Crime and Policing Expenditures Exploratory Questions

In this exercise we'll be examining the relationship between crime and policing expenditures using county-level data from Massachusetts. In particular, we're hoping to answer the question "Is there a substantial relationship between crime and policing expenditures?"

Exercises

Exercise 1

Begin by downloading the data for this exercise from https://github.com/nickeubank/MIDS_Data/blob/master/descriptive_exercise/crime_expend_ (just go to `github.com/nickeubank/MIDS_Data`, then go to `descriptive_exercise` and get `crime_expend_MA.csv` if you don't want to type all that).

Exercise 2

This data includes monthly data on both each county's policing expenditures (`policeexpenditures` as share of county budget) and an index of crime (`crimeindex`, scaled 0-100) from 1990 to late 2001.

In these exercises, we'll be focusing on just two counties -- `county_code` 4 and 10.

First, for each of these two counties, calculate the mean expenditure level and mean crimeindex score (i.e. calculate both means separately for each county).

Just to make sure we're practicing applied skills—use a loop to calculate your means and print your results nicely! So you should get output like this (though obviously with different numbers—I'm not gonna give you the answer!):

```
for county 4, average policing expenditure is 23.7 and
average crime index is 75.83
for county 10, average policing expenditure is 62.15 and
average crime index is 55.88
```

In [ ]:
```
import pandas as pd
import numpy as np

bd = pd.read_csv(
    "https://github.com/nickeubank/MIDS_Data/raw/master/descriptive_exercise
```

```
)
bd_s = bd[bd["county_code"].isin([4, 10])]
```

In [ ]:
```
for i in bd_s["county_code"].unique():
    bd_loop = bd_s[bd_s["county_code"] == i]
    mean_police = bd_loop["policeexpenditures"].mean()
    mean_crime = bd_loop["crimeindex"].mean()
    print(
        f"for county {i}, average policing expenditure is {mean_police:.2f}
    )
```

```
for county 10, average policing expenditure is 54.24 and average crime index
is 47.77
for county 4, average policing expenditure is 54.26 and average crime index
is 47.83
```

Exercise 3

Now calculate the standard deviation of both expenditures and crime for these two counties.

In [ ]:
```
for i in bd_s["county_code"].unique():
    bd_loop = bd_s[bd_s["county_code"] == i]
    sd_police = bd_loop["policeexpenditures"].std()
    sd_crime = bd_loop["crimeindex"].std()
    print(
        f"for county {i}, the standard deviation of policing expenditure is
    )
```

```
for county 10, the standard deviation of policing expenditure is 16.68 and t
he standard deviation of crime index is 27.00
for county 4, the standard deviation of policing expenditure is 16.77 and th
e standard deviation of crime index is 26.94
```

Exercise 4

Now calculate the correlation between `policeexpenditures` and `crimeindex` for both of these counties (again, output the correlations with nicely formatted and labelled statements!)

In [ ]:
```
for i in bd_s["county_code"].unique():
    bd_loop = bd_s[bd_s["county_code"] == i]
    corr_pc = bd_loop["policeexpenditures"].corr(bd_loop["crimeindex"])
    print(
        f"for county {i}, the correlation of policing expenditure and crime
    )
```

```
for county 10, the correlation of policing expenditure and crime index is -
0.06
for county 4, the correlation of policing expenditure and crime index is -0.
06
```

In [ ]:
```
bd_s[bd_s["county_code"] == 4]
```

Out[ ]:

	months	county_code	crimeindex	policeexpenditures	month	year
7	0	4	97.1795	55.3846	1	1990
20	1	4	96.0256	51.5385	2	1990
33	2	4	94.4872	46.1538	3	1990
46	3	4	91.4103	42.8205	4	1990
59	4	4	88.3333	40.7692	5	1990
...	...	...	...	...	...	...
1788	137	4	25.3846	39.4872	6	2001
1801	138	4	41.5385	91.2821	7	2001
1814	139	4	95.7692	50.0000	8	2001
1827	140	4	95.0000	47.9487	9	2001
1840	141	4	92.6923	44.1026	10	2001

142 rows × 6 columns

Exercise 5

Based on your results up to this point, what would you guess about whether policing reduces crime? (I know -- this is just a descriptive statistics, and correlation does not imply causality. But what would you infer if this was all you knew?

> The correlation for both counties is similar with a negative value of -0.06, from this negative relation we can imply that as policy expenditure increases, crime rate diminishes. However, as the correlation is very low: -0.06 very small so this impact may not be singificant.

Exercise 6

Given what you've seen up till now, would you infer that county 4 and county 10 have a similar relationship between crime and police expenditures?

> Having almost the same correlation indicates that the linear relationship between the crime and policy moves in the same way, indicating a similar relationship. However, although the mean and standard deviations for both counties are similar, it is necessary to visualize the distribution of the values to understand if there are any key differences.

Exercise 7

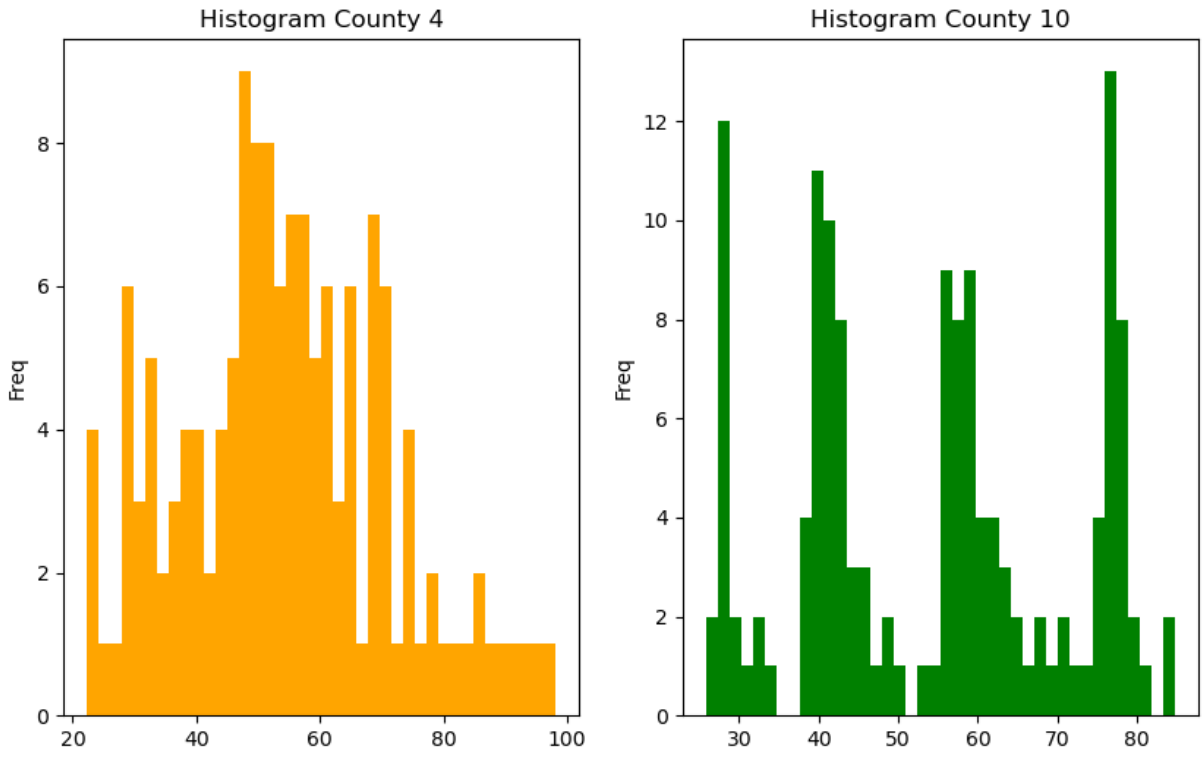
Now plot histograms of `policeexpenditures` for both county 4 and county 10. Do the results change you impression of the similarity of county 4 and county 10?

