In [1]:
```
# Importing necessary modules to clean and merge data
import numpy as np
import pandas as pd
```

In [2]:
```
crime2015 = pd.read_excel('crime2015.xls', header=0)
```

In [28]:
```
#fillna with the value of the row before it
crime2015.fillna(method = 'ffill', inplace = True)
crime2015.head()
```

Out[28]:

er and gligent ughter	Rape (revised definition)2	Rape (legacy definition)3	Robbery	Aggravated assault	Property crime	Burglary	Larceny-theft	Moto vehicl thef
	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	1496.0	1080.0	4080	12263	113447	27208	78141.0	8098.0
	1537.0	1111.0	4138	12538	115920	27852	79799.0	8269.0
	1537.0	1111.0	4138	12538	115920	27852	79799.0	8269.0
	311.0	218.0	354	2481	20873	4438	15106.0	1329.0

In [21]:
```python
#Delete all rows except for the ones that contain state total
crime2015[crime2015['Area'] == 'State Total']
crime2015.reset_index(drop = True)
```

Out[21]:

	State	Area	Unnamed: 2	Population	Violent crime1	Murder and nonnegligent manslaughter	(re definir
0	ALABAMA	Metropolitan Statistical Area	NaN	3708033	NaN	NaN	NaN
1	ALABAMA	Metropolitan Statistical Area	Area actually reporting	0.971	18122	283	1496.
2	ALABAMA	Metropolitan Statistical Area	Estimated total	1	18500	287	1537.
3	ALABAMA	Cities outside metropolitan areas	Estimated total	522241	18500	287	1537.
4	ALABAMA	Cities outside metropolitan areas	Area actually reporting	0.974	3178	32	311.0
5	ALABAMA	Cities outside metropolitan areas	Estimated total	1	3240	33	315.0
6	ALABAMA	Nonmetropolitan counties	Estimated total	628705	3240	33	315.0
7	ALABAMA	Nonmetropolitan counties	Area actually reporting	0.994	1205	28	186.0
8	ALABAMA	Nonmetropolitan counties	Estimated total	1	1212	28	187.0
9	ALABAMA	State Total	Estimated total	4858979	22952	348	2039.
10	ALABAMA	State Total	Rate per 100,000 inhabitants	4858979	472.4	7.2	42.0
11	ALASKA	Metropolitan Statistical Area	Rate per 100,000 inhabitants	351857	472.4	7.2	42.0
12	ALASKA	Metropolitan Statistical Area	Area actually reporting	1	3464	27	541.0
13	ALASKA	Cities outside metropolitan areas	Area actually reporting	127499	3464	27	541.0

	State	Area	Unnamed: 2	Population	Violent crime1	Murder and nonnegligent manslaughter	(re defini
14	ALASKA	Cities outside metropolitan areas	Area actually reporting	0.972	886	10	178.0
15	ALASKA	Cities outside metropolitan areas	Estimated total	1	913	10	184.0
16	ALASKA	Nonmetropolitan counties	Estimated total	259076	913	10	184.0
17	ALASKA	Nonmetropolitan counties	Area actually reporting	1	1015	22	176.0
18	ALASKA	State Total	Area actually reporting	738432	5392	59	901.0
19	ALASKA	State Total	Rate per 100,000 inhabitants	738432	730.2	8	122.0
20	ARIZONA	Metropolitan Statistical Area	Rate per 100,000 inhabitants	6479810	730.2	8	122.0
21	ARIZONA	Metropolitan Statistical Area	Area actually reporting	0.997	24725	280	2805.
22	ARIZONA	Metropolitan Statistical Area	Estimated total	1	24781	280	2811.
23	ARIZONA	Cities outside metropolitan areas		124240	24781		2811.
24	ARIZONA	Cities outside metropolitan areas	Area actually reporting	0.939	2678	24	269.0
25	ARIZONA	Cities outside metropolitan areas	Estimated total	1	2851	26	286.0
26	ARIZONA	Nonmetropolitan counties		224015	2851	26	286.0
27	ARIZONA	Nonmetropolitan counties	Area actually reporting	1	380	3	11.0

	State	Area	Unnamed: 2	Population	Violent crime1	Murder and nonnegligent manslaughter	(re definit
28	ARIZONA	State Total	Area actually reporting	6828065	28012	309	3108.
29	ARIZONA	State Total	Rate per 100,000 inhabitants	6828065	410.2	4.5	45.5
...	...	...	...	...	...	...	...
472	WEST VIRGINIA	Metropolitan Statistical Area	Area actually reporting	0.898	3557	45	402.0
473	WEST VIRGINIA	Metropolitan Statistical Area	Estimated total	1	3887	47	433.0
474	WEST VIRGINIA	Cities outside metropolitan areas	Estimated total	189216	3887	47	433.0
475	WEST VIRGINIA	Cities outside metropolitan areas	Area actually reporting	0.53	280	4	23.0
476	WEST VIRGINIA	Cities outside metropolitan areas	Estimated total	1	519	8	33.0
477	WEST VIRGINIA	Nonmetropolitan counties	Estimated total	518050	519	8	33.0
478	WEST VIRGINIA	Nonmetropolitan counties	Area actually reporting	0.731	1325	11	142.0
479	WEST VIRGINIA	Nonmetropolitan counties	Estimated total	1	1825	15	206.0
480	WEST VIRGINIA	State Total	Estimated total	1844128	6231	70	672.0
481	WEST VIRGINIA	State Total	Rate per 100,000 inhabitants	1844128	337.9	3.8	36.4
482	WISCONSIN6	Metropolitan Statistical Area	Rate per 100,000 inhabitants	4268682	337.9	3.8	36.4

	State	Area	Unnamed: 2	Population	Violent crime1	Murder and nonnegligent manslaughter	(re definii
483	WISCONSIN7	Metropolitan Statistical Area	Area actually reporting	0.965	15597	207	1368.
484	WISCONSIN8	Metropolitan Statistical Area	Estimated total	1	15770	209	1396.
485	WISCONSIN9	Cities outside metropolitan areas	Estimated total	639613	15770	209	1396.
486	WISCONSIN10	Cities outside metropolitan areas	Area actually reporting	0.938	1015	9	189.0
487	WISCONSIN11	Cities outside metropolitan areas	Estimated total	1	1082	10	201.0
488	WISCONSIN12	Nonmetropolitan counties	Estimated total	863042	1082	10	201.0
489	WISCONSIN13	Nonmetropolitan counties	Area actually reporting	1	795	21	183.0
490	WISCONSIN14	State Total	Area actually reporting	5771337	17647	240	1780.
491	WISCONSIN15	State Total	Rate per 100,000 inhabitants	5771337	305.8	4.2	30.8
492	WYOMING	Metropolitan Statistical Area	Rate per 100,000 inhabitants	179738		4.2	30.8
493	WYOMING	Metropolitan Statistical Area	Area actually reporting	1	366	8	50.0
494	WYOMING	Cities outside metropolitan areas	Area actually reporting	245549	366	8	50.0
495	WYOMING	Cities outside metropolitan areas	Area actually reporting	0.967	629	5	82.0

	State	Area	Unnamed: 2	Population	Violent crime1	Murder and nonnegligent manslaughter	(re defini
496	WYOMING	Cities outside metropolitan areas	Estimated total	1	649	5	83.0
497	WYOMING	Nonmetropolitan counties	Estimated total	160820	649		83.0
498	WYOMING	Nonmetropolitan counties	Area actually reporting	0.916	262	3	35.0
499	WYOMING	Nonmetropolitan counties	Estimated total	1	287	3	40.0
500	WYOMING	State Total	Estimated total	586107	1302	16	173.0
501	WYOMING	State Total	Rate per 100,000 inhabitants	586107	222.1	2.7	29.5

502 rows × 14 columns

```
In [8]:  #import data from other years:
         crime2014 = pd.read_excel('crime2014.xls', header=0)
         crime2013 = pd.read_excel('crime2013.xls', header=0)
         crime2012 = pd.read_excel('crime2012.xls', header=0)
         crime2011 = pd.read_excel('crime2011.xls', header=0)
         crime2010 = pd.read_excel('crime2010.xls', header=0)
```

