PROJECT 2: CRIME RATES IN THE US

BACKGROUND

- This data set displays the violent crime rates per 100,000 residents by US State in 1973.
- There are 4 variables with 50 states in this data set:
  - murder
  - assault
  - rape
  - urban population

THE QUESTIONS

Is there a correlation between the violent crime rates and each other?

- rape vs murder
- murder vs assault
- assault vs rape

Is there a correlation between the violent crime and urban population?

In [1]:
```
#import necessary libraries
!pip install beautifulsoup4
import requests
from bs4 import BeautifulSoup
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
```

```
/usr/lib/python3/dist-packages/secretstorage/dhcrypto.py:15: Cryptogra
phyDeprecationWarning: int_from_bytes is deprecated, use int.from_byte
s instead
  from cryptography.utils import int_from_bytes
/usr/lib/python3/dist-packages/secretstorage/util.py:19: CryptographyD
eprecationWarning: int_from_bytes is deprecated, use int.from_bytes in
stead
  from cryptography.utils import int_from_bytes
Requirement already satisfied: beautifulsoup4 in /home/nbgrader/fall2
3/student-accounts/jtowgood/.local/lib/python3.8/site-packages (4.12.
2)
Requirement already satisfied: soupsieve>1.2 in /home/nbgrader/fall23/
student-accounts/jtowgood/.local/lib/python3.8/site-packages (from bea
utifulsoup4) (2.5)
```

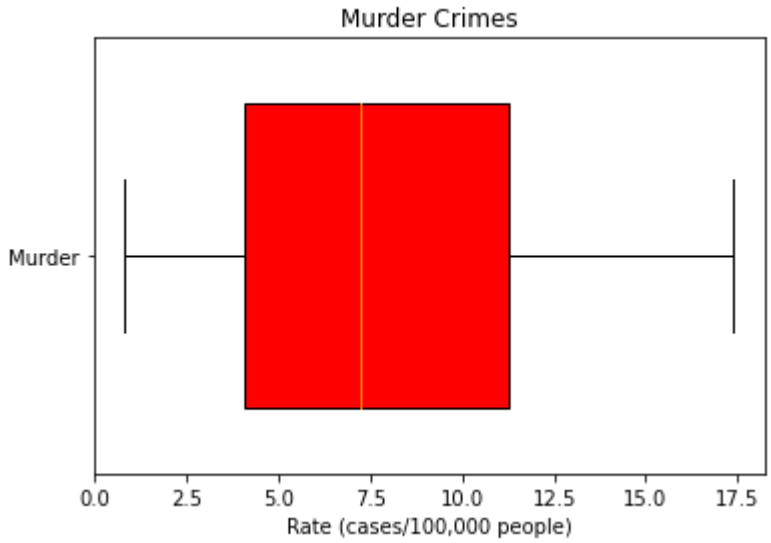
In [2]:
```
from statsmodels.datasets import get_rdataset
USArrests = get_rdataset('USArrests').data
print(USArrests.describe())
```

```
         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
```

SUMMARY OF VIOLENT CRIMES

MURDER

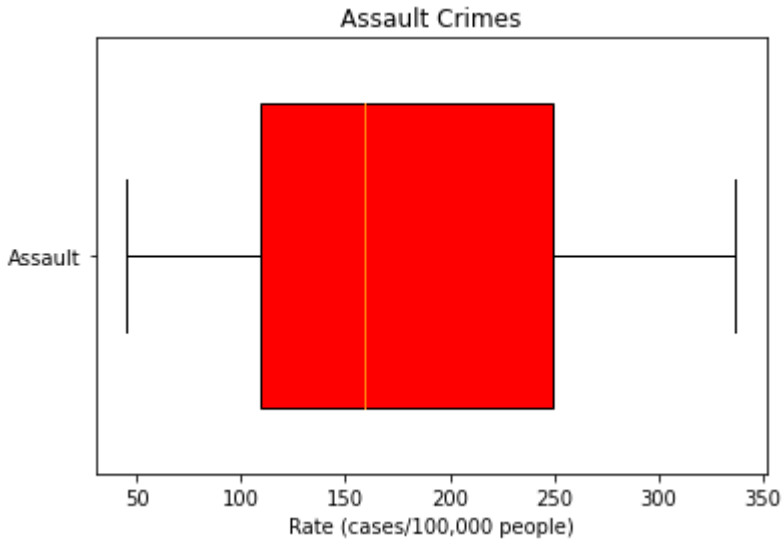
In [3]:
```
murder_rates = USArrests['Murder']

plt.boxplot(murder_rates, vert=False, labels=['Murder'], widths=0.7, p
atch_artist=True, boxprops=dict(facecolor='red'))

plt.xlabel('Rate (cases/100,000 people)')
plt.title('Murder Crimes')

plt.show()
print(murder_rates.describe())
```




```
count    50.00000
mean      7.78800
std       4.35551
min       0.80000
25%       4.07500
50%       7.25000
75%      11.25000
max      17.40000
Name: Murder, dtype: float64
```

ASSAULT

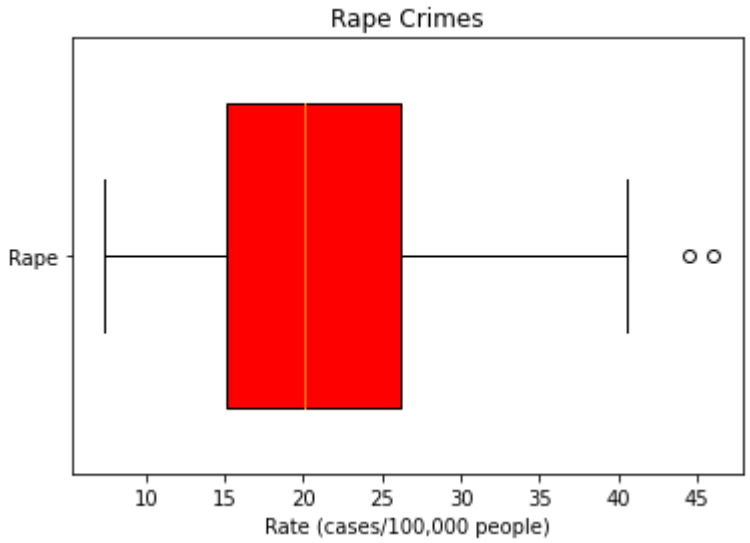
In [4]:
```
assault_rates = USArrests['Assault']

plt.boxplot(assault_rates, vert=False, labels=['Assault'], widths=0.7,
patch_artist=True, boxprops=dict(facecolor='red'))

plt.xlabel('Rate (cases/100,000 people)')
plt.title('Assault Crimes')

plt.show()
print(assault_rates.describe())
```




```
count     50.000000
mean     170.760000
std       83.337661
min       45.000000
25%      109.000000
50%      159.000000
75%      249.000000
max      337.000000
Name: Assault, dtype: float64
```

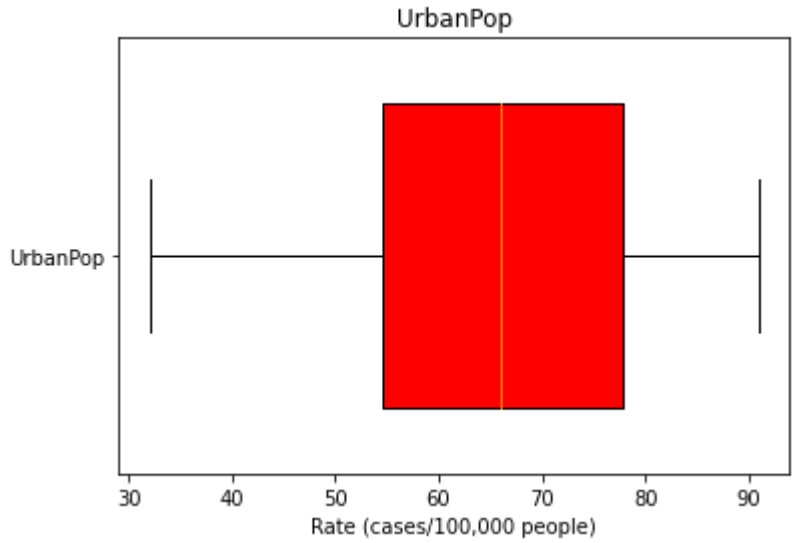
RAPE

In [5]:
```
rape_rates = USArrests['Rape']

plt.boxplot(rape_rates, vert=False, labels=['Rape'], widths=0.7, patch
_artist=True, boxprops=dict(facecolor='red'))

plt.xlabel('Rate (cases/100,000 people)')
plt.title('Rape Crimes')

plt.show()
print(rape_rates.describe())
```




```
count    50.000000
mean     21.232000
std       9.366385
min       7.300000
25%      15.075000
50%      20.100000
75%      26.175000
max      46.000000
Name: Rape, dtype: float64
```

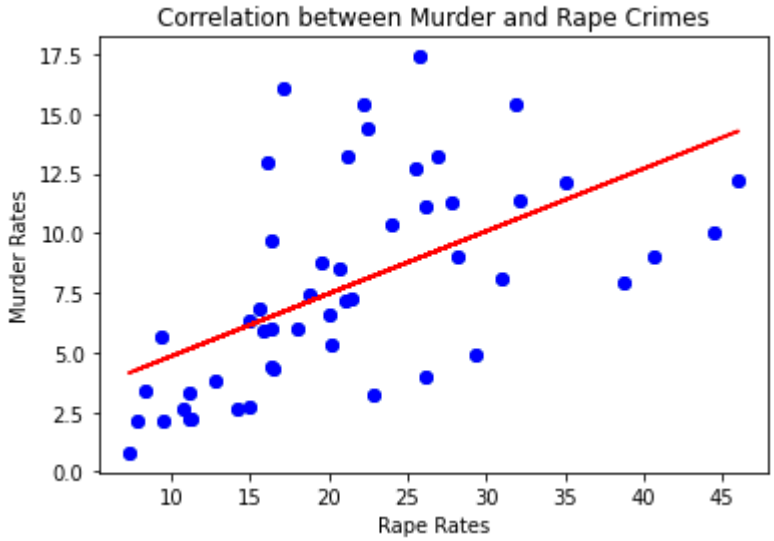
In [6]:
```
urbanPop_rates = USArrests['UrbanPop']

plt.boxplot(urbanPop_rates, vert=False, labels=['UrbanPop'], widths=0.
7, patch_artist=True, boxprops=dict(facecolor='red'))

plt.xlabel('Rate (cases/100,000 people)')
plt.title('UrbanPop')

plt.show()
print(urbanPop_rates.describe())
```




```
count    50.000000
mean     65.540000
std      14.474763
min      32.000000
25%      54.500000
50%      66.000000
75%      77.750000
max      91.000000
Name: UrbanPop, dtype: float64
```

RAPE VS MURDER

In [7]:
```
plt.scatter(USArrests['Rape'], USArrests['Murder'])

rape_rates = USArrests['Rape']
murder_rates = USArrests['Murder']

coefficients = np.polyfit(rape_rates, murder_rates, 1)
line = np.polyval(coefficients, rape_rates)

plt.scatter(rape_rates, murder_rates, color='blue')
plt.plot(rape_rates, line, color='red')

plt.xlabel('Rape Rates')
plt.ylabel('Murder Rates')
plt.title('Correlation between Murder and Rape Crimes')

plt.show()
```



This scatter plot displays a positive correlation between the rate of rape and murder crimes.

MURDER VS ASSAULT

In [ ]:
```
plt.scatter(USArrests['Murder'], USArrests['Assault'])

murder_rates = USArrests['Murder']
assault_rates = USArrests['Assault']

coefficients = np.polyfit(murder_rates, assault_rates, 1)
line = np.polyval(coefficients, murder_rates)

plt.scatter(murder_rates, assault_rates, color='blue')
plt.plot(murder_rates, line, color='red')

plt.xlabel('Murder Rates')
plt.ylabel('Assault Rates')
plt.title('Correlation between Assault and Murder Crimes')

plt.show()
```

This scatter plot displays a positive correlation between the rate of assault and murder crimes.

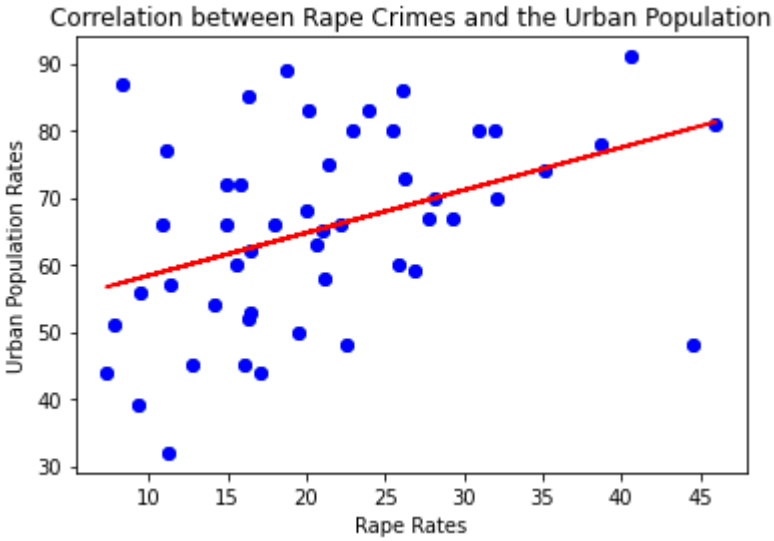
ASSAULT VS RAPE

In [ ]:
```
plt.scatter(USArrests['Assault'], USArrests['Rape'])

assault_rates = USArrests['Assault']
rape_rates = USArrests['Rape']

coefficients = np.polyfit(assault_rates, rape_rates, 1)
line = np.polyval(coefficients, assault_rates)

plt.scatter(assault_rates, rape_rates, color='blue')
plt.plot(assault_rates, line, color='red')

plt.xlabel('Assault Rates')
plt.ylabel('Rape Rates')
plt.title('Correlation between Assault and Rape Crimes')

plt.show()
```

This scatter plot displays a positive correlation between the rate of assault and rape crimes.

EFFECT OF RAPE ON THE URBAN POPULATION

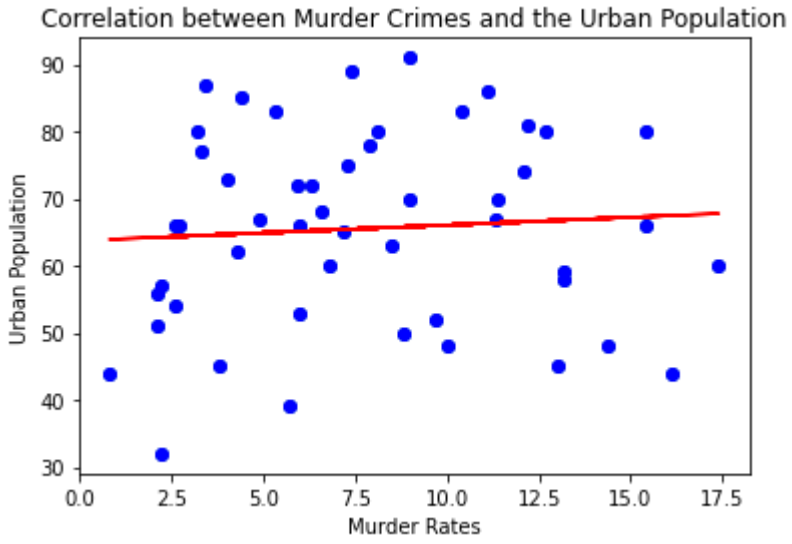
In [10]:
```
plt.scatter(USArrests['Rape'], USArrests['UrbanPop'])

rape_rates = USArrests['Rape']
urbanPop_rates = USArrests['UrbanPop']

coefficients = np.polyfit(rape_rates, urbanPop_rates, 1)
line = np.polyval(coefficients, rape_rates)

plt.scatter(rape_rates, urbanPop_rates, color='blue')
plt.plot(rape_rates, line, color='red')

plt.xlabel('Rape Rates')
plt.ylabel('Urban Population Rates')
plt.title('Correlation between Rape Crimes and the Urban Population')

plt.show()
```



This scatterplot doesn't display a negative correlation between rape crimes and the urban population but it is also not very strong.

EFFECT OF MURDER ON THE URBAN POPULATION

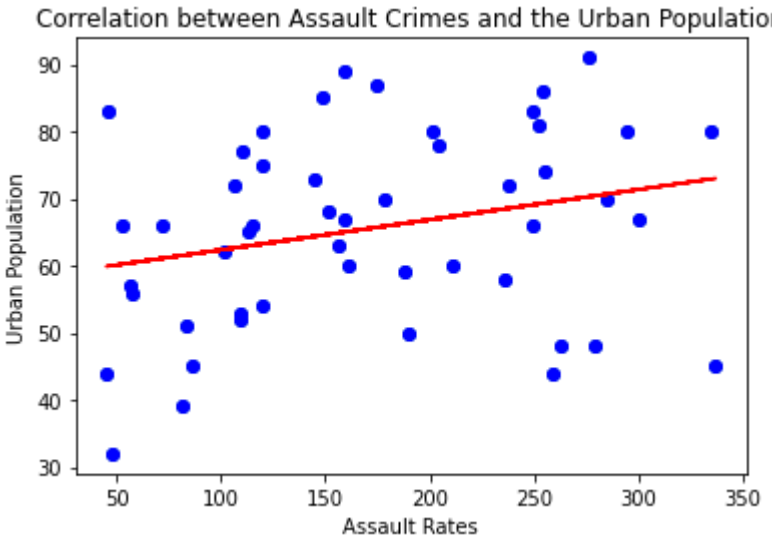
In [11]:
```
plt.scatter(USArrests['Murder'], USArrests['UrbanPop'])

murder_rates = USArrests['Murder']
urbanPop_rates = USArrests['UrbanPop']

coefficients = np.polyfit(murder_rates, urbanPop_rates, 1)
line = np.polyval(coefficients, murder_rates)

plt.scatter(murder_rates, urbanPop_rates, color='blue')
plt.plot(murder_rates, line, color='red')

plt.xlabel('Murder Rates')
plt.ylabel('Urban Population')
plt.title('Correlation between Murder Crimes and the Urban Populatio
n')

plt.show()
```



This scatterplot doesn't show a strong correlation between murder crimes and urban population.

EFFECT OF ASSAULT ON THE URBAN POPULATION

In [12]:
```
plt.scatter(USArrests['Assault'], USArrests['UrbanPop'])

assault_rates = USArrests['Assault']
urbanPop_rates = USArrests['UrbanPop']

coefficients = np.polyfit(assault_rates, urbanPop_rates, 1)
line = np.polyval(coefficients, assault_rates)

plt.scatter(assault_rates, urbanPop_rates, color='blue')
plt.plot(assault_rates, line, color='red')

plt.xlabel('Assault Rates')
plt.ylabel('Urban Population')
plt.title('Correlation between Assault Crimes and the Urban Populatio
n')

plt.show()
```



This scatterplot seems to have a weak correlation between assualt crimes and the urban population.

CONCLUSION

There seems to be a positive correlation between all crimes in this data set, which answers my first question of whether each violent crime has some type of correlation with one another.

And the correlation between the urban population and each violent crime individually does not seem to have much of a correlation positive or negative.