Crime and Policing Expenditures Descriptive Analysis

Jiaxin Ying, Yuanjing Zhu

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

Exercise 1: Load the dataset

```
In [ ]: df = pd.read_csv("https://media.githubusercontent.com/media/nickeubank/MIDS_Data/"
                         "master/descriptive_exercise/crime_expend_MA.csv")
        df.head()
```

Out[ ]:

	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: calculate the mean expenditure level and mean crimeindex score for county_code 4 and 10.

```
In [ ]: county = [4,10]
        df_county=df[df['county_code'].isin(county)]
        df_county.head()
```

Out[ ]:

	months	county_code	crimeindex	policeexpenditures	month	year
1	0	10	92.779361	59.342067	1	1990
7	0	4	97.179500	55.384600	1	1990
14	1	10	93.633645	60.638669	2	1990
20	1	4	96.025600	51.538500	2	1990
27	2	10	91.531520	58.727422	3	1990

```
In [ ]: for code in county:
            mean_policeexpenditures=round(df[df['county_code']==code]['policeexpenditures'].mean(),2)
            mean_crimeindex=round(df[df['county_code']==code]['crimeindex'].mean(),2)
            print('For county ' + str(code)+', average policing expenditure is '\
                + str(mean_policeexpenditures) + ' and average crime index is ' + str(mean_crimeindex))

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

Exercise 3: calculate the standard deviation of both expenditures and crime for county_code 4 and 10.

```
In [ ]: for code in county:
            std_policeexpenditures=round(df[df['county_code']==code]['policeexpenditures'].std(),2)
            std_crimeindex=round(df[df['county_code']==code]['crimeindex'].std(),2)
            print('For county ' + str(code)+', the standard deviation of policing expenditure is ' \
                +str(std_policeexpenditures)+' and the standard deviation of crime index is '+str(std_crimeindex))

        For county 4, the standard deviation of policing expenditure is 16.77 and the standard deviation of crime index
        is 26.94
        For county 10, the standard deviation of policing expenditure is 16.68 and the standard deviation of crime inde
        x is 27.0
```

Exercise 4: calculate the correlation between policeexpenditures and crimeindex for county_code 4 and 10.

```
In [ ]: for code in county:
            corr = df[df['county_code']==code]['crimeindex'].corr(df[df['county_code']==code]['policeexpenditures'])
            corr = round(corr,4)
            print('For county ' + str(code)+', the correlation between policeexpenditures and crime index is '\
                + str(corr))
```

```
For county 4, the correlation between policeexpenditures and crime index is -0.0645
For county 10, the correlation between policeexpenditures and crime index is -0.056
```

Exercise 5: what would you guess about whether policing reduces crime?

Based on the results up to this point, increasing the policy expenditure dose not reduce the crime index; in contrast, it has a inreasing trend as policy expenditure increses.

Exercse 6: would you infer that county 4 and county 10 have a similar relationship between crime and police expenditures?

Based on the results up to the point, we think that county 4 and county 10 are likely to have a similiar relationship between crime and policy expenditure because they both have similar mean, standard deviation and correlation.

Exercise 7: plot histograms of policeexpenditures for both county 4 and county 10.

