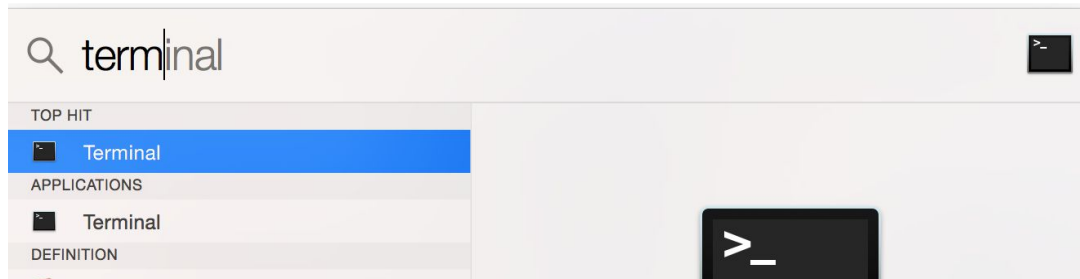# Sochiatrist Extraction Guide

The following document contains a list of instructions to guide you through the data extraction process. Happy data-digging!

If this is your first time running the extraction, follow the Setup.
If not, jump to Part I (on page 5).

1. The first thing you will need to do is **download the Sochiatrist zip file**.

2. Open the zip file, and save it as a folder on your *Desktop*. Make sure to rename the folder to just "***Sochiatrist"* (without quotes).**

Open Terminal (the easiest way to do this is just by searching for it through Finder).

Part I: Setup (now optional)
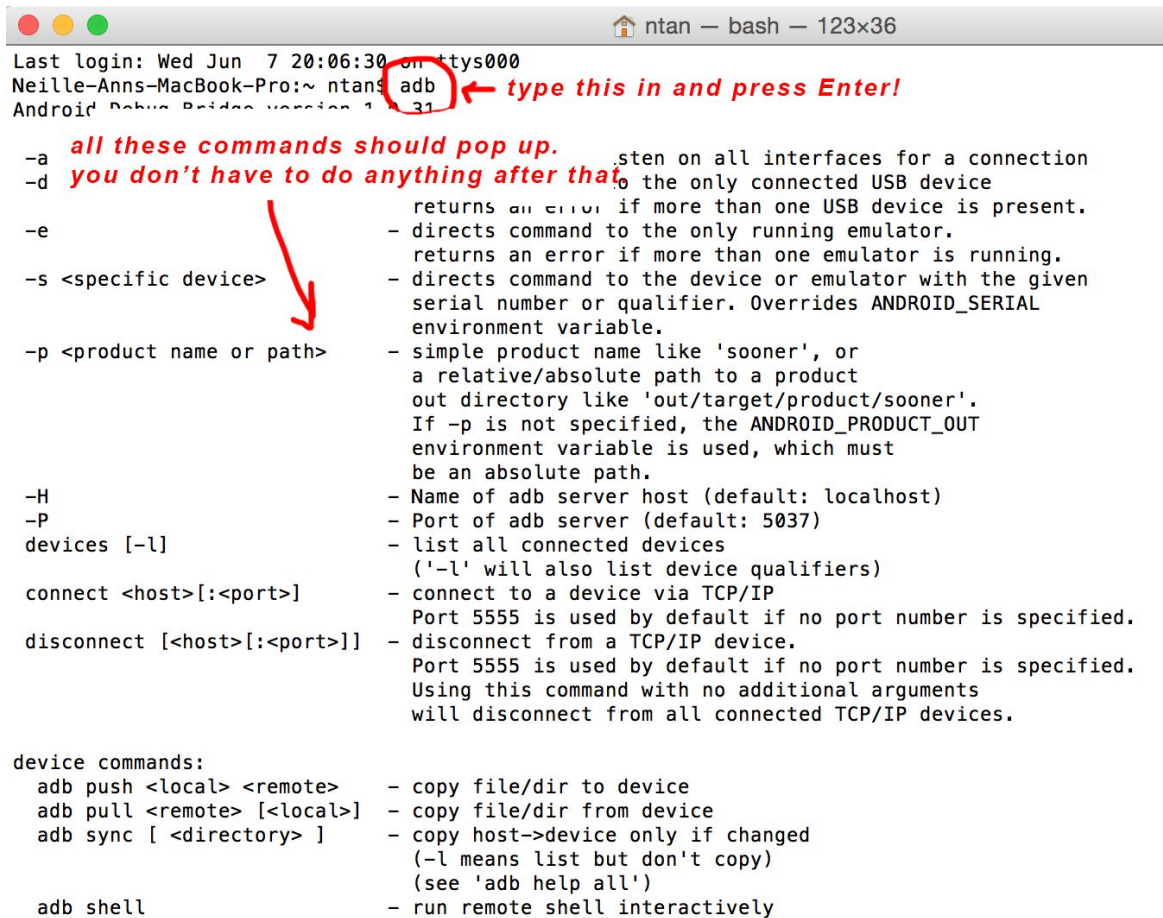
In your terminal, run **python ~/Desktop/Sochiatrist/sochiatrist-social-data-extractor/installer.py**
This should download and install all necessary software, which allows all the OSN extraction to run smoothly.

