

Assignment #1

Dominic LaRoche

August 29, 2014

I will include embedded code for all analyses in this assignment but please let me know if you would like me to just include one chunk of code at the end of the document. I first read in the data from the .csv file.

```
rm(list=ls())  
crime<-read.csv("C:/Classes/GeoStat/crime.csv")
```

Now we can do some exploration of the variables to see how they are distributed and what the relationships might be between them. Figure 1 shows the distribution of the crime variable. We see some depatures from normality, although it is not too severe. Figure 2 shows the distribtion of the variable “HOVAL”, which has a much more severe depature from normality. Finally, figure 3 gives the sample distribution of the variable “TNC”, which also has a strong right skew (similar to “HOVAL”) and does not appear normally distributed.

```
par(mfrow=c(1,2))
hist(crime$CRIME,breaks="Scott")
qqnorm(crime$CRIME)
qqline(crime$CRIME)
```

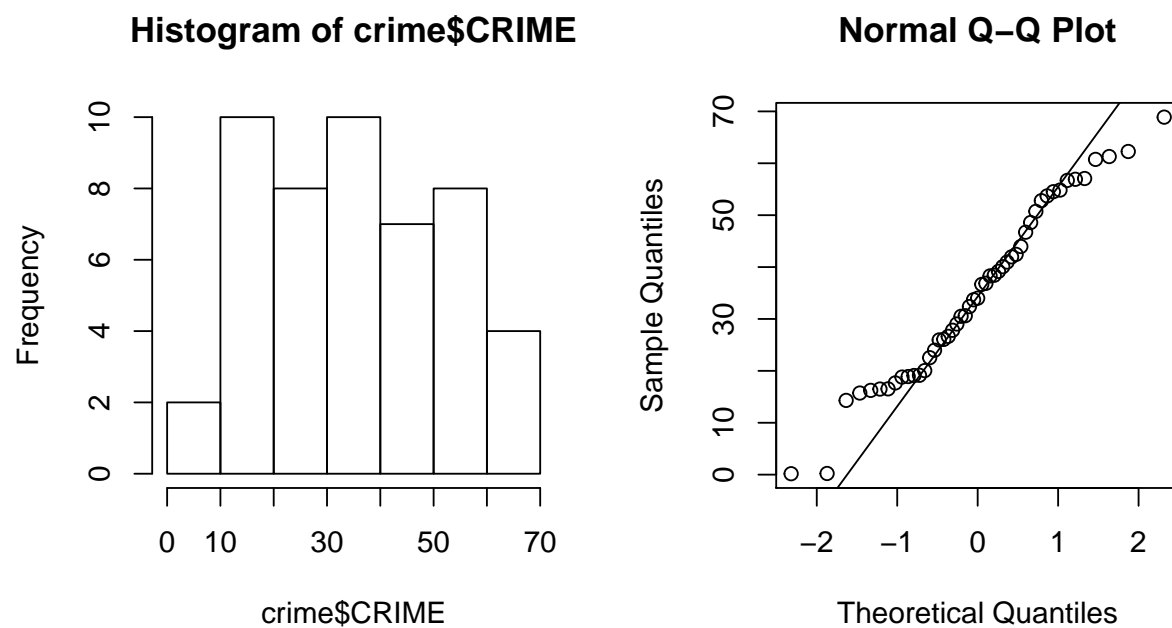


Figure 1: Histogram and qqplot for the CRIME variable showing departure from normality

```
par(mfrow=c(1,2))
hist(crime$HOVAL,breaks="Scott")
qqnorm(crime$HOVAL)
qqline(crime$HOVAL)
```