In [ ]:
```
import seaborn as sns
import matplotlib.pyplot as plt
import warnings

warnings.simplefilter(action="ignore", category=FutureWarning)

fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(10, 6))  # nrows, ncols of axe
plt.sca(ax1)
plt.hist(bd_s[bd_s["county_code"] == 4]["policeexpenditures"], bins=40, colo
plt.ylabel("Freq")
plt.title("Histogram County 4")
plt.sca(ax2)
plt.hist(bd_s[bd_s["county_code"] == 10]["policeexpenditures"], bins=40, col
plt.ylabel("Freq")
plt.title("Histogram County 10")
```

Out[ ]:
```
Text(0.5, 1.0, 'Histogram County 10')
```



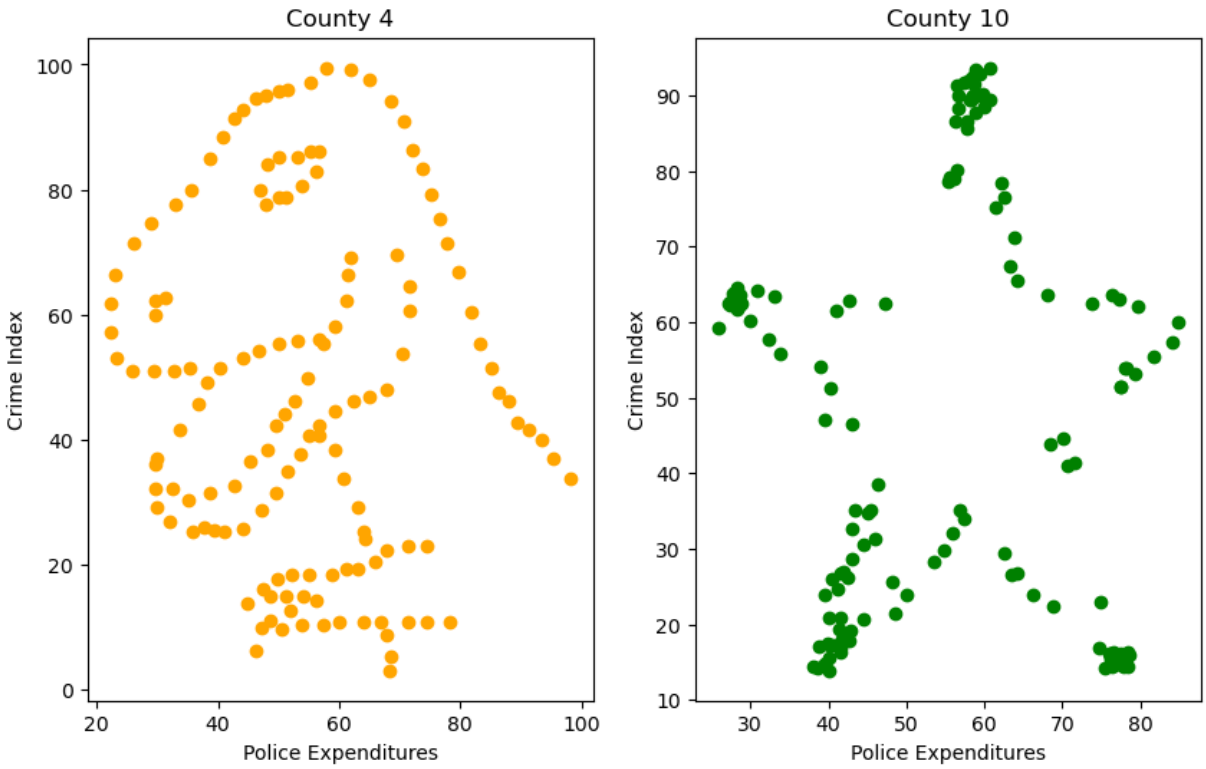
> yes

Exercise 8

Finally, create a scatter plot of the relationship between crime and police expenditures for each county (e.g. crime on one axis, police expenditures on the other). Does this change your sense of how similar these are?

In [ ]:
```
fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(10, 6))  # nrows, ncols of axe
plt.sca(ax1)
plt.scatter(
    bd_s[bd_s["county_code"] == 4]["policeexpenditures"],
    bd_s[bd_s["county_code"] == 4]["crimeindex"],
    color="orange",
)
plt.xlabel("Police Expenditures")
plt.ylabel("Crime Index")
plt.title("County 4")
plt.sca(ax2)
plt.scatter(
    bd_s[bd_s["county_code"] == 10]["policeexpenditures"],
    bd_s[bd_s["county_code"] == 10]["crimeindex"],
    color="green",
)
plt.xlabel("Police Expenditures")
plt.ylabel("Crime Index")
plt.title("County 10")
```

