Contents

Introduction 2

Requirements 2

Installation 3

Running the pipeline 5

Introduction

This software was developed as part of the **UNITE** (***Un****derstanding l****i****nks between social determinants and firearm violence in California communi****ti******es***) project, in order to facilitate mapping of addresses associated with protected health records to census tracts in a decentralized surveillance network. It is essentially a user-friendly graphical user interface to the DeGAUSS offline geocoder (v 3.0), using R and Shiny through a Docker image, which provides a reproducible workflow across machines and operating systems with minimal set-up.

Requirements

- **Docker desktop**
- **Internet browser** (required to open graphical user interface running on local machine)

Installation

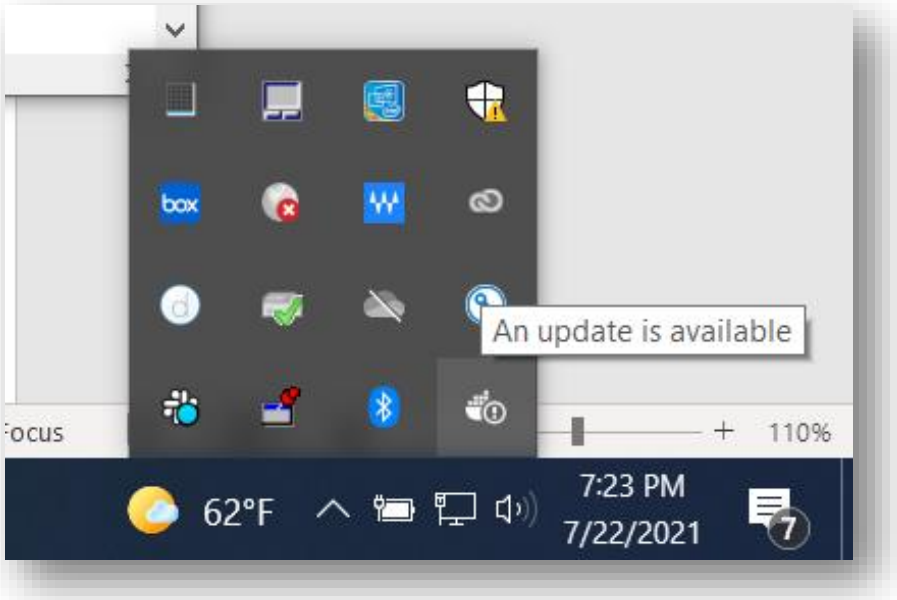
1. Install and start Docker desktop (detailed instructions for Windows, Mac)

2. ***Especially if you are running the pipeline on a secure computer that is not normally connected to the internet***, install the pipeline Docker Image by running install.bat (Windows) or install.command (Mac)

