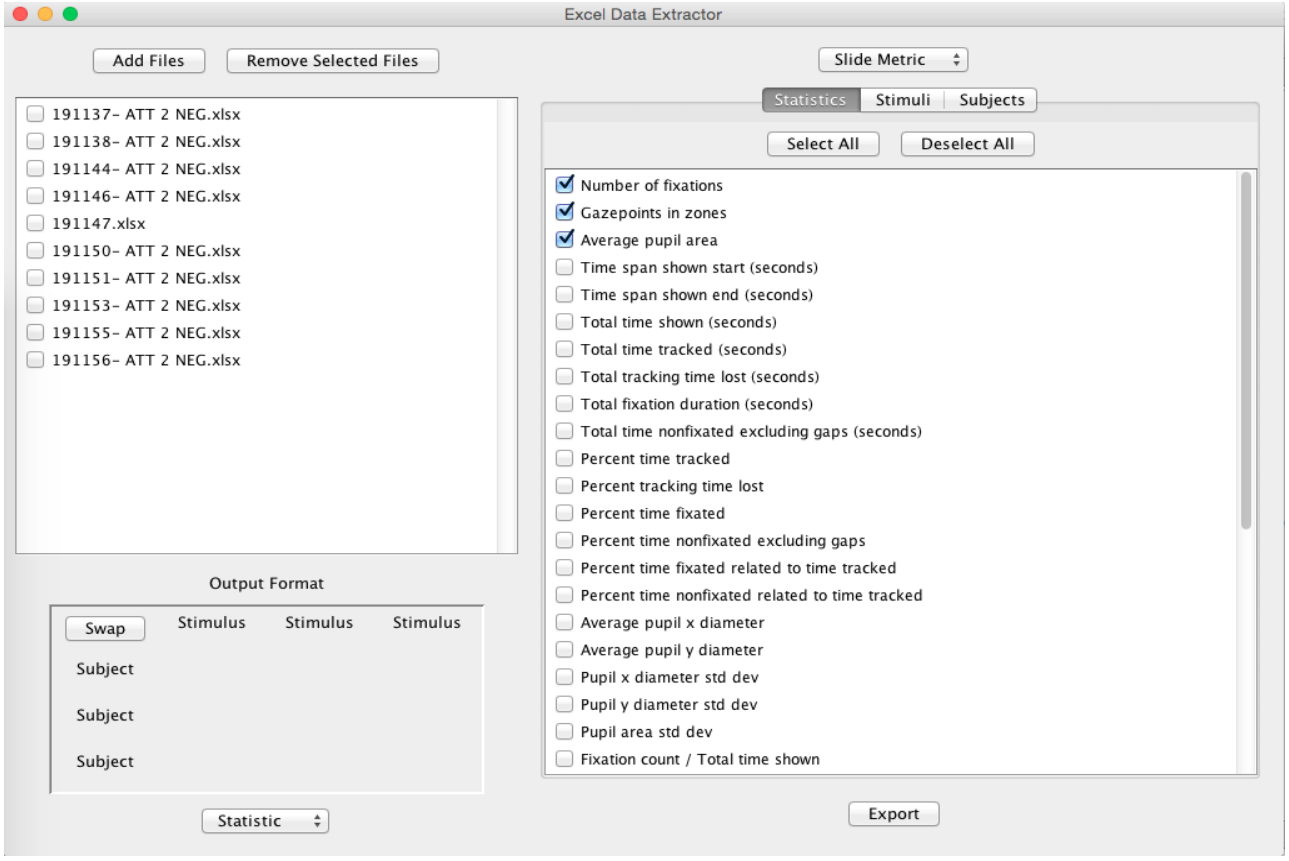
Output Format

The bottom left portion of the window allows the user to configure the format of their output file. This defaults to the standard configuration, where all statistics are added as individual sheets, with subject names on the y axis and stimuli names on the x axis. Users can both swap axis and change the data type for each sheet to create any of six

desired output configurations.