Out[ ]:
```
Text(0.5, 1.0, 'County 10')
```



> yes

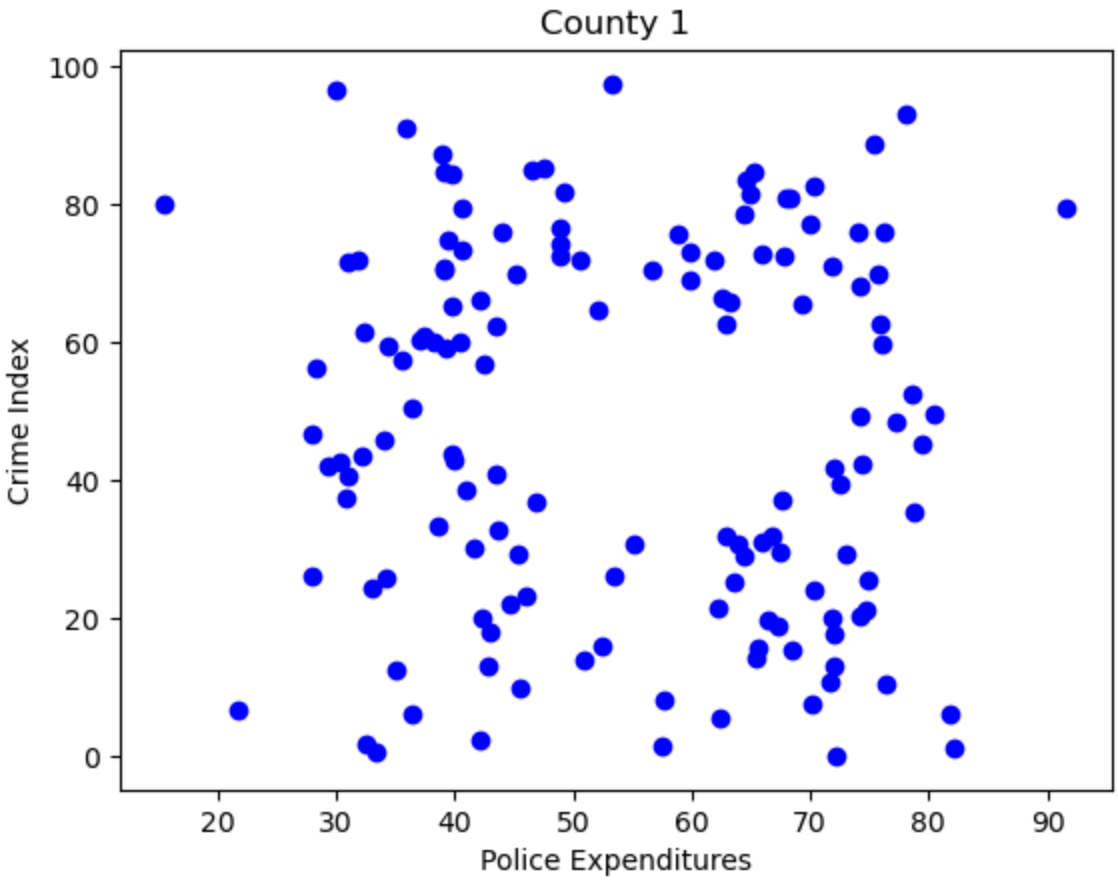
After you have answered...

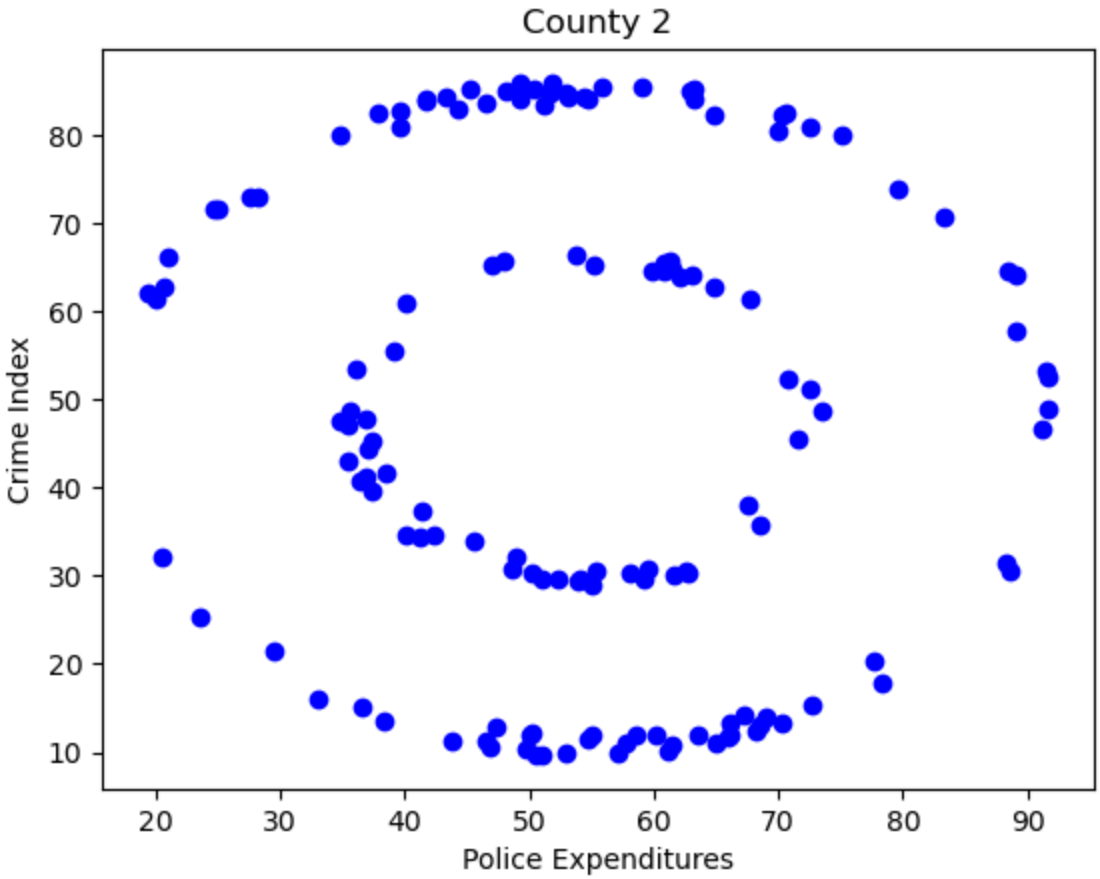
Read this discussion page.

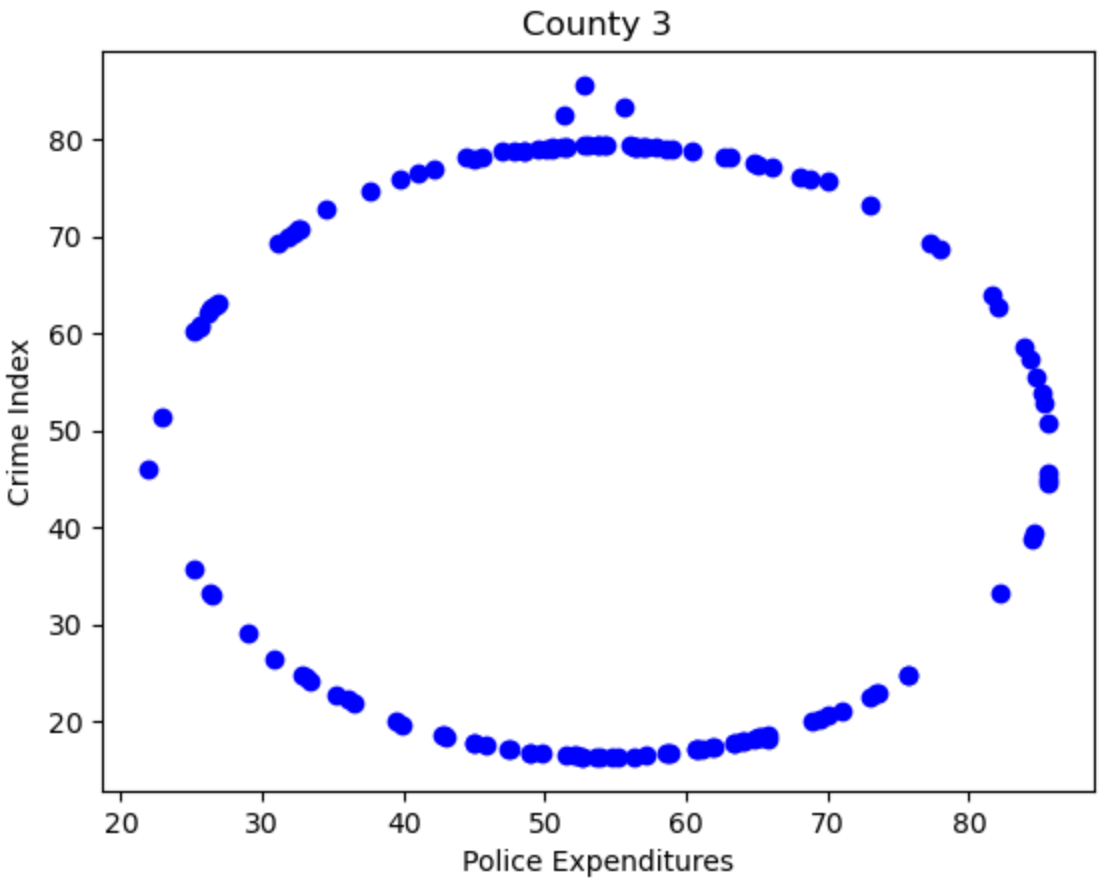
Exercise 9

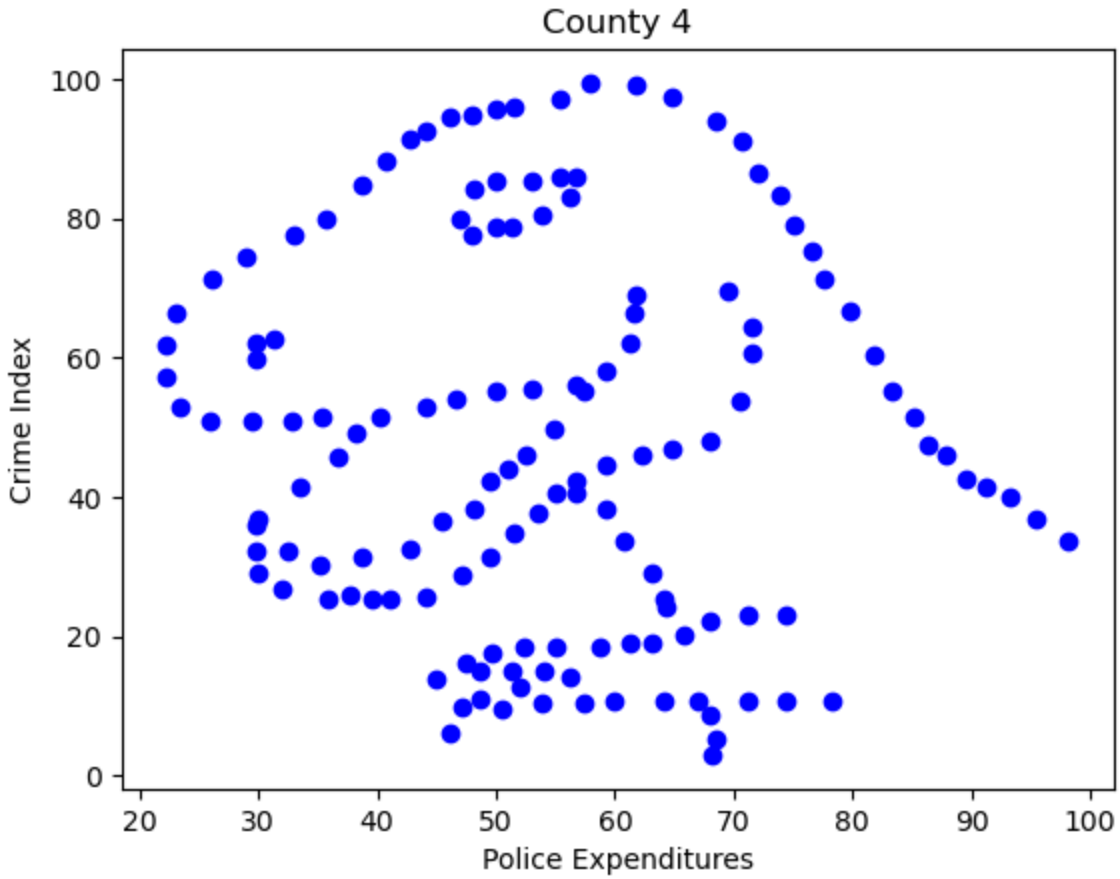
Write a loop that plots the relationship between police expenditures and crimeindex for all 13 counties in the dataset separately!

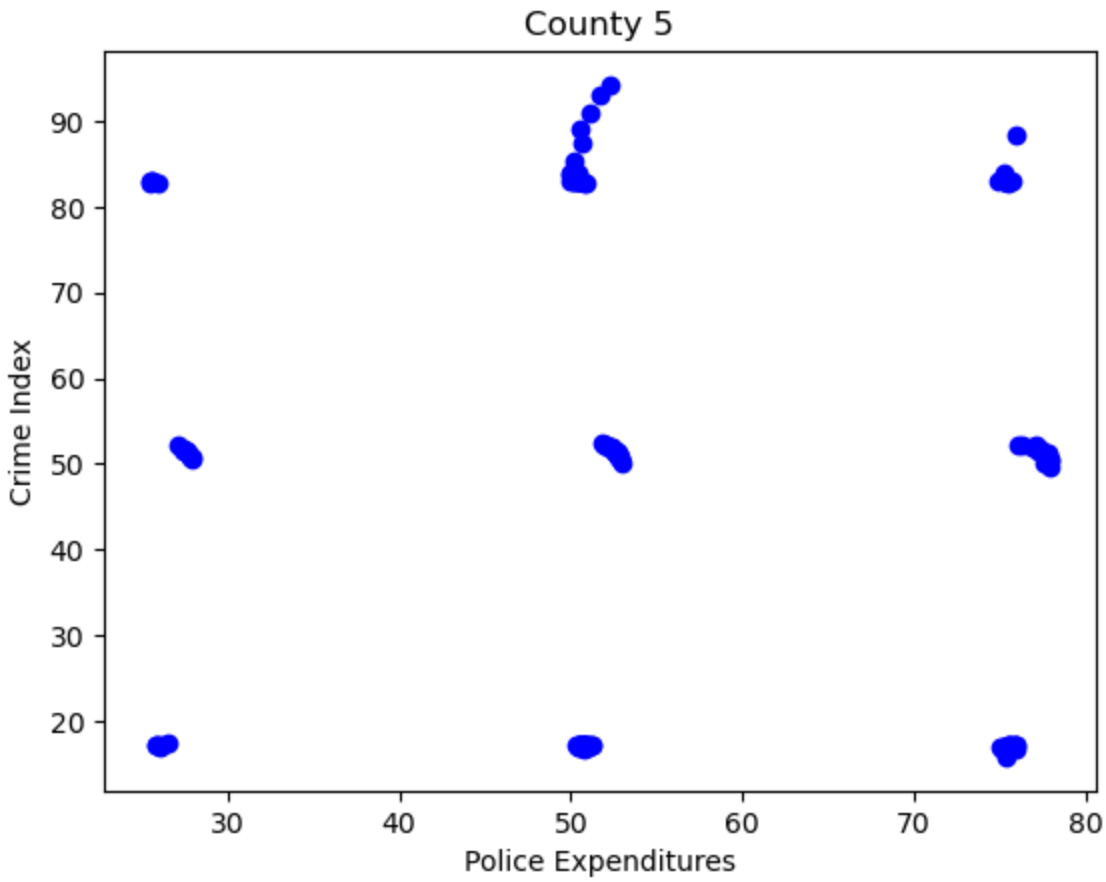
In [ ]:
```
for i in range(1, 14):
    plt.figure()
    plt.scatter(
        bd[bd["county_code"] == i]["policeexpenditures"],
        bd[bd["county_code"] == i]["crimeindex"],
        color="blue",
    )
    plt.xlabel("Police Expenditures")
    plt.ylabel("Crime Index")
    plt.title(f"County {i}")
plt.show()
```