```
In [ ]: import matplotlib.pyplot as plt
        for code in county:
            plt.figure(figsize=(7, 5))
            plt.hist(df[df['county_code']==code]['policeexpenditures'], bins=30, edgecolor='black', linewidth=1.2)
            plt.xlabel('Police expenditures')
            plt.ylabel('Counts')
            plt.title('Hitogram plot of policy expenditures for county '+str(code))
            plt.show()
```


Hitogram plot of policy expenditures for county 4

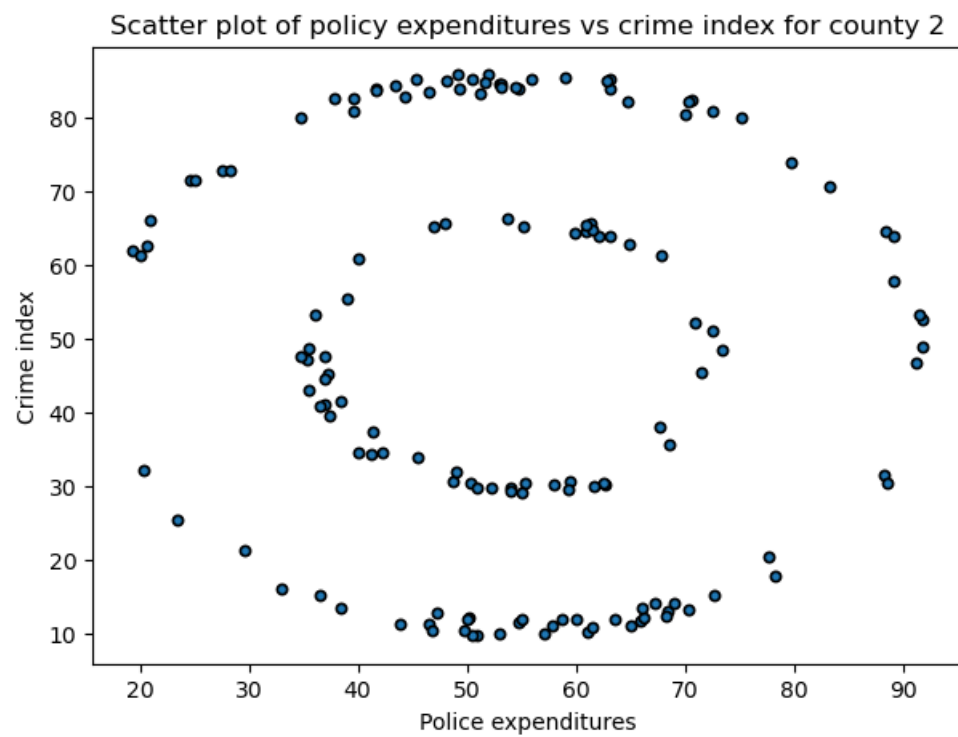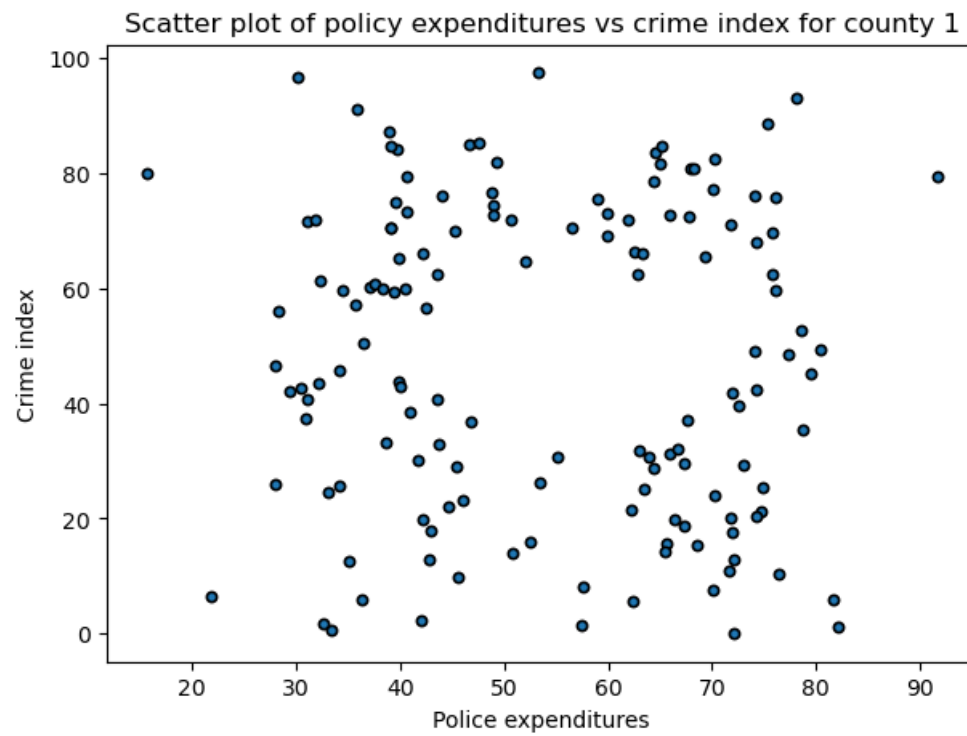Hitogram plot of policy expenditures for county 10
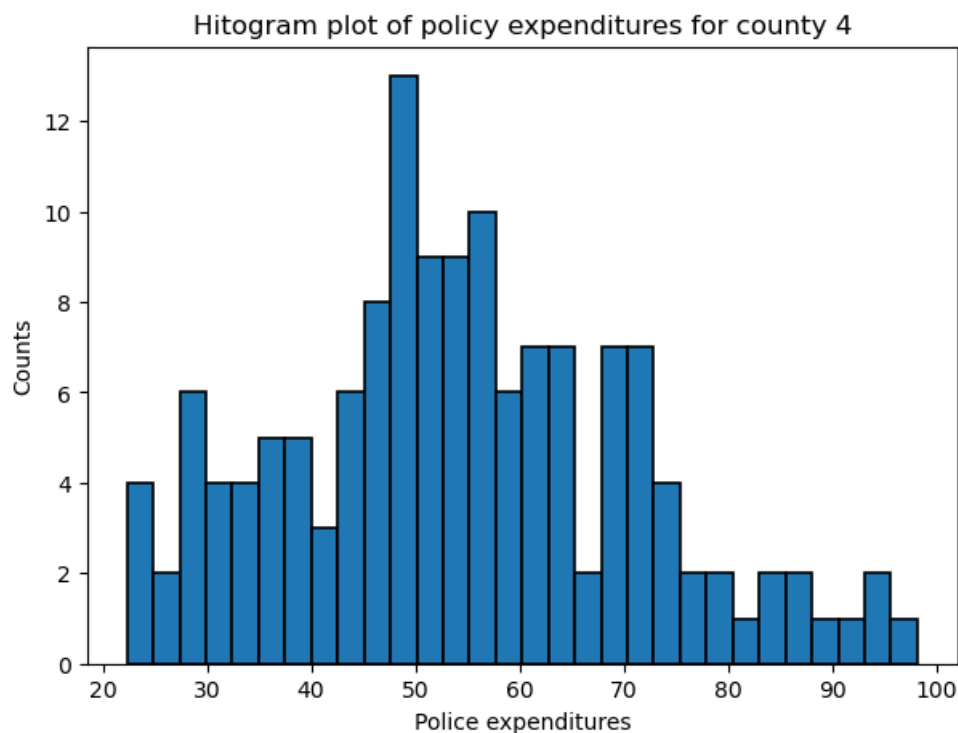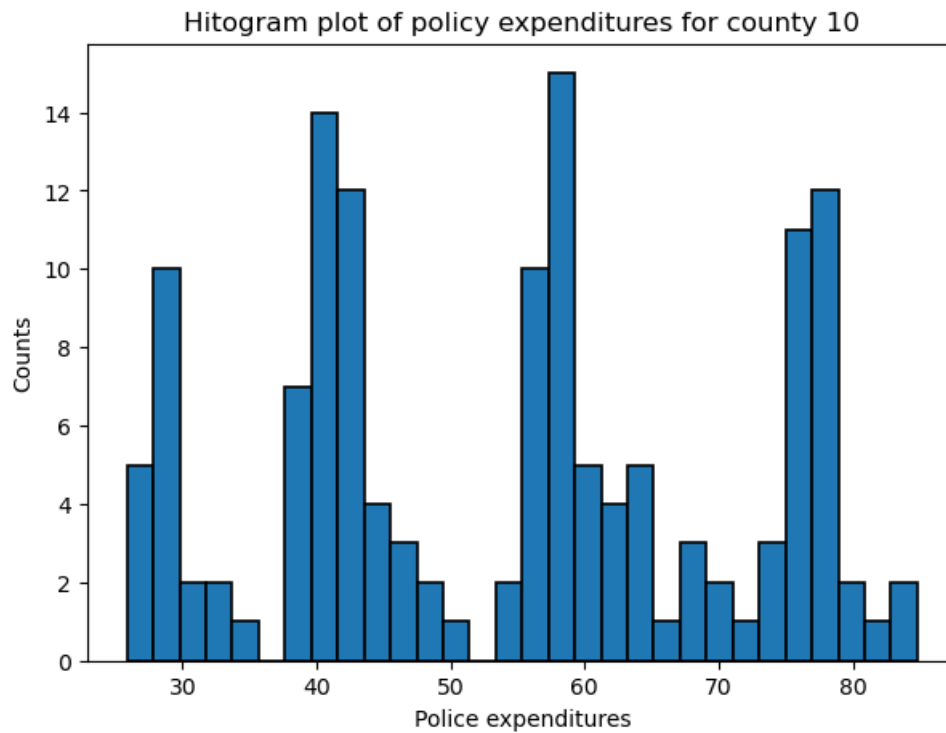
The histograms of county 4 and county 10 indicate a dissimilar distribution of police expenditures. Despite some similarities in their descriptive statistics, the distribution is vastly different.

Exercise 8: create a scatter plot of the relationship between crime and police expenditures for country_code 4 and 10.

```python
for code in county:
    plt.figure(figsize=(7, 5))
    plt.scatter(df[df['county_code']==code]['policeexpenditures'],df[df['county_code']==code]['crimeindex'],\
                s = 20, edgecolors='black', linewidth=1.2)
    plt.xlabel('Police expenditures')
    plt.ylabel('Crime index')
    plt.title('Scatter plot of policy expenditures vs crime index for county '+str(code))
    plt.show()
```

Scatter plot of policy expenditures vs crime index for county 4

Scatter plot of policy expenditures vs crime index for county 10
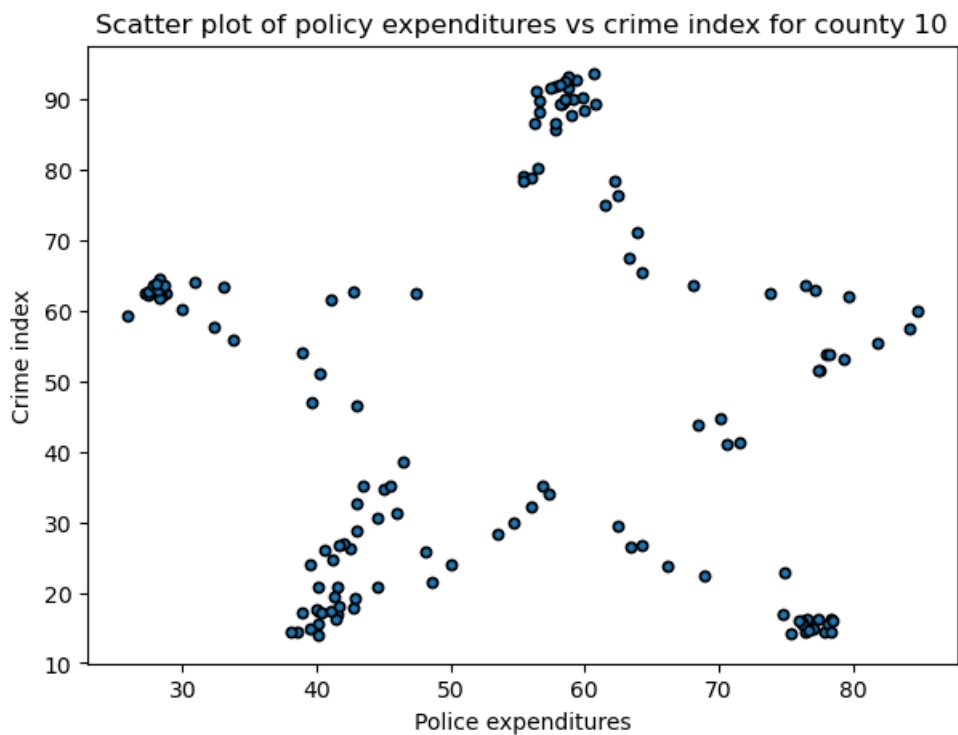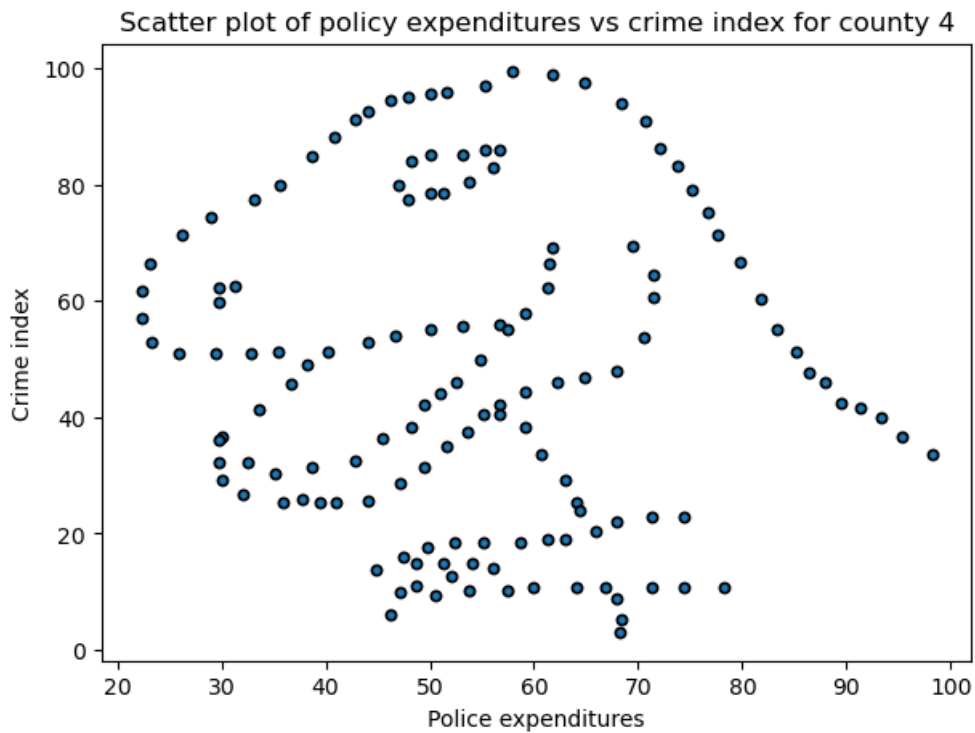
The scatter plot completely changes our perception of the similarity in the relationship between crime and police expenditures for county 4 and county 10. Even though their correlation score is similar, the relationship and distribution of crime index and police expenditures show distinct patterns.

Exercise 8: plot the relationship between policeexpenditures and crimeindex for all 13 counties