(press Enter after this!)

… which will bring you to a screen that looks like this. **Enter your laptop password -- note that this will be blank when you type.**

Restart your terminal by closing out of it and opening a new terminal window. To check if adb is running correctly, type in "adb" into your Terminal. A list of commands (see following image below) should appear after you type in "adb" and press enter.


```
Last login: Wed Jun  7 20:06:30 on ttys000
Neille-Anns-MacBook-Pro:~ ntan$ adb          ← type this in and press Enter!
Android Debug Bridge version 1.0.31

 -a                           ...sten on all interfaces for a connection
 -d                           ...o the only connected USB device
                              returns an error if more than one USB device is present.
 -e                           - directs command to the only running emulator.
                              returns an error if more than one emulator is running.
 -s <specific device>         - directs command to the device or emulator with the given
                              serial number or qualifier. Overrides ANDROID_SERIAL
                              environment variable.
 -p <product name or path>    - simple product name like 'sooner', or
                              a relative/absolute path to a product
                              out directory like 'out/target/product/sooner'.
                              If -p is not specified, the ANDROID_PRODUCT_OUT
                              environment variable is used, which must
                              be an absolute path.
 -H                           - Name of adb server host (default: localhost)
 -P                           - Port of adb server (default: 5037)
 devices [-l]                 - list all connected devices
                              ('-l' will also list device qualifiers)
 connect <host>[:<port>]      - connect to a device via TCP/IP
                              Port 5555 is used by default if no port number is specified.
 disconnect [<host>[:<port>]] - disconnect from a TCP/IP device.
                              Port 5555 is used by default if no port number is specified.
                              Using this command with no additional arguments
                              will disconnect from all connected TCP/IP devices.

device commands:
  adb push <local> <remote>   - copy file/dir to device
  adb pull <remote> [<local>] - copy file/dir from device
  adb sync [ <directory> ]    - copy host->device only if changed
                              (-l means list but don't copy)
                              (see 'adb help all')
  adb shell                   - run remote shell interactively
```

all these commands should pop up.
you don't have to do anything after that

If you are doing an extraction with an Android phone, type in "adb devices" and look for a screen that looks like


```
Last login: Tue Jul 11 00:36:21 on ttys000
Neille-Anns-MacBook-Pro:~ ntan$ adb devices
List of devices attached
0025c7b79451e6dc          device

Neille-Anns-MacBook-Pro:~ ntan$ ▋
```

If you don't see a screen like that, that's okay! All you need to do is make sure your phone is connected to your computer and unlocked. Once you have confirmed this, swipe down from the top of your Android screen, and click "USB Charging. Then click the Image Transfer option. Then hit the Media Transfer or File Transfer option. Then type in "**adb devices**" again and the device should show up.

Part II: Running the sochiatrist-social-data-extractor script

3. **Open your Terminal**. From there, copy and paste the script below (go ahead and insert the participant's number where it says "[participant number]").

   python ~/Desktop/Sochiatrist/sochiatrist-social-data-extractor/sochiatrist-social-data-extractor.py [participant number]

   If this is the first time running the tool, the script should download and install all necessary software, which allows all the OSN extraction to run smoothly.

