MSiA 490 – Winter 2017

Project # 1

DUE: 2/13/2017

Exercise 1 – BABY NAMES BY STATE and YEAR

Write a python code to do the following:

1. Download and compile **baby names by state** dataset.
2. Create a library **class** called **BabyNames:** a The class should offer the following interfaces:
    a. **Constructor** where you pass the file location to create Pandas DataFrame.
    b. **Count** (state='', year=''): returns the total number of births. A blank state or year should return all births of the empty input.
    c. **Top10BabyNames**(state='IL', year=2015): your output should look like(empty state means all, same for year):

Rank	Male	Female
1	Noah	Emma
…	…	…

    d. **ChangeOfPopularity** (fromYear=2014, toYear=2015, top=10): This function should lists of baby names (male or female) that showed change in popularity as follows:
        i. **Names that increased in popularity**
        ii. **Names that decreased in popularity**
        iii. **Names having the popularity**

e. **Top5NamesPerYear**(year=2015,sex=''): Returns a table that show the five most frequent given names, by State, for male, female, or both in a given year. The number to the right of each name is the number of occurrences in the data. (see Figure for format)

Top Five Female Names for Births in 2015										
State	**Rank 1**	**Num**	**Rank 2**	**Num**	**Rank 3**	**Num**	**Rank 4**	**Num**	**Rank 5**	**Num**
Alabama	Ava	297	Emma	285	Olivia	258	Harper	213	Elizabeth	186
Alaska	Olivia	56	Emma	49	Aurora	46	Amelia	39	Ava	39

f. **NamePopularityPlot** (name='Jim', yearRange=(2000,2015), state='IL', sex='M'): This function will create a plot file that shows the name popularity changes over the year. (popularity is based on the proportional use of the name within a state and year)

g. **NameFlip**(n=10): List top n names that filliped over the years. (i.e. from boy name to girl or the reverse). Provide a plot of the names showing the year.

3. **Make sure to document your class and follow the python standards.**
4. **Extra Credit:** Tell a story that was not told in the class (of course from Baby Names dataset). Support your story with plots.

Exercise 2 – Pandas

Using **MySQL** and the sample database "**WORLD**", your tasks are:

1. Translate the following SQL commands into Pandas library commands.
2. Make sure to compare results
3. Provide a description of what each SQL statement intends to do.

1.

```
select *
from country
where population > 50000000
order by population DESC limit 10
```

2.

```
select Continent, count(*) As Number_Countries, sum(population) As Population
from country
where population > 0
group by Continent
order by 1 ASC
```

3.

```
select city.Name As City, city.population
from city
inner join country ON city.CountryCode = country.code
where country.code = 'USA'
order by city.population DESC limit 10
```

4.

```
select country.Name, Language, (Percentage * population) / 100
from countrylanguage
inner join country on countrylanguage.CountryCode = country.code
where IsOfficial = True
order by 3 DESC limit 10
```

5.

```
select Language, sum((Percentage * population) / 100)
from countrylanguage
inner join country ON countrylanguage.CountryCode = country.code
group by Language
order by 2 desc limit 5
```