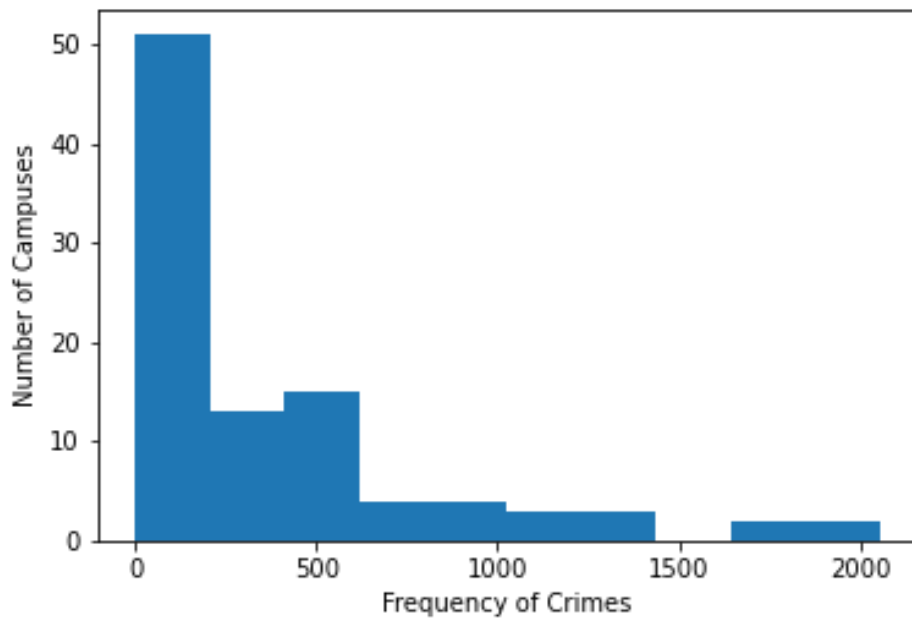


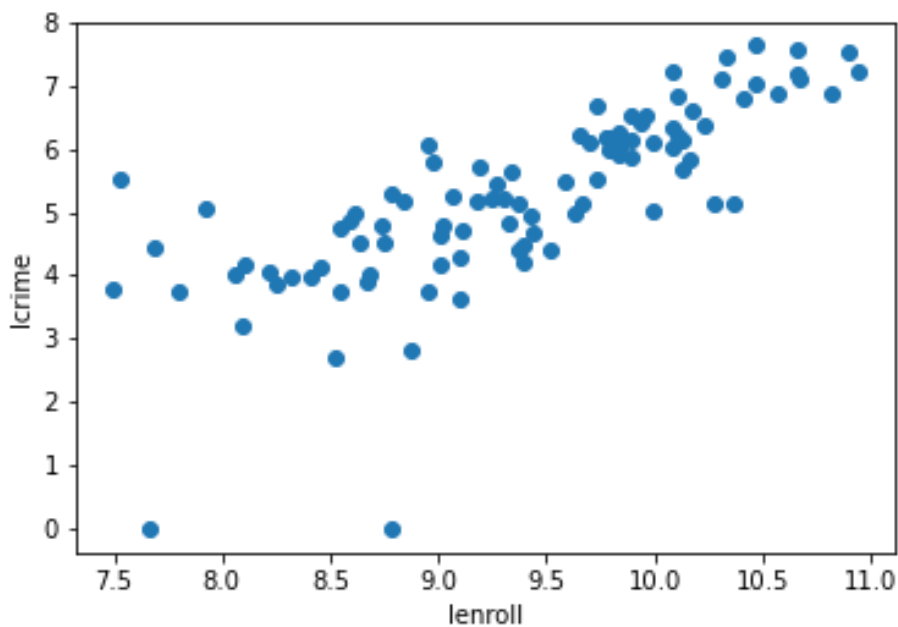
## ECON 406 Problem Set 4

1.1



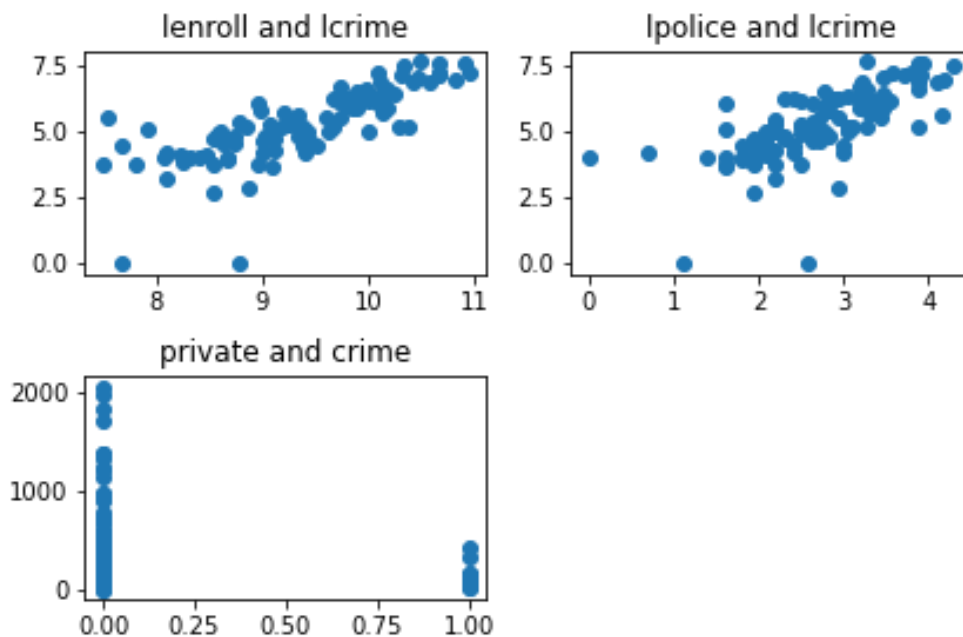
Comment: The histogram shows a unimodal distribution that skews to the right. It also shows that