```
In [ ]: for code in range(1,14):
            plt.figure(figsize=(7, 5))
            plt.scatter(df[df['county_code']==code]['policeexpenditures'],df[df['county_code']==code]['crimeindex'],\
                        s = 20, edgecolors='black', linewidth=1.2)
            plt.xlabel('Police expenditures')
            plt.ylabel('Crime index')
```

```
plt.title('Scatter plot of policy expenditures vs crime index for county '+str(code))
plt.show()
```

Scatter plot of policy expenditures vs crime index for county 1

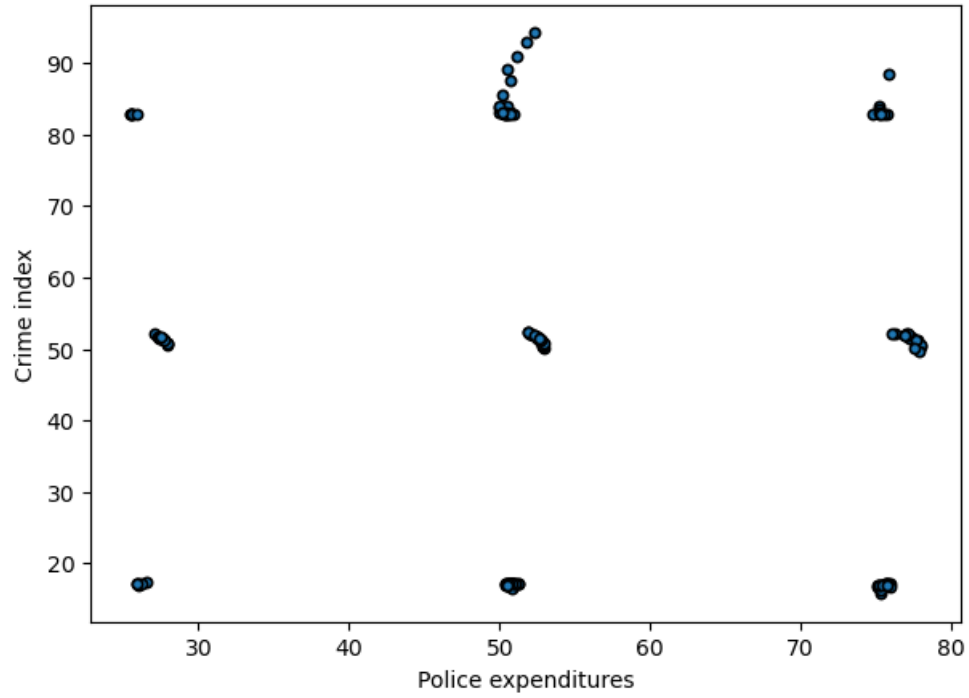Scatter plot of policy expenditures vs crime index for county 2

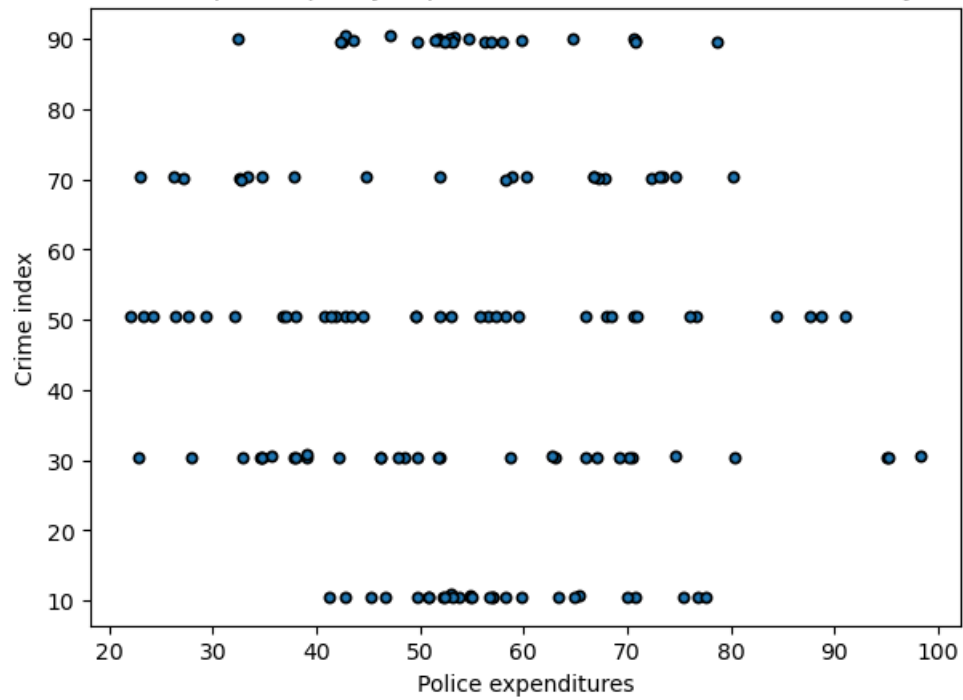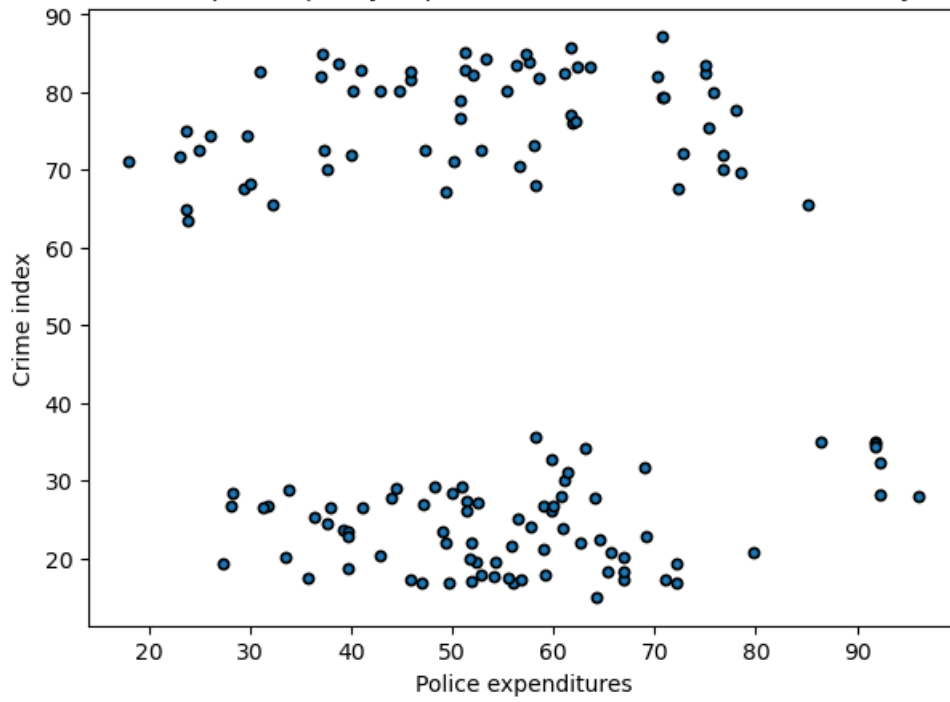Scatter plot of policy expenditures vs crime index for county 3
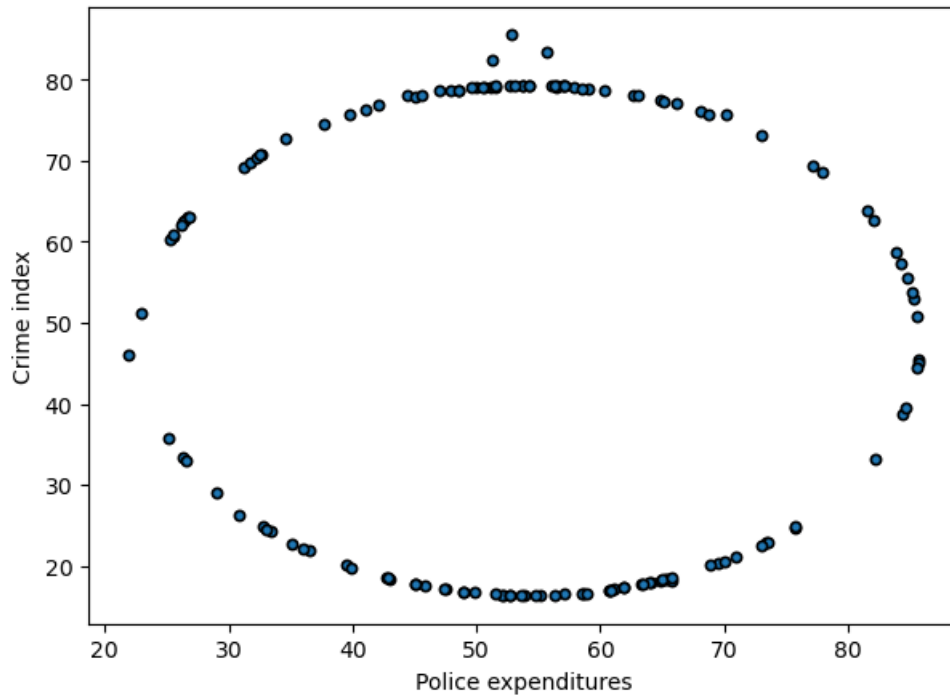
Scatter plot of policy expenditures vs crime index for county 4