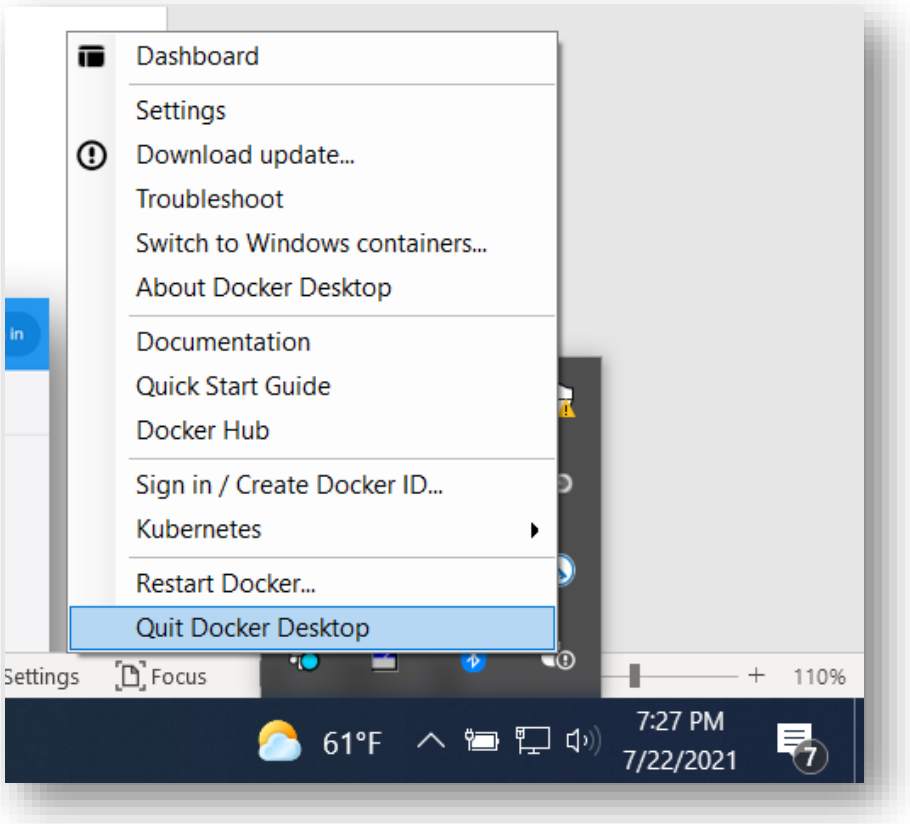
    a. If it's acceptable for the host computer to maintain an internet connection, installation of the Docker Image can be handled by running *UNITE_Geocoding_Pipeline.bat* (Windows) or *UNITE_Geocoding_Pipeline.command* (Mac). If the image is not present on the system yet, it will be installed the first time these scripts are run. The installation scripts are simply a means to install the image without running the rest of the pipeline

    b. After the installation completes, it is recommended to exit and restart Docker Desktop in order to release some RAM arrogated during the image installation process (this seems to be mainly a Windows problem where Docker build and installation steps do a poor job of releasing memory back to the system…)

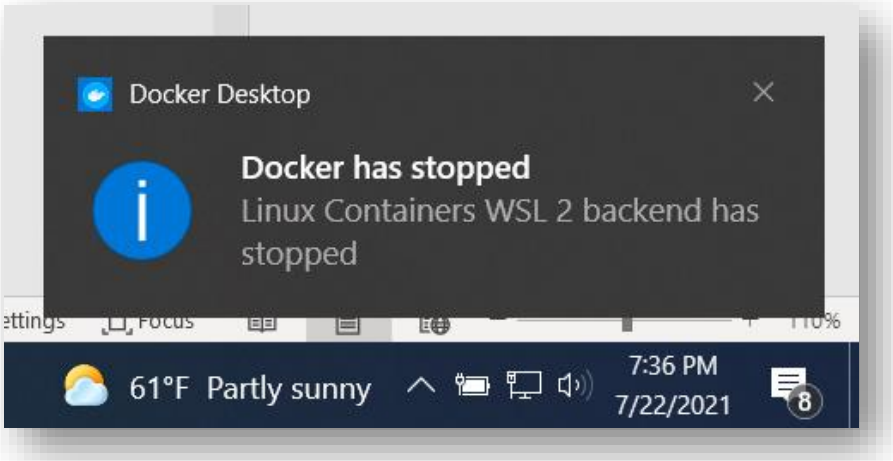
        i. Find Docker Desktop via the System Tray (a white whale icon, which may have an exclamation point button if there is a notification from the service)



ii. Right click to open the context menu and select “Quit Docker Desktop”



iii. Wait for the notification that Docker Desktop has stopped running, then restart it from the start menu or search bar