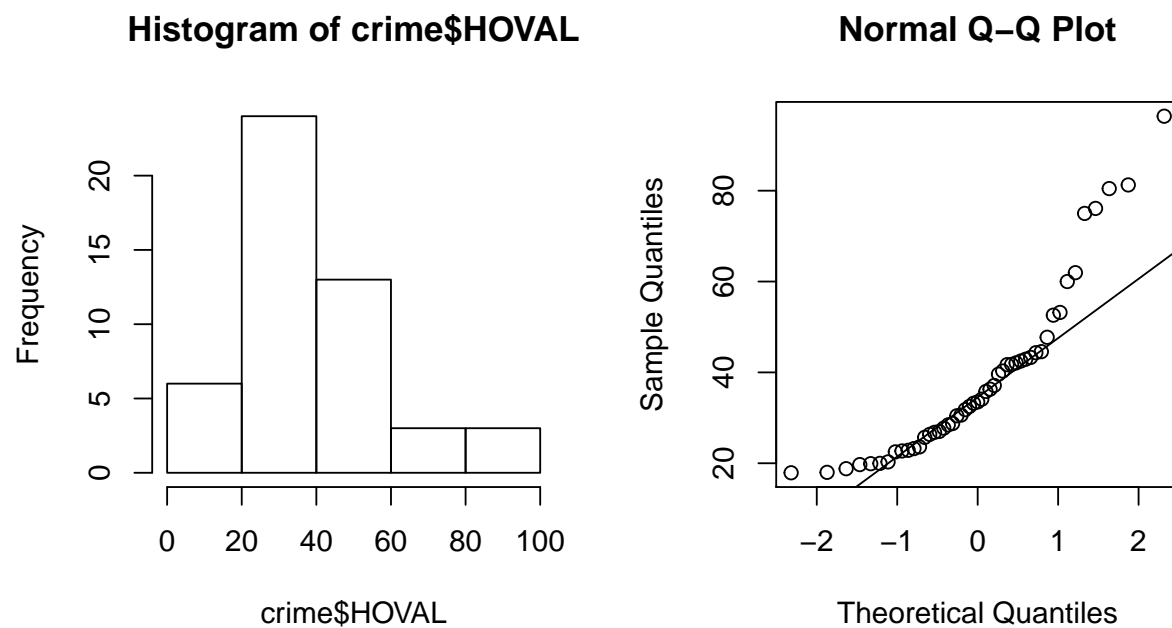
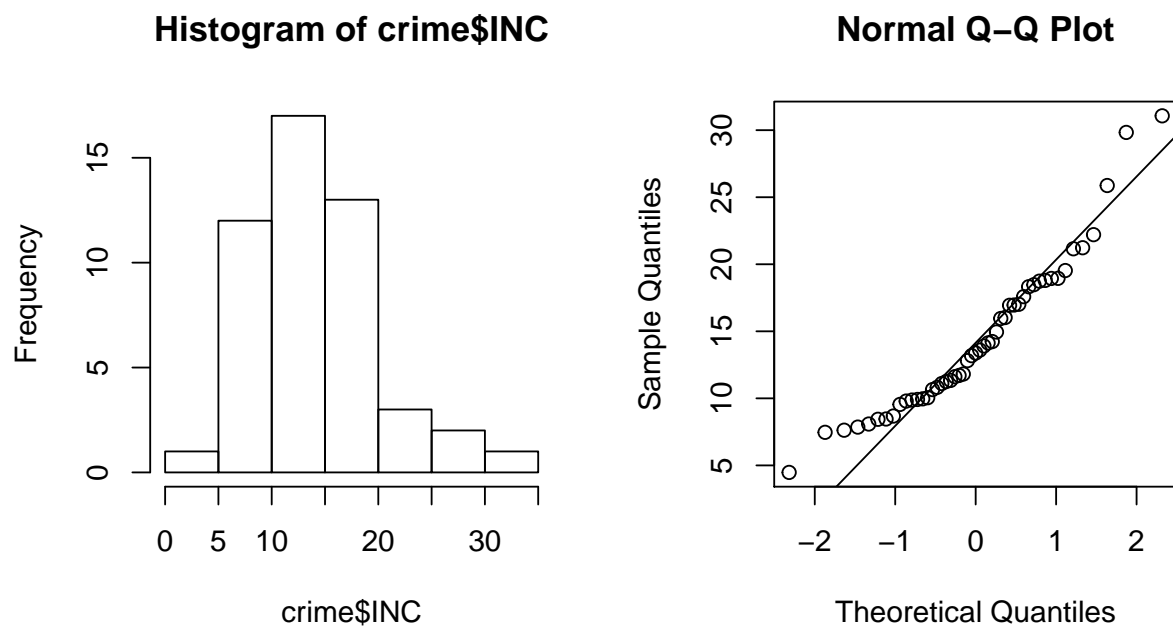


Figure 2: Histogram and qqplot for the variable “HOVAL” showing severe departure from normality with a strong right skew.

```
par(mfrow=c(1,2))
hist(crime$INC,breaks="Scott")
qqnorm(crime$INC)
qqline(crime$INC)
```



```
par(mfrow=c(1,1))
```

Figure 3: Histogram and qqplot for the variable “INC” showing the strong right skew and departure from normality