most college campuses have low levels of crimes.

1.2



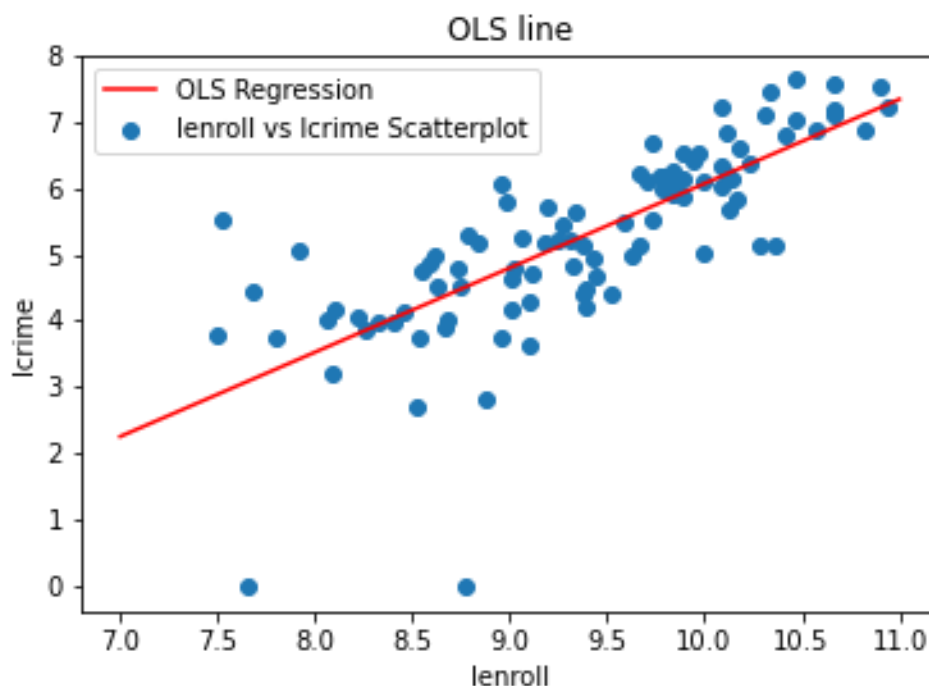
Comment: This scatter plot shows a positive, moderate to strong correlation between log of number of students enrolled on campus and log of levels of crime.

1.3

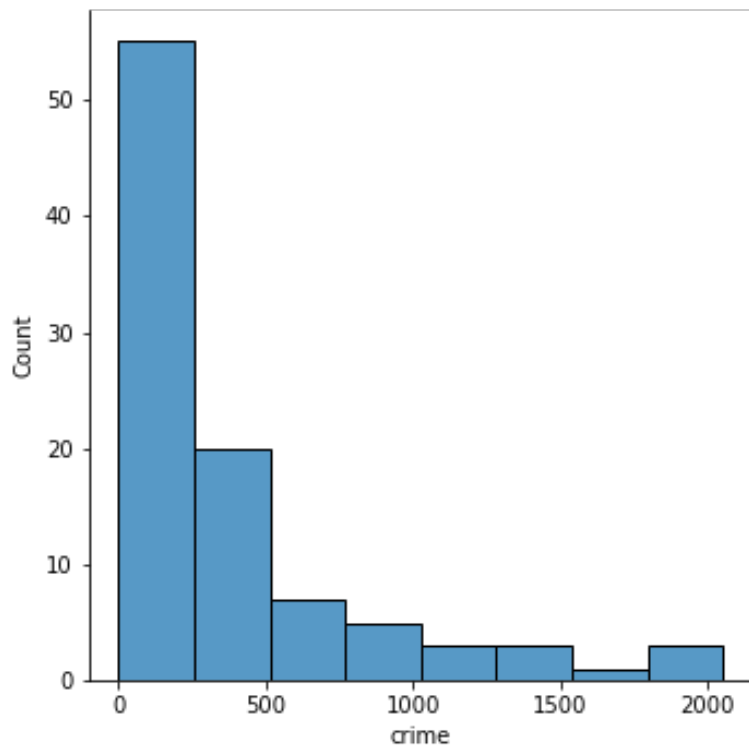


Comment: The scatter plot of log of enroll and log of crime shows their positive, moderate to strong relationship as stated in the previous problem. The plot of log of police and log of crime also shows a positive, moderate to strong correlation between each other, suggesting that campuses with more law enforcement have higher levels of crime. The private vs crime plot shows that private schools typically have lower levels of crime comparing to public schools.