   Enter your laptop password to install the dependencies when prompted -- note that this will be blank when you type.

   If this is not your first time using the tool, it will ask you if you want to force update. The program will automatically update every week!

4. You will receive a prompt asking you to type in your Participant ID - type in the same ID as what was given to you by the researcher.

5. After this, the script itself should start giving you commands on what to do. This process will extract all of the messaging data from each social media network, and place it into a separate CSV file. Follow what it says in the terminal, and type in yes where it is applicable to the participant (i.e. Do you have Kik? **yes**).

   * If the participant's Twitter is set to private, you will not be able to extract their Tweets. You can, however, extract their Twitter DMs (this is separate from the tweets, since the participant gives their password for authorization).

When you're done, you should see something like this:

```
 BeautifulSoup(YOUR_MARKUP})

to this:

 BeautifulSoup(YOUR_MARKUP, "lxml")

  markup_type=markup_type))
Parse done.
Loading into thread...
Load done. Pickling...
Pickle saved. Done.
first pass for name threads
second pass for id threads
third pass for group message threads
making CSV right now
Is the participant sharing Facebook timeline posts? If so, type in yes.

yes
Finished all extraction, now please run the CSV Time Tool to anonymize the messa
ges, and get the messages from the specific date ranges you want.

Neille-Anns-MBP:~ ntan$ python ~/Desktop/sochiatrist/social-media-data-grabber/C
SV_Time_Tool/csvtimetool.py█
```

Part III: Running the Time Tool

6. Now that the social media data has been extracted from each social media network, you will need to run a script to combine all the data into one CSV, and then limit the data to the date range you need. To do this, copy and paste the script below. **Please make sure not to have a '_' in the file name you choose. Make a note of the file names you have chosen.**
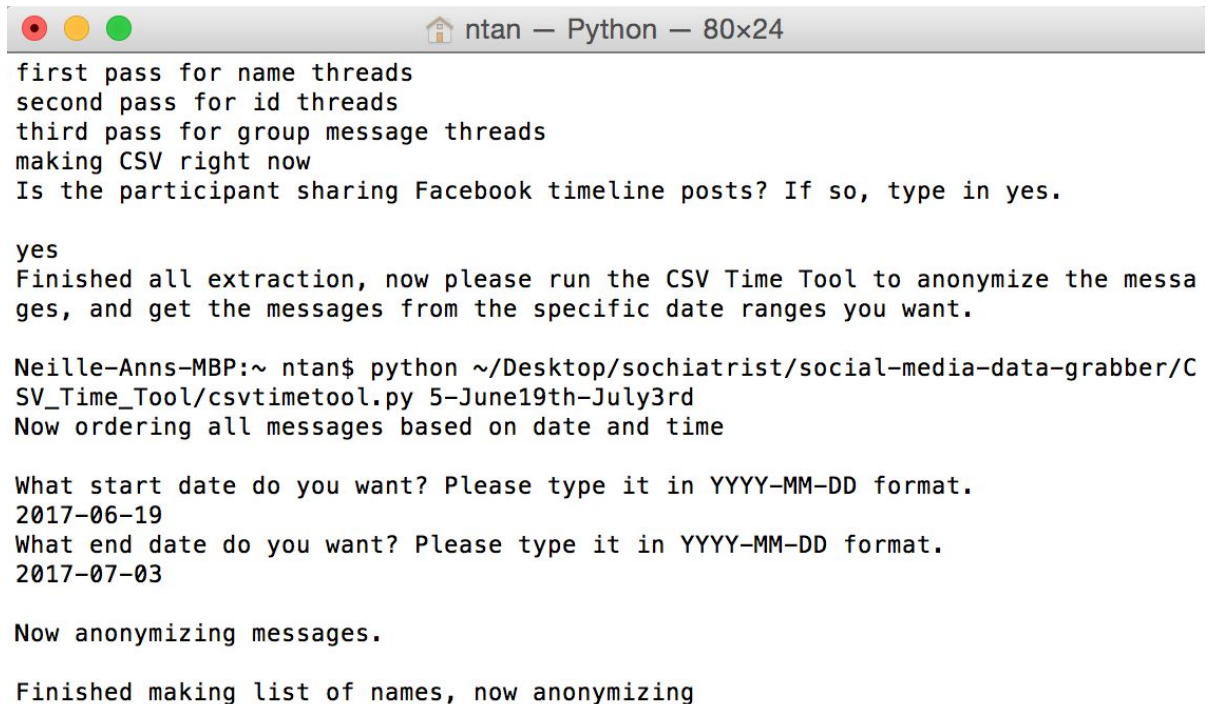
   python ~/Desktop/Sochiatrist/sochiatrist-social-data-extractor/timeanonymizer.py [participant number and date range]

