USArrests

September 7, 2023

1 USArrests Data Analysis

Time period: 1973

Murder, Assault, and Rape arrests per 100,000 people for each state

Tasks performed:

```
* Loaded pydataset: USArrests
* Changed index to a column: State
* Re-indexed dataframe
* Created categorization field for State
* Created summation variable: TotalArrests
* Analyzed data with graphs by region
* Looked at correlation
* Analyzed data with graphs by state
* Summary
```

```python
[1433]: import pandas as pd
        import numpy as np
        from matplotlib import pyplot as plt
        import seaborn as sns
        %matplotlib inline

        plt.style.use('ggplot')
```

```python
[1434]: # Load USArrests dataset
        from pydataset import data

        # Murder, Assault, Rape arrests per 100,000 people
        # Percentage of population residing in an urban area
        # Display first 20 rows
        df = data('USArrests')
        df
```

```
[1434]:          Murder  Assault  UrbanPop  Rape
        Alabama    13.2      236        58  21.2
        Alaska     10.0      263        48  44.5
        Arizona     8.1      294        80  31.0
```

Arkansas	8.8	190	50	19.5
California	9.0	276	91	40.6
Colorado	7.9	204	78	38.7
Connecticut	3.3	110	77	11.1
Delaware	5.9	238	72	15.8
Florida	15.4	335	80	31.9
Georgia	17.4	211	60	25.8
Hawaii	5.3	46	83	20.2
Idaho	2.6	120	54	14.2
Illinois	10.4	249	83	24.0
Indiana	7.2	113	65	21.0
Iowa	2.2	56	57	11.3
Kansas	6.0	115	66	18.0
Kentucky	9.7	109	52	16.3
Louisiana	15.4	249	66	22.2
Maine	2.1	83	51	7.8
Maryland	11.3	300	67	27.8
Massachusetts	4.4	149	85	16.3
Michigan	12.1	255	74	35.1
Minnesota	2.7	72	66	14.9
Mississippi	16.1	259	44	17.1
Missouri	9.0	178	70	28.2
Montana	6.0	109	53	16.4
Nebraska	4.3	102	62	16.5
Nevada	12.2	252	81	46.0
New Hampshire	2.1	57	56	9.5
New Jersey	7.4	159	89	18.8
New Mexico	11.4	285	70	32.1
New York	11.1	254	86	26.1
North Carolina	13.0	337	45	16.1
North Dakota	0.8	45	44	7.3
Ohio	7.3	120	75	21.4
Oklahoma	6.6	151	68	20.0
Oregon	4.9	159	67	29.3
Pennsylvania	6.3	106	72	14.9
Rhode Island	3.4	174	87	8.3
South Carolina	14.4	279	48	22.5
South Dakota	3.8	86	45	12.8
Tennessee	13.2	188	59	26.9
Texas	12.7	201	80	25.5
Utah	3.2	120	80	22.9
Vermont	2.2	48	32	11.2
Virginia	8.5	156	63	20.7
Washington	4.0	145	73	26.2
West Virginia	5.7	81	39	9.3
Wisconsin	2.6	53	66	10.8
Wyoming	6.8	161	60	15.6

1.1 Data Cleaning

```
[1436]:  # dataframe shape
         df.shape
```

```
[1436]:  (50, 4)
```

```
[1437]:  # dataframe columns, data type, # non-null values
         df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 50 entries, Alabama to Wyoming
Data columns (total 4 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Murder    50 non-null     float64
 1   Assault   50 non-null     int64
 2   UrbanPop  50 non-null     int64
 3   Rape      50 non-null     float64
dtypes: float64(2), int64(2)
memory usage: 2.0+ KB
```

```
[1438]:  # dataframe describe.  Basic statistics
         df.describe()
```

```
[1438]:            Murder      Assault    UrbanPop        Rape
         count  50.00000   50.000000   50.000000   50.000000
         mean    7.78800  170.760000   65.540000   21.232000
         std     4.35551   83.337661   14.474763    9.366385
         min     0.80000   45.000000   32.000000    7.300000
         25%     4.07500  109.000000   54.500000   15.075000
         50%     7.25000  159.000000   66.000000   20.100000
         75%    11.25000  249.000000   77.750000   26.175000
         max    17.40000  337.000000   91.000000   46.000000
```

```
[1439]:  # Create column, State
         df['State'] = df.index
         df.head(10)
```

```
[1439]:              Murder  Assault  UrbanPop  Rape         State
         Alabama       13.2      236        58  21.2      Alabama
         Alaska        10.0      263        48  44.5       Alaska
         Arizona        8.1      294        80  31.0      Arizona
         Arkansas       8.8      190        50  19.5     Arkansas
         California     9.0      276        91  40.6   California
         Colorado       7.9      204        78  38.7     Colorado
         Connecticut    3.3      110        77  11.1  Connecticut
         Delaware       5.9      238        72  15.8     Delaware
```

3

```
Florida          15.4        335        80  31.9        Florida
Georgia          17.4        211        60  25.8        Georgia
```

[1440]: `# reset index`
```
df=df.reset_index()
df.head(10)
```

[1440]:
```
         index  Murder  Assault  UrbanPop  Rape        State
0      Alabama    13.2      236        58  21.2      Alabama
1       Alaska    10.0      263        48  44.5       Alaska
2      Arizona     8.1      294        80  31.0      Arizona
3     Arkansas     8.8      190        50  19.5     Arkansas
4   California     9.0      276        91  40.6   California
5     Colorado     7.9      204        78  38.7     Colorado
6  Connecticut     3.3      110        77  11.1  Connecticut
7     Delaware     5.9      238        72  15.8     Delaware
8      Florida    15.4      335        80  31.9      Florida
9      Georgia    17.4      211        60  25.8      Georgia
```

[1441]: `# Drop columm Index`
```
df = df.drop(['index'], axis = 1)
df.head(10)
```

[1441]:
```
   Murder  Assault  UrbanPop  Rape        State
0    13.2      236        58  21.2      Alabama
1    10.0      263        48  44.5       Alaska
2     8.1      294        80  31.0      Arizona
3     8.8      190        50  19.5     Arkansas
4     9.0      276        91  40.6   California
5     7.9      204        78  38.7     Colorado
6     3.3      110        77  11.1  Connecticut
7     5.9      238        72  15.8     Delaware
8    15.4      335        80  31.9      Florida
9    17.4      211        60  25.8      Georgia
```

[1442]: `# Check data types`
```
df.dtypes
```

[1442]:
```
Murder      float64
Assault       int64
UrbanPop      int64
Rape        float64
State        object
dtype: object
```

[1443]: `# Create new column that sums all the arrest categories`
```
df['TotalArrests'] = df['Murder'] + df['Rape'] + df['Assault']
```

```
df.head(10)
```

[1443]:

	Murder	Assault	UrbanPop	Rape	State	TotalArrests
0	13.2	236	58	21.2	Alabama	270.4
1	10.0	263	48	44.5	Alaska	317.5
2	8.1	294	80	31.0	Arizona	333.1
3	8.8	190	50	19.5	Arkansas	218.3
4	9.0	276	91	40.6	California	325.6
5	7.9	204	78	38.7	Colorado	250.6
6	3.3	110	77	11.1	Connecticut	124.4
7	5.9	238	72	15.8	Delaware	259.7
8	15.4	335	80	31.9	Florida	382.3
9	17.4	211	60	25.8	Georgia	254.2

[1444]:
```python
# Create Categorization field Region for State field
df.loc[df['State'].isin (['Connecticut', 'Maine', 'Massachusetts', 'New
 ↪Hampshire', 'Rhode Island', 'Vermont']), 'Region'] = 'New England'

df.loc[df['State'].isin (['New Jersey', 'New York', 'Pennsylvania']), 'Region']
 ↪= 'Middle Atlantic'

df.loc[df['State'].isin (['Illinois', 'Indiana', 'Michigan', 'Ohio',
 ↪'Wisconsin']), 'Region'] = 'East North Central'

df.loc[df['State'].isin (['Iowa', 'Kansas', 'Minnesota', 'Missouri',
 ↪'Nebraska', 'North Dakota', 'South Dakota']), 'Region'] = 'West North
 ↪Central'

df.loc[df['State'].isin (['Delaware', 'Florida', 'Georgia', 'Maryland', 'North
 ↪Carolina', 'South Carolina', 'Virginia', 'Washington D.C.', 'West
 ↪Virginia']), 'Region'] = 'South Atlantic'

df.loc[df['State'].isin (['Alabama', 'Kentucky', 'Mississippi', 'Tennessee']),
 ↪'Region'] = 'East South Central'

df.loc[df['State'].isin (['Arkansas', 'Louisiana', 'Oklahoma', 'Texas']),
 ↪'Region'] = 'West South Central'

df.loc[df['State'].isin (['Arizona', 'Colorado', 'Idaho', 'Montana', 'Nevada',
 ↪'New Mexico', 'Utah',  'Wyoming']), 'Region'] = 'Mountain'

df.loc[df['State'].isin (['Alaska', 'California', 'Hawaii', 'Oregon',
 ↪'Washington']), 'Region'] = 'Pacific'

df.head(10)
```

5

```
[1444]:      Murder  Assault  UrbanPop  Rape        State  TotalArrests  \
       0      13.2      236        58  21.2      Alabama         270.4
       1      10.0      263        48  44.5       Alaska         317.5
       2       8.1      294        80  31.0      Arizona         333.1
       3       8.8      190        50  19.5     Arkansas         218.3
       4       9.0      276        91  40.6   California         325.6
       5       7.9      204        78  38.7     Colorado         250.6
       6       3.3      110        77  11.1  Connecticut         124.4
       7       5.9      238        72  15.8     Delaware         259.7
       8      15.4      335        80  31.9      Florida         382.3
       9      17.4      211        60  25.8      Georgia         254.2

                      Region
       0  East South Central
       1             Pacific
       2            Mountain
       3  West South Central
       4             Pacific
       5            Mountain
       6         New England
       7      South Atlantic
       8      South Atlantic
       9      South Atlantic
```

```
[1445]: # Confirm region was applied to all rows
        df.groupby('Region')['Murder']\
                .agg(['count'])\
                .sort_values(by='Region',ascending=False)
```

```
[1445]:                     count
        Region
        West South Central      4
        West North Central      7
        South Atlantic          8
        Pacific                 5
        New England             6
        Mountain                8
        Middle Atlantic         3
        East South Central      4
        East North Central      5
```

```
[1446]: # Check data type of each field
        df.dtypes
```

```
[1446]: Murder        float64
        Assault         int64
        UrbanPop        int64
```

```
Rape            float64
State            object
TotalArrests    float64
Region           object
dtype: object
```

[1447]:
```python
# Change State from object to string
df['State'] = df['State'].astype("string")

# Change Region from object to string
df['Region'] = df['Region'].astype("string")
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50 entries, 0 to 49
Data columns (total 7 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Murder        50 non-null     float64
 1   Assault       50 non-null     int64
 2   UrbanPop      50 non-null     int64
 3   Rape          50 non-null     float64
 4   State         50 non-null     string
 5   TotalArrests  50 non-null     float64
 6   Region        50 non-null     string
dtypes: float64(3), int64(2), string(2)
memory usage: 2.9 KB
```

1.2 Regions

South Atlantic: 'Delaware', 'Florida', 'Georgia', 'Maryland', 'North Carolina', 'South Carolina', 'Virginia', 'Washington D.C.', 'West Virginia'

New England: 'Connecticut', 'Maine', 'Massachusetts', 'New Hampshire', 'Rhode Island', 'Vermont'

Middle Atlantic: 'New Jersey', 'New York', 'Pennsylvania'

East North Central: 'Illinois', 'Indiana', 'Michigan', 'Ohio', 'Wisconsin'

West North Central: 'Iowa', 'Kansas', 'Minnesota', 'Missouri', 'Nebraska', 'North Dakota', 'South Dakota'

East South Central: 'Alabama', 'Kentucky', 'Mississippi', 'Tennessee'

West South Central: 'Arkansas', 'Louisiana', 'Oklahoma', 'Texas'

Mountain: 'Arizona', 'Colorado', 'Idaho', 'Montana', 'Nevada', 'New Mexico', 'Utah', 'Wyoming'

Pacific: 'Alaska', 'California', 'Hawaii', 'Oregon', 'Washington'

1.2.1 Q1: Which region has the most arrests for murder? least?

```
[1450]:  # Murders by Region
         ax=df.groupby('Region')['Murder']\
                 .agg(['sum'])\
                 .sort_values('sum',ascending=False) \
                 .plot(kind='barh', figsize=(10,5), title = ' Murder Arrests per 100K⌴
           ↪By Region')
         ax.set_xlabel('Murders')
         ax.bar_label(ax.containers[0])
         plt.show()
```
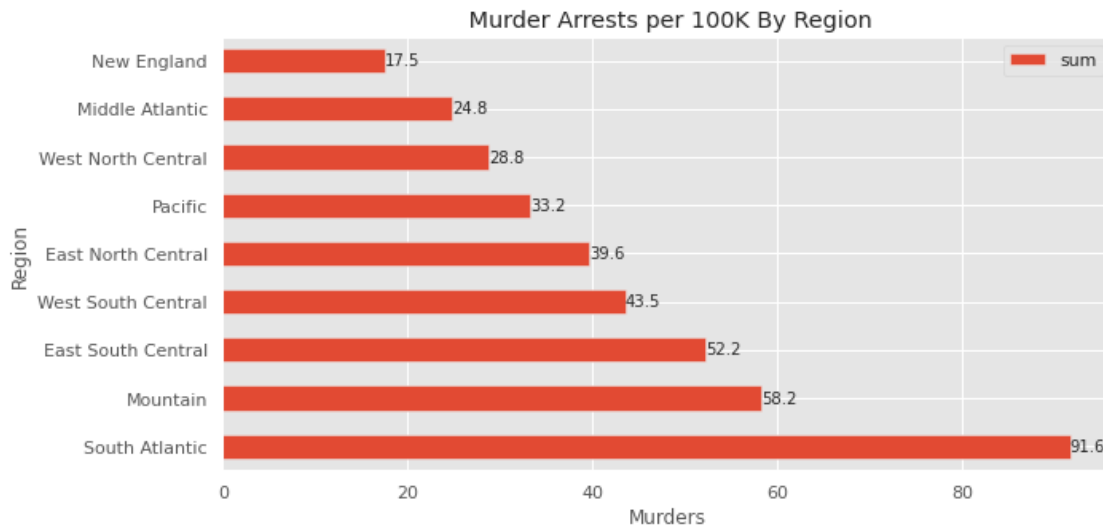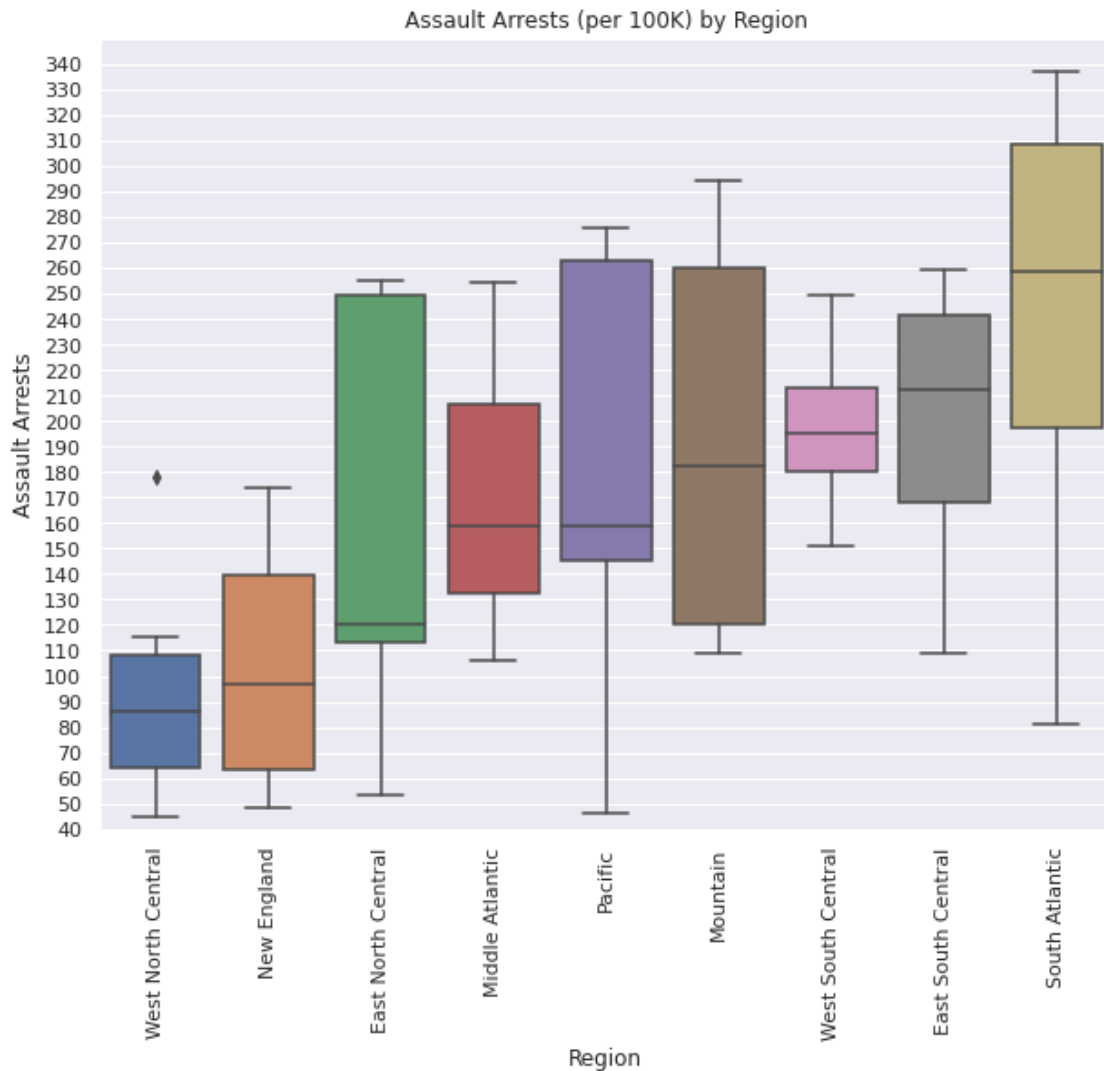

Murders per 100K by Region

The above graph shows the South Atlantic region has the highest amount of arrests for murder.
The New England region has the lowest amount of arrests for murder.

```
[1452]:  #boxplot.  Murders
         mean_by_region = df.groupby(['Region'])['Murder'].mean().
           ↪sort_values(ascending=True)
         sns.set(rc={"figure.figsize":(10,5)})
         ax=sns.boxplot(x='Region',
                     y="Murder", data=df, order=mean_by_region.index)
         ax.set(ylabel='Murder', title='Total Murder Arrests (per 100K) by⌴
           ↪Region',yticks=np.arange(19),
                 ylim=(0.0, 18.0) )
         plt.xticks(rotation='vertical')
         plt.show()
```
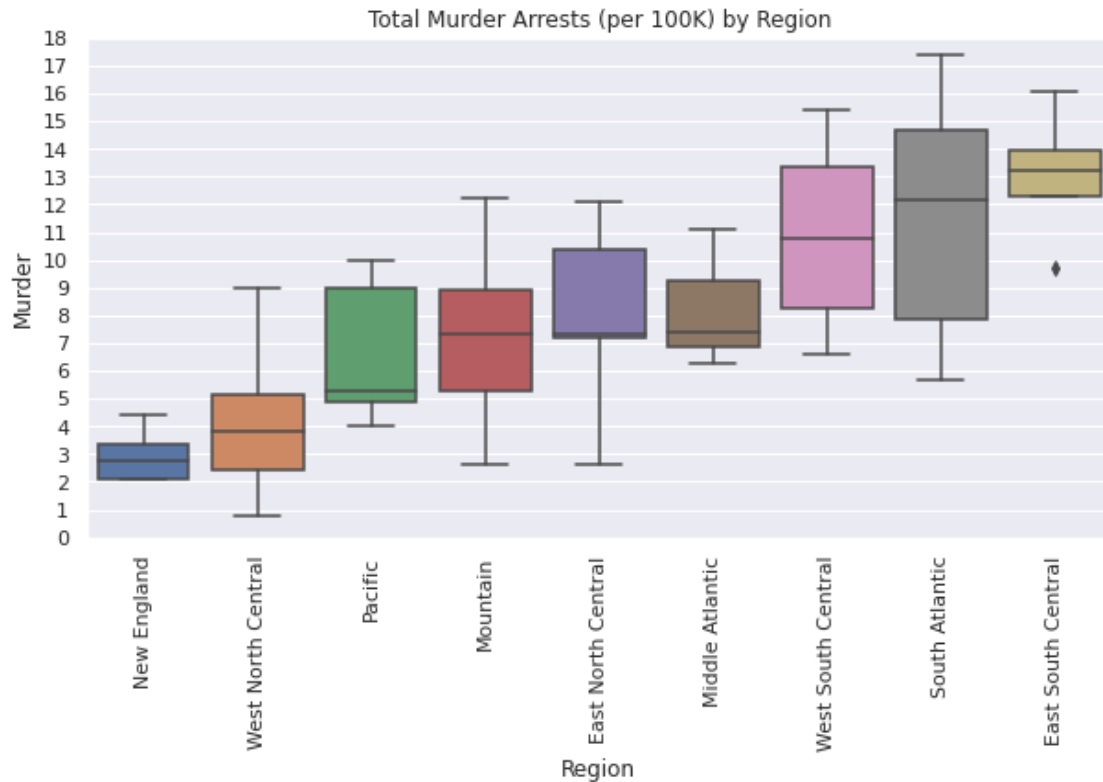
Total Murder Arrests (per 100K) by Region

The above graphs look at distribution of values. The East South Central region has the highest median murder arrest rate per 100K. The New England region has the lowest median arrest rate per 100K.

Answer Q1:

The South Atlantic region has the most arrests for murders. The New England region has the fewest arrests for murder. The East South Central is the most dangerous region as its minimum arrest rate per 100K is greater than the median arrest rate for the South Atlantic region.

1.2.2 Q2: Which region has the most arrests for assault? least?

```
[1456]:  # Assault by Region
         ax=df.groupby('Region')['Assault']\
              .agg(['sum'])\
              .sort_values('sum',ascending=False) \
              .plot(kind='barh', figsize=(10,5), title = ' Assault Arrests per 100K␣
           ↪by Region')
         ax.set_xlabel('Assault')
         ax.bar_label(ax.containers[0])
         plt.show()
```
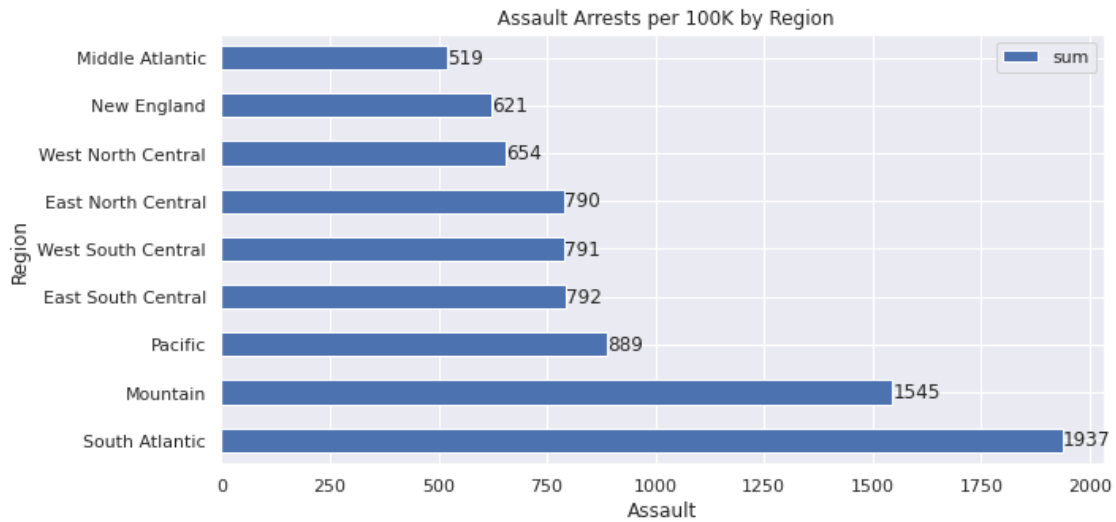
Assault Arrests per 100K by Region

Assaults per 100K by Region

The above graph shows the South Atlantic region has the highest amount of arrests for assaults. The Middle Atlantic region has the lowest amount of arrests for assaults.

```
[1458]: #boxplot.  Assaults
        assault_mean = df.groupby(['Region'])['Assault'].mean().
         ↪sort_values(ascending=True)
        sns.set(rc={"figure.figsize":(10,8)})
        ax=sns.boxplot(x='Region',
                    y="Assault", data=df, order=assault_mean.index)
        ax.set(ylabel='Assault Arrests', title='Assault Arrests (per 100K) by␣
         ↪Region',yticks=np.arange(40,350, 10),
                ylim=(40.0, 350.0))
        plt.xticks(rotation='vertical')

        plt.show()
```

Assaults per 100K by Region

The above graphs look at distribution of values. The South Atlantic region has the highest median arrest rate per 100K. The West North Central region has the lowest median arrest rate per 100K for assaults.

Answer Q2:

South Atlantic region has the most arrests. The Middle Atlantic region has the fewest arrests. The West North Central region has the lowest median arrest rate for assault.

1.2.3   Q3: Which region has the most arrests for rape? least?

```
[1462]: # Rape by Region
        ax=df.groupby('Region')['Rape']\
               .agg(['sum'])\
               .sort_values('sum',ascending=False) \
               .plot(kind='barh', figsize=(10,5), title = ' Rape Arrests per 100K by␣
          ↪Region')
        ax.set_xlabel('Rape')
        ax.bar_label(ax.containers[0])
        plt.show()
```
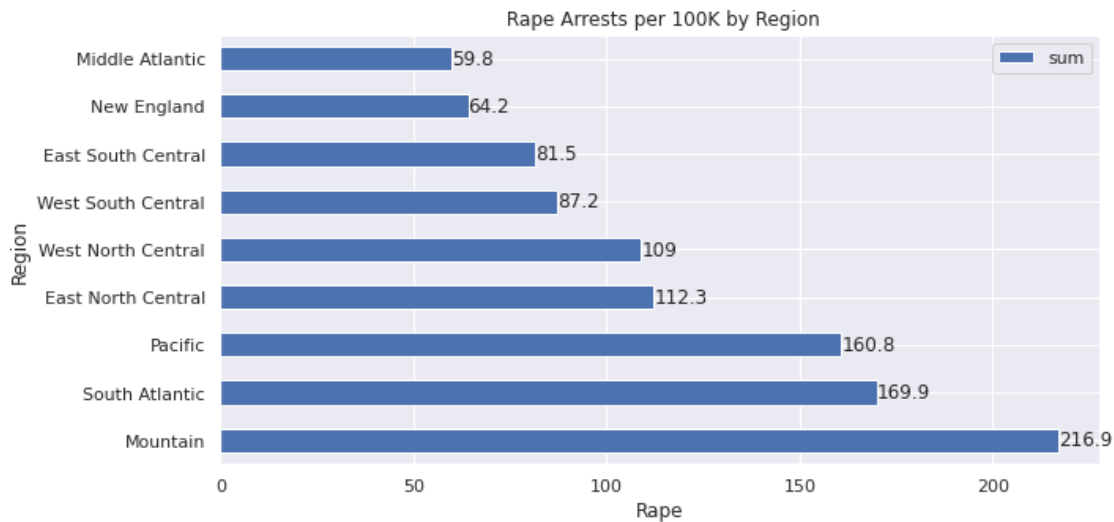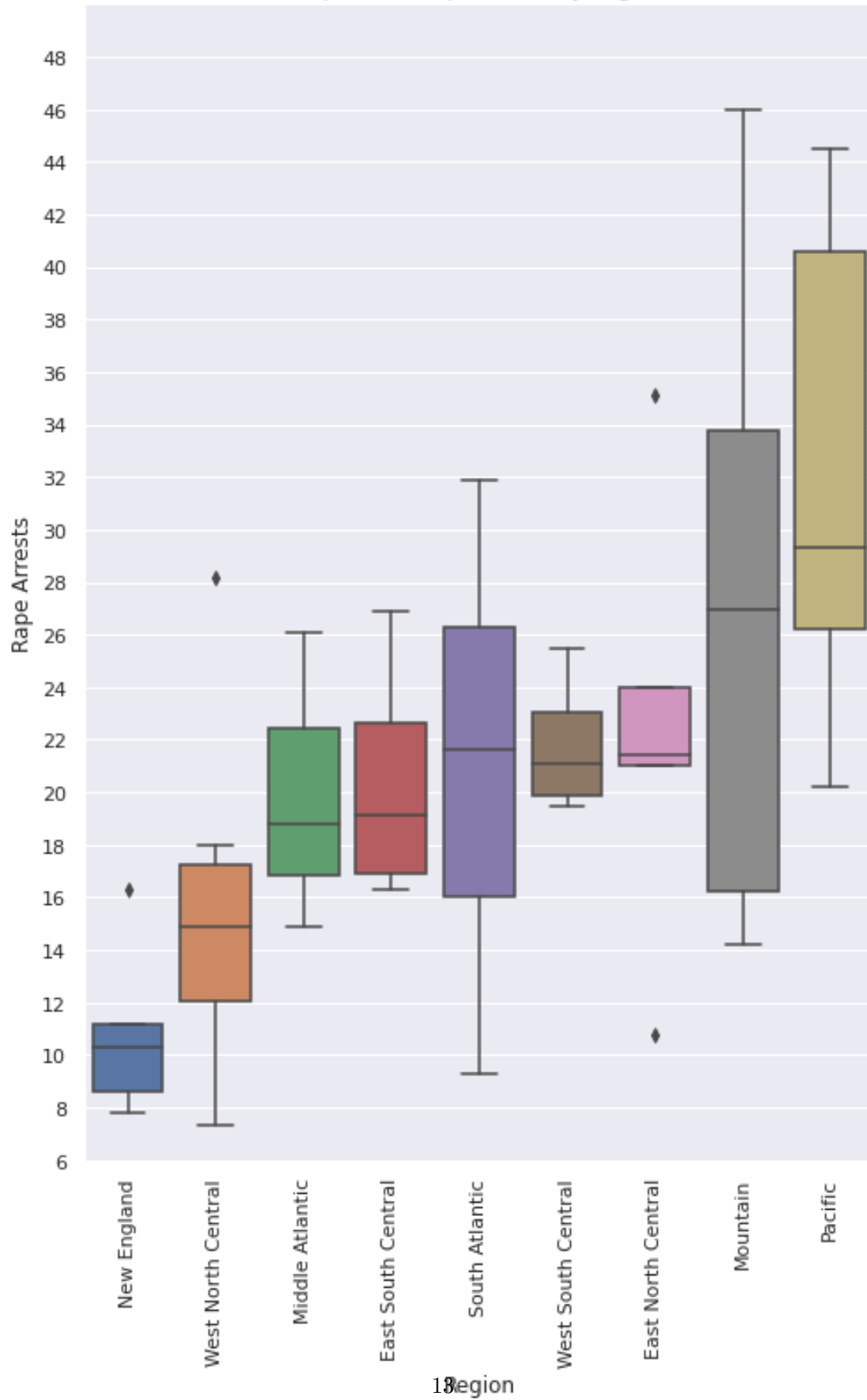

Rape Arrests per 100K by Region

The above graph shows the Mountain region has the highest amount of arrests for rape. The Middle
Atlantic region has the lowest amount of arrests for rape.

```
[1464]: #boxplot.  Total Arrests
        rape_mean = df.groupby(['Region'])['Rape'].mean().sort_values(ascending=True)
        sns.set(rc={"figure.figsize":(8,12)})
        ax=sns.boxplot(x='Region',
                      y="Rape", data=df, order=rape_mean.index)
        ax.set(ylabel='Rape Arrests', title='Rape Arrests (per 100K) by␣
          ↪Region',yticks=np.arange(0,50, 2),
                 ylim=(6.0, 50.0))
        plt.xticks(rotation='vertical')

        plt.show()
```

Rape Arrests (per 100K) by Region

Total Rape Arrests (per 100K) by Region

The above graphs look at distribution of values. The New England region has the lowest median arrest rate per 100K for rape. The Pacific region has the highest median arrest rate for rape.

Answer Q3:

The Mountain region has the most arrests for rape. The Middle Atlantic region has the fewest arrests for rape. The New England region has lowest median arrest rate per 100K for rape. The Pacific region has the highest median arrest rate per 100K for rape.

1.2.4 Q4: Which region has the most arrests for violent crime? least?

```python
# Total Arrests by Region
ax=df.groupby('Region')['TotalArrests']\
        .agg(['sum'])\
        .sort_values('sum',ascending=False) \
        .plot(kind='barh', figsize=(10,5), title = ' Total Arrests for Violent␣
    ↪Crime (per 100K) by Region')
ax.set_xlabel('Total Arrests')
ax.bar_label(ax.containers[0])
plt.show()
```
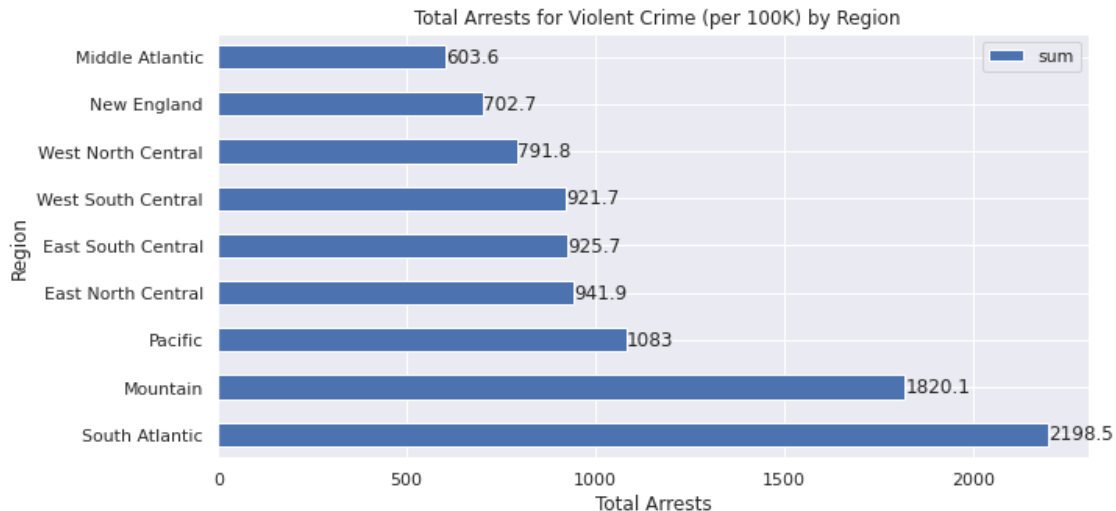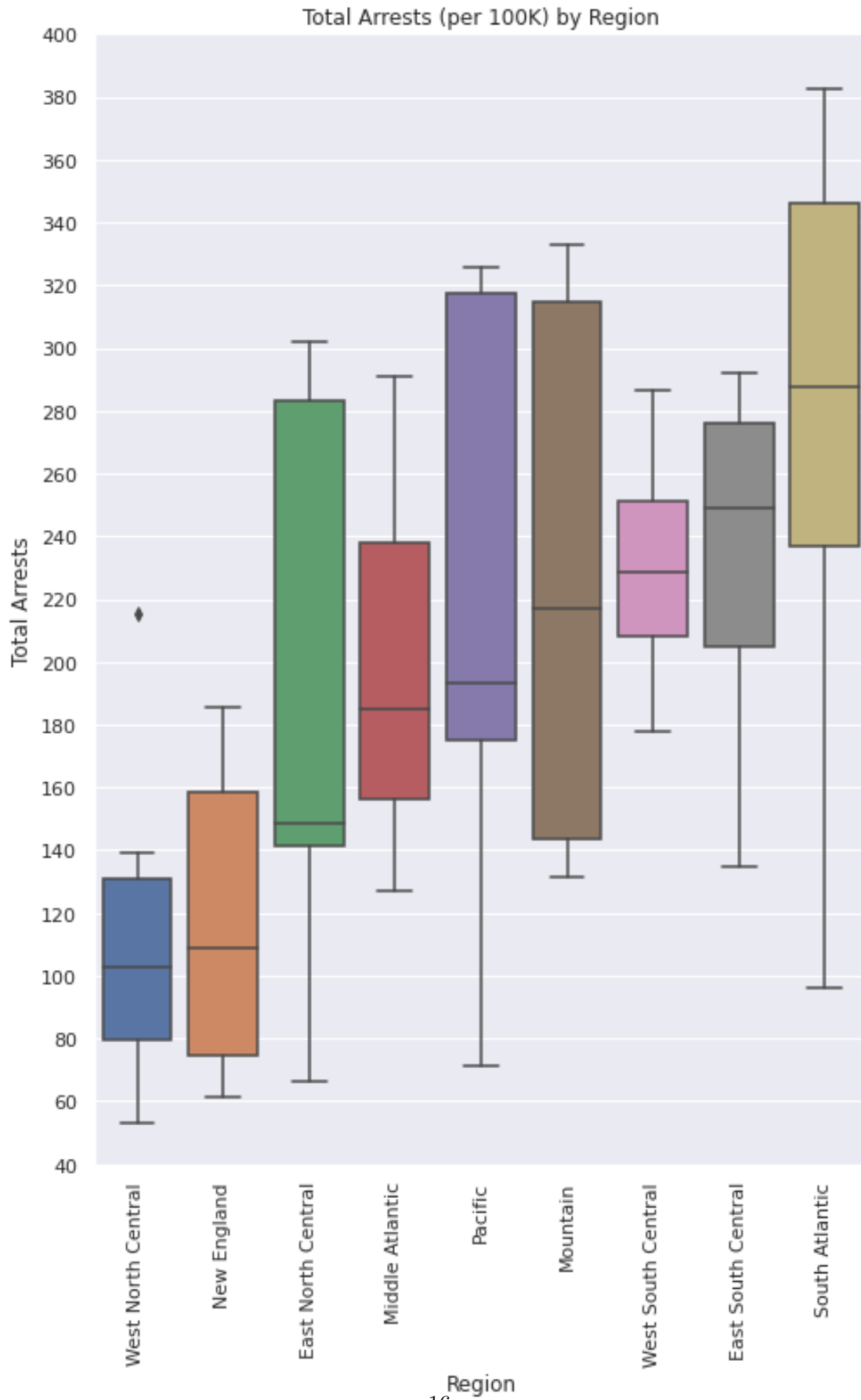

Total Arrests for Violent Crime by Region

The above graph shows the South Atlantic region has the highest amount of arrests for violent crime. The Middle Atlantic region has the lowest amount of arrests for violent crime.

14

```
[1470]:  #boxplot.   Total Arrests
         totarrest_mean = df.groupby(['Region'])['TotalArrests'].mean().
          ↪sort_values(ascending=True)
         sns.set(rc={"figure.figsize":(8,12)})
         ax=sns.boxplot(x='Region',
                         y="TotalArrests", data=df, order=totarrest_mean.index)
         ax.set(ylabel='Total Arrests', title='Total Arrests (per 100K) by␣
          ↪Region',yticks=np.arange(40,4000,20),
                 ylim=(40.0, 400.0))
         plt.xticks(rotation='vertical')

         plt.show()
```

Total Arrests (per 100K) by Region

Total Arrests by Region (per 100K) - Boxplot

The above graphs show distribution of values. The South Atlantic region has the highest median arrest rate for violent crime. The West North Central region has the lowest median arrest rate for violent crime.

Answer Q4:

The South Atlantic region has the most arrests for violent crime. The Middle Atlantic has the fewest arrests for violent crime. The West North Central region has the lowest median arrest rate per 100K.

1.3 Correlation

```
*   A positive correlation indicates when one crime increases the other crime increases.
*   A negative correlation indicated when one crime increases the other crime decreases.
*   Correlation indicates relationship not causation.
```

1.3.1 Q5: Is there a correlation between the types of crimes?

1.3.2 Q6: Does having a high percentage of the population living in an urban area increase crime?

```
[1475]: #Correlation
        df_corr = df[['Murder','Rape', 'Assault', 'UrbanPop']].dropna().corr()
        df_corr
```

```
[1475]:             Murder       Rape    Assault  UrbanPop
        Murder    1.000000  0.563579  0.801873  0.069573
        Rape      0.563579  1.000000  0.665241  0.411341
        Assault   0.801873  0.665241  1.000000  0.258872
        UrbanPop  0.069573  0.411341  0.258872  1.000000
```

```
[1476]: # Correlation shown as a graph
        mask = np.zeros_like(df_corr)
        mask[np.triu_indices_from(mask)]=True
        sns.heatmap(df_corr, annot=True,center=0,fmt='.3f', square=True, linewidth=3,␣
          ↪mask=mask)
```

```
[1476]: <AxesSubplot:>
```
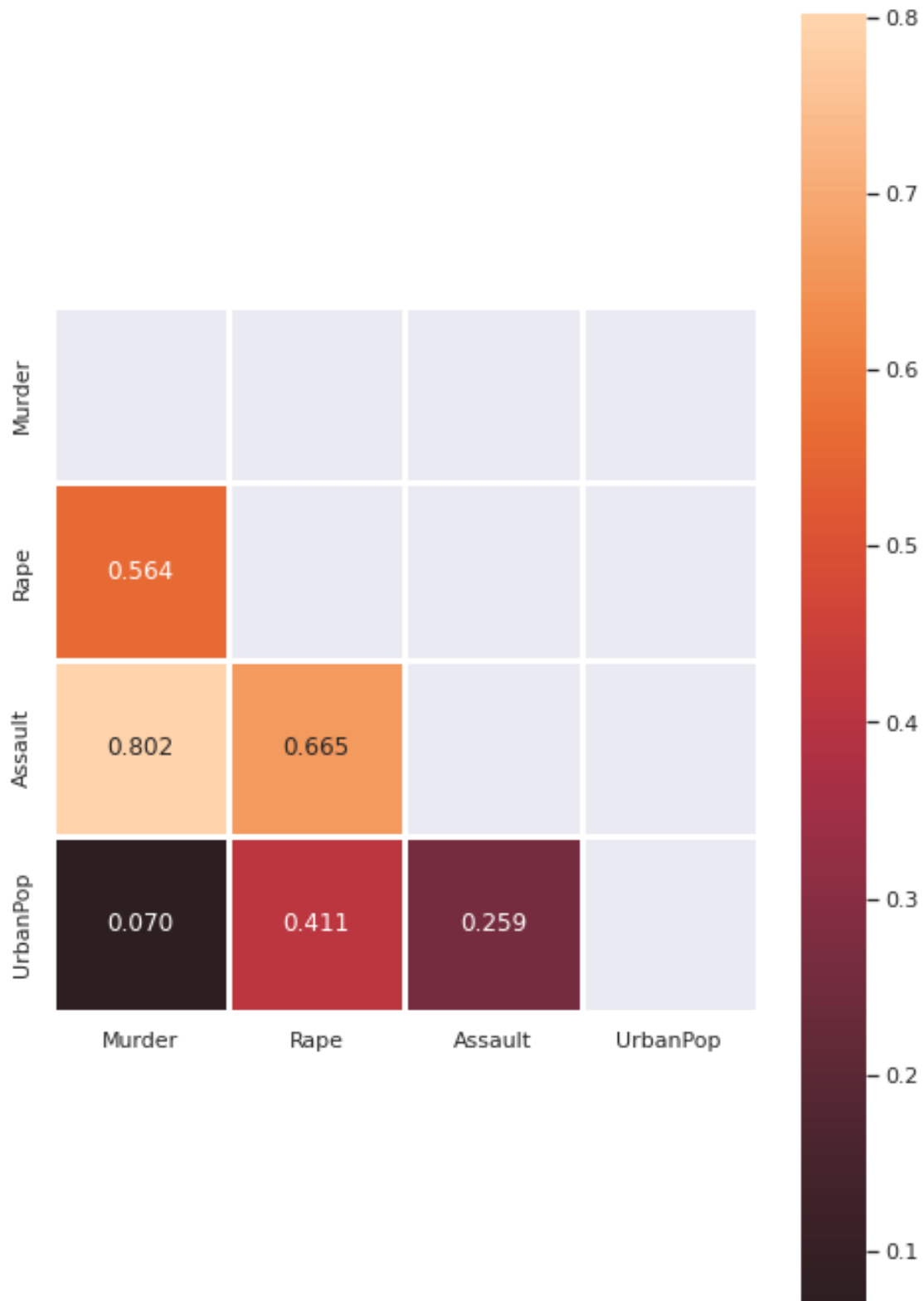
The above graph shows correlation between the types of crimes.

* Murder arrests and assault arrests have a strong positive correlation.
* Rape arrests have a moderate positive relationship to murder and assault.
* All of the crimes have a weak correlation to percent of population living in an urban area.

Answer Q5:

Murder and assault arrests have a strong positive correlation. If murders increase, assaults increase.

Answer Q6:

There isn't a correlation between percent of population living in an urban area and number of arrests.

## 1.4 States

```
[1480]: # Basic statistics
        df.describe()
```

```
[1480]:          Murder      Assault    UrbanPop         Rape  TotalArrests
        count  50.00000   50.000000   50.000000   50.000000     50.000000
        mean    7.78800  170.760000   65.540000   21.232000    199.780000
        std     4.35551   83.337661   14.474763    9.366385     93.372967
        min     0.80000   45.000000   32.000000    7.300000     53.100000
        25%     4.07500  109.000000   54.500000   15.075000    128.250000
        50%     7.25000  159.000000   66.000000   20.100000    185.200000
        75%    11.25000  249.000000   77.750000   26.175000    285.800000
        max    17.40000  337.000000   91.000000   46.000000    382.300000
```

1.4.1 Q7: Which state has the most arrests for murder? least?

```
[1482]: # Horizontal bar chart.  Sort values descending.  Add labels to bars
        sns.set(rc={"figure.figsize":(8,10)})
        ax=sns.barplot(x='Murder',
                    y="State", data=df,
                    order=df.sort_values('Murder', ascending=False).State)
        ax.set(xlabel='Murder per 100K', title='Murder Arrests per 100K by State \n␣
         ↪(Avg = 7.89)')
        ax.bar_label(ax.containers[0])
        plt.show()
```

Murder Arrests per 100K by State
(Avg = 7.89)

[ ]:
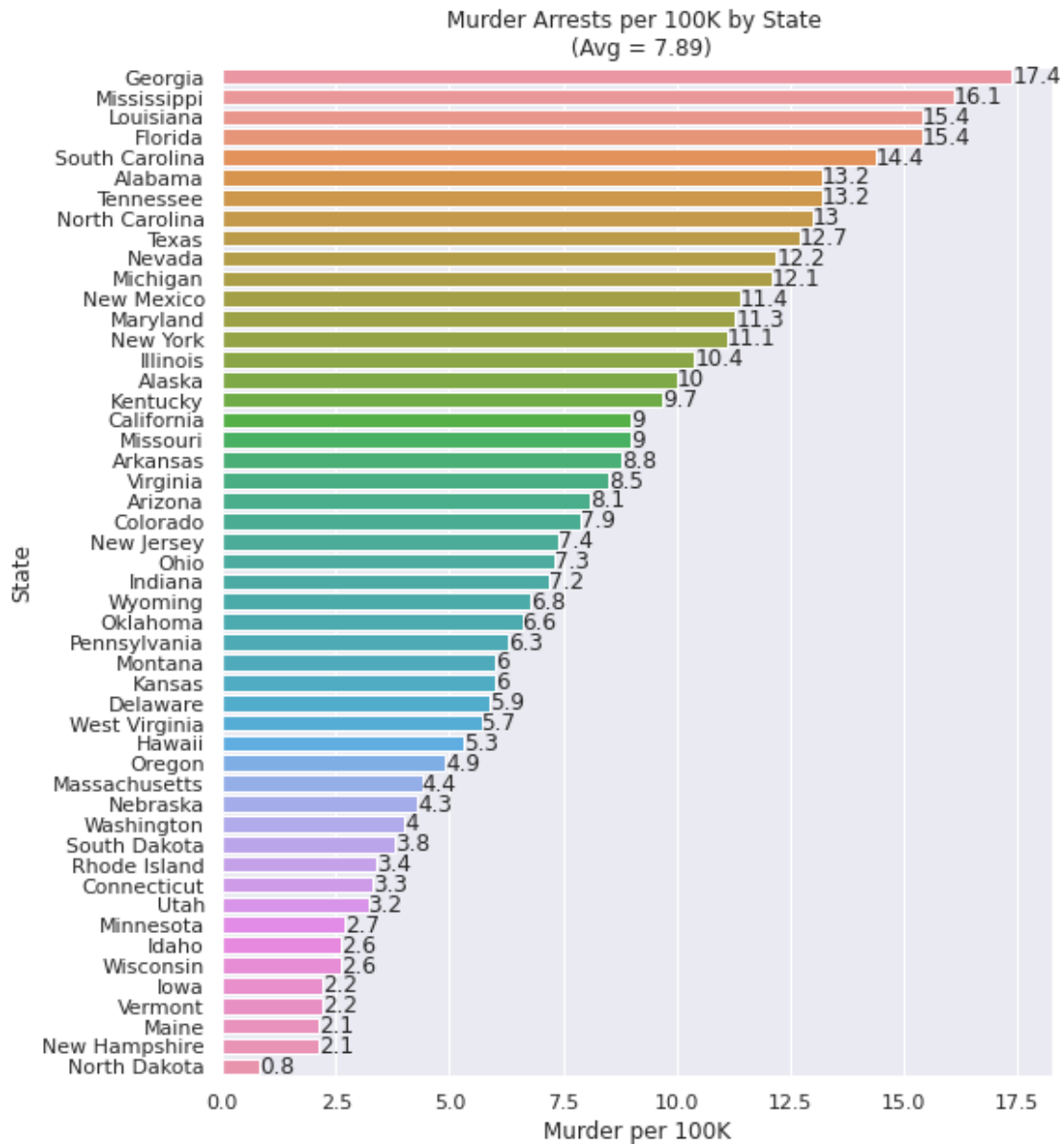
Murder Arrests per 100K by State

The above graphs show the number of murder arrests for each state in descending order. Georgia has the most murder arrests. North Dakota has the least murder arrests.

Answer Q7:

Georgia has the most murder arrests. North Dakota has the fewest murder arrests.

1.4.2 Q8: Which state has the most assault arrests? least?

```
[1486]:  ## Horizontal bar chart.  Sort values descending.  Add labels to bars
         sns.set(rc={"figure.figsize":(8,10)})
         ax=sns.barplot(x='Assault',
                     y="State", data=df,
                     order=df.sort_values('Assault', ascending=False).State)
         ax.set(xlabel='Assault Arrests', title='Assault Arrests per 100K by State \n␣
          ↪(Avg = 170.76)')
         ax.bar_label(ax.containers[0])
         plt.show()
```
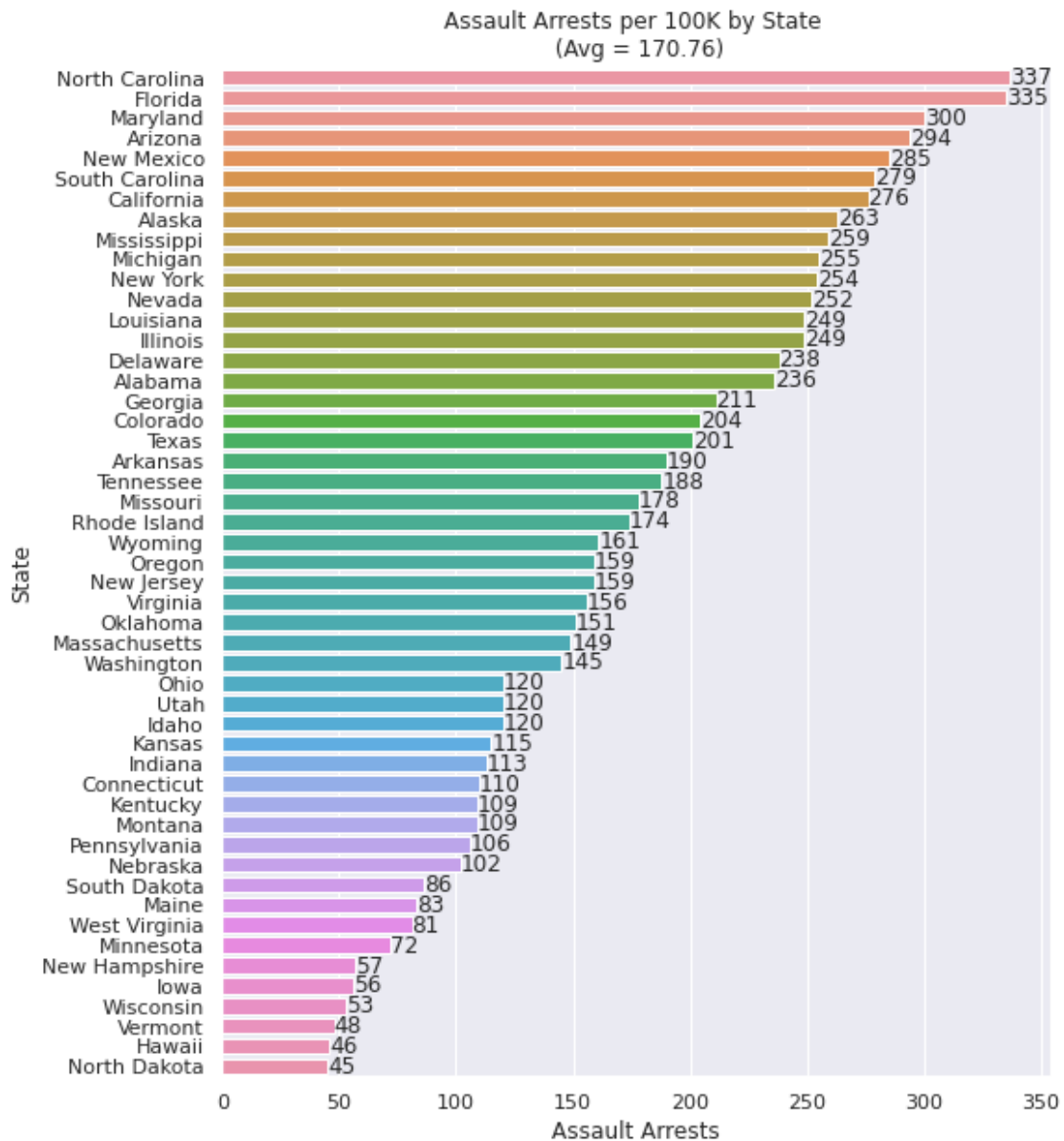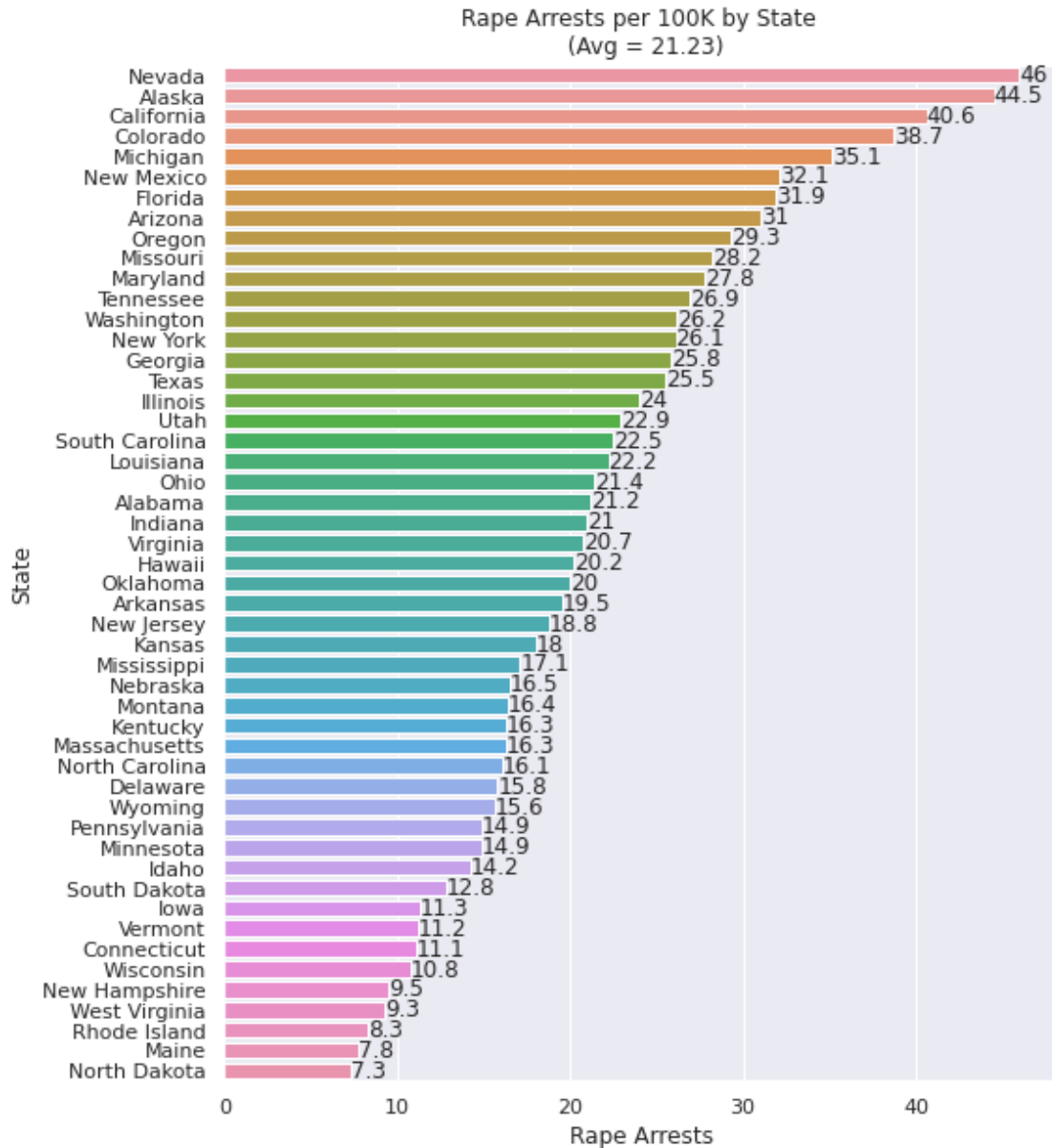
Assaults per 100K by State

The above graphs shows number of assault arrests by state. North Carolina has the most assault arrests. North Dakota and Hawaii have the fewest assault arrests.

Answer Q8:

North Carolina has the most arrests for assaults. North Dakota has the fewest arrests for assault.

1.4.3 Q9: Which state has the most arrests for rape? least?

```
[1490]:  ## Horizontal bar chart.  Sort values descending.  Add labels to bars
         sns.set(rc={"figure.figsize":(8,10)})
         ax=sns.barplot(x='Rape',
                     y="State", data=df,
                     order=df.sort_values('Rape', ascending=False).State)
         ax.set(xlabel='Rape Arrests', title='Rape Arrests per 100K by State \n (Avg =␣
           ↪21.23)')
         ax.bar_label(ax.containers[0])
         plt.show()
```

Rape Arrests per 100K by State
(Avg = 21.23)
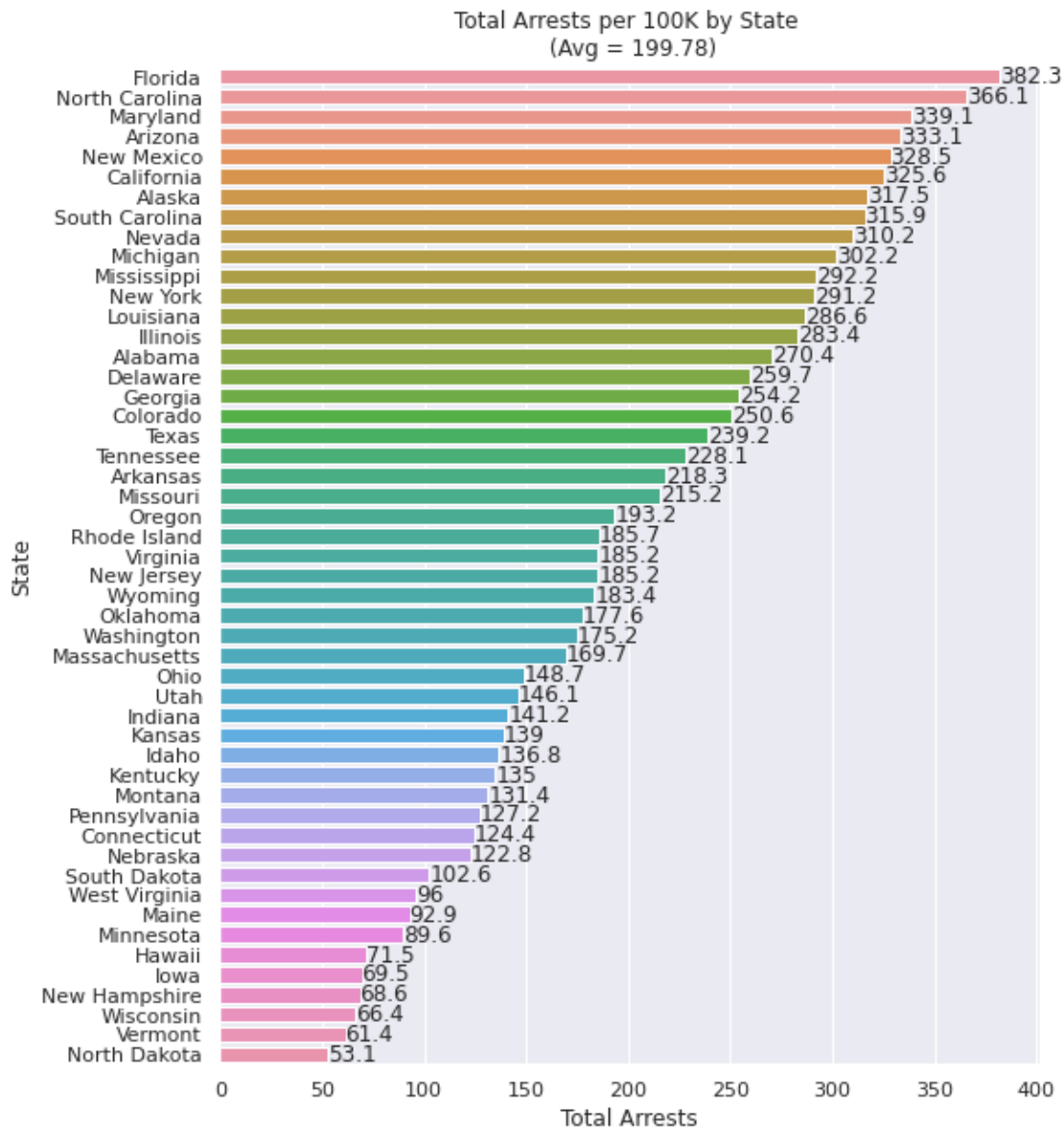
Rape Arrests per 100K by State

The above graph shows Nevada and Alaska have the most arrests for rape. North Dakota and Maine have the fewest arrests for rape.

Answer Q9:

Nevada and Alaska have the highest amount of arrests for rape. North Dakota and Maine have the fewest amount of arrests for rape.

1.4.4 Q10: Which state has the most arrests for violent crimes? least?

```
[1494]:  ## Horizontal bar chart.  Sort values descending.  Add labels to bars
         sns.set(rc={"figure.figsize":(8,10)})
         ax=sns.barplot(x='TotalArrests',
                     y="State", data=df,
                     order=df.sort_values('TotalArrests', ascending=False).State)
         ax.set(xlabel='Total Arrests', title='Total Arrests per 100K by State \n (Avg =␣
           ↪199.78)')
         ax.bar_label(ax.containers[0])
         plt.show()
```

Total Arrests per 100K by State
(Avg = 199.78)

State	Total Arrests
Florida	382.3
North Carolina	366.1
Maryland	339.1
Arizona	333.1
New Mexico	328.5
California	325.6
Alaska	317.5
South Carolina	315.9
Nevada	310.2
Michigan	302.2
Mississippi	292.2
New York	291.2
Louisiana	286.6
Illinois	283.4
Alabama	270.4
Delaware	259.7
Georgia	254.2
Colorado	250.6
Texas	239.2
Tennessee	228.1
Arkansas	218.3
Missouri	215.2
Oregon	193.2
Rhode Island	185.7
Virginia	185.2
New Jersey	185.2
Wyoming	183.4
Oklahoma	177.6
Washington	175.2
Massachusetts	169.7
Ohio	148.7
Utah	146.1
Indiana	141.2
Kansas	139
Idaho	136.8
Kentucky	135
Montana	131.4
Pennsylvania	127.2
Connecticut	124.4
Nebraska	122.8
South Dakota	102.6
West Virginia	96
Maine	92.9
Minnesota	89.6
Hawaii	71.5
Iowa	69.5
New Hampshire	68.6
Wisconsin	66.4
Vermont	61.4
North Dakota	53.1

Total Arrests by State

The above graphs shows the state with the most arrests for violent crime is Florida. The state with the fewest arrests is North Dakota.

Answer Q10:

Florida has the most arrests for violent crime. North Dakota has the fewest arrest for violent crime.

1.5 Summary

The USArrests dataset presents data for 1973.

Based on the analysis of the USArrests dataset, the data presents the following:

```
1.  A high urban population percentage is not an indicator for high crime.
2.  The deadliest state is Georgia.  It has the most arrests for murder.
3.  The safest state is North Dakota.  It has the fewest amount of arrests for each category.
4.  The state with the most crime is Florida.  It has the most arrests for violent crime.
5.  The most dangerous area of the country is the South Atlantic region.
6.  The safest area of the country is the West North Central region.
7.  The safest states are New Hampshire, Iowa, Wisconsin, Vermont, and North Dakota.
```

Note:

South Atlantic: Delaware, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, Washington D.C., West Virginia

West North Central: 'Iowa', 'Kansas', 'Minnesota', 'Missouri', 'Nebraska', 'North Dakota', 'South Dakota'

```python
[1498]: # Write cleaned USArrests to csv
        df.to_csv('USArrests_cleaned.csv', index=False)
```

```
[ ]:
```