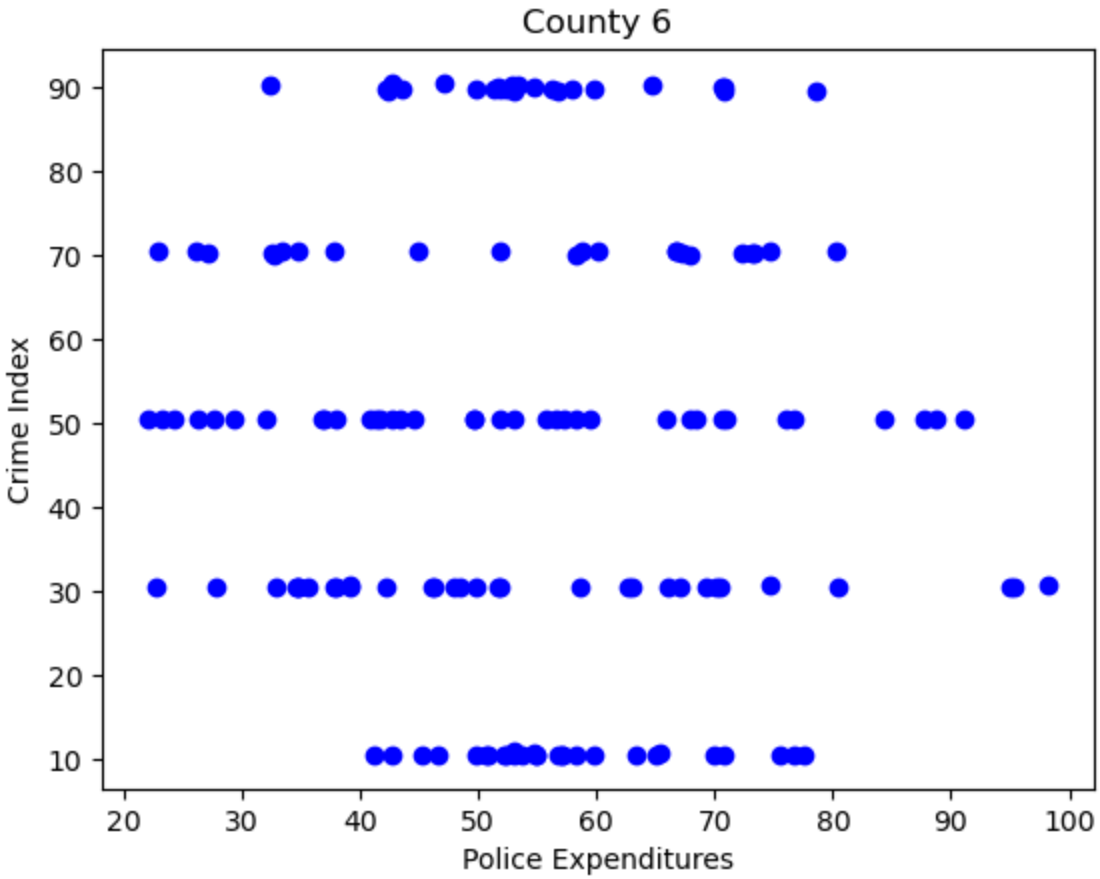


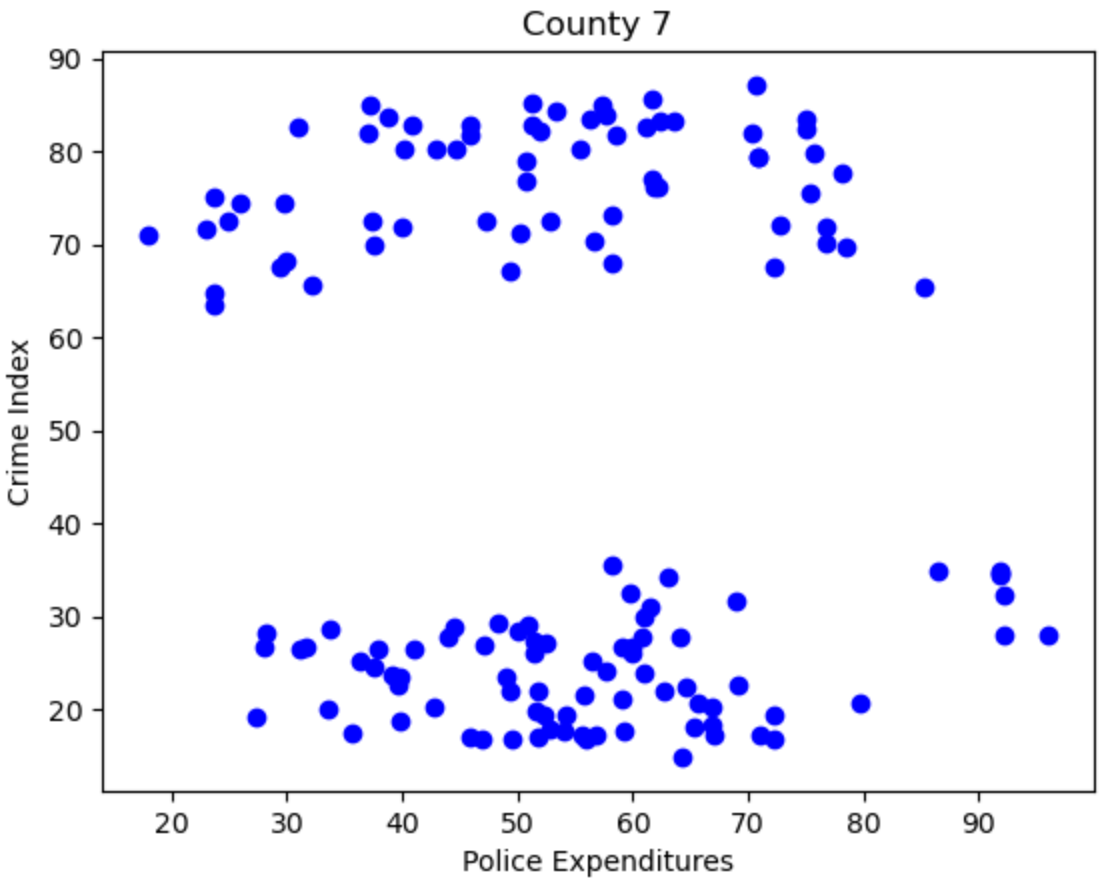

County 2
Crime Index
Police Expenditures
10
20
30
40
50
60
70
80
20
30
40
50
60
70
80
90

