

## Contents

Introduction.....	2
Requirements .....	2
Installation.....	3
Running the pipeline .....	5

## Introduction

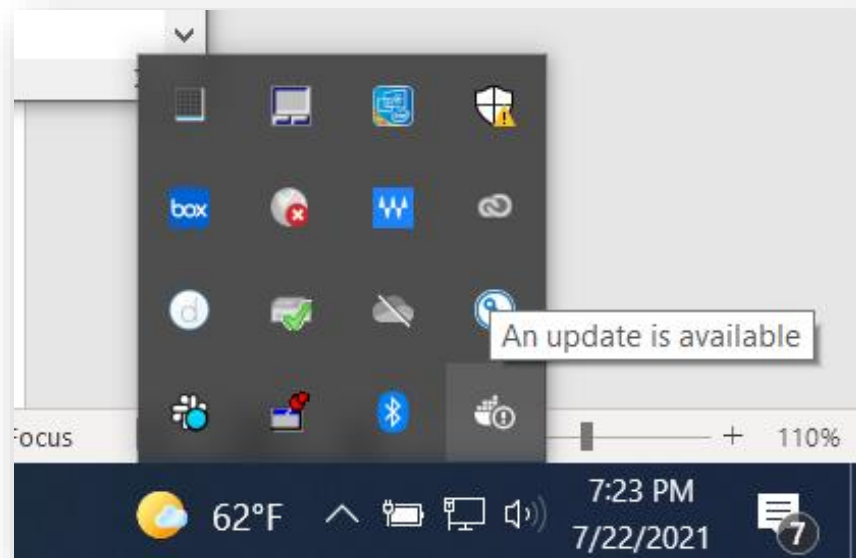
This software was developed as part of the **UNITE** (*Understanding links between social determinants and firearm violence in California communities*) project, in order to facilitate mapping of addresses associated with protected health records to census tracts in a decentralized surveillance network. 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

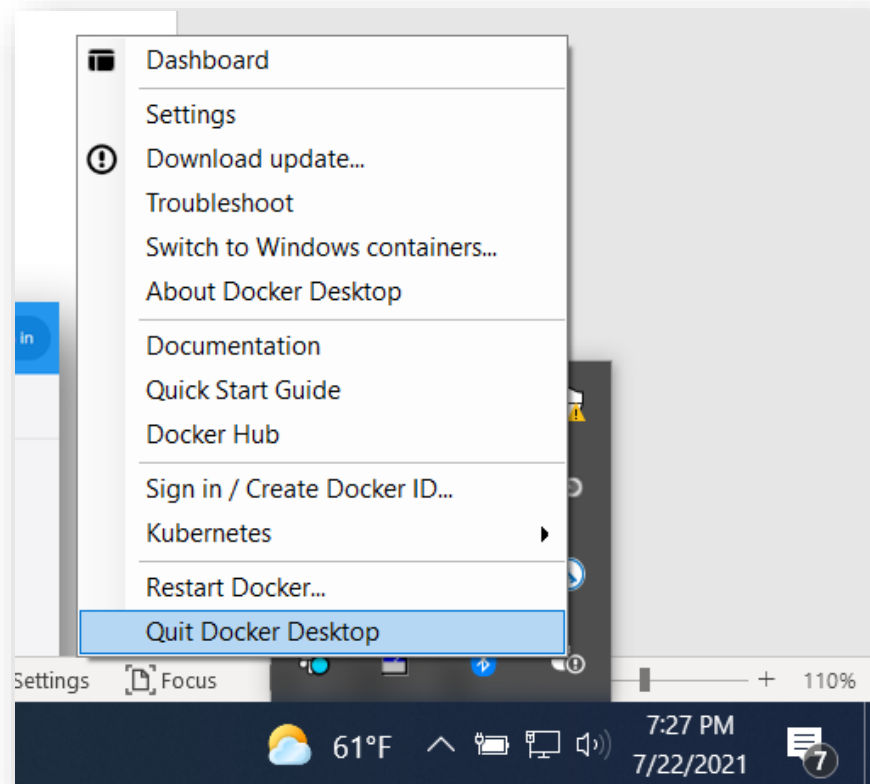
- **Docker desktop**
- **Internet browser** (required to open graphical user interface running on local machine)