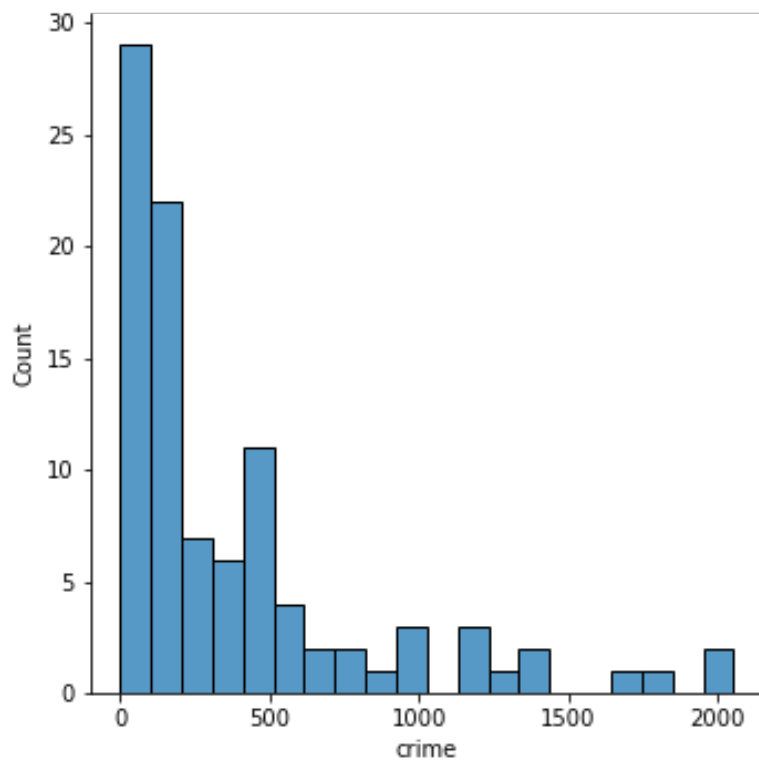
1.4



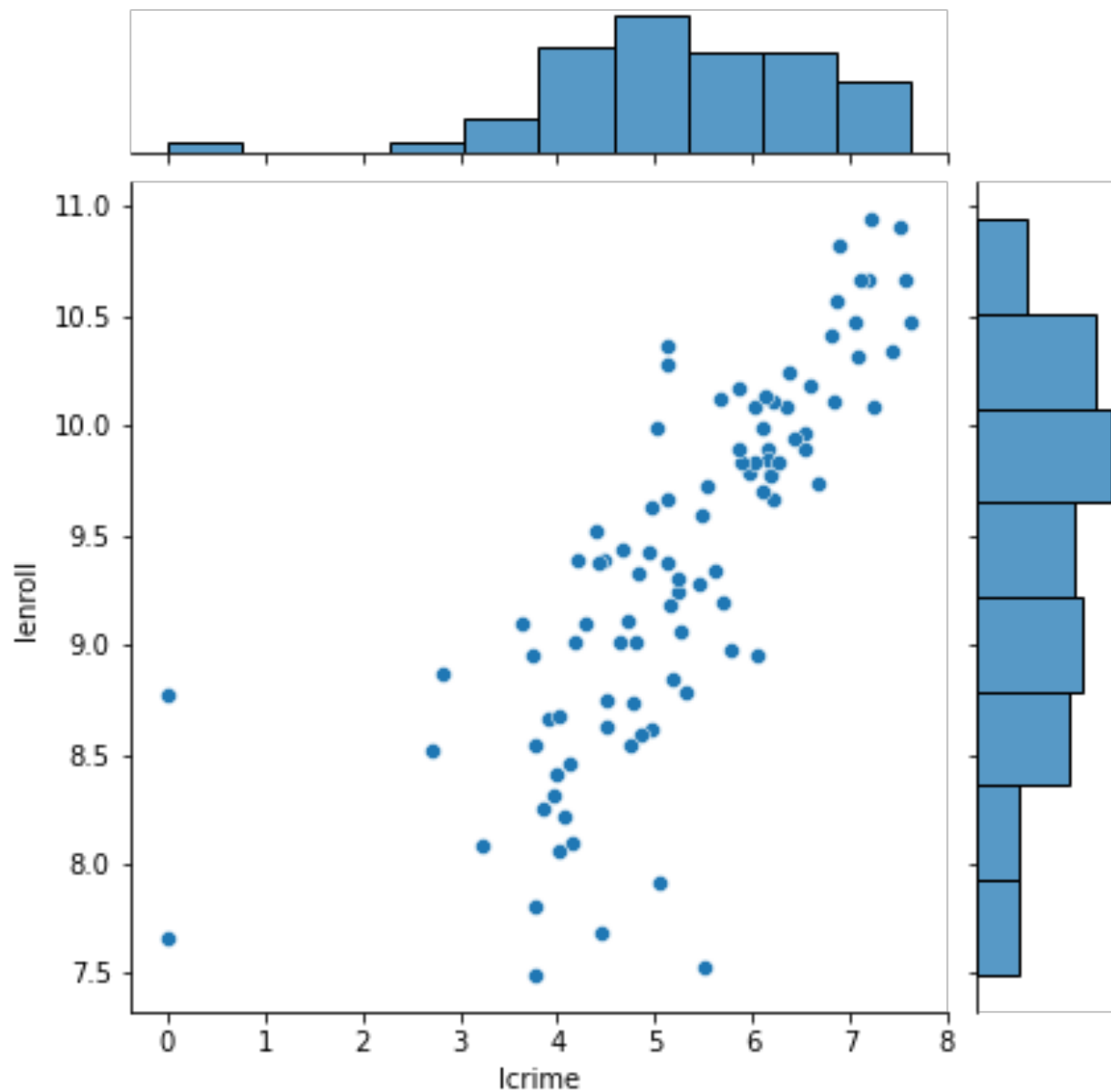
2.1A



2.1B

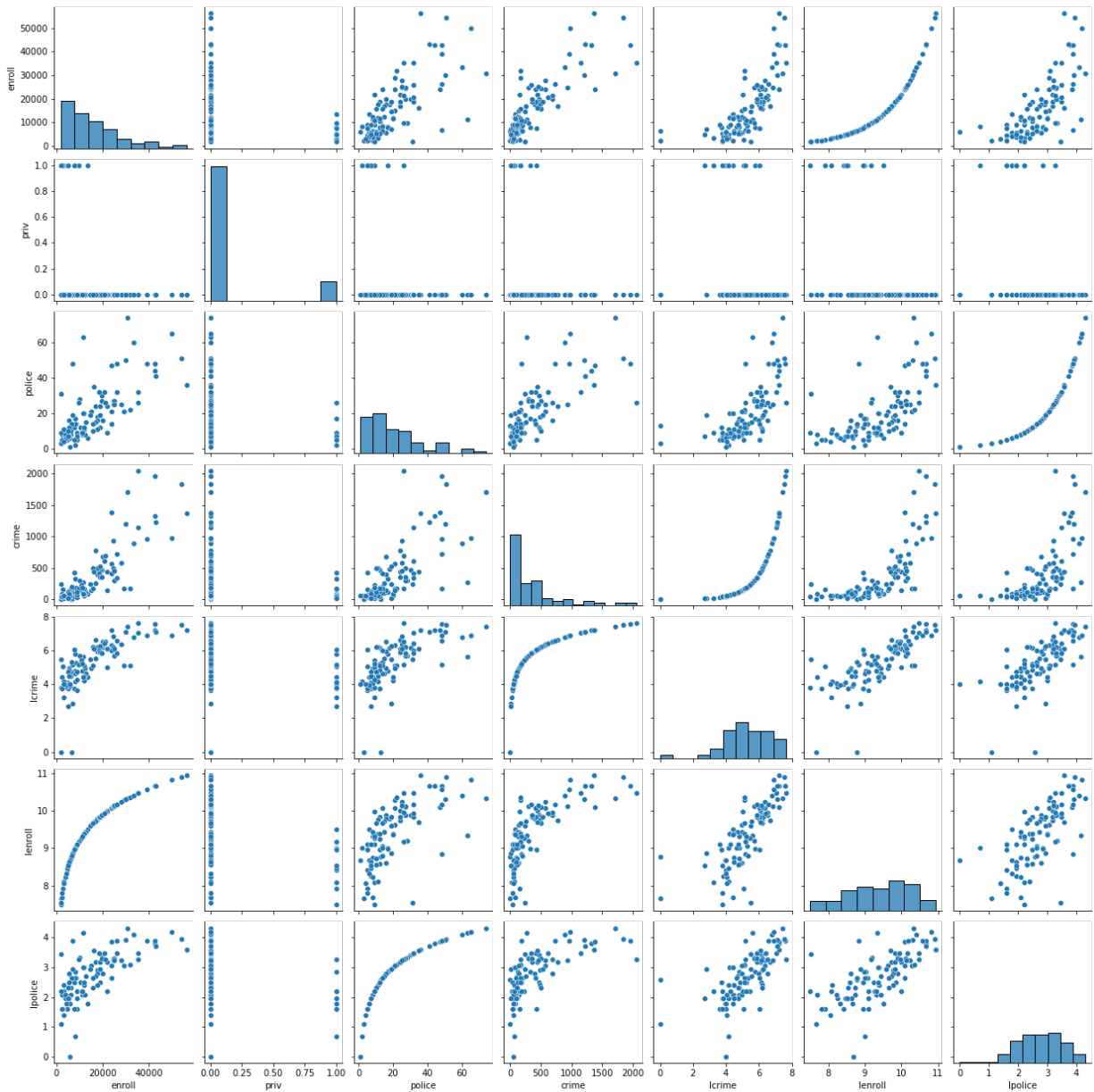


## 2.2



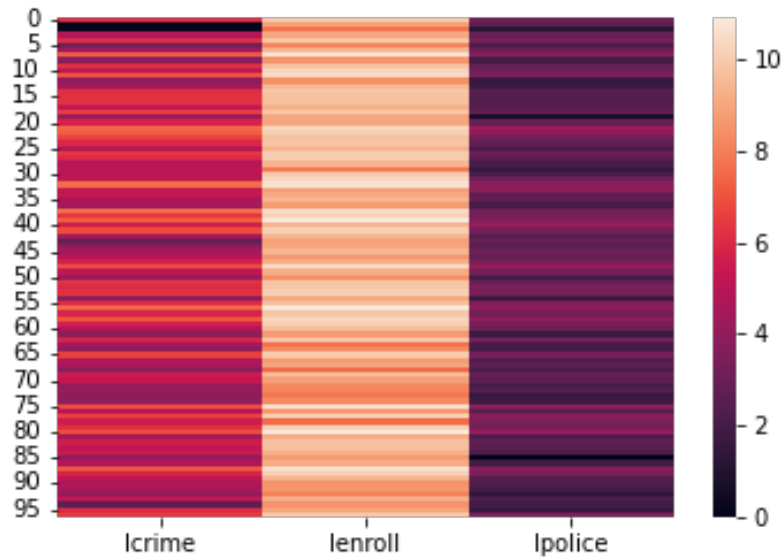
Comment: The plot shows the joint distribution between log of crime and log of enroll through the scatter plot, which shows that when there's higher levels of crime, there's also higher levels of enrollment. The plot also shows the marginal distributions of log of crime and log of enrollment through the histograms that are both unimodal and skewed to the left.

## 2.4



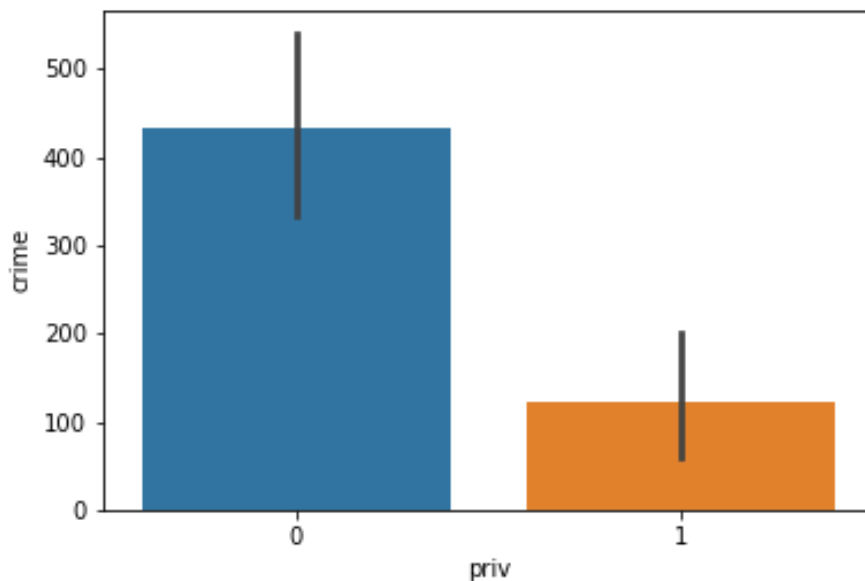
Comment: The entire figure shows a comprehensive overview of the relationships between variables enroll, priv, police, crime and their logarithmic versions. It also shows a histogram of each variable's marginal distributions, and we can make certain claims: Firstly, there's a positive, moderate to strong association between levels of crime and amount of student enrollment (size of the school), as well as the size of law enforcement; Secondly, there's a clear difference between crime levels of private schools comparing to public schools where private schools have significantly lower levels of crime.

2.5



Comment: The heat plot shows that log of crime is between 4-6, log of enroll is between 8-10 and log of police is between 0-2.

2.6



Comment: The box plot shows that when priv=0 (which means the school is public), levels of crime are higher. When priv=1 (which means the school is private), levels of crime are lower. Therefore, the boxplot shows that whether the school is public or private is a significant indicator to levels of crime.

Code:

```
"""This is the code file for the matplotlib exercise portion of problem set 4."""
```

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
```

```
# 1.1
```

```
df = pd.read_csv("campus.csv")
fig1 = plt.figure(1)
plt.hist(df["crime"])
plt.xlabel("Frequency of Crimes")
plt.ylabel("Number of Campuses")
fig1.savefig("1_1.png")
```

```
#1.2
```

```
fig2 = plt.figure(2)
plt.scatter(df["lenroll"], df["lcrime"])
plt.xlabel("lenroll")
plt.ylabel("lcrime")
fig2.savefig("1_2.png")
```

```
#1.3
```

```
fig3 = plt.figure(3)
ax1 = fig3.add_subplot(2, 2, 1)
ax2 = fig3.add_subplot(2, 2, 2)
ax3 = fig3.add_subplot(2, 2, 3)
ax1.scatter(df["lenroll"], df["lcrime"])
ax1.set_title("lenroll and lcrime")
ax2.scatter(df["lpolice"], df["lcrime"])
ax2.set_title("lpolice and lcrime")
ax3.scatter(df["priv"], df["crime"])
ax3.set_title("private and crime")
fig3.tight_layout()
fig3.savefig("1_3.png")
```

```
#1.4
BETA_0 = -6.63136926
BETA_1 = 1.26976026
fig4 = plt.figure(4)
plt.scatter(df["lenroll"], df["lcrime"], label="lenroll vs lcrime Scatterplot")
plt.xlabel("lenroll")
plt.ylabel("lcrime")
xl=np.linspace(7,11,100)
yl=BETA_0 + BETA_1*xl
plt.plot(xl,yl,color="red", label="OLS Regression")
plt.legend(loc="best")
plt.title("OLS line")
fig4.savefig("1_4.png")
```

```
#2.1A
sns.displot(df, x="crime", bins=8) # Part A
fig5 = plt.figure(5)
fig5.savefig("2_1A.png")
```

```
#2.1B
sns.displot(df, x="crime", bins=20, kde=False) #Part B
fig6 = plt.figure(6)
fig6.savefig("2_1B.png")
```

```
#2.2
sns.jointplot(data=df, x="lcrime", y="lenroll")
fig7 = plt.figure(7)
fig7.savefig("2_2.png")
```

```
#2.3
sns.lmplot(x="lcrime", y="lenroll", data=df)
fig8 = plt.figure(8)
fig8.savefig("2_3.png")
```

```
#2.4
sns.pairplot(df)
fig9 = plt.figure(9)
fig9.savefig("2_4.png")
```



#2.5

```
fig10 = plt.figure(10)
print(df[["lcrime", "lenroll", "lpolice"]].corr())
sns.heatmap(df[["lcrime", "lenroll", "lpolice"]])
fig10.savefig("2_5.png")
```

#2.6

```
fig11 = plt.figure(11)
sns.barplot(x='priv', y='crime', data=df)
fig11.savefig("2_6.png")
```