   An example name is "5-June19th-July3rd" (without quotes) for participant 5's data from June 19th to July 3rd.

   If you need to get data from multiple date ranges, just repeat this process with the date ranges you want to get data from.

   After pulling the messages, it will ask if you want to anonymize the data. Enter yes if you do.

Once you're done, you should get something that looks like the screenshot below.


```
● ● ●                    🏠 ntan — Python — 80×24
first pass for name threads
second pass for id threads
third pass for group message threads
making CSV right now
Is the participant sharing Facebook timeline posts? If so, type in yes.

yes
Finished all extraction, now please run the CSV Time Tool to anonymize the messa
ges, and get the messages from the specific date ranges you want.

Neille-Anns-MBP:~ ntan$ python ~/Desktop/sochiatrist/social-media-data-grabber/C
SV_Time_Tool/csvtimetool.py 5-June19th-July3rd
Now ordering all messages based on date and time

What start date do you want? Please type it in YYYY-MM-DD format.
2017-06-19
What end date do you want? Please type it in YYYY-MM-DD format.
2017-07-03

Now anonymizing messages.

Finished making list of names, now anonymizing
```

This part can take a while, so be patient.

7. Now that the social media data has been anonymized, you need to delete all of the files that are not being used for the study. First, go to the CSVs folder on the Desktop, and delete all the files that do not have "anonymized-" at the beginning of their name.

   THIS IS SUPER IMPORTANT!!! If you do not delete the files from the CSVs folder, the next time you run the sochiatrist-social-data-extractor script it will not work. The CSVs folder needs to be completely empty. Beware!