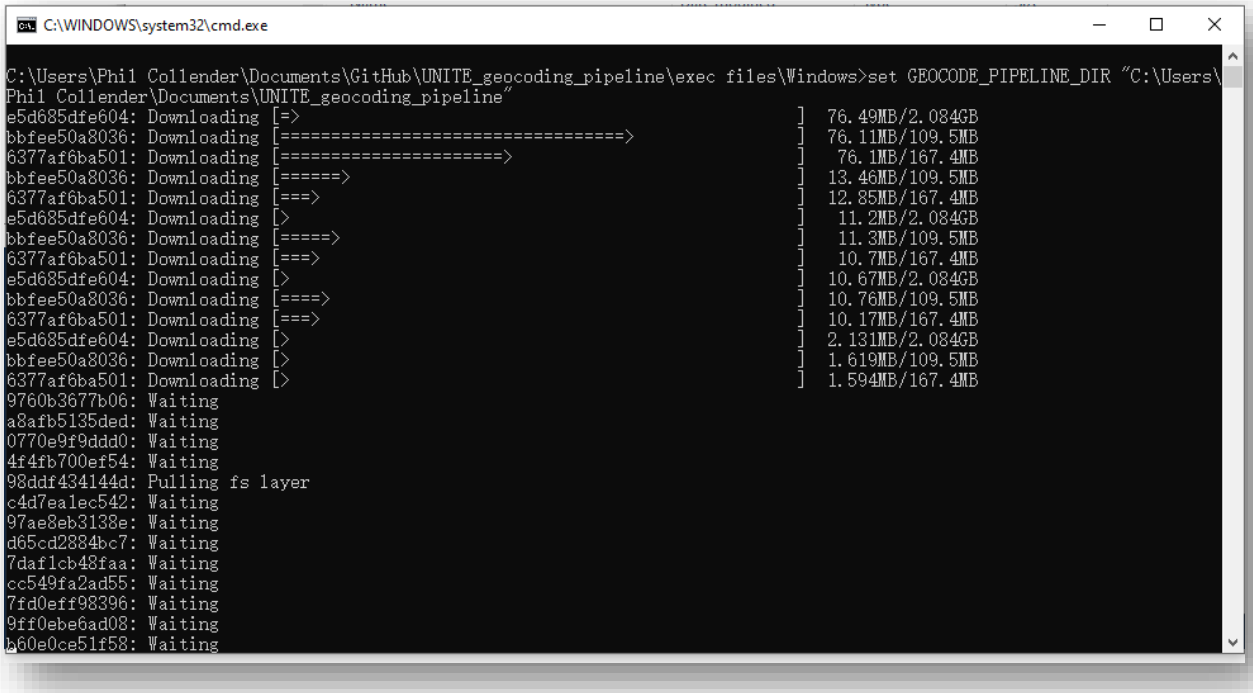


Running the pipeline

1. **Make sure Docker Desktop is running**

2. Start the pipeline by double clicking *UNITE_Geocoding_Pipeline.bat* (Windows) or *UNITE_Geocoding_Pipeline.command* (Mac)

3. A terminal window will open, showing the commands run in the script. If you have not previously installed the pipeline Docker Image, you can monitor installation progress in this window. Otherwise, it can be ignored (unless you want to view the shell code being run)