To see the relationship between these variables we can look at a scatterplot matrix (figure 4). We can see from this matrix that the income is positively correlated with home value and both home value and income are negatively correlated with crime.

```
plot(crime[,c("INC", "HOVAL", "CRIME")])
```

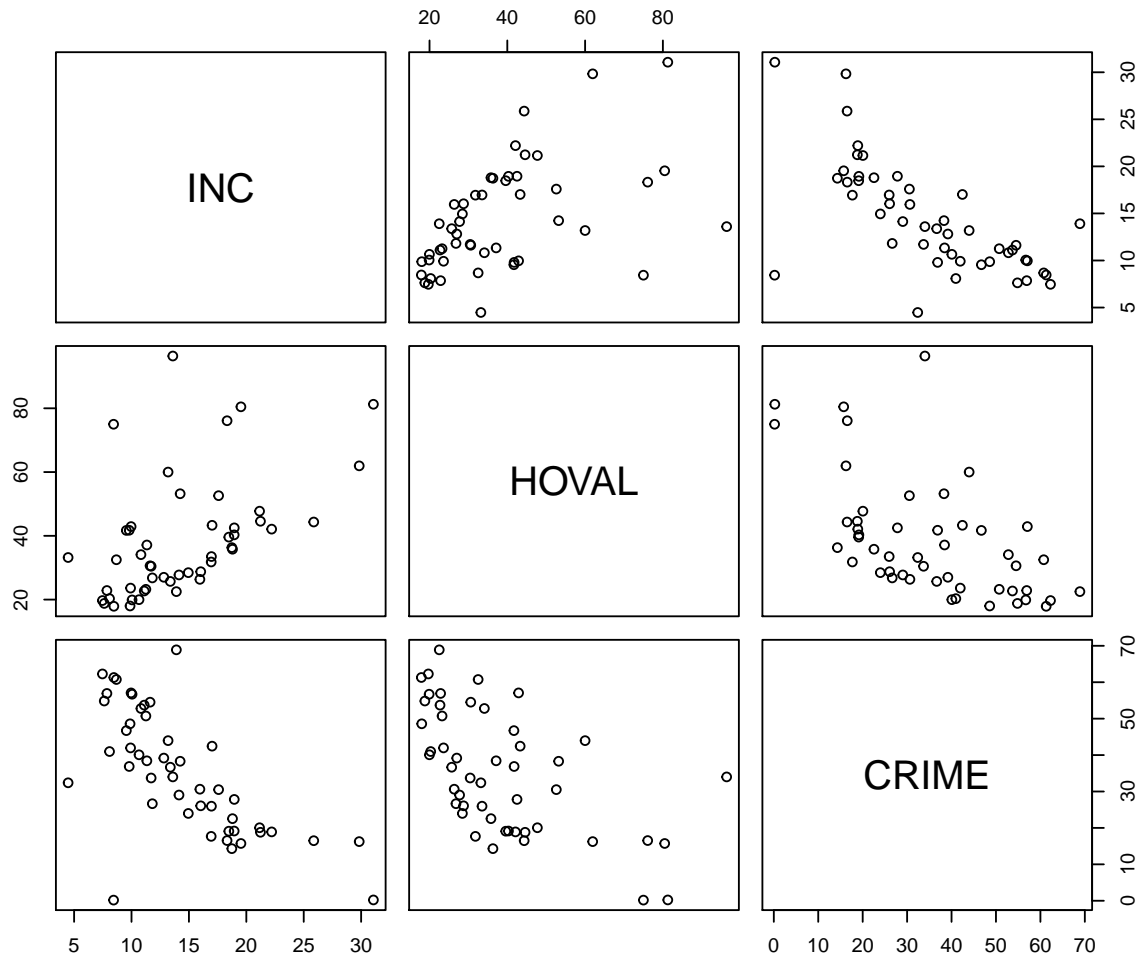


Figure 4: Scatterplot matrix of variables “INC”, “HOVAL”, and “CRIME” showing a positive relationship between “INC” and “HOVAL” and negative relationships between “INC” and “CRIME” as well as “CRIME” and “HOVAL”.

```
m1<-lm(CRIME~INC+HOVAL,data=crime)
r1<-residuals(m1)
f1<-fitted(m1)
```

We can model crime as a function of home value and income and get the following fitted model using

maximum likelihood methods (equivalent to least-squares in this case).