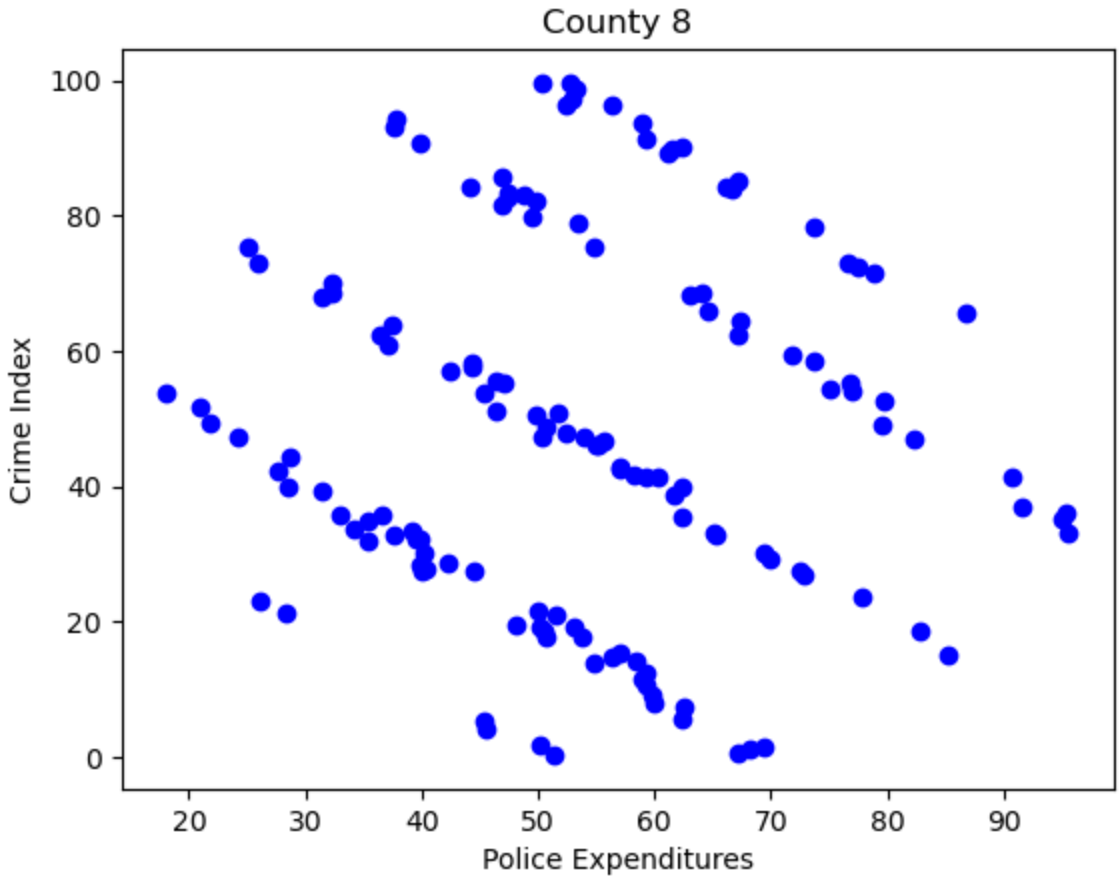

County 3
Crime Index
Police Expenditures
20
30
40
50
60
70
80
20
30
40
50
60
70
80

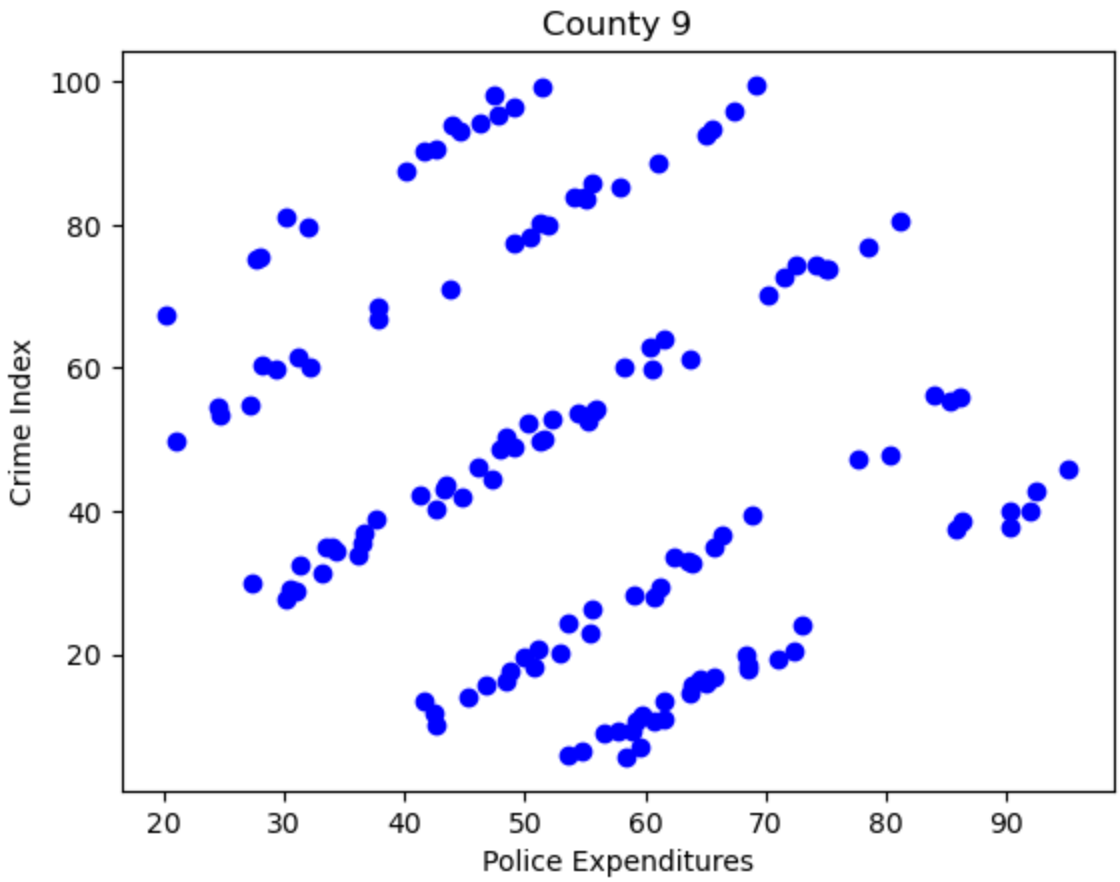

County 4
Crime Index
Police Expenditures
0
20
40
60
80
100
20
30
40
50
60
70
80
90
100

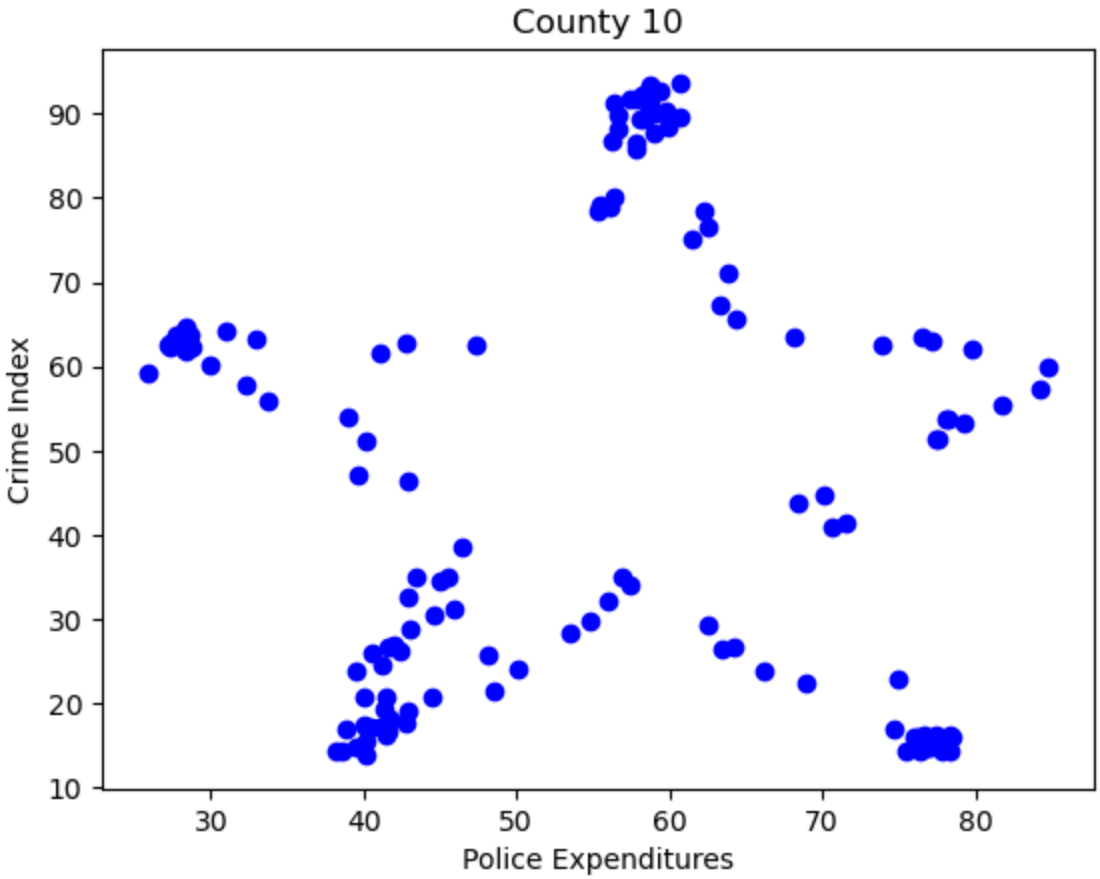

County 5
Crime Index
Police Expenditures
20
30
40
50
60
70
80
90
30
40
50
60
70
80

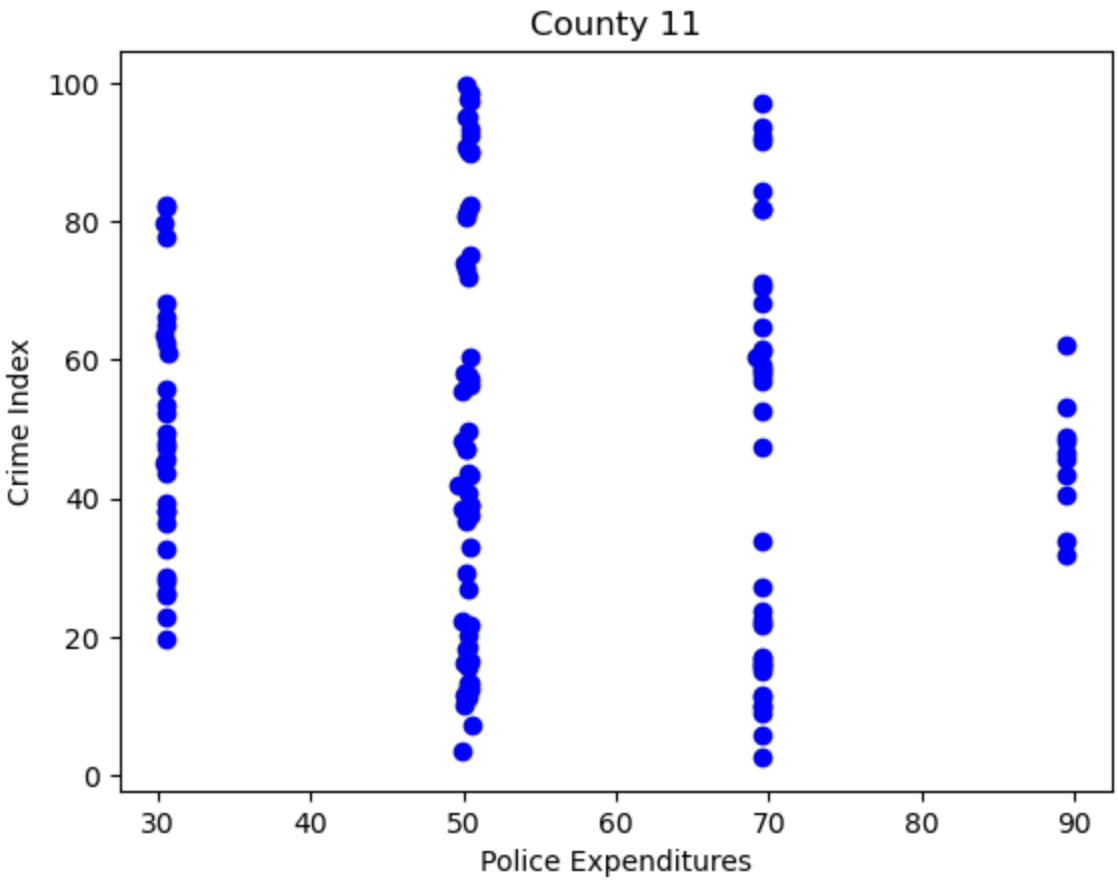

County 6
Crime Index
Police Expenditures
90
80
70
60
50
40
30
20
10
20
30
40
50
60
70
80
90
100

