Harmonization Installation (Dec. 2018)

# ASSUMPTIONS

* PostgreSQL is installed
* OMOP is installed in that instance of PostgreSQL
* Python 3 is installed
* Required packages are installed into Python

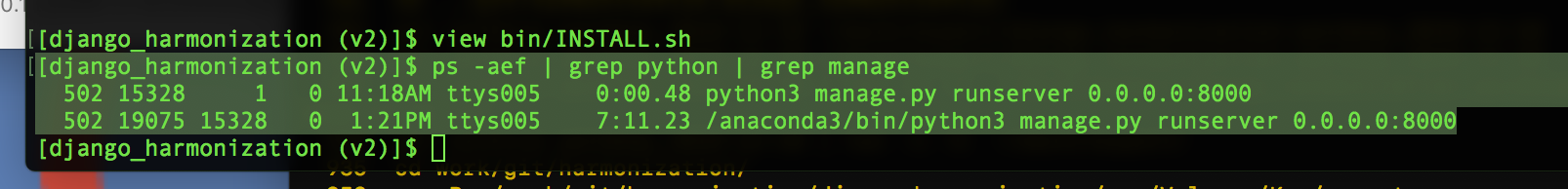
# PRE-STEPS

1. Identify the user that has permissions to log into postgres.
   1. Most probably your user name, PSJ, croeder, etc.
   2. In a terminal window “echo $PGUSER”. If this agrees with the above idea, go with it. If not, go with it unless it’s either croeder or christopherroeder.
2. Kill any running Django instances (more detail below in the issues section). Pid is from the second column in the ps display. And ForceQuit is no help.
   1. ps -aef | grep python | grep manage
   2. kill <pid>
   3. run ps again, sometimes they are stubborn or there are two.

# STEPS

1. Create a directory somewhere…in your home directory /Users/croeder for example, call it “harmonization”
   1. mkdir ~/harmonization
2. Copy the zip file into that directory
   1. cp /some/other/place/harm\_2018-12-12\_16\_53.zip ~/harmonization
3. Change into that directory
   1. cd ~/harmonization
4. Unzip the zip file
   1. unzip harm\_2018-12-12\_16:53.zip
5. Edit the bin/env.sh file to have the correct PGUSER value as in step 1.
   1. Run textedit on ~/harmonization/bin/env.sh and replace the croeder in “PGUSER=croeder” with what we have from step 1.
6. Run the install script. Follow along the directions there, hitting return as steps complete. Following this item are additional details worth reading before you start.  
   Sudo will prompt for the password you use to unlock this mac.
   1. sudo bin/INSTALL.sh
7. At one point the script will stop and ask you to copy the data files into /opt/local: “Pausing for file copy”. At this point, the place for your studies’ data files has been created, so it’s an appropriate time to copy files into the directories under /opt/local/harmonization/deployment/studies/CORONA/data. Files exist with the names test.csv and sample.csv, as they were created for each study. The files are there are bogus and have no content. They are just there to mark the correct place.
   1. Copy the real data files in their place, AND USE THE SAME NAMES: test.csv or sample.csv. The code is written to look for those. I assume the real data files have different names. If we need to distinguish them, we can use the “wc” command to count the number of lines. The real files of course have more lines.

# Potential Issues

* New python packages need to be installed: install as needed with (sudo) pip3.
* Installing different studies than what is prepared for: edit the load\_studies.py
* A previously running instance of Django needs to be killed
  + ps -aef | grep python | grep manage
  + use the numbers in the second column to kill
  + 
    - kill 19075; kill 15328
* Wrong input data files. The blanks in the zip file are not meant to work.
* Wrong extraction (categorization) configuration selected.