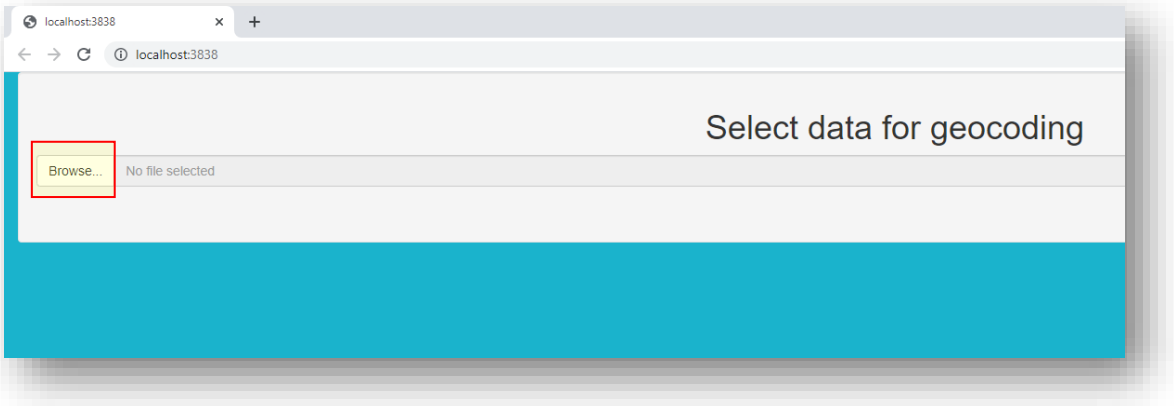


Terminal window during image installation

4. Once the image is available, a browser window will launch with the URL http:\\localhost:3838, which is a port on your computer hosting the pipeline application. The window may briefly display a 'page not found' message if it opens before the application is up and running, but this should resolve within a couple of seconds.

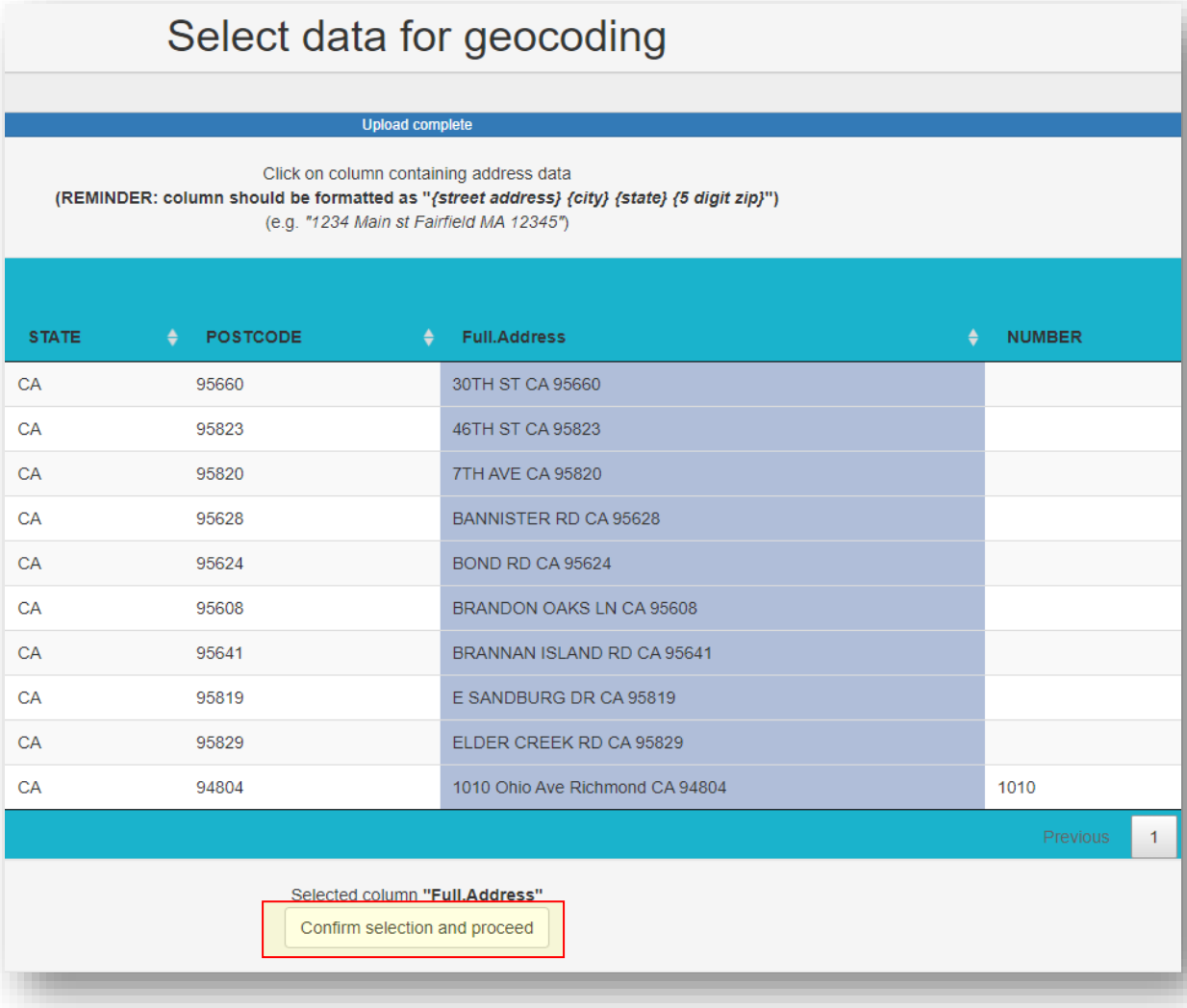
5. Click the **Browse** button to open a file selection window and select a *.csv* file containing your address data. If you mess up and select the wrong file, you can click the **Browse** button again. To test the pipeline, feel free to use the packaged *example_data.csv*, as shown in screenshots below.

