The Columbia University Built Environment and Health GIS team is happy to share a python script [<https://github.com/nygeog/mappluto_to_fgdb/blob/master/download_all_mappluto_and_merge_into_file_geodatabase.py>] that downloads and compiles all of the current and archived New York City (NYC) Department of City Planning’s (DCP) MapPluto [<http://www.nyc.gov/html/dcp/html/bytes/applbyte.shtml#pluto>] versions into a single file geodatabase with feature datasets for each year-version. The purpose of the script was to save the BEH GIS group time and effort of downloading, unzipping and merging all this data by hand. It is also our hope that it may save person-time for anyone who wishes to compile all this data as well. It is based off an in-house urllib script for mining tract shapefiles by state from the US Census Bureau.

In July of 2013 NYC DCP announced that the most recent version of MapPluto (combination of DCP’s map geometry of tax block and lots and Primary Land Use Tax Lot Output from the NYC Department of Finance) would be free to the public, after nearly 10 years of licensing the data by borough for a fee [<http://www.socrata.com/blog/new-york-city-opens-pluto-tax-lot-dataset/>]. This was a huge leap forward and an important milestone in NYC's progress towards open data. While many data sets were free for public consumption in the past, MapPluto was often the most coveted [<http://spatialityblog.com/2013/04/04/a-modest-proposal-for-nyc-tax-parcel-data/>]. In the days and weeks that followed the release of the most recent MapPluto data set many questions regarding the opening of previous versions of MapPluto emerged. MapPluto data license holders wondered if earlier versions could be shared freely as well.

On Friday December 6th, it became apparent to the the NYC open data community that MapPluto data from 2003 to present day was now freely available to use and share [<http://spatialityblog.com/2013/12/06/when-it-rains-it-pours-nyc-gis-data-floodgates-opened/>].

The data is available by year and version as individual .zip files. Within these .zip files were folders that contained shapefiles for each borough. From 2003 to 2013 there are 16 versions of the data, consisting of 80 shapefiles. This data is fairly easy to download and start using for one borough for one year. However, the time and effort it would take to download and merge all this data manually is no small task. One could easily spend the better part of a day or two downloading, unzipping, merging these data sets into a file format that could hold them. The file size limitations of shapefiles cannot contain the whole data set of a version of merged boroughs. Hence, I decided, rather than immediately grab all the data in the normal point and click fashion and merge datasets using the traditional ArcGIS UI, to create a distributable script to share on GitHub, where a user need only define a working directly and the downloading, unzipping and merging of the data into a single file geodatabase (with feature datasets for each version) may all be done in an automated fashion with a single python script. This script may accessed here on GitHub: [<https://github.com/nygeog/mappluto_to_fgdb/blob/master/download_all_mappluto_and_merge_into_file_geodatabase.py>] or below [if wordpress allows for coding blocks]. The read me file may be accessed here with some more detailed instructions [<https://github.com/nygeog/mappluto_to_fgdb>]. So far users have reported script download and completion times anywhere from under 1 hour to 12 hours, depending on internet connection and processing bandwith.

Since this script is Esri-based, users of QGIS and other open-source GIS software will not be able to execute it as it utilizes the arcpy python module. The next step is to generate an accompanying script that allows the user to download and compile everything into an open source GIS format such as geoJSON.