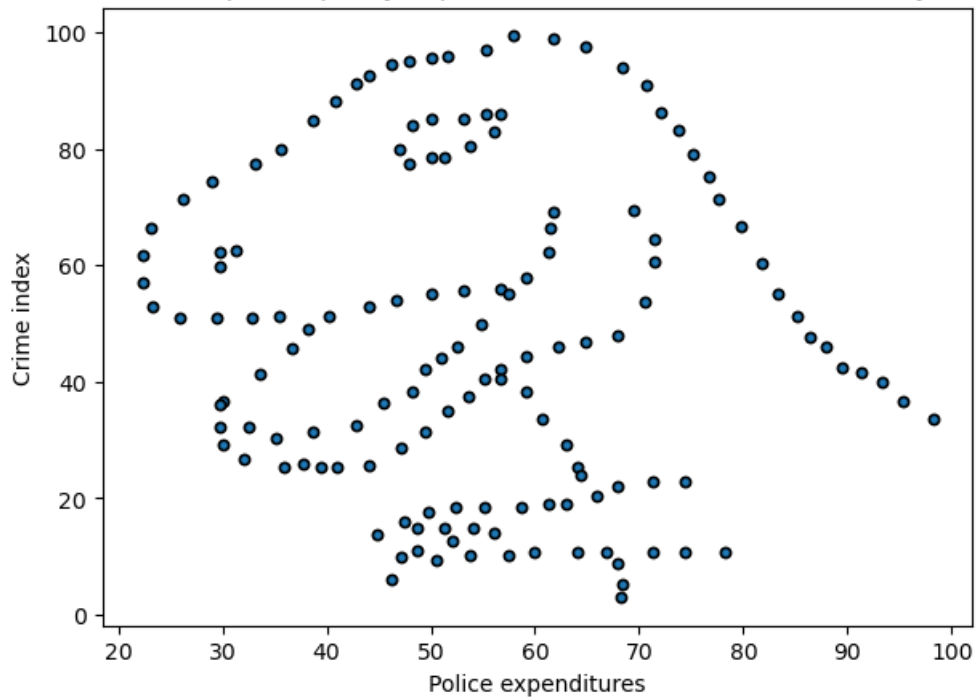
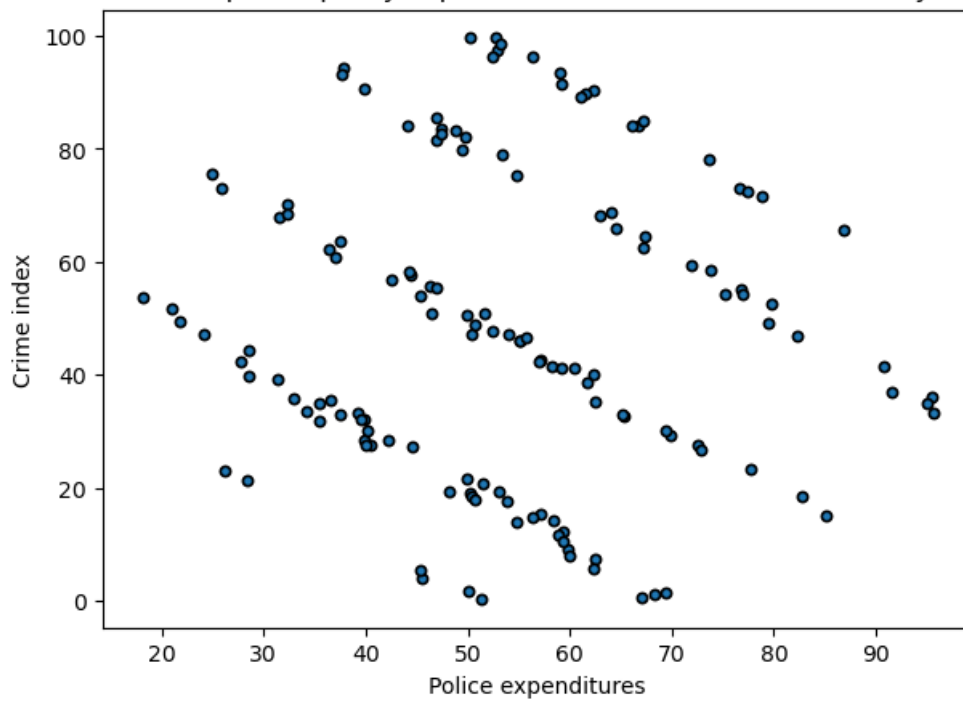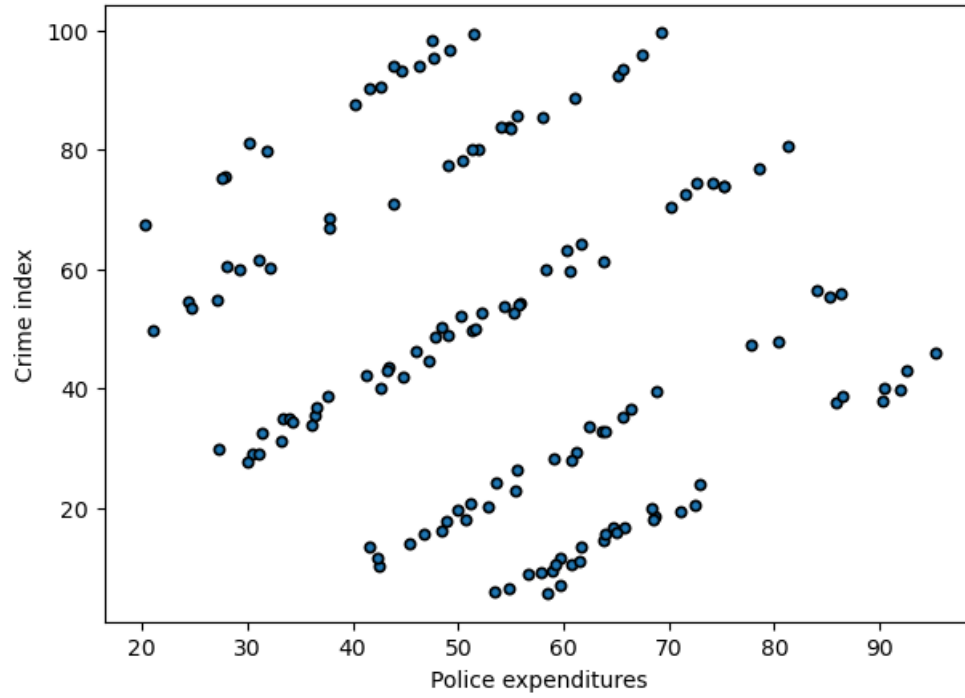Scatter plot of policy expenditures vs crime index for county 5

Scatter plot of policy expenditures vs crime index for county 6

Scatter plot of policy expenditures vs crime index for county 7

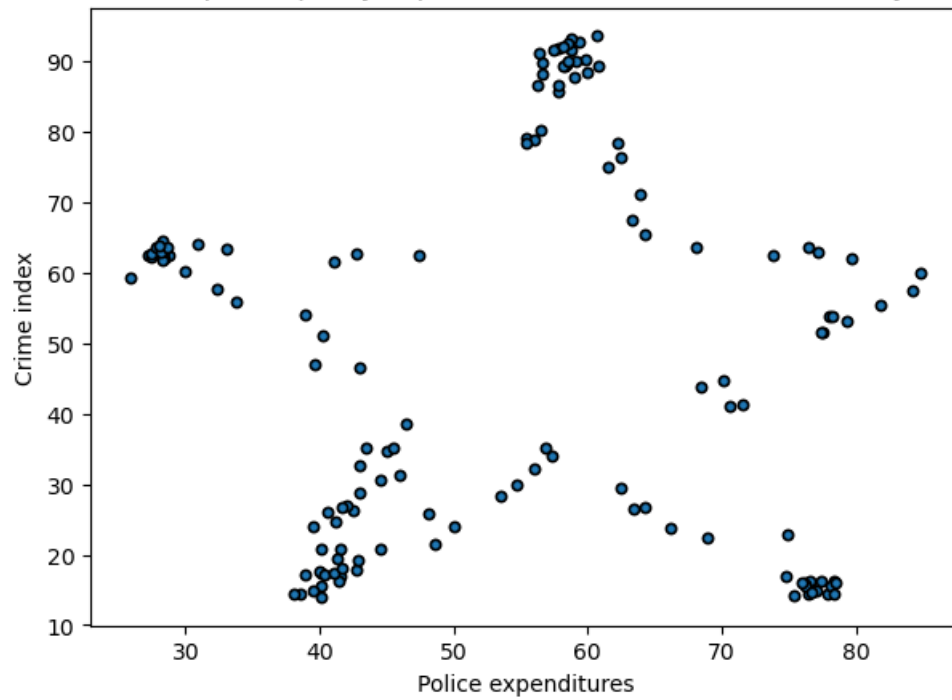Scatter plot of policy expenditures vs crime index for county 8
