Milestone 3

Crime Statistics

```
Brandon Sams
```

03May2020

Scraped from https://www.ucrdatatool.gov/Search/Crime/State/RunCrimeStatebyState.cfm

```
In [1]: import bs4 as bs
          import pandas as pd
          import codecs
          import numpy as np
 In [2]: pd.set_option('display.max_columns', None)
 In [3]: f_hand = 'Uniform Crime Reporting Statistics.html'
          f=codecs.open(f_hand, 'r')
          soup = bs.BeautifulSoup(f)
 In [4]: table = soup.find all("table")[4]
 In [5]: df = pd.DataFrame(columns=range(0,22),index = range(60))
 In [6]: row_marker = 0
          for row in table.find_all('tr'):
              column marker = 0
              columns = row.find all('td')
              for column in columns:
                  df.iat[row_marker,column_marker] = column.get_text()
                  column marker += 1
             row marker +=1
 In [7]: # Rename Columns
          cols = [
              'Year',
              'Population',
              'Violent Crime Total',
              'Murder and Nonnegligent Manslaughter',
              'Legacy Rape',
              'Revised Rape',
              'Robbery',
              'Aggravated Assault',
              'Property Crime Total',
              'Burglary',
              'Larceny-Theft',
              'Motor Vehicle Theft',
              'Violent Crime Rate',
              'Murder and Nonnegligent Manslaughter Rate',
              'Legacy Rape Rate',
              'Revised Rape Rate',
              'Robbery Rate',
              'Aggravated Assault Rate',
              'Property Crime Rate',
              'Burglary Rate',
              'Larceny-Theft Rate',
              'Motor Vehicle Theft Rate'
 In [8]: df.columns = cols
 In [9]: # Remove junk rows
          df.drop([0,1,2,3,59],inplace = True)
In [10]: # Index by year
         df.set_index('Year',inplace = True)
In [11]: # Fill Revised Rape columns with NaN
         df = df.replace('',np.NaN)
In [12]: # Convert all columns to numeric type
          for col in df:
             df[col] = df[col].str.replace(',','').replace(' ','')
In [13]: df = df.apply(pd.to_numeric)
In [14]: df.dtypes
Out[14]: Population
                                                          int64
         Violent Crime Total
                                                          int64
         Murder and Nonnegligent Manslaughter
                                                          int64
         Legacy Rape
                                                          int64
         Revised Rape
                                                       float64
         Robbery
                                                          int64
                                                          int64
         Aggravated Assault
         Property Crime Total
                                                          int64
                                                          int64
         Burglary
         Larceny-Theft
                                                         int64
         Motor Vehicle Theft
                                                         int64
                                                       float64
         Violent Crime Rate
         Murder and Nonnegligent Manslaughter Rate
                                                       float64
                                                       float64
         Legacy Rape Rate
                                                       float64
         Revised Rape Rate
         Robbery Rate
                                                       float64
         Aggravated Assault Rate
                                                       float64
                                                       float64
         Property Crime Rate
         Burglary Rate
                                                       float64
                                                       float64
         Larceny-Theft Rate
         Motor Vehicle Theft Rate
                                                       float64
         dtype: object
In [15]: df
Out[15]:
                                                                                                               Murder and
                                                                                                                         Legacy Re
```

NaN

NaN

NaN

354772

355051

345095

325802

752423

762009

726575

741291

9052743 2185140

9001992 2109932

8277829 1729806

1931835

8650761

6151095

6168874

6018632

5858496

716508

723186

700294

689527

387.1

387.8

379.1

375.7

4.7

4.7

4.5

4.5

27.0

27.1

25.9

26.4

14661

14866

14319

14249

84175

85141

82109 113695.0

84041 116645.0

In []: In []: **2011** 311587816 1206031

2012 313873685 1217067

2013 316497531 1199684

2014 318857056 1197987