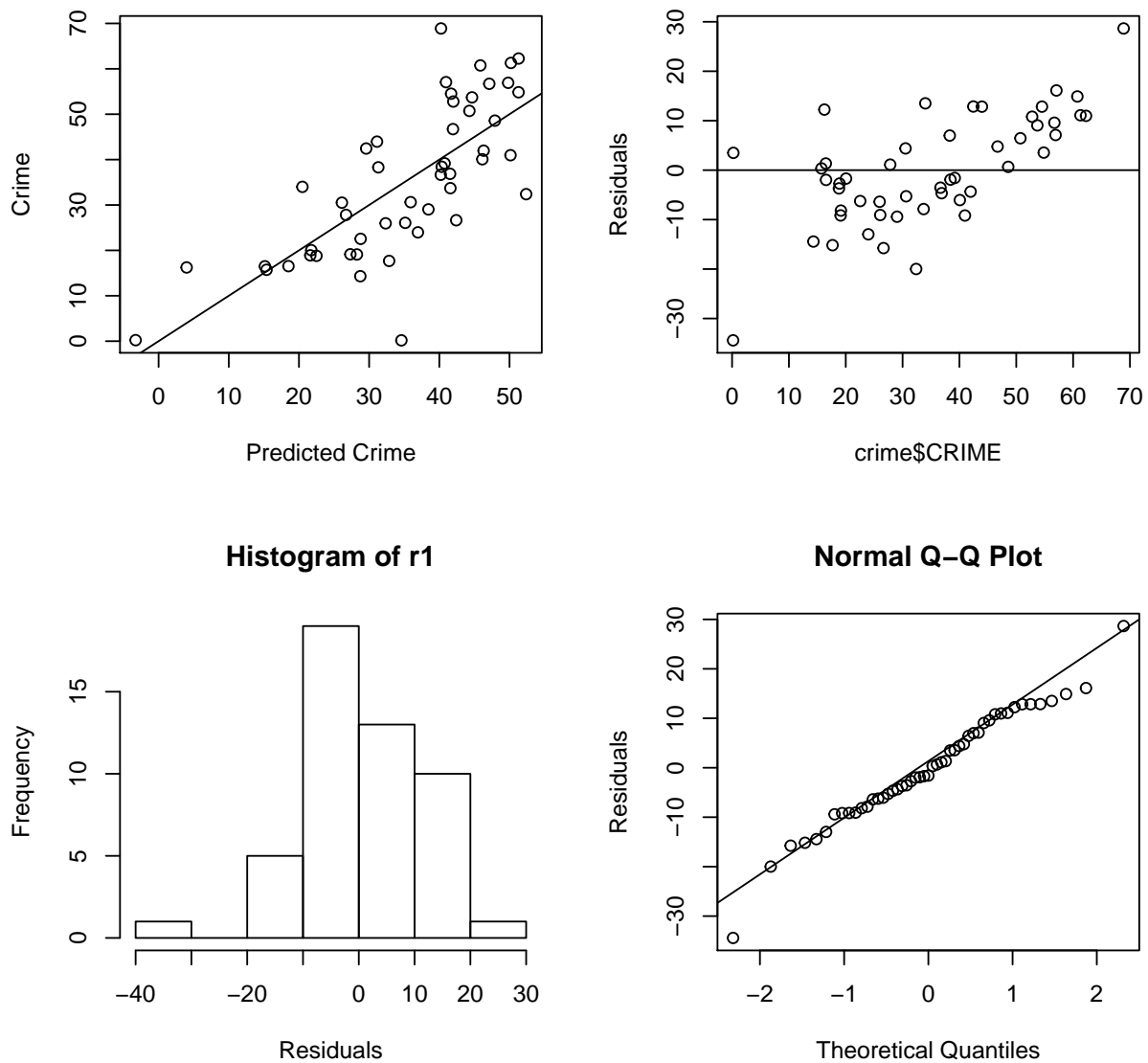
$$CRIME = 68.619 + -1.5973 \times INC + -0.2739 \times HOVAL + \epsilon$$

We must check to see if the fitted model violates the normality assumption of the linear regression, i.e. that the residuals (ϵ) are normally distributed.

```

par(mfrow=c(2,2))
plot(crime$CRIME~f1, ylab="Crime", xlab="Predicted Crime")
abline(0,1)
plot(crime$CRIME, r1, ylab="Residuals")
abline(h=0)
hist(r1, xlab="Residuals")
qqnorm(r1, ylab="Residuals")
qqline(r1)

```



```

par(mfrow=c(1,1))

```

Figure 5: Inspection of predicted vs observed values as well as the residuals indicates that the normality assumption generally holds for this model