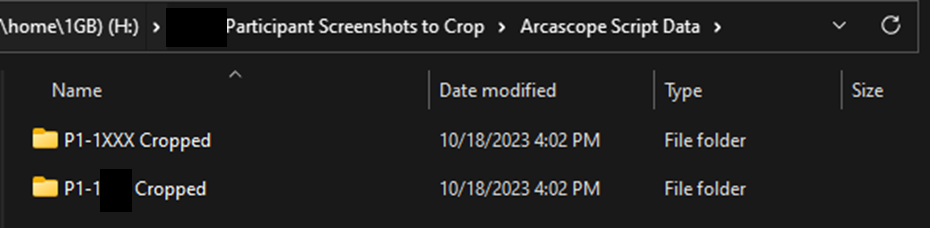
Extracting Hourly Screen Use and App Names from iOS Screenshots with Arcascope's Script

**Remember to only use scripts with copied and pasted versions of the original data.**

**NEVER use the original data with scripts. NEVER cut and paste the original data.**

For the iOS screenshots to be processed by the script, they **may** first need to be cropped and then copied and pasted into a data folder for Arcascope’s script. It should look something like this:



If you have not cropped the images yet and it is necessary to do so, please see the procedure named “Cropping and Removing PHI from iOS Screenshots.”

If you have not deleted duplicate and/or extra daily usage screenshots from the cropped images yet, you must do that before running Arcascope’s script. See the procedure named “Cropping and Removing PHI from iOS Screenshots” for details on that as well.

Installation Requirements and Path Variables

1. Arcascope’s latest script is available publicly on GitHub here:

<https://github.com/Arcascope/screen-scrape/>

1. You must install pytesseract in your Python environment using pip:

<https://pypi.org/project/pytesseract/>

1. You must also install Tesseract on your system, and you can see the instructions to install it here:

<https://tesseract-ocr.github.io/tessdoc/Installation.html>

1. Look for your operating system (Linux, macOS, Windows) by scrolling down.
2. On Windows, Tesseract’s path must be set up properly before the commands in Arcascope’s script that reference it can work.
3. You can do this the simple way by opening “main.py” in an IDE and adding this line to the top after the imports, if it’s not already there:

**pytesseract.pytesseract.tesseract\_cmd =** **r”C:\Users\(YOUR USERNAME WITHOUT PARENTHESES)\AppData\Local\Programs\Tesseract-OCR”**

1. If this does not work, then you can try editing the “Path/PATH” environment variable instead.
2. First, open “Search” (Windows key + S) and type “Edit the system environment variables,” then click it
3. In the window that comes up, click “Environment Variables…” on the bottom right
4. Double click on “Path” in the bottom or top box
5. Click on “New” and add your Tesseract path. This will usually be:

**”C:\Users\(YOUR\_USERNAME\_WITHOUT\_PARENTHESES)\AppData\Local\Programs\Tesseract-OCR”**

1. Then press “Enter” and click “OK.”

Configuring and Running the Script

1. You will now need to open Arcascope’s script in an IDE, if not already open.
2. You may also need to reformat the names of the participant folders stored within the data folder to match the script’s expected format (Main Folder -> Participant Folders -> Day/Date folders with date in the format M.D.Y or M-D-Y) or alter the script to match your format.
3. After you run the script, you will see a GUI with the main components described below:

A screenshot of a video game

Description automatically generated

* 1. This is where you select your main folder, which contains the participant folders, which contain the folders of screenshots for each day/date.









* 1. As shown above, right below the folder selection is where you will see prompts depending on the state of the current image. The first two messages are to remind you to check and reselect the graph, but only if necessary. The other two messages appear in the case of an error and will force you to manually reselect or enter text.



* 1. The buttons at the bottom of the GUI, “Next/Save” and “Previous” are used to navigate through the images. The “Next/Save” button will go to the next image AND save the data from the parsed graph from the current image. Do not press this button if you do not want to save the data from the current image, just press the “Skip” button as described below. The “Previous” button will allow you to go back to the previous image. However, if you make any changes and press “Next/Save” again, you will create a duplicate entry in the output CSV file. You can select the checkbox shown below to automatically remove duplicate entries with the same full image path. Only the last updated entry will be kept for each full image path.
  2. (Not displayed here) The left-hand side of the image will display the currently selected screenshot.
  3. (Not displayed here) The middle of the image will display the image of the graph extracted from the currently selected screenshot and the parsing of that graph into actual values, if applicable. This is where you will verify that the graph is parsed correctly and reselect the graph if they do not match. If they match, you will click “Next/Save”.

A screenshot of a computer

Description automatically generated

* 1. As shown above, the right hand side of the GUI displays the name of the current displayed screenshot, the app name extracted from it, and a few other options.
     1. The “Automatically snap to grid” checkbox currently does nothing.
     2. The “Automatically process images” will minimize the GUI and process the output for images based on the graph that it detects automatically, until it encounters an error. At this point it will bring up the GUI again for manual input. It is recommended to only do this if you will manually check the output later.
     3. The “Remove duplicate entries” button will remove any duplicates in the output CSV if the full image paths of two or more entries match, keeping the last updated entry.
     4. The “Skip” button will skip the current image in case you don’t want to save the data from it, or if the GUI will not let you press “Next/Save” without selecting a graph and the image does not have a graph.

1. If the script has an error not described above, you will have to debug it. Common errors will be related to the folders or an issue with a screenshot.
2. Ask research staff for assistance if you are unable to prevent errors from occurring, or contact Arcascope if truly necessary.
3. Once the script runs successfully, it will take a few minutes to finish. You will find the outputs for each participant folder in the form of Excel files within the folder you previously set as output.
4. Double check that the usage in the columns of the Excel sheet closely matches the usage in the screenshots. If the usage is incorrect by a few minutes, make sure that it’s only underestimating rather than overestimating. Make sure that no usage is ignored.
5. Finally, you can postprocess the output files and/or share them for any reports that need to be made. You must also double check the postprocessed data as well to make sure that it matches up with the original screenshot usage.