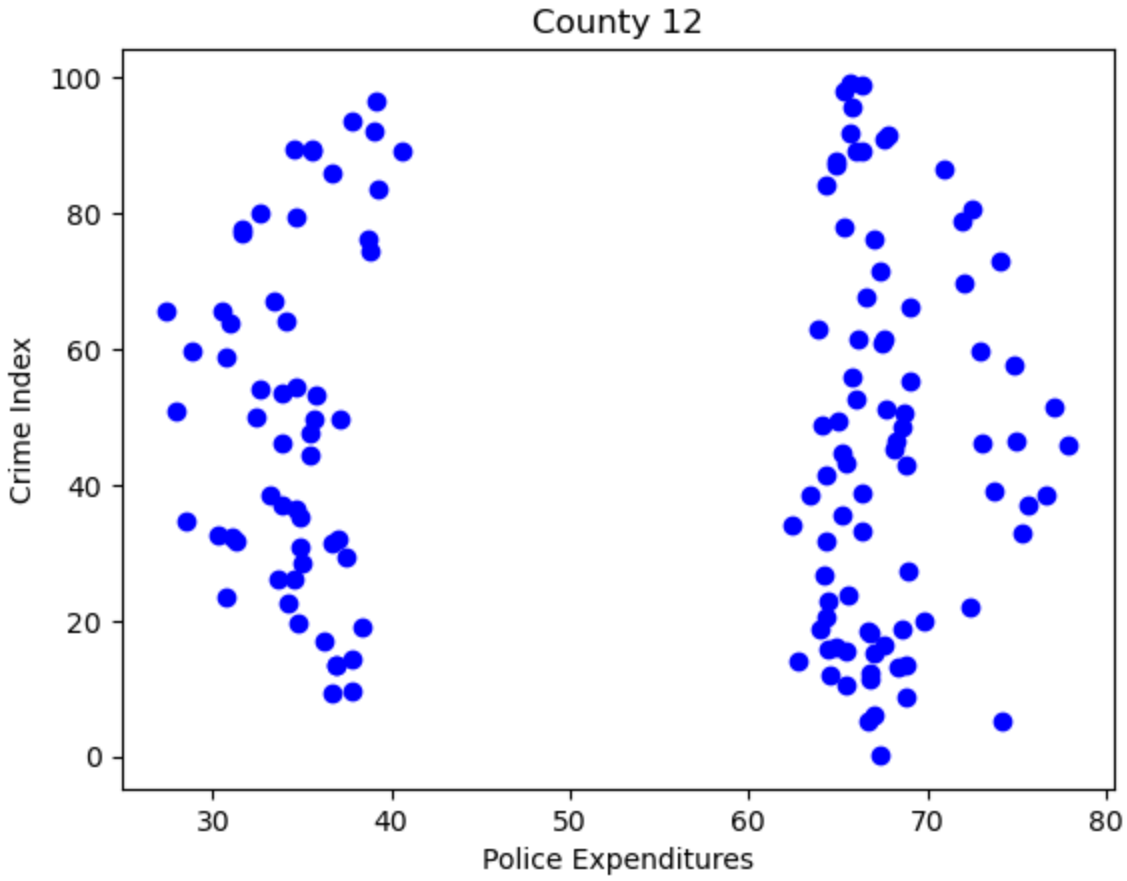

County 7
Crime Index
Police Expenditures
90
80
70
60
50
40
30
20
20
30
40
50
60
70
80
90

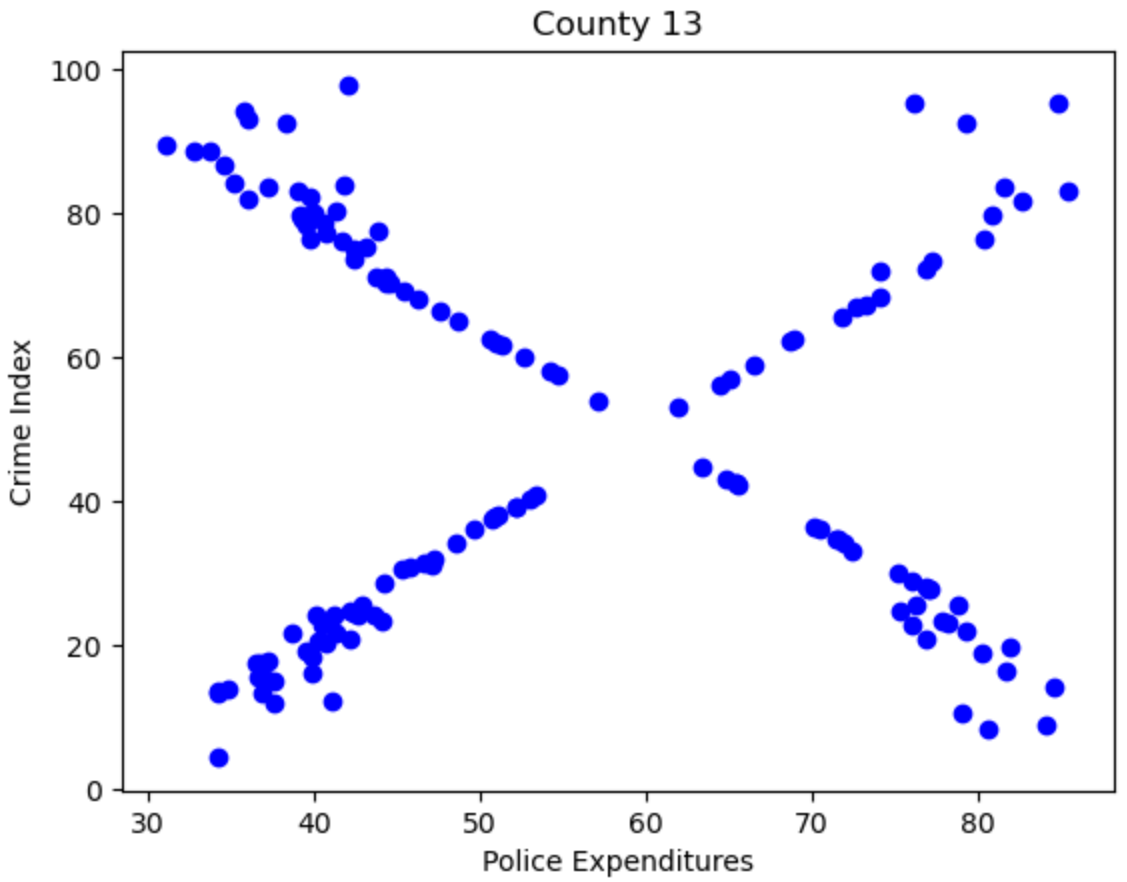

County 8
Crime Index
Police Expenditures
0
20
40
60
80
100
20
30
40
50
60
70
80
90


County 9
Crime Index
Police Expenditures
20
40
60
80
100
20
30
40
50
60
70
80
90


County 10
Crime Index
Police Expenditures
10
20
30
40
50
60
70
80
90
30
40
50
60
70
80


County 11
Crime Index
Police Expenditures
0
20
40
60
80
100
30
40
50
60
70
80
90


County 12
Crime Index
Police Expenditures
0
20
40
60
80
100
30
40
50
60
70
80


County 13
Crime Index
Police Expenditures
0
20
40
60
80
100
30
40
50
60
70
80