i. As per the DeGAUSS documentation[1], addresses should be in a single column formatted as
<Street Number and Street Name> <City> <State> <Zip Code>
Separated by single spaces
ii. Do not include apartment numbers or second address lines
iii. Zip codes should be 5 digits, without "plus four" digits
iv. Use Arabic numerals instead of written numbers



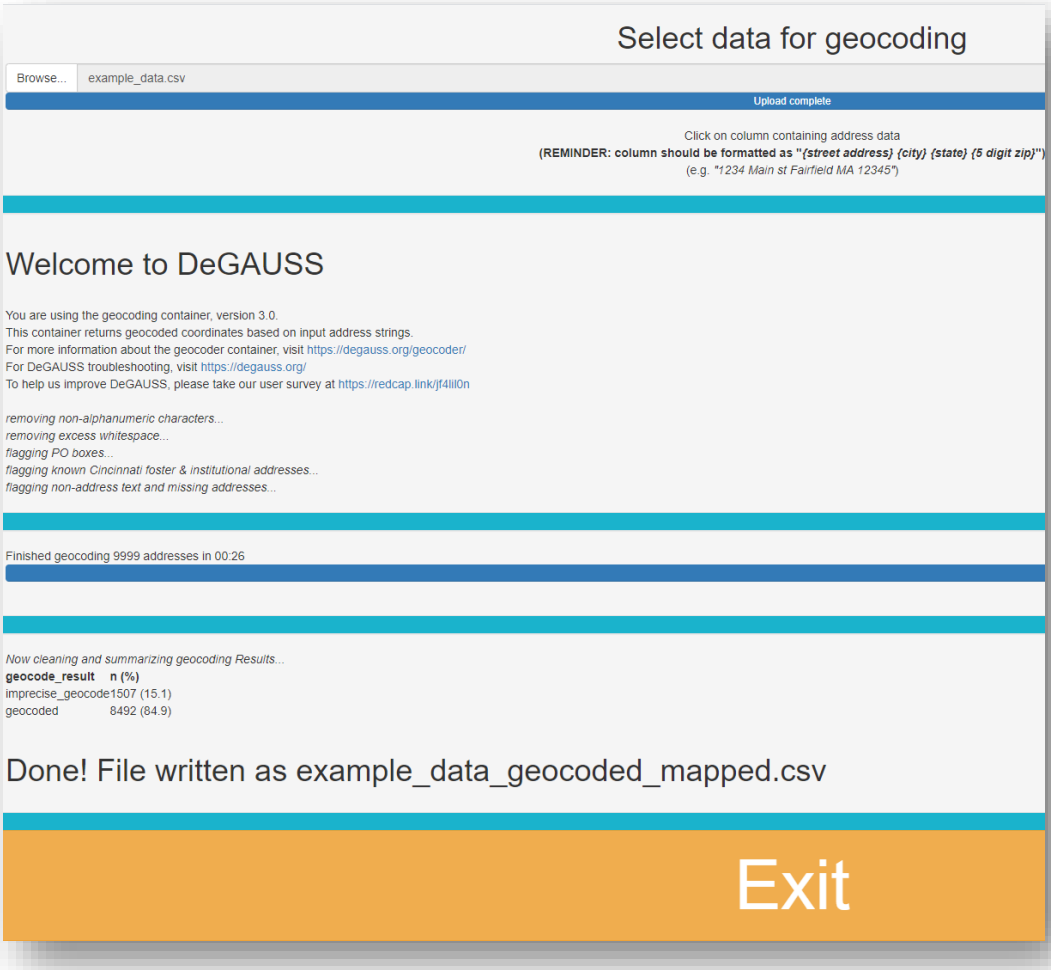
[1] One thing that is different from stock DeGAUSS is that the column containing address data does not have to be named "**address**"

