Assignment

June 27, 2016

1 Reading and Writing Data

In this assignment you will be reading and writing data. In this folder are 3 included data files ending in `csv`, `json` and `pkl`.

- data.csv
- data.json
- data.pkl

These are different file formats that exist. You can run the following command(s) to see what is in each file:

```
head data.csv # or pkl # or json
```

You'll see that there is some method to the madness but that each file has its peculiarities. Each file contains a portion of the total dataset that consists of 100 records, so you will need to read in all of the files and combine them into some standard format with which you are comfortable. Aim for something standard where each "row" is the same format.

After you've standardized all of the data, report the following bits of information by writing them to a csv file labelled `question_1.csv`, `question_2.csv` etc. In addition, show all your work in an iPython notebook.

1. What are the unique countries in the dataset?
2. What are the unique email domains in the dataset?
3. What are the first names of everyone that does not have a P.O. Box address?
4. What are the names of the first 5 people when you sort the data by Country?
5. What are the names of the first 5 people when you sort the data by phone number?

1.0.1 Restrictions

You should use these standard library imports

```
import json
import csv
import pickle
```

Some of you may be familiar with a Python package called `pandas` which would greatly speed up this sort of file processing. The point of this homework is to do the work manually. You can use `pandas` to independently check your work if you are so inclined. Don't worry if you are not familiar with `pandas`. We will do this homework as a class exercise using `pandas` in the near future.

1.0.2 Comments

- You may use regular expressions if you wish to extract data from each row. You do not need to use them if you do not want to or see a need to. The Python regular expression module is called `re`.
- You may want to use the `operator` module to help in sorting.
- There are many data structures and formats that you might use to solve this problem. You will have to decide if you want to keep the information for each person together as one record or all the information for each of the fields together.

In [ ]: