Real World Python

for map makers

Objective

Create a database of Target store locations in the U.S.

Steps

1. Find Target store location info on the web
2. Develop a strategy
3. Get our python environment ready
4. Write python code to parse and store post office location data

Finding Target store location info

Target has a Store Locator. Promising!

But...

No map.
You cannot download a file.
And what's worse:
There are no coordinates attached to the locations, only addresses.

all locations

All Target Locations

Select a State*

We have a store near you.

Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
Florida
Georgia
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
North Dakota
Ohio
Oklahoma
Oregon
Pennsylvania
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Virginia
Washington
Washington DC
West Virginia
Wisconsin
Wyoming

*Currently, there are no Target stores in Vermont.

So we need to find ways to:

1) scrape the 'locations' from the web site
2) geocode these 'locations'.

Python is great for this!

Scraping

Sucking information from web sites that are not set up to give it to you the way you want it.

An astounding amount of (spatial) information is hidden away on the web that way.

Python has awesome tools to help you get to it.

Our python toolbelt

'Requests is an elegant and simple HTTP library for Python, built for human beings.'

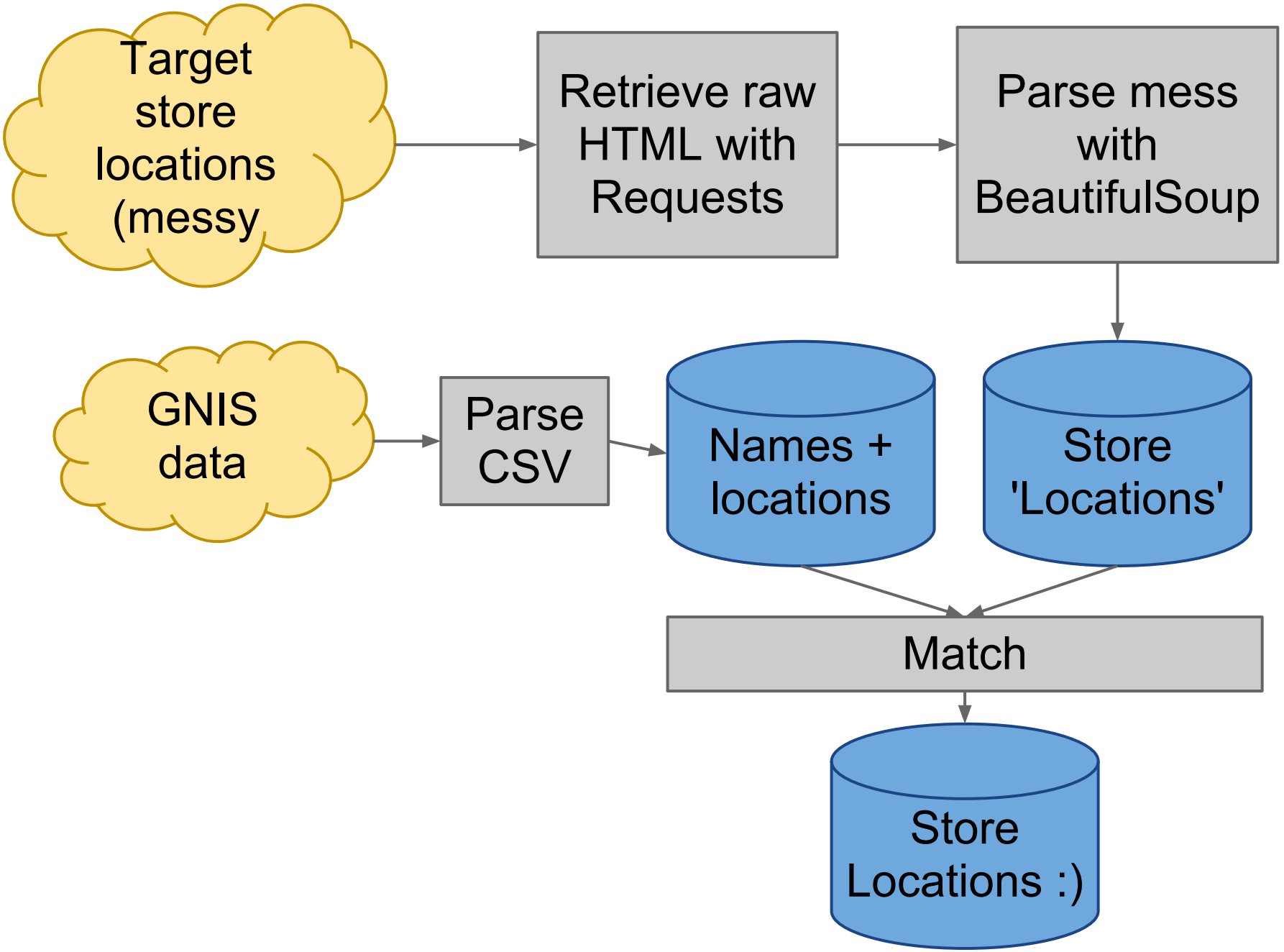
'You didn't write that awful page. You're just trying to get some data out of it. Beautiful Soup is here to help'

Geocoding

Two strategies

1) Use an online geocoding service (through geopy for example)
2) Match with USGS GNIS data

Option 2 is better because the data is free (as in beer and as in speech).


Target store locations (messy
Retrieve raw HTML with Requests
Parse mess with BeautifulSoup
GNIS data
Parse CSV
Names + locations
Store 'Locations'
Match
Store Locations :)

Python Environment

First, we will create and activate a python **virtual environment**.

This is a fully self-contained python environment that we can use as a sandbox without messing up our systemwide python environment (which may be used for other, more important tasks)

```
C:\Users\mvexel>virtualenv class
New python executable in class\Scripts\python.exe
Installing setuptools.................done.
Installing pip...................done.

C:\Users\mvexel>class\Scripts\activate.bat
(class) C:\Users\mvexel>
```

Python environment

Next, we will install the modules we will use. This is made easy by the Python Package Index (PyPI) and the pip tool.

You'll find the package names in their installation instructions.

```
(class) C:\Users\mvexel>pip install beautifulsoup4
Downloading/unpacking beautifulsoup4
  Downloading beautifulsoup4-4.1.3.tar.gz (131Kb): 131Kb downloaded
  Running setup.py egg_info for package beautifulsoup4

Installing collected packages: beautifulsoup4
  Running setup.py install for beautifulsoup4

Successfully installed beautifulsoup4
Cleaning up...
```

Write Python code

1. import modules
2. load GNIS populated places
3. for each state:
    a. Get web page
    b. Find the relevant elements on the page (place names)
    c. for each placename found:
        i. Match to GNIS place names
        ii. add (lon,lat) from matched GNIS place to record
        iii. write to file

What can you do with this?

The resulting file can be loaded directly into ArcMap (Display XY data...)

Instead of CSV, we could also have written the output as GeoJSON, KML or Shapefile for further GIS processing.

And even make an interactive web map using Leaflet

Try it yourself

Rite-Aid Locations
Chase locations
CVS locations
Tea Party groups locations
Five Guys locations

.....more? just google