Automate cleaning for each excel data set:

```
In [22]:  def clean(df):
              #fill na values with the value of the row above
              df.fillna(method = 'ffill', inplace = True)
              #select the rows where the column 'Area' = 'State Total'
              df = df[df['Area'] == 'State Total']
              #return cleaned table with index reset to 0 start
              return df.reset_index(drop=True).head()
```

In [23]: `clean(crime2010)`

Out[23]:

	State	Area	Unnamed: 2	Population	Violent crime	Murder and nonnegligent manslaughter	Forcible rape	Robbery	Agg
0	ALABAMA	State Total	Estimated total	4779736	18056	273	1349	4761	116
1	ALABAMA	State Total	Rate per 100,000 inhabitants	4779736	377.8	5.7	28.2	99.6	244
2	ALASKA	State Total	Area actually reporting	710231	4537	31	533	594	337
3	ALASKA	State Total	Rate per 100,000 inhabitants	710231	638.8	4.4	75	83.6	475
4	ARIZONA	State Total	Estimated total	6392017	26085	409	2165	6937	165

In [24]: `clean(crime2011)`

Out[24]:

	State	Area	Unnamed: 2	Population	Violent crime	Murder and nonnegligent manslaughter	Forcible rape	Robbery	Ag
0	ALABAMA1	State Total	Estimated total	4802740	20174	301	1371	4910	13
1	ALABAMA1	State Total	Rate per 100,000 inhabitants	4802740	420.1	6.3	28.5	102.2	28
2	ALASKA	State Total	Area actually reporting	722718	4383	29	420	576	33
3	ALASKA	State Total	Rate per 100,000 inhabitants	722718	606.5	4	58.1	79.7	46
4	ARIZONA	State Total	Estimated total	6482505	26311	405	2264	7127	16

In [25]: `clean(crime2012)`

Out[25]:

	State	Area	Unnamed: 2	Population	Violent crime	Murder and nonnegligent manslaughter	Forcible rape	Robbery	Agg
0	ALABAMA	State Total	Estimated total	4822023	21693	342	1296	5020	150
1	ALABAMA	State Total	Rate per 100,000 inhabitants	4822023	449.9	7.1	26.9	104.1	311
2	ALASKA	State Total	Area actually reporting	731449	4412	30	583	630	316
3	ALASKA	State Total	Rate per 100,000 inhabitants	731449	603.2	4.1	79.7	86.1	433
4	ARIZONA	State Total	Area actually reporting	6553255	28108	358	2277	7386	180

In [26]: `clean(crime2013)`

Out[26]:

	State	Area	Unnamed: 2	Population	Violent crime1	Murder and nonnegligent manslaughter	Rape (revised definition)2	Ra (lega definitio
0	ALABAMA4	State Total	Estimated total	4833722	20826	347	2044.0	1428.0
1	ALABAMA4	State Total	Rate per 100,000 inhabitants	4833722	430.8	7.2	42.3	29.5
2	ALASKA4	State Total	Area actually reporting	735132	4708	34	922.0	644.0
3	ALASKA4	State Total	Rate per 100,000 inhabitants	735132	640.4	4.6	125.4	87.6
4	ARIZONA5,6	State Total	Area actually reporting	6626624	27599	357	3050.0	2343.0

In [27]: `clean(crime2014)`

Out[27]:

	State	Area	Unnamed: 2	Population	Violent crime1	Murder and nonnegligent manslaughter	Rape (revised definition)2	Rape (legacy definition
0	ALABAMA4	State Total	Estimated total	4849377	20727	276	2005.0	1436.0
1	ALABAMA4	State Total	Rate per 100,000 inhabitants	4849377	427.4	5.7	41.3	29.6
2	ALASKA4	State Total	Area actually reporting	736732	4684	41	771.0	555.0
3	ALASKA4	State Total	Rate per 100,000 inhabitants	736732	635.8	5.6	104.7	75.3
4	ARIZONA4	State Total	Area actually reporting	6731484	26916	319	3378.0	2464.0

In [ ]: