Crime and Policing Expenditures Descriptive Analysis

Unifying Data Science

1/27/20

In [2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

Exercise 1

In [3]:
```
# Load data
crime = pd.read_csv('https://media.githubusercontent.com/media/nickeubank/MIDS_Data/master/descriptive_exercise/crime_expend_MA.csv')
```

In [4]:
```
crime.head()
```

Out[4]:

	months	county_code	crimeindex	policeexpenditures	month	year
0	0	1	61.411101	32.331110	1	1990
1	0	10	92.779361	59.342067	1	1990
2	0	11	93.222701	50.481508	1	1990
3	0	12	95.588374	65.815540	1	1990
4	0	13	92.472719	38.337757	1	1990

Exercise 2 A

In [5]:
```
# Create filter for counties of interest
county_filter = [4, 10]
```

In [19]:
```
# Print average statistics
# Repeat for each county of interest
for c in county_filter:
    df_sub = crime[crime['county_code']==c] # Create subset of crime df
    print(f'for county {c}, average policing expenditure is {df_sub.policeexpe
nditures.mean():.1f}'
          f' and average crime index is {df_sub.crimeindex.mean():.2f}')
```

```
for county 4, average policing expenditure is 54.3 and average crime index is
47.83
for county 10, average policing expenditure is 54.2 and average crime index i
s 47.77
```

Exercise 2 B

In [18]:
```
# Print standard deviation statistics
# Repeat for each county of interest
for c in county_filter:
    df_sub = crime[crime['county_code']==c] # Create subset of crime df
    print(f'for county {c}, standard deviation policing expenditure is '
          f'{df_sub.policeexpenditures.std():.1f} and standard deviation crime
index is '
          f'{df_sub.crimeindex.std():.2f}')
```

```
for county 4, standard deviation policing expenditure is 16.8 and standard de
viation crime index is 26.94
for county 10, standard deviation policing expenditure is 16.7 and standard d
eviation crime index is 27.00
```

Exercise 3

In [20]:
```
# Print correlation between policing expenditures and crime rate
# Repeat for each county of interest
for c in county_filter:
    # Create subset of crime df
    df_sub = crime[crime['county_code']==c]
    # Calculate correlations
    county_corr = df_sub[['policeexpenditures', 'crimeindex']].corr().iloc[0,1
]
    print(f'for county {c}, the correlation between crime and police expenditu
res is {county_corr:.3f}')
```

```
for county 4, the correlation between crime and police expenditures is -0.064
for county 10, the correlation between crime and police expenditures is -0.05
6
```

Exercise 4

Based on our basic descriptive analysis, there does not seem to be a strong relationship between policing expenditures and crime rates. There is a very small negative correlation between the two, indicating that higher police expenditures is **slightly** associated to lower crime rates. However this correlation is so small that this finding is almost negligible.

Exercise 5

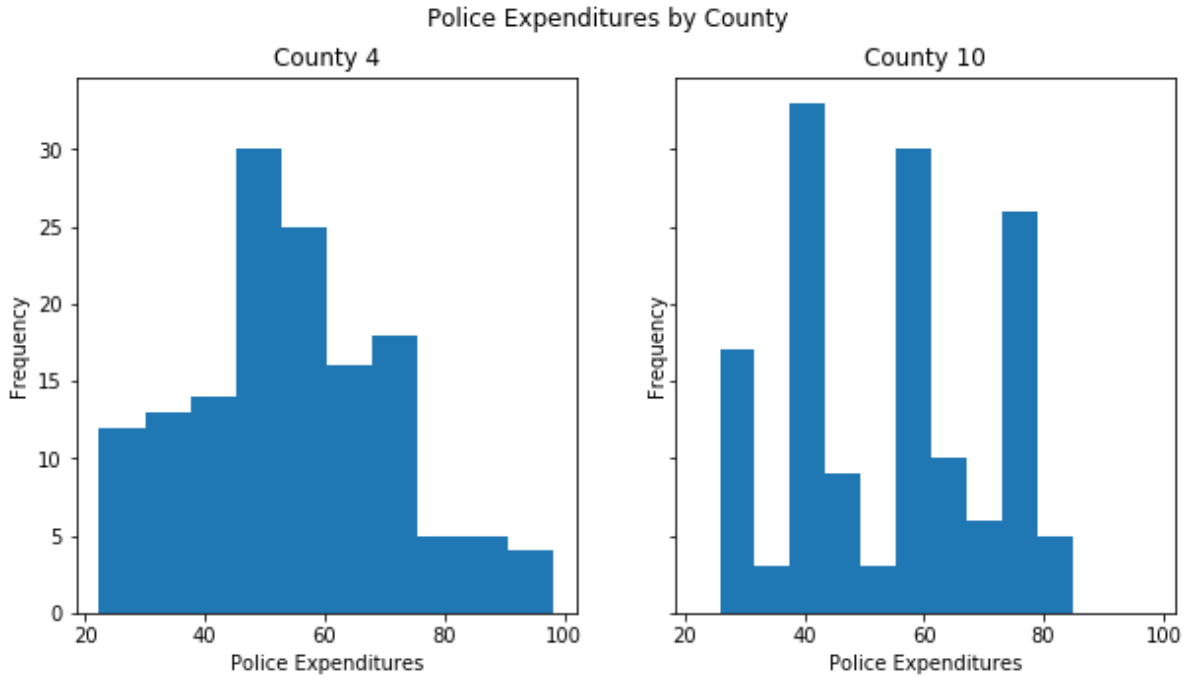
Again, based on our basic descriptive analysis, both crime rates and police expenditures seem to be very similar between country 4 and 10. Both the means and standard deviations of these variables are very comparable across counties. Additionally, the relationship between crime rates and police expenditures is very similar across counties. Given what we have seen so far, there is no evidence that the relationship betwen crime and police expenditures are different between counties.

Exercise 6

In [21]:
```
# Histograms of policing expenditures by county (4 and 10)
# Create subplot space
fig, ax = plt.subplots(1,2, sharex=True, sharey=True, figsize=(10,5))

# Fill subplot with county-speific data
for i, c in enumerate(county_filter):
    ax[i].hist('policeexpenditures', data = crime[crime['county_code']==c]);
    ax[i].set_title(f'County {c}')
    ax[i].set_xlabel('Police Expenditures')
    ax[i].set_ylabel('Frequency')

# Add figure attributes
plt.suptitle('Police Expenditures by County')
plt.show()
```



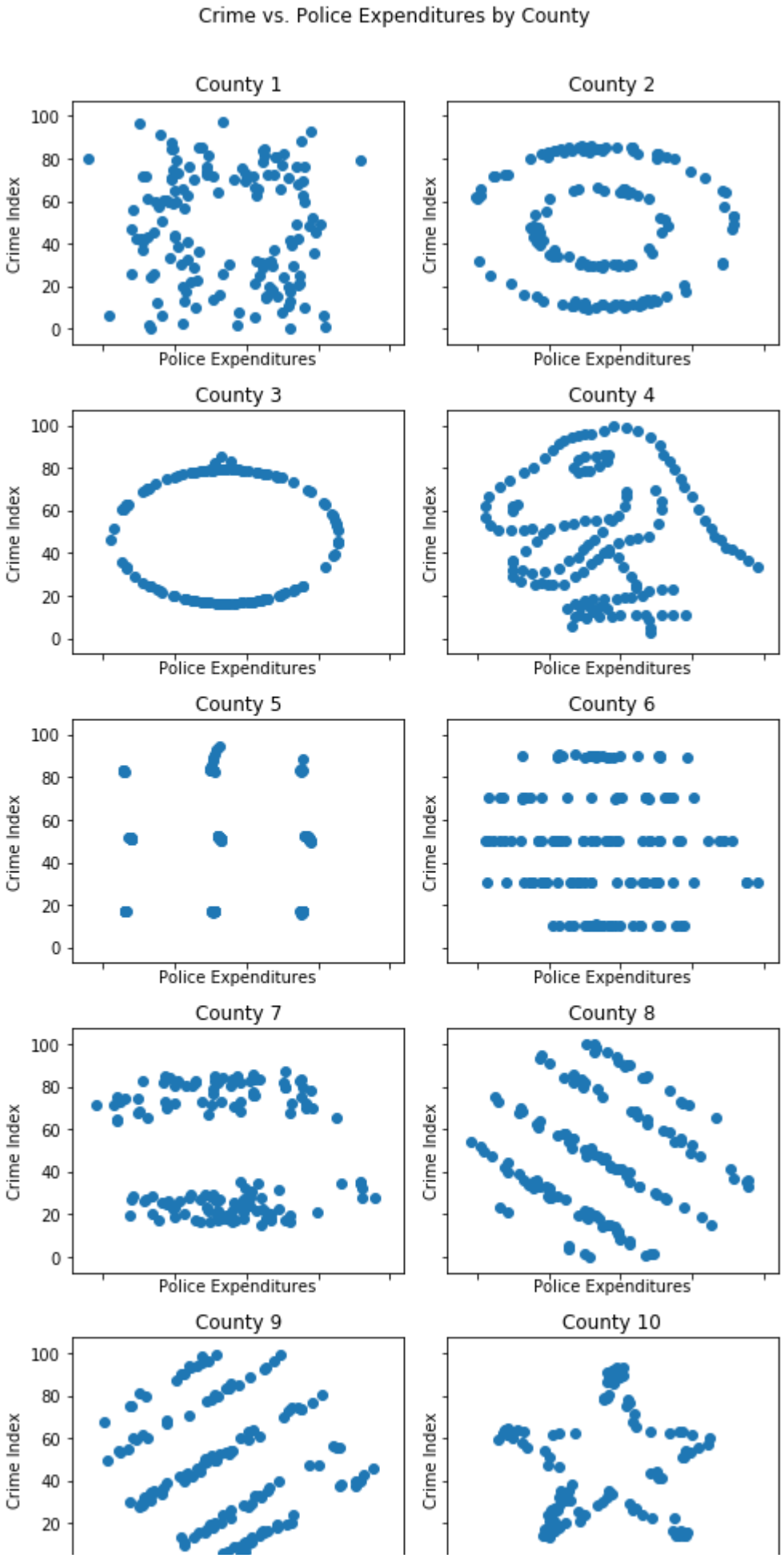
These results **do** change our impression of the similarity of country 4 and 10. County 4 seems to have a relatively normal distribution of police expenditures whereas county 10 seems to have police stations that fall into four buckets of police expenditure levels.

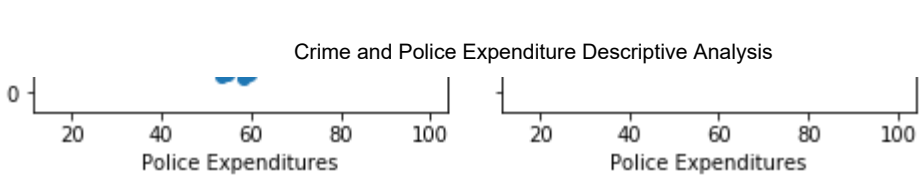
Exercise 7

In [22]:
```
# Scatterplots of policing expenditures and crime by county
# Create subplot space
fig, ax = plt.subplots(5,2, sharex=True, sharey=True, figsize=(7,15))

# Fill subplot with county-specific data
# All counties included
for i, c in enumerate(list(range(1,11))):
    row, col = int(i/2), i%2
    df_sub = crime[crime['county_code']==c] # Create subset of crime df
    ax[row, col].scatter(df_sub['policeexpenditures'], df_sub['crimeindex'])
    ax[row, col].set_title(f'County {c}')
    ax[row, col].set_xlabel('Police Expenditures')
    ax[row, col].set_ylabel('Crime Index')

# Add figure attributes
plt.suptitle('Crime vs. Police Expenditures by County')
plt.tight_layout(rect=[0, 0.03, 1, .95])
plt.show()
```


Crime vs. Police Expenditures by County
County 1
County 2
County 3
County 4
County 5
County 6
County 7
County 8
County 9
County 10
Crime Index
Police Expenditures
0
20
40
60
80
100



This absolutely changes our impression of how similar these two counties are. The relationship between police expenditures and crime are completely different between the two counties. County 4 shows a standard Tyrannosaurus relationship, where as county 10 shows a rather celestial one. Like we said, completely different. The plots for the other counties are also shown to highlight the differences between all the counties.