Start time: 5:30 p.m.

Writing a python script to read the shape file and assign unique numerical indices for every assembly.

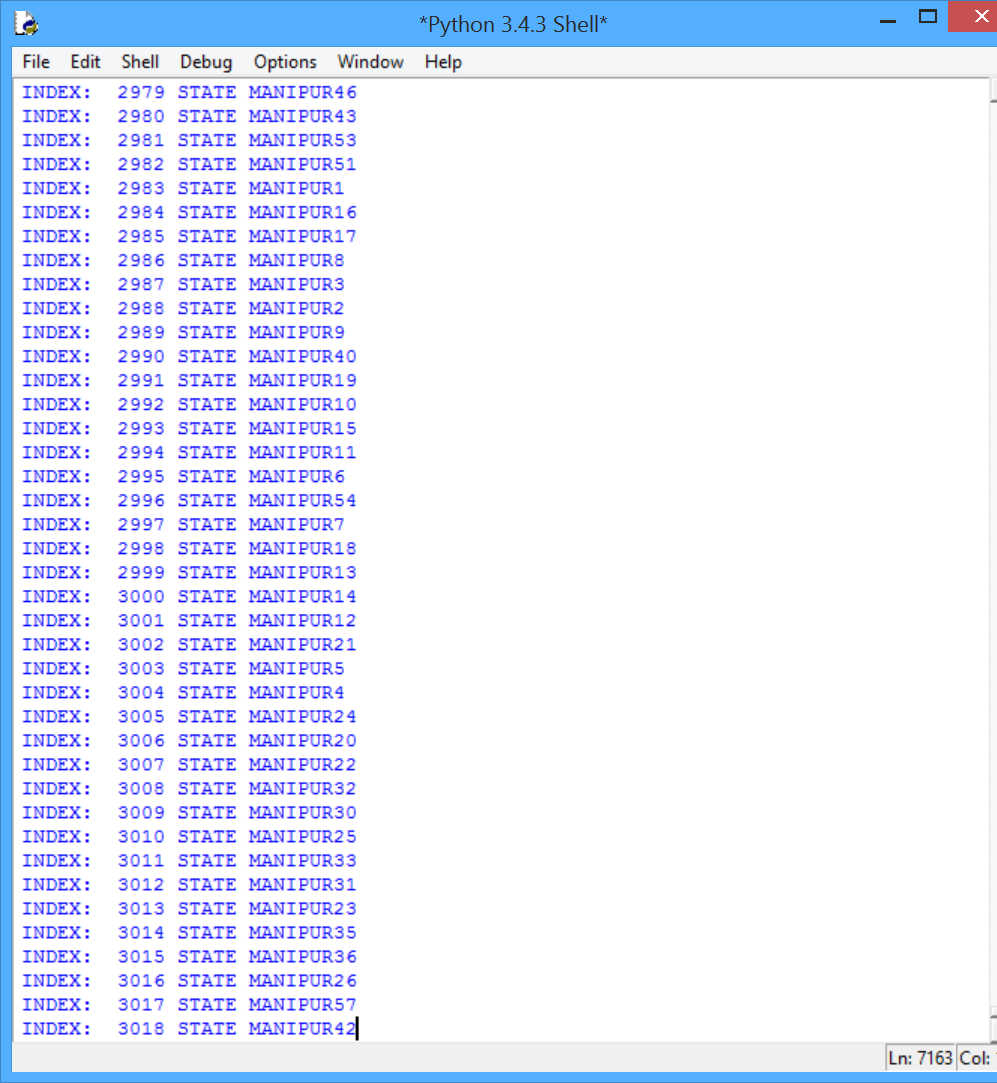
Unique ID (Index), State:Constituency

Writing a python script to reorganize the data into the desired format. No need for headers in the reorganized data, since it will correspond with the ordering of states in the shapefile records, although it might be convenient to have an external list of unique ID and state/constituency in case it’s ever needed.

One slightly inconvenient issue is that in the CSV, the names are formatted differently from the shapefile records, e.g. State – AC\_ID in the csv as opposed to StateAC\_ID in the shapefile. Also, AC\_ID 0 is not reserved for the state data, which is present in the CSV. Instead, state data uses State – NA in the CSV. Not necessarily a huge issue since I don’t think the shapefiles include state boundaries. In the formatted data, the state data will probably be left out entirely.

Conveniently, there is a constituency true/false (0/1) column. I’ll use this to filter out state averaged data.

Current output:



Current code:

1. **import** shapefile
2. sf = shapefile.Reader("C:/Users/Young/Documents/Github/Research-2015-Energy-Data/data/AC\_India\_Geo/AC\_ALLINDIA")
3. **print**("Record format:", sf.fields)
4. unique\_id = 0;
5. **for** r **in** sf.iterRecords():
6. **print**("INDEX: ", unique\_id, "STATE",  r[4]);
7. unique\_id += 1

Next up, I’m going to add a delimiter between “state” and “constituency#” in the output. Having a delimiter will be important, especially if India ever starts using numbers in state names.

Break: 6:50 p.m.

Resume: 8:00 p.m.

Using regexps to parse into ‘state’ and ‘id’

Had to add ‘ ‘ and ‘&’ to regexp in addition to all alpha characters, as some state names have spaces and ampersands.

On a side note, Uttarakhand is lower case while everything else is capitalized. Intentional?

Adding file writing to turn the prints into a csv file.

Finish: 9:00 p.m.

Total: 2.3 hours