Keep this, but delete everything else in this folder for privacy.
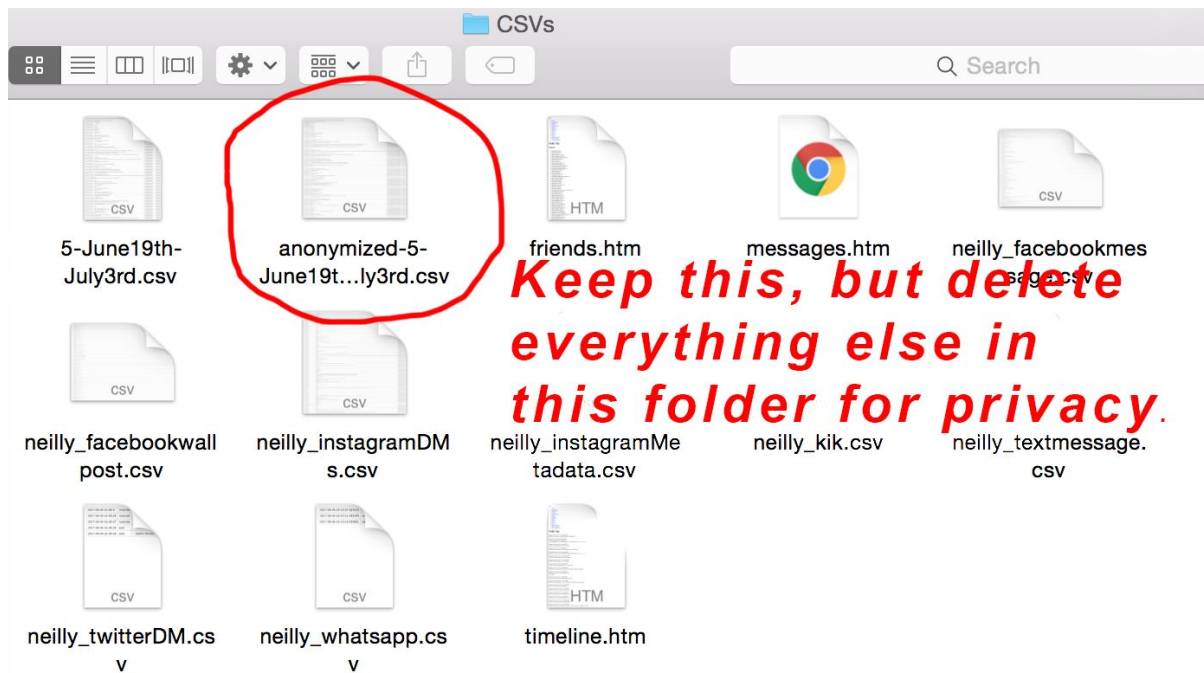
Delete EVERYTHING in the CSVs folder, and move the anonymized data to a different location for permanent storage (as seen below).
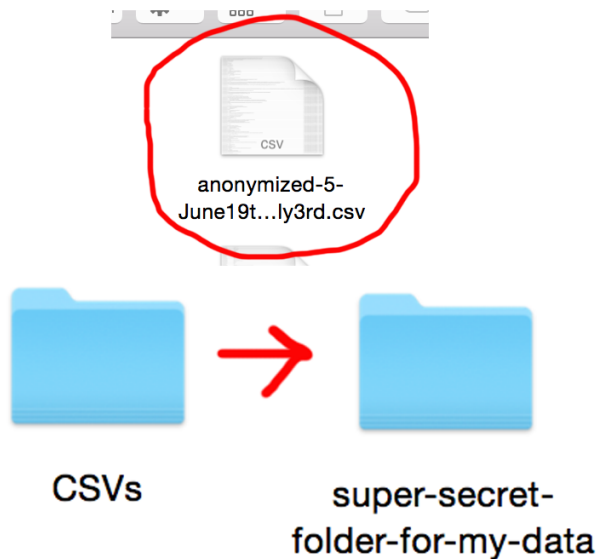
8. Next, delete any participant Facebook data from the Downloads folder. There will be both a .zip file and a folder.
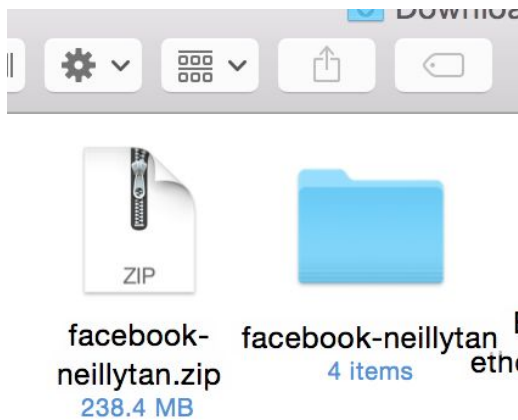
Next, if the participant had an iPhone, run

python ~/Desktop/Sochiatrist/sochiatrist-social-data-extractor/resetScript.py

^ Copy and paste this into your Terminal. This will delete their iPhone backup.

9. You're done!

FAQ:

1. **What if the participant forgets their password or types their password in wrong?**
   - That's okay! Take note of which social network the participant wasn't able to extract, continue on with the script, and stop before step 6 (Running the Time Tool). Then, depending on the social network, run one of these scripts:

   1. iPhone
      a. Text messages ---

         python ~/Desktop/sochiatrist/sochiatrist-social-data-extractor/iMessages/getiOSMessages.py participantID

b. Facebook Messages --- python ~/Desktop/sochiatrist/sochiatrist-social-data-extractor/Facebook_Messages/getFacebookMessages.py ~/Desktop/CSVs/messages.htm participantID

    i. **Before running this script, make sure that there is a messages.htm file in the CSVs folder on the Desktop. If there isn't, go to the Facebook data .zip in the Downloads folder, extract it, and from the extracted folder, move messages.htm, timeline.htm, and friends.htm to the CSVs folder on the Desktop.**

c. Facebook Timeline Posts --- python ~/Desktop/sochiatrist/sochiatrist-social-data-extractor/Facebook_Wallposts/getFacebookWallposts.py ~/Desktop/CSVs/timeline.htm participantID

    i. **Before running this script, make sure that there is a timeline.htm file in the CSVs folder on the Desktop. If there isn't, go to the Facebook data .zip in the Downloads folder, extract it, and from the extracted folder, move messages.htm, timeline.htm, and friends.htm to the CSVs folder on the Desktop.**

d. WhatsApp Messages --- python ~/Desktop/sochiatrist/sochiatrist-social-data-extractor/WhatsApp/getWhatsAppiOS.py participantID

e. Twitter Tweets --- python ~/Desktop/sochiatrist/sochiatrist-social-data-extractor/Twitter/getTweets.py participantID

f. Twitter Direct Messages --- python ~/Desktop/sochiatrist/sochiatrist-social-data-extractor/Twitter_DMs/getTwitterMessages.py participantID

g. Kik Messages --- python ~/Desktop/sochiatrist/sochiatrist-social-data-extractor/Kik/getKikMessagesiOS.py 31

h. Instagram Direct Messages --- python ~/Desktop/sochiatrist/sochiatrist-social-data-extractor/Instagram_DMs/getInstagramDMs.py participantID

i. Instagram Post Metadata --- python ~/Desktop/sochiatrist/sochiatrist-social-data-extractor/Instagram_Metadata/getInstagramMetadata.py participantID

2. Android

   a. Text messages --- python ~/Desktop/sochiatrist/sochiatrist-social-data-extractor/Android_SMS/getTextMessagesAndroid.py participantID

   b. Facebook Messages --- python ~/Desktop/sochiatrist/sochiatrist-social-data-extractor/Facebook_Messages/getFacebookMessages.py ~/Desktop/CSVs/messages.htm participantID

      i. **Before running this script, make sure that there is a messages.htm file in the CSVs folder on the Desktop. If there isn't, go to the Facebook data .zip in the Downloads folder, extract it, and from the extracted folder, move messages.htm, timeline.htm, and friends.htm to the CSVs folder on the Desktop.**

   c. Facebook Timeline Posts --- python ~/Desktop/sochiatrist/sochiatrist-social-data-extractor/Facebook_Wallposts/getFacebookWallposts.py ~/Desktop/CSVs/timeline.htm participantID

      i. **Before running this script, make sure that there is a timeline.htm file in the CSVs folder on the Desktop. If there isn't, go to the Facebook data .zip in the Downloads folder, extract it, and from the extracted folder, move messages.htm, timeline.htm, and friends.htm to the CSVs folder on the Desktop.**

    d. WhatsApp Messages --- python ~/Desktop/sochiatrist/sochiatrist-social-data-extractor/WhatsApp/android_whatsapp/getWhatsAppMessagesAndroid.py participantID

    e. Twitter Tweets --- python ~/Desktop/sochiatrist/sochiatrist-social-data-extractor/Twitter/getTweets.py participantID

    f. Twitter Direct Messages --- python ~/Desktop/sochiatrist/sochiatrist-social-data-extractor/Twitter_DMs/getTwitterMessages.py participantID

    g. Kik Messages --- python ~/Desktop/sochiatrist/sochiatrist-social-data-extractor/Kik/kik_android/getKikMessagesAndroid.py participantID

    h. Instagram Direct Messages --- python ~/Desktop/sochiatrist/sochiatrist-social-data-extractor/Instagram_DMs/getInstagramDMs.py participantID

    i. Instagram Post Metadata --- python ~/Desktop/sochiatrist/sochiatrist-social-data-extractor/Instagram_Metadata/getInstagramMetadata.py participantID

After you have finished this, proceed from step 6.

2. What if the participant decides, last minute, that they don't want to include social media data from a specific social media network?
   - If you have passed step 8, you will have to redo the process. If you have not passed step 8, just delete the file in the folder CSVs that has the name of the social network that the participant does not want to share in it. Then begin again from step 6 onward.

Helpful tips and tricks:
- Remember to wait for all your backups (if you have an iPhone) to complete downloading.
- Make sure that your Facebook .zip file stays in the Downloads folder, and does not extract automatically (it needs to stay compressed as a .zip file)