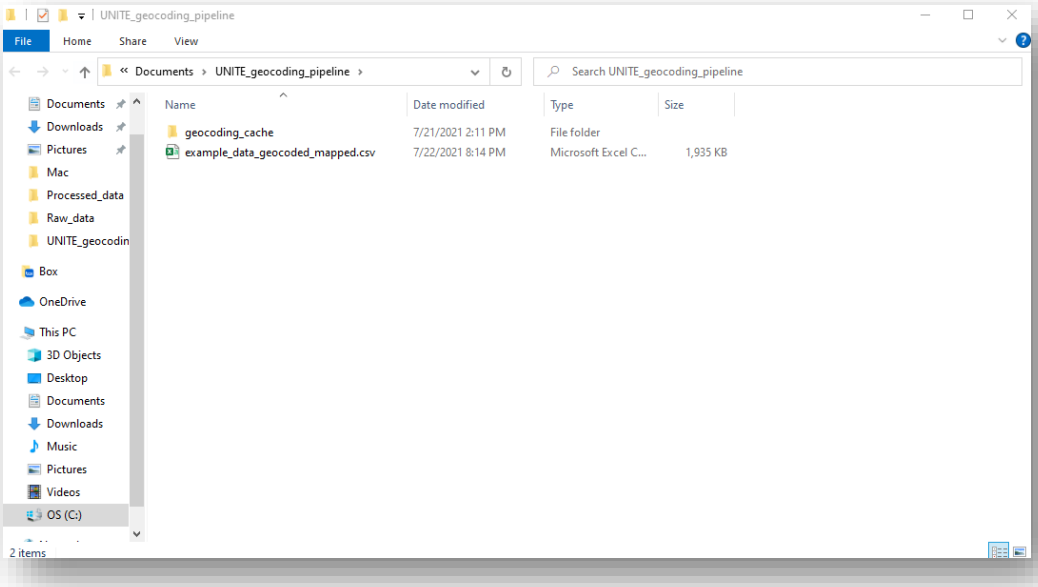
6. Single click to select the column containing your address data and, if satisfied with your selection, click the **Confirm selection and proceed** button



7. The pipeline will now run, printing status and progress messages along the way. Note that the initial setup of the geocoding cache can take some time before progress of the geocoding process is apparent.

8. Once finished, a message will announce the name of the file holding your geocoded and mapped data (it will have the format *{original_file_name}_geocoded_mapped.csv*). You may now hit the larger **Exit** button at the bottom of the window or close the browser window, at which point the process will clean up temporary files, launch a window to the directory containing the output file (*Documents/Unite_geocoding_pipeline*) , and close itself.





Output directory. The geocoding_cache folder contains temporary results of calculations which can speed up repeat geocoding, e.g. if the process dies for some reason or is taking too long and has to be shut down before finishing