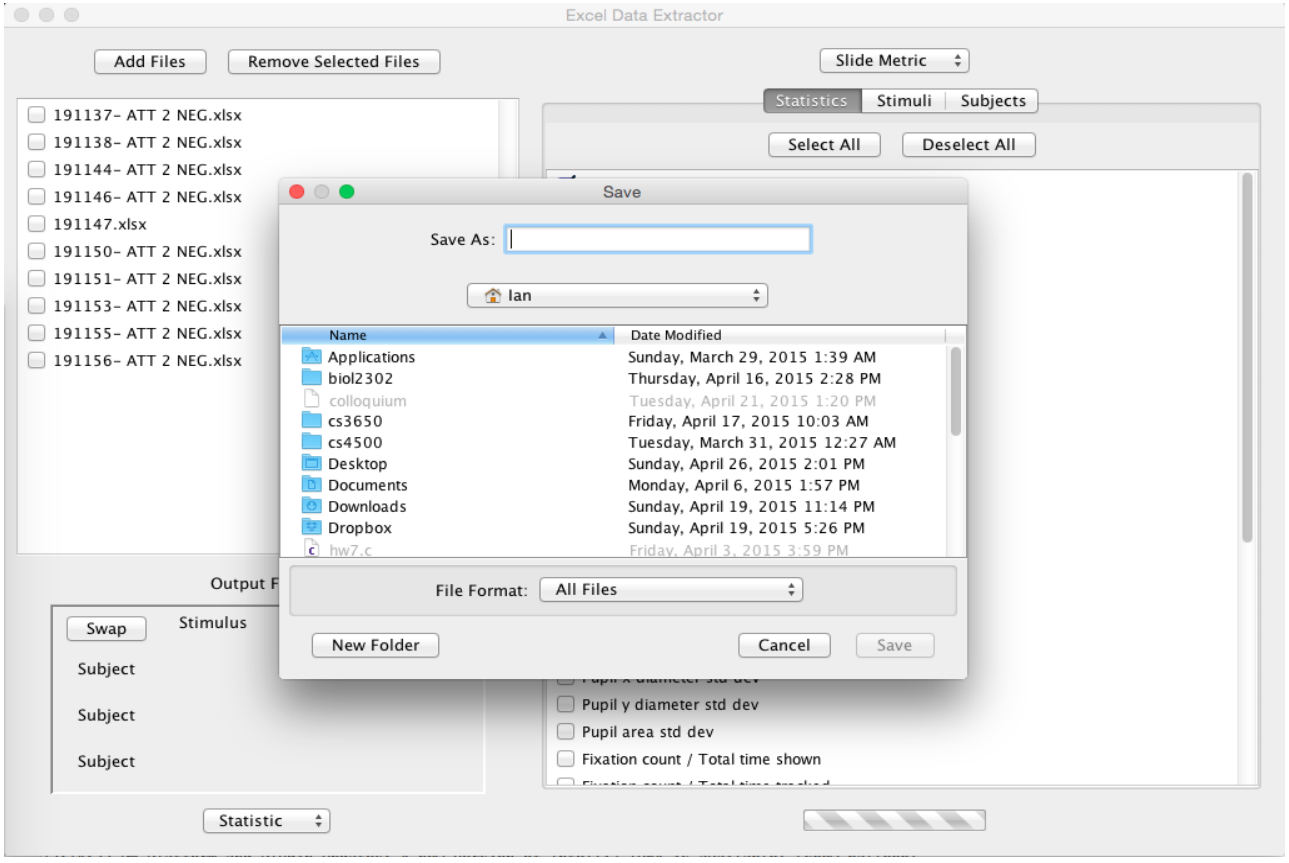
Included Data

Additionally, users can specify what data they would like to include in their output file, and in what order. Each data set for statistics, stimuli, and subjects supports drag-and-drop functionality to specify the order of data in the output file. Only selected data will be included; select all and deselect all buttons near the top of the window provide simple bulk operations for inclusion or exclusion. For this example, we will include three statistics - Average pupil area, Number of Fixations, and Gazepoints in zones, altering the order of the three.

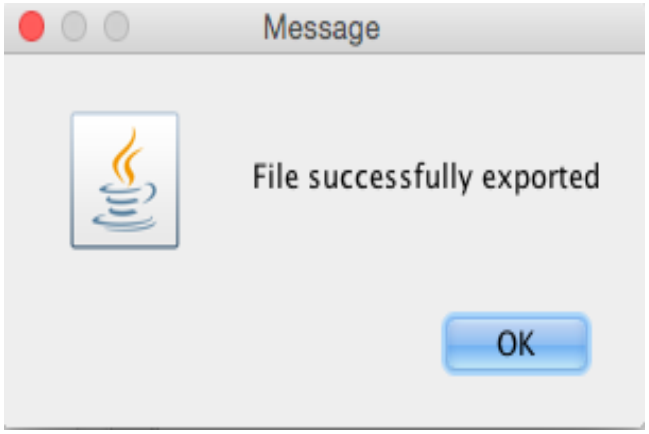


Export

Once you are satisfied with the state of your output file, click “export” in the bottom right corner of the window. This will bring up a file navigation window, where you can name the file and choose its output destination. All files will be given a .xlsx extension if not explicitly included by the user.



Select an appropriate name and output destination, and click “save.” A loading bar will appear in the main window, and disappear upon export completion. Additionally, a confirmation dialog will appear, specifying the success or failure of the export.



The application remains open after export regardless of success, allowing easy access to the data for additional fixes or exports. The output directory will be saved after each export, so there is no need to navigate to it each time.

Data Assumptions

- Sheet names adhere to the tuple pattern “NAME - G”, “NAME - F”, “NAME - STAT”. The stat page may be omitted; if so all data for this page will be blank.
- Cell A1 of “NAME - G” represents the name of the stimulus.
    - For a video study, this takes the form NAME;… where additional text may follow the semicolon.
    - For an image study, this takes the form identical to the sheet name (“NAME - G”)
- STAT pages take on the following form:
    - Row 3 represents the start of slide metric data, and slide metric data continues until a blank row is reached.
    - Slide metric statistic names are in column A, and corresponding numerical data values are in column F
        - Two rows after the blank row, the look zone data begins. This is a set of N data blocks of the following form:
        - Row B contains the name of the look zone
        - For all but the final block, the cell in column F of the following row contains a description of the look zone. This field is optional.
        - Until a blank line is reached, data exists in an identical pattern to that for slide metric (column A is

statistic name, column F is value)

- All data referring to the same field (statistic, stimuli, etc) is expected to be named identically. The software will treat all unmatched strings as new fields

Final Words

This project was developed by Ian Redpath, Teddy Stoddard, Anthony Vatousiou, and Qimeng Song as part of CS4500: Software Development at Northeastern University. This project is no longer in active development, but feel free to contact us with any questions or issues.

Special thanks to our mentor Akshay Raje and Professor Magy Seif El-Nasr for their feedback and help with this project.