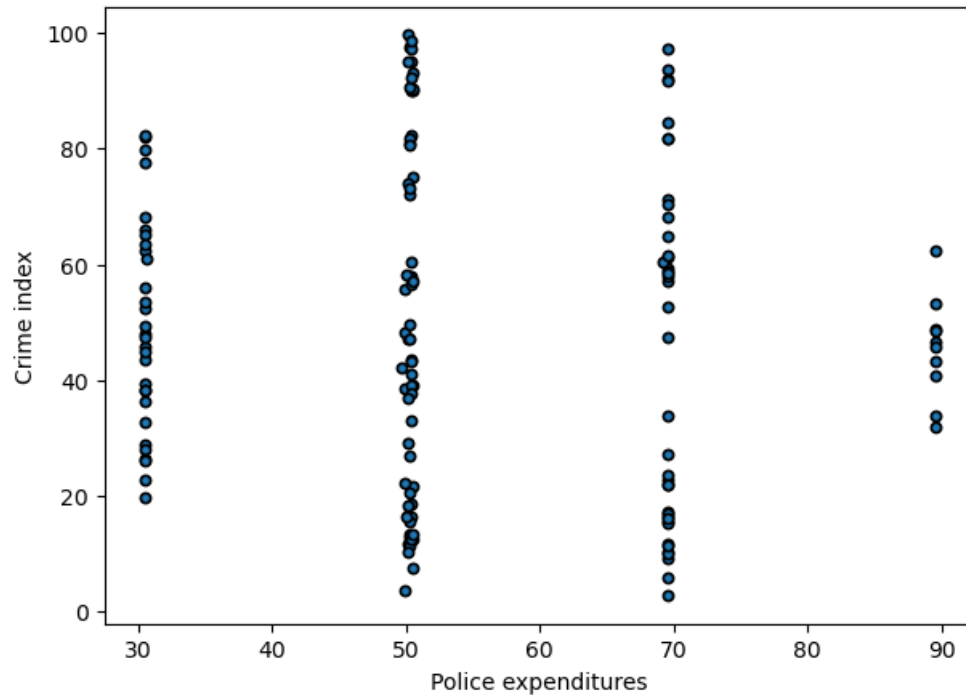
Scatter plot of policy expenditures vs crime index for county 9
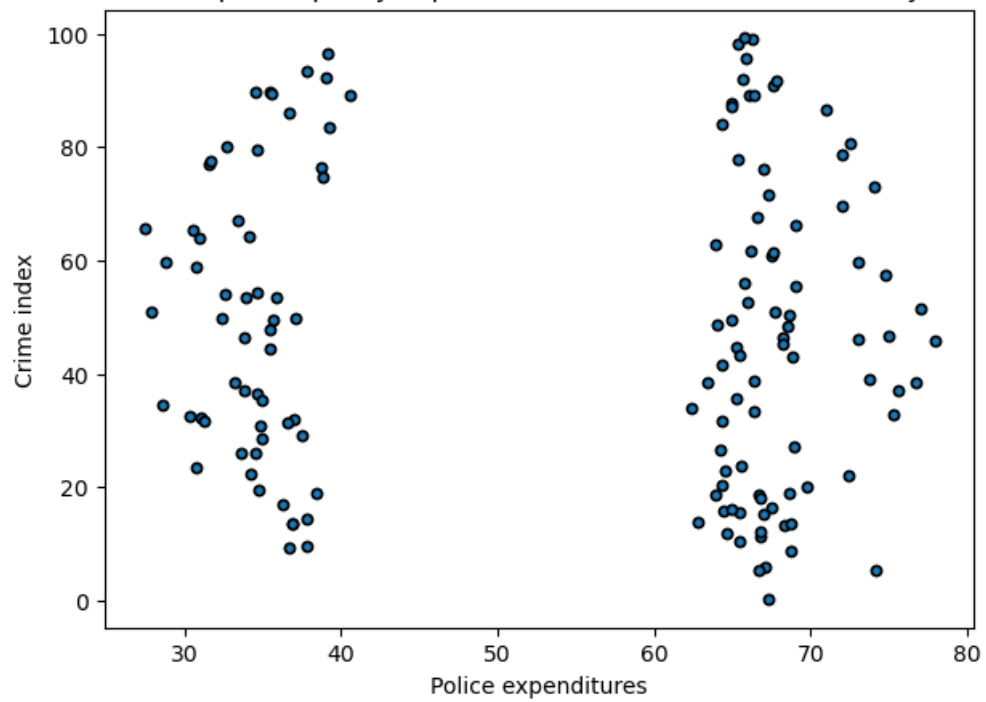

Scatter plot of policy expenditures vs crime index for county 10

Scatter plot of policy expenditures vs crime index for county 11

Scatter plot of policy expenditures vs crime index for county 12

Scatter plot of policy expenditures vs crime index for county 13