## Installation

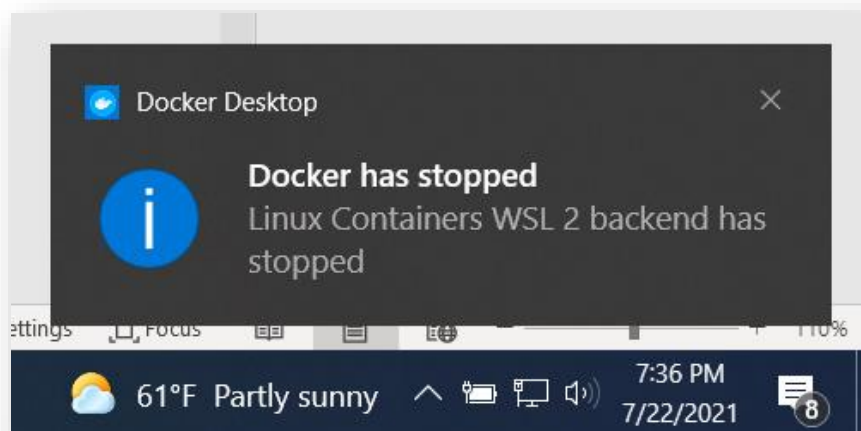
1. Install and start Docker desktop (detailed instructions for [Windows](#), [Mac](#))
2. ***Especially if you are running the pipeline on a secure computer that is not normally connected to the internet***, install the pipeline Docker Image by running `install.bat` (Windows) or `install.command` (Mac)
  - 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”

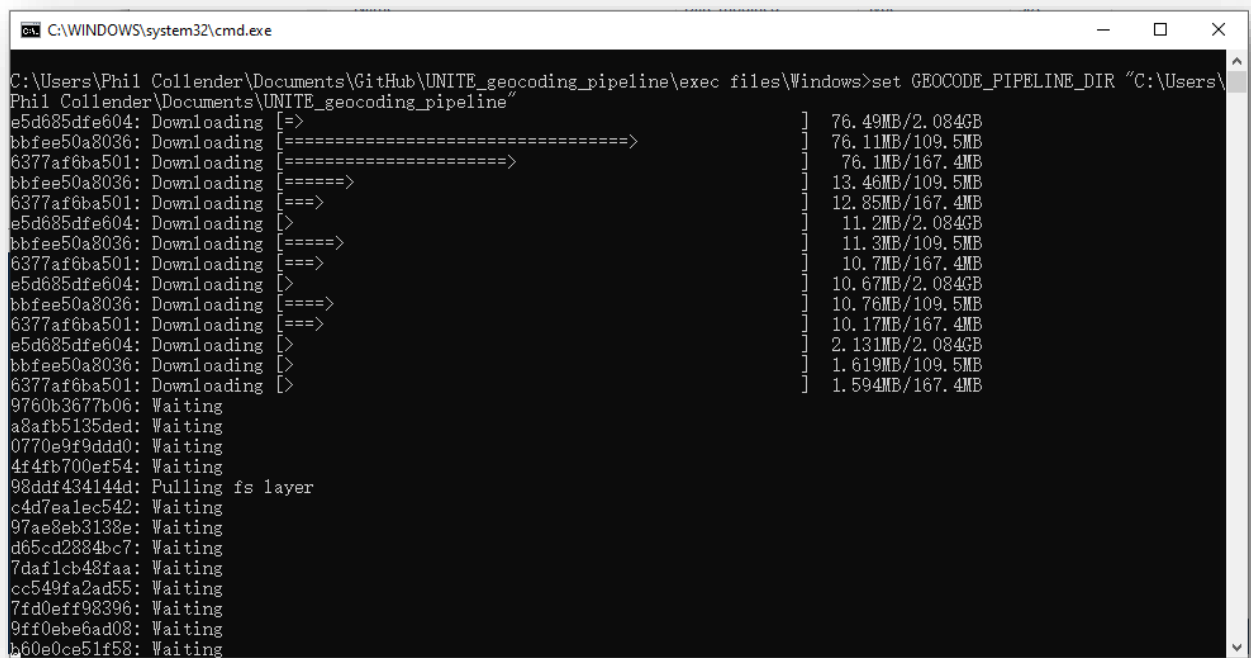


- iii. Wait for the notification that Docker Desktop has stopped running, then restart it from the start menu or search bar



## 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)

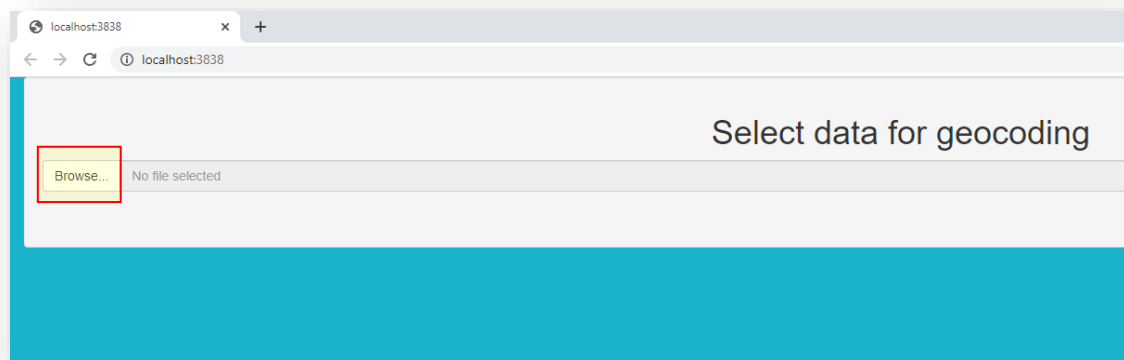


```
C:\WINDOWS\system32\cmd.exe
C:\Users\Phil Collender\Documents\GitHub\UNITE_geocoding_pipeline>set GEOCODE_PIPELINE_DIR "C:\Users\Phil Collender\Documents\GitHub\UNITE_geocoding_pipeline"
e5d685dfe604: Downloading [==>] 76.49MB/2.084GB
bbfee50a8036: Downloading [=====>] 76.11MB/109.5MB
6377af6ba501: Downloading [=====>] 76.1MB/167.4MB
bbfee50a8036: Downloading [=====>] 13.46MB/109.5MB
6377af6ba501: Downloading [====>] 12.85MB/167.4MB
e5d685dfe604: Downloading [====>] 11.2MB/2.084GB
bbfee50a8036: Downloading [=====>] 11.3MB/109.5MB
6377af6ba501: Downloading [====>] 10.7MB/167.4MB
e5d685dfe604: Downloading [====>] 10.67MB/2.084GB
bbfee50a8036: Downloading [=====>] 10.76MB/109.5MB
6377af6ba501: Downloading [====>] 10.17MB/167.4MB
e5d685dfe604: Downloading [====>] 2.131MB/2.084GB
bbfee50a8036: Downloading [====>] 1.619MB/109.5MB
6377af6ba501: Downloading [====>] 1.594MB/167.4MB
9760b3677b06: Waiting
a8afb5135ded: Waiting
0770e9f9ddd0: Waiting
4f4fb700ef54: Waiting
98ddf434144d: Pulling fs layer
c4d7ealec542: Waiting
97ae8eb3138e: Waiting
d65cd2884bc7: Waiting
7daf1cb48faa: Waiting
cc549fa2ad55: Waiting
7fd0eff98396: Waiting
9ff0ebe6ad08: Waiting
b60e0ce51f58: Waiting
```

*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](#)<sup>1</sup>, 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



---

<sup>1</sup> One thing that is different from stock DeGAUSS is that the column containing address data does not have to be named “**address**”

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

## Select data for geocoding

Upload complete

Click on column containing address data  
(REMINDER: column should be formatted as "{street address} {city} {state} {5 digit zip}")  
(e.g. "1234 Main st Fairfield MA 12345")

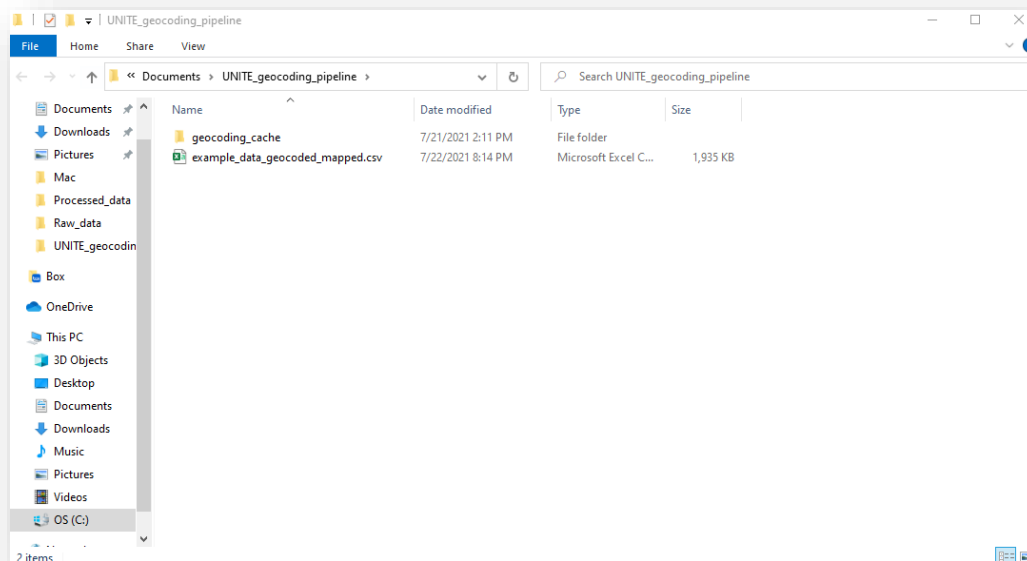
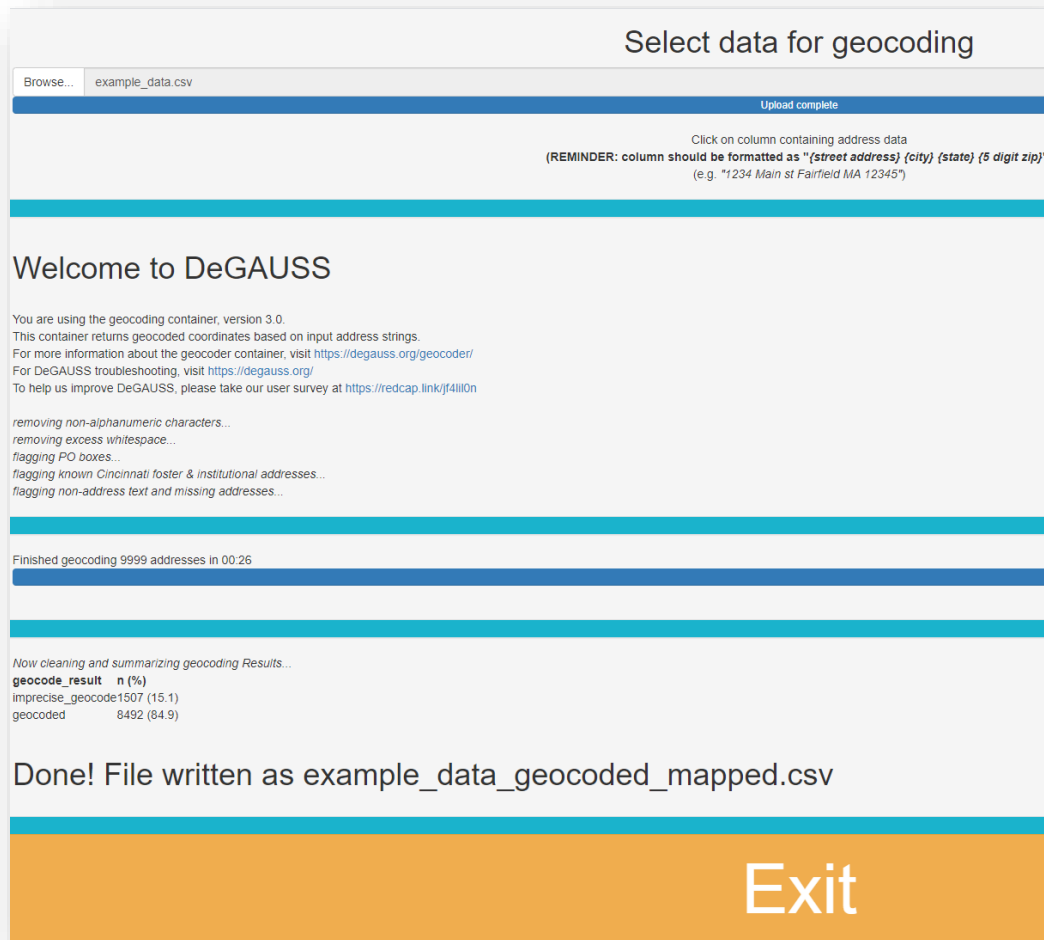
STATE	POSTCODE	Full.Address	NUMBER
CA	95660	30TH ST CA 95660	
CA	95823	46TH ST CA 95823	
CA	95820	7TH AVE CA 95820	
CA	95628	BANNISTER RD CA 95628	
CA	95624	BOND RD CA 95624	
CA	95608	BRANDON OAKS LN CA 95608	
CA	95641	BRANNAN ISLAND RD CA 95641	
CA	95819	E SANDBURG DR CA 95819	
CA	95829	ELDER CREEK RD CA 95829	
CA	94804	1010 Ohio Ave Richmond CA 94804	1010

Previous1

Selected column "**Full.Address**"

Confirm selection